

# Συστήματα και Τεχνολογίες Γνώσης

## Θέμα 1β

Στο θέμα αυτό θα εμπλουτίσετε σημασιολογικά το σύνολο δεδομένων μουσικών κομματιών το οποίο θα χρησιμοποιήσουμε στα θέματα.

### Σημασιολογικός εμπλουτισμός συνόλου δεδομένων μέσω πληθοπορισμού (crowdsourcing)

Τα τελευταία χρόνια η αύξηση της παραγωγής ποσότητας δεδομένων σε συνδυασμό με την τεχνολογική και επιστημονική πρόοδο για την αποθήκευση και επεξεργασία τους έχουν συντελέσει στο να αναβαθμιστεί η σημασία των καλά δομημένων και σημασιολογικά συνεπών μεγάλων συνόλων δεδομένων.

Οι εφαρμογές τεχνολογιών στα δεδομένα και κατ' επέκταση η χρησιμότητά τους φαίνεται σε πολλές πτυχές της ζωής. Αρχικά, η καθημερινότητα του ανθρώπου έχει αλλάξει σε μεγάλο βαθμό αφού πλέον με την πρόσβασή του σε χρήσιμες πληροφορίες μπορεί να αποφεύγει καθημερινά εμπόδια, όπως για παράδειγμα η κίνηση στους δρόμους. Η εφαρμογή αυτή μπορεί να επιτευχθεί και μέσω μιας διαδικασίας crowdsourcing. Συλλέγεται δηλαδή πληροφορία από ένα μεγάλο πλήθος κόσμου προκειμένου να λυθεί ένα πρόβλημα της καθημερινότητας. Αν και χρήσιμο ορισμένες φορές, υπάρχουν “κενά” σε αυτήν την διαδικασία. Διαβάστε [εδώ](#).

Οι εφαρμογές δεν σταματάνε στην καθημερινότητα - έχουν και μεγάλη επιρροή στην επιστήμη, όπου ο διαμοιρασμός γνώσης και αποτελεσμάτων μεταξύ ερευνητών έχουν οδηγήσει στην ραγδαία εξέλιξή της ([εδώ](#)). Επίσης, εμφανή είναι και τα αποτελέσματα στον πολιτισμό: χαρακτηριστικά παραδείγματα είναι η ψηφιοποίηση μουσείων ή η συλλογή δεδομένων για την πολιτιστική κληρονομιά ενός τόπου. Ένα τέτοιο παράδειγμα είναι η ιστοσελίδα της [Europeana](#) η οποία προσφέρει ψηφιακή πρόσβαση σε υλικό ευρωπαϊκής πολιτιστικής κληρονομιάς.

Τα προβλήματα που προκύπτουν από την συλλογή ενός συνόλου δεδομένων είναι αρκετά. Πολλές φορές τα δεδομένα δεν είναι δομημένα και αυτό έχει ως αποτέλεσμα να μην είναι φιλικά προς τον άνθρωπο, αλλά και να καθίσταται δύσκολη η εύρεση και η ανάλυσή τους με την χρήση υπολογιστή. Μια λύση σε αυτό είναι η συγκρότηση σημασιολογικών βάσεων δεδομένων, κάτι το οποίο θα μελετήσουμε περαιτέρω στα πλαίσια του μαθήματος.

Ένα άλλο πρόβλημα είναι ότι πολλές φορές τα δεδομένα δεν είναι αρκετά για την ανάπτυξη μιας ενδιαφέρουσας εφαρμογής. Ένας τρόπος για την αντιμετώπιση αυτού το προβλήματος είναι το crowdsourcing. Η λογική του crowdsourcing βασίζεται στο ότι μπορείς να εμπλουτίσεις τα δεδομένα σου με μικρή προσπάθεια πολλών ανθρώπων. Συγκεκριμένα λειτουργεί με καμπάνιες (campaigns) οι οποίες καλούν τους χρήστες να επισημειώσουν (annotate) τα μεταδεδομένα ενός συνόλου δεδομένων έτσι ώστε αυτό να εμπλουτιστεί σημασιολογικά.

Σε αυτή τη σειρά των εργαστηριακών θεμάτων θα μελετήσουμε την αξία της σημασιολογικά δομημένης και συνεπούς πληροφορίας, καθώς και των τεχνικών εμπλουτισμού και αναζήτησής της. Συγκεκριμένα, προχωρήσαμε στη συγκρότηση ενός συνόλου δεδομένων μουσικών κομματιών μέσω της ιστοσελίδας της Europeana στο οποίο θα βασιστούν οι

## Η πλατφόρμα CrowdHeritage

Καλείστε λοιπόν να συμμετέχετε σε αυτή την διαδικασία στην οποία θα παίζετε τον ρόλο του επισημειωτή (annotator) εμπλουτίζοντας το σύνολο δεδομένων με ετικέτες οι οποίες θα αφορούν το μουσικό είδος του τραγουδιού, το συναίσθημα που σας προκαλεί καθώς και τα μουσικά όργανα που αναγνωρίζετε σε αυτό. Οι ετικέτες συνδέονται με καθιερωμένα λεξιλόγια ([wikidata](#)) και αντιστοιχίζονται με όρους που χαρακτηρίζονται από ένα Universal Resource Identifier (URI). Η πληροφορία αυτή θα σας φανεί ιδιαίτερα χρήσιμη σε παρακάτω στάδια της άσκησης: όσο πιο πλούσια και δομημένα τα δεδομένα που έχετε, τόσο πιο ενδιαφέροντα συμπεράσματα θα μπορέσετε να εξαγάγετε με αυτόματο τρόπο και τόσο πιο αποτελεσματικό θα είναι το σύστημα συστάσεων που θα κληθείτε να υλοποιήσετε. Επιπλέον, με τον τρόπο αυτό συμβάλετε στη δημιουργία μιας εμπλουτισμένης βάσης δεδομένων που μπορεί να αξιοποιηθεί περαιτέρω για ερευνητικούς σκοπούς στον τομέα της ανάκτησης μουσικής πληροφορίας ([Music Information Retrieval - MIR](#)).



# Οδηγίες για τη συμμετοχή στο crowdsourcing campaign

## Εγγραφή στην πλατφόρμα CrowdHeritage

Η εγγραφή που θα κάνετε στην [πλατφόρμα](#) πρέπει να περιλαμβάνει το ονοματεπώνυμό σας, ώστε να είναι δυνατή η αξιολόγηση της συμμετοχής σας. Για username προσθέστε τον αριθμό μητρώου σας.

## Περιγραφή της διαδικασίας του annotation

Συμμετέχοντας στο [campaign](#) καλείστε να ακούσετε μια σειρά από μουσικά κομμάτια που έχουν συλλεχθεί από την Europeana και να τα κάνετε annotate σε τρία πεδία (emotions, genres, instruments). Τα annotation tags βασίζονται σε τρία λεξιλόγια (vocabularies), ένα για κάθε πεδίο ([Δείτε εδώ](#)). Πατώντας στο κάθε πεδίο, σας εμφανίζονται όλα τα δυνατά annotation tags. Επιλέγοντάς κάποια, τα προσθέτετε στο μουσικό κομμάτι και γίνονται αυτόματα upvote. Αν συμφωνείτε με ένα annotation tag που έχει προταθεί ήδη, αρκεί μόνο να το υπερψηφίσετε (upvote). Αν θεωρείτε ότι ένα annotation tag που έχει προταθεί είναι άστοχο μπορείτε να το καταψηφίσετε (downvote). Επίσης, έχουμε χωρίσει το σύνολο δεδομένων σε 8 υποομάδες, έτσι ώστε να γίνουν annotate όλα τα αντικείμενα στον ίδιο βαθμό. [Εδώ](#) σας δίνεται ο αριθμός (1-8) ο οποίος αντιστοιχεί στην υποομάδα του συνόλου δεδομένων που καλείστε να κάνετε annotate.

Πιο αναλυτικά για το κάθε πεδίο:

Emotions:

Εδώ καλείστε να επιλέξετε ποιο συναίσθημα σας προκαλεί η ακρόαση του μουσικού κομματιού. Μπορείτε να προσθέσετε ή να κάνετε upvote μέχρι δύο annotation tags. Οι ετικέτες των συναισθημάτων έχουν βασιστεί στον κύκλο του [Russell](#).

Genres:

Εδώ καλείστε να επιλέξετε το μουσικό είδος ή τα μουσικά είδη στα οποία θεωρείτε ότι ανήκει το μουσικό κομμάτι. Μπορείτε να προσθέσετε ή να κάνετε upvote μέχρι δύο annotation tags.

Instruments:

Εδώ καλείστε να επιλέξετε ποια μουσικά όργανα ακούτε στο μουσικό κομμάτι. Αν ακούτε ολόκληρη κλασική ορχήστρα, επιλέξτε μόνο την ετικέτα Orchestra. Σε άλλη περίπτωση δεν υπάρχει όριο στο πόσα annotation tags μπορείτε να προσθέσετε ή να κάνετε upvote.

## Commenting και Favorites

Θα έχετε δύο ακόμα επιλογές κατά την διαδικασία του annotation. Η πρώτη είναι το commenting, με το οποία θα μπορείτε να γράψετε στα αγγλικά τον συνειρμό που σας προκάλεσε η ακρόαση του μουσικού κομματιού. Η δεύτερη είναι το favorite, όπου θα μπορείτε να επιλέξετε τα αγαπημένα σας μουσικά κομμάτια. Με αυτά τα χαρακτηριστικά, στα επόμενα εργαστηριακά θέματα, θα μπορέσουμε να αξιολογήσουμε συστήματα συστάσεων (recommender systems) που θα κατασκευάσουμε καθώς και να εξάγουμε χρήσιμη γνώση για την περαιτέρω κατηγοριοποίηση των μουσικών μας κομματιών με τεχνικές επεξεργασίας φυσικής γλώσσας (NLP - Natural Language Processing).



[Εδώ](#) σας δίνεται ένα μικρό tutorial για το πώς να κάνετε annotate τα items - μουσικά κομμάτια.

## Αξιολόγηση και βαθμολογία

Η αξιολόγηση θα γίνει με βάση τον αριθμό και την ποιότητα των annotations που θα κάνετε. Προκειμένου να λάβετε την πλήρη βαθμολογία του αρχικού αυτού βήματος θα πρέπει να έχετε κάνει up-votes σε annotations tags και των τριών πεδίων (emotions, genres, instruments) σε τουλάχιστον 100 items - μουσικά κομμάτια της υποομάδας στην οποία ανήκετε (1-8). Είναι σημαντικό για την ποιότητα του σημασιολογικού εμπλουτισμού η διαδικασία του annotation να μην γίνεται τυχαία. Η τυχαιότητα θα φανεί από τα πόσα downvotes έχει μαζέψει ένα annotation tag που έχετε προτείνει. Το πόσα items - μουσικά κομμάτια έχετε κάνει annotate μπορείτε να το δείτε από το προφίλ σας στο πεδίο statistics. Εκεί μπορείτε να δείτε και άλλες πληροφορίες για την συμμετοχή σας στο campaign όπως το πόσα annotation tags έχετε κάνει upvote, downvote και έχετε προσθέσει.

## Συμπλήρωση ερωτηματολογίου

Μετά το τέλος της διαδικασίας του crowdsourcing καλείστε να συμπληρώσετε προαιρετικά ένα ερωτηματολόγιο ([εδώ](#)) για τις δυσκολίες που μπορεί να συναντήσατε κατά τη διαδικασία του crowdsourcing ή για τυχόν δικές σας παρατηρήσεις και προτάσεις. Η διαδικασία αυτή είναι ανώνυμη. Η πληροφορία που θα συγκεντρωθεί μέσω του ερωτηματολογίου αποσκοπεί στην βελτίωση της crowdsourcing πλατφόρμας CrowdHeritage και θα αξιοποιηθεί για ερευνητικούς σκοπούς στα πλαίσια του Erasmus έργου CitizenHeritage - Citizen Science Practices in Cultural Heritage: towards a Sustainable Model in Higher Education. Στο έργο συμμετέχουν εκπαιδευτικοί και πολιτιστικοί οργανισμοί από διάφορες χώρες, συμπεριλαμβανομένου και του ΕΜΠ, με σκοπό την διερεύνηση του ρόλου των πολιτών στην επιστήμη. Προτού ξεκινήσετε τη συμπλήρωση του ερωτηματολογίου θα πρέπει να συναινέσετε στην Πολιτική Προστασίας Προσωπικών Δεδομένων.

