



**Πανεπιστήμιο Πειραιώς ΠΜΣ Διδακτική της Τεχνολογίας
και Ψηφιακών Συστημάτων**

Δικτυοκεντρικά Συστήματα

**Εξαγωγή Χωροχρονικής Πληροφορίας από Κοινωνικά Δίκτυα και
Ανάπτυξη Εφαρμογής**

Βοντετσιάνου Μαρία

Επιβλέπων Καθηγητής

Δουλκερίδης Χρήστος

Πειραιάς 2015

Περίληψη

Ο τεράστιος όγκος δεδομένων μαζί με την ανάγκη για αξιοποίηση της πληροφορίας αυτής δημιουργούν την ανάγκη νέων τεχνολογιών και μεθόδων. Η βελτίωση των μεθόδων ή ανακάλυψη νέων για την ανάκτηση πληροφορίας από μεγάλες βάσεις δεδομένων αποτελεί πρόκληση. Η ραγδαία ανάπτυξη των τεχνολογιών και των μεθόδων ανάκτησης/εξαγωγής μεγάλου όγκου δεδομένων, οδηγεί ταυτόχρονα στην αύξηση των αναγκών για την εύρεση βελτιωμένων τεχνικών, όσον αφορά τον χρόνο ανταπόκρισης αλλά και την απόδοση. Η παρούσα μεταπτυχιακή εργασία στοχεύει στην εξαγωγή πληροφορίας από κοινωνικά δίκτυα, όπως το Flickr, και την αποθήκευσή τους σε βάση δεδομένων. Η πληροφορία που θα εξαχθεί θα ικανοποιεί κάποιους περιορισμούς, για παράδειγμα θα αφορά κάποιες συγκεκριμένες γεωγραφικές θέσεις ή θα αφορά μια προκαθορισμένη χρονική περίοδο. Για το σκοπό αυτό έχει μελετηθεί το API του κοινωνικού δικτύου και έχουν καταγραφεί οι δυνατότητες που παρέχει. Για την εξαγωγή πληροφορίας και την αποθήκευσή αυτής, χρησιμοποιήθηκε η NoSQL βάση δεδομένων MongoDB. Στόχος της εργασίας είναι να αναπτυχθεί web εφαρμογή, η οποία θα δίνει τη δυνατότητα υποβολής ερωτήσεων που συνδυάζει αναζήτηση με λέξεις κλειδιά με γεωγραφικούς ή χρονικούς περιορισμούς.

Abstract

A big amount of data along with the need for retrieval of this information, creates the need for new technologies and methods. Improved methods or discovering new for information retrieval from large databases is a challenge. The rapid development of technologies and methods to retrieve / export large amounts of data simultaneously leads to increased needs for finding improved techniques, in terms of response time and performance. This thesis aims at extracting information from social networks such as Flickr, and its storage in a database. The exported information meets some requirements, for example, will cover some specific locations or a predetermined time period. For the purpose of the thesis, we studied the Flickr API and analyzed the potentials it offers to the user. In order to export and storage the exported data, a NoSQL database, MongoDB, a cross-platform document-oriented database is used. We also created a web application which gives the opportunity to combine a simple search with keywords with geographical or time constraints.

Περιεχόμενα

Περίληψη.....	2
Abstract	3
Κεφάλαιο 1	5
1.1 Εισαγωγή.....	5
1.2 Σκοπός.....	6
Κεφάλαιο 2	7
2.1 Flickr	7
2.2 Flickr API.....	8
2.3 API Methods.....	9
Κεφάλαιο 3	16
3.1 Βάσεις Δεδομένων.....	16
3.2 Σύστημα Διαχείρισης Βάσεων Δεδομένων (ΣΔΒΔ).....	16
3.3 Σχεσιακή Βάση Δεδομένων.....	17
3.4 ACID	18
3.5 Μη Σχεσιακές Βάσεις Δεδομένων - Solr.....	20
Κεφάλαιο 4	29
4.1 Αρχιτεκτονική Εφαρμογής.....	29
4.2 Ιεραρχικό Διάγραμμα	32
Κεφάλαιο 5	33
5.1 Κώδικας.....	33
5.2 Εγκατάσταση.....	35
Κεφάλαιο 6	37
Δοκιμές – Αποτελέσματα	37
Κεφάλαιο 7	45
Οδηγίες Εφαρμογής (Manual).....	45
Κεφάλαιο 8	59
Συμπεράσματα.....	59
Βιβλιογραφία	62

Κεφάλαιο 1

1.1 Εισαγωγή

Ο συνεχώς αυξανόμενος όγκος δεδομένων που υπάρχει στο διαδίκτυο, τα τελευταία χρόνια έχει δημιουργήσει την ανάγκη για νέα συστημάτων βάσεων δεδομένων. Οι παραδοσιακές σχεσιακές βάσεις πλέον αδυνατούν να ανταποκριθούν όταν πρόκειται να διαχειριστούν μεγάλο όγκο δεδομένων. Για να γίνει δυνατή η διαχείρισή και η εξαγωγή χρήσιμων πληροφοριών - συμπερασμάτων από τα δεδομένα αυτά, τα νέα συστήματα βάσεων δεδομένων διαφοροποιούνται από τα παραδοσιακά, προκειμένου να εξασφαλίσουν τη δυνατότητα οριζόντιας κλιμάκωσης και κατανομής των δεδομένων σε πολλαπλούς κόμβους για παράλληλη αποθήκευση και επεξεργασία. Στην παρούσα διπλωματική εργασία το πρόβλημα που μελετάται είναι η εξαγωγή δεδομένων από κοινωνικό δίκτυο (Flickr) με χρήση μεταδεδομένων, όπως tags, geotags, κτλ, η αποθήκευση αυτών σε βάση δεδομένων, και στη συνέχεια η ανάπτυξη web εφαρμογής που επιτρέπει επερωτήσεις που δεν είναι δυνατό να γίνουν χρησιμοποιώντας το API του κοινωνικού δικτύου (Flickr). Ο εντοπισμός δεδομένων με βάση το περιεχόμενό τους γίνεται με χρήση των metadata, αρχείων που περιέχουν την περιγραφή του περιεχομένου τους. Τα αρχεία στα οποία θα εκτελεστεί η αναζήτηση συνδυάζονται με metadata. Ευρέως χρησιμοποιούνται σκληροί δίσκοι και σχεσιακές βάσεις δεδομένων για την διατήρηση των metadata, πρακτικές που παρουσιάζουν διάφορα προβλήματα στην ταχύτητα εκτέλεσης της αναζήτησης, όπως και για την εξέταση των metadata, όπου απαιτούνται εργαλεία XML Parsing και εφαρμογή πολλών πολύπλοκων JOINS. Η διπλωματική αυτή προσπάθει να λύσει το παραπάνω πρόβλημα με εφαρμογή εναλλακτικών τρόπων αποθήκευσης. Καθώς τα πρότυπα μέχρι σήμερα επί τω πλείστον βασίζονται σε σχεσιακές βάσεις δεδομένων, όπου απαιτείται η χρήση πολλών διαφορετικών tables και διατήρηση πολλών relations για την πλήρη αναπαράσταση των μεταξύ τους σχέσεων.

1.2 Σκοπός

Η συγκεκριμένη εργασία έχει σκοπό την εξαγωγή δεδομένων από το Flickr. Την αποθήκευση τους, και στην συνέχεια την αναζήτηση αυτών, μέσω ερωτημάτων, μέσα από την εφαρμογή που υλοποιήθηκε για αυτόν τον σκοπό. Γενικότερα στα κοινωνικά δίκτυα υπάρχει μεγάλος όγκος χρηστο-παραγόμενων δεδομένων (user-generated) που αυξάνεται με ταχύτατους ρυθμούς καθημερινά. Αυτό που θέλουμε είναι να αναλύσουμε αυτά τα δεδομένα, όμως το API που μας παρέχεται έχει περιορισμούς, Αρχική ιδέα είναι να εξάγουμε τα πρωτογενή δεδομένα, να τα αποθηκεύσουμε κάπου, και να μπορέσουμε εν συνεχεία να τα αναζητήσουμε και να τα αναλύσουμε. Πιο αναλυτικά, να μπορέσουμε να εξάγουμε πληροφορίες από το Flickr, δηλαδή φωτογραφίες που ανεβάζουν οι χρήστες του, να τις αποθηκεύσουμε στη βάση δεδομένων της Mongo DB και μετά να μπορέσουμε να εξάγουμε αυτήν την πληροφορία μέσω της εφαρμογής που θα φτιαχτεί για τον σκοπό αυτό. Η MongoDB απέδειξε ότι είναι ιδανική σε τέτοιες περιπτώσεις, εξαγωγής μεγάλου όγκου δεδομένων και συγκεκριμένα φωτογραφιών. Στην ουσία η συγκεκριμένη εργασία δεν έχει ως στόχο μόνο την εξαγωγή πληροφορίας και στην συνέχεια την αναζήτηση της, αλλά την διεκπεραίωση της όλης διαδικασίας σε εύλογο χρόνο. Μετά από έρευνα διαπιστώθηκε ότι μία σχεσιακή βάση δεν θα ήταν καθόλου καλή επιλογή. Καθώς σε τέτοιες περιπτώσεις εξαγωγής αδόμητων δεδομένων από το Flickr, δηλαδή φωτογραφίες μαζί με σχόλια, views, χρήστες, likes, γεωγραφική θέση και η εξαγωγή δεδομένων δεν μπορεί να εφαρμοστεί σε μία σχεσιακή βάση δεδομένων, λόγω των πολλαπλών πινάκων (tables) που χρησιμοποιεί και της σχέσης ανάμεσα στα δεδομένα.

Κεφάλαιο 2

2.1 Flickr

Τα τελευταία χρόνια έχει παρατηρηθεί μια αξιοσημείωτη ανάπτυξη σε ιστοτόπους κοινωνικών δικτύων καθώς και σε online εφαρμογές τους όπου “χρήστες διαμοιράζονται πόρους, δημιουργούν, αναθέτουν ετικέτες σε περιεχόμενο και το αξιολογούν. Τέτοιου είδους εφαρμογές ποικίλλουν σε μεγάλο βαθμό καθώς κυμαινόμαστε από ιστοτόπους όπου ο χρήστης μπορεί να προσθέσει, να οργανώσει και να μοιραστεί bookmarks, ακαδημαϊκές αναφορές και φωτογραφίες (π.χ. Flickr.com). Ένα πράγμα το οποίο έχουν ως κοινό αυτοί οι ιστότοποι είναι η έμφαση που δίνουν στην online συνεργασία και στο διαμοιρασμό των πόρων μεταξύ των χρηστών. Το **Flickr** είναι μια ιστοσελίδα, η οποία δημιουργήθηκε για να φιλοξενεί φωτογραφίες και βίντεο από τους χρήστες του. Δημιουργήθηκε αρχικά από την εταιρία Ludicorp και ύστερα εξαγοράστηκε από την Yahoo. Είναι ένας δημοφιλής ιστότοπος κοινωνικού χαρακτήρα στον οποίο επιτελείται online διαχείριση εικόνων και εφαρμογών διαμοιρασμού τους. Το Flickr επιτρέπει στους χρήστες να κάνουν upload, να αποθηκεύσουν και να μοιραστούν ψηφιακές εικόνες οι οποίες υπάρχει η δυνατότητα να έχουν χαρακτηριστεί από ετικέτες (ελεύθερα επιλεγμένες λέξεις-κλειδιά) προκειμένου να βοηθήσουν την οργάνωση και την ανάκτηση τους. Τέτοιου είδους εικόνες υπάρχει η δυνατότητα να χαρακτηριστούν ως ιδιωτικές, να είναι ορατές σε επιλεγμένους φίλους και στην οικογένεια μόνο ή μπορούν να είναι ορατές από τη συνολική κοινότητα χρηστών του συστήματος. Η υπηρεσία χρησιμοποιείται συχνά από bloggers για να ενσωματώσουν τις φωτογραφίες τους στα blogs τους. Τον Σεπτέμβριο του 2010, το Flickr έφτασε τα 5 δις. φωτογραφιών. Η εγγραφή στο Flickr μπορεί να είναι είτε δωρεάν είτε επί πληρωμή, η οποία παρέχει περισσότερα δικαιώματα στους χρήστες του.



2.2 Flickr API

Προκύπτει από την συντομία (application program interface) διεπαφή προγραμματισμού εφαρμογών, είναι ένα σύνολο από ρουτίνες, πρωτόκολλα και εργαλεία για τη δημιουργία εφαρμογών λογισμικού. Ένα καλό API καθιστά ευκολότερη την ανάπτυξη ενός προγράμματος, παρέχοντας όλα τα δομικά στοιχεία. Τα περισσότερα λειτουργικά περιβάλλοντα, όπως το MS-Windows, παρέχουν ένα API, έτσι ώστε οι προγραμματιστές να μπορούν να γράψουν εφαρμογές σύμφωνα με το περιβάλλον λειτουργίας της κάθε εφαρμογής. Αν και τα APIs έχουν σχεδιαστεί για τους προγραμματιστές, είναι τελικά καλό και για τους χρήστες, επειδή εγγυάται ότι όλα τα προγράμματα που χρησιμοποιούν ένα κοινό API θα έχουν παρόμοιες διασυνδέσεις. Έτσι οι χρήστες μπορούν ευκολότερα να μάθουν νέα προγράμματα.

Για να μπορέσει κάποιος να χρησιμοποιήσει το Flickr API χρειάζεται να διαθέτει ένα κλειδί της εφαρμογής. Αυτό χρησιμοποιείται από τους διαχειριστές του Flickr για να εντοπίζουν τη χρήση του API. Για να κάνουμε αναζήτηση των φωτογραφιών που πληρούν τα κριτήρια των διαφόρων σεναρίων που εφαρμόσαμε χρησιμοποιήσαμε τη συνάρτηση `flickr.photos.search` του Flickr API. Η συγκεκριμένη συνάρτηση επιστρέφει μια λίστα φωτογραφιών οι οποίες πληρούν κάποια κριτήρια. Θα επιστραφούν μόνο φωτογραφίες στις οποίες ο συγκεκριμένος χρήστης που κάνει την αναζήτηση έχει δικαιώματα πρόσβασης. Παρακάτω θα περιγράψουμε κάποια από τα ορίσματα που μπορεί να δεχτεί η συνάρτηση. Ο αριθμός και το είδος των ορισμάτων ποικίλλει αναλόγως με το πόσο εξειδικευμένη αναζήτηση θέλει να κάνει ο συγκεκριμένος χρήστης. Τα ορίσματα που περιγράφουμε παρακάτω είναι αυτά τα οποία χρησιμοποιήθηκαν και στα διάφορα σενάρια που υλοποιήσαμε και των οποίων τα αποτελέσματα θα παρουσιάσουμε σε επόμενο κεφάλαιο. Μερικά χαρακτηριστικά του API είναι:

api_key: Το κλειδί της API εφαρμογής.

tags: Η λίστα των ετικετών. Ο διαχωρισμός μεταξύ τους γίνεται με κόμματα. Θα επιστραφούν φωτογραφίες που έχουν χαρακτηριστεί με τουλάχιστον μία από τις ετικέτες που ανήκουν στη συγκεκριμένη λίστα.

tag_mode: Το συγκεκριμένο όρισμα μπορεί να πάρει δύο τιμές. Η τιμή 'any' σημαίνει ότι θα επιστραφούν φωτογραφίες που θα περιέχουν τουλάχιστον μία από τις ετικέτες που περιλαμβάνονται στη λίστα. Αν το όρισμα πάρει την τιμή 'all' σημαίνει ότι θα επιστραφούν φωτογραφίες που θα έχουν χαρακτηριστεί με το σύνολο των ετικετών που ανήκουν στη λίστα.

page: Η σελίδα των αποτελεσμάτων που επιστρέφονται. Εάν το όρισμα αυτό παραλειφθεί τότε η προεπιλεγμένη τιμή είναι 1. - **per_page:** Ο αριθμός των

φωτογραφιών που επιστρέφονται ανά σελίδα. Εάν αυτό το όρισμα παραλειφθεί έχει ως προεπιλεγμένη το 100. Η μέγιστη επιτρεπόμενη τιμή είναι το 500.

lat: Μία έγκυρη τιμή για το γεωγραφικό πλάτος σε δεκαδική μορφή, για να υπάρχει η δυνατότητα να τεθούν ερωτήματα στη βάση δεδομένων με επιπλέον κριτήριο την ακτίνα από ένα συγκεκριμένο γεωγραφικό σημείο.

long: Μία έγκυρη τιμή για το γεωγραφικό μήκος σε δεκαδική μορφή.

radius: Μία έγκυρη τιμή για την ακτίνα μεγαλύτερο από το μηδέν και μικρότερο από 20 μίλια ή 32 χιλιόμετρα. Χρησιμοποιείται ως παράμετρος για ερωτήματα με κριτήριο τις γεωγραφικές συντεταγμένες ενός συγκεκριμένου σημείου. Η προεπιλεγμένη τιμή είναι 5 km.

2.3 API Methods

Ακολουθεί μια σύντομη περιγραφή μερικών από των μεθόδων που χρησιμοποιεί το Flickr API.

activity

- flickr.activity.userComments

Επιστρέφει μια λίστα με πρόσφατη δραστηριότητα στις φωτογραφίες που έχουν σχόλια από τον καλούντα χρήστη. Η μέθοδος αυτή καλείται κάθε μία ώρα.

people

- flickr.people.findByEmail
- flickr.people.findByUsername
- flickr.people.getInfo
- flickr.people.getPhotos

photos

- flickr.photos.addTags

Προσθήκη ετικέτας σε μια φωτογραφία

- flickr.photos.getContactsPhotos
- flickr.photos.getContactsPublicPhotos
- flickr.photos.getContext
- flickr.photos.getCounts

Εμφάνιση λίστας με τις πρόσφατες φωτογραφίες από τις επαφές των καλούντων χρηστών.

- flickr.photos.getFavorites

Επιστρέφει τη λίστα των ατόμων που έχουν επιλέξει ως αγαπημένη μια συγκεκριμένη φωτογραφία.

photos.comments

- flickr.photos.comments.addComment
- flickr.photos.comments.deleteComment
- flickr.photos.comments.editComment
- flickr.photos.comments.getList

Επιστρέφει τα σχόλια για μια φωτογραφία

photos.people

- flickr.photos.people.add
- flickr.photos.people.delete
- flickr.photos.people.getList

Επιστρέφει μια λίστα των ανθρώπων σε μια συγκεκριμένη φωτογραφία.

photosets

- flickr.photosets.addPhoto
- flickr.photosets.create
- flickr.photosets.delete
- flickr.photosets.editPhotos

photosets.comments

- flickr.places.getPlaceTypes
- flickr.places.getShapeHistory

Επιστρέφει το top 100 geotagged για μια μέρα.

tags

- flickr.tags.getClusterPhotos
- flickr.tags.getClusters

Επιστρέφει μια λίστα των σχετικών ετικετών για μια συγκεκριμένη ετικέτα.

- flickr.tags.getListPhoto

Επιστρέφει τη λίστα ετικετών για μια δεδομένη φωτογραφία.

Πιο συγκεκριμένα, υπάρχουν πάρα πολλές πληροφορίες που μπορούμε να εξάγουμε. Κάποιες από αυτές αναλύονται παρακάτω. Οι πληροφορίες δραστηριότητας έχουν να κάνουν με την πρόσφατη δραστηριότητα του χρήστη. Δηλαδή πρόσφατα σχόλια ή φωτογραφίες του χρήστη. Πληροφορίες για blog, αν κάποιος χρήστης έχει δικό του blog, και πληροφορίες γύρω από αυτό. Πληροφορίες για κάμερες, όπως μάρκες και μοντέλα που έχουν χρησιμοποιηθεί για την λήψη των φωτογραφιών. Πληροφορίες για τις επαφές ενός χρήστη, είτε ως λίστα, είτε ως ετικέτα σε φωτογραφία, είτε ως προτεινόμενες επαφές για έναν χρήστη. Πληροφορίες για «*favorites*», δηλαδή τις φωτογραφίες που ένας χρήστης έχει μαρκάρει ως αγαπημένες. Πληροφορίες για τις συλλογές φωτογραφιών του χρήστη, οι οποίες μπορεί να είναι κατά ημερομηνία ή θεματολογία. Πληροφορίες για ομάδες του **Flickr**, διάφορα groups και υποομάδες, ή ακόμα και τα θέματα συζητήσεων των ομάδων και τα μέλη τους. Πληροφορίες για ψηφοφορίες από συγκεκριμένες ομάδες. Πληροφορίες για την πιο προσφάτως ενδιαφέρουσα φωτογραφία ή για μια συγκεκριμένη ημερομηνία. Πληροφορίες για «*machinetags*». Τα machinetags έχουν να κάνουν με τις πληροφορίες για μια φωτογραφία, είτε η πληροφορία αυτή έχει να κάνει με γεωγραφική τοποθεσία είτε πρόκειται για απλή περιγραφή της φωτογραφίας. Πληροφορίες για μια ξεχωριστή ομάδα που ονομάζεται «**Flickr Pandas**», και έχει να κάνει με 3 υποομάδες που σχετίζονται με τα συγκεκριμένα ζώα. Στην συνέχεια έχουμε πληροφορίες για τον χρήστη με βάση πχ το email του, το username του, τις φωτογραφίες που έχει ανεβάσει ή τις ομάδες στις οποίες ανήκει, και τέτοιου τύπου πληροφορίες γύρω από ένα συγκεκριμένο λογαριασμό. Πληροφορίες για φωτογραφίες όπως την ετικέτα μιας φωτογραφίας, αν έχει μαρκαριστεί ως αγαπημένη, φωτογραφίες με γεωγραφική τοποθεσία, με πληροφορίες ημερομηνίας και ώρας, με συγκεκριμένο μέγεθος ή θέμα κλπ. Πληροφορίες για σχόλια φωτογραφιών, πχ τα πιο πρόσφατα σχόλια ενός συγκεκριμένου χρήστη. Πληροφορίες για γεωγραφικές τοποθεσίες είτε έχουν να κάνουν με συγκεκριμένη φωτογραφία, είτε με φωτογραφία που ανήκει σε συγκεκριμένο χρήστη. Πληροφορίες για συλλογές συγκεκριμένου θέματος που ανήκουν σε συγκεκριμένο χρήστη και τα σχόλια για κάθε συλλογή. Πληροφορίες για μέρη/τοποθεσίες με βάση ένα url ή συγκεκριμένη θεματολογία, ακόμα και ιστορικής σημασίας μέρη, είτε σχετίζονται με ένα συγκεκριμένο χρήστη ή την ετικέτα μιας φωτογραφίας. Πληροφορίες στατιστικού περιεχομένου όπως φωτογραφίες που ανέβηκαν σε μια συγκεκριμένη ημερομηνία ή πόσα views έχει φωτογραφία μια συγκεκριμένη ημερομηνία, ή ακόμα τις φωτογραφίες με τα περισσότερα σχόλια. Πληροφορίες για τις ετικέτες φωτογραφιών, την πιο δημοφιλή θεματολογία και φωτογραφίες σχετικές με την συγκεκριμένη θεματολογία ή παρόμοιες.

Το Flickr API είναι ένας ισχυρός τρόπος αλληλεπίδρασης με τους Flickr λογαριασμούς. Με το API, μπορούμε να διαβάσουμε σχεδόν όλα τα δεδομένα που σχετίζονται με τις εικόνες. Το πρώτο πράγμα που χρειαζόμαστε για να εξάγουμε τις πληροφορίες που θέλουμε είναι ένα κλειδί API (API key) . Με το κλειδί API το Yahoo παρακολουθεί την δραστηριότητα που σχετίζεται με κάθε κλειδί API. Πρόκειται ουσιαστικά για το όνομα χρήστη (username) για το Flickr API. Μπορούμε να προμηθευτούμε ένα τέτοιο κλειδί αρκετά εύκολα, ανάλογα με το είδος του έργου που θέλουμε να φτιάξουμε. Αν πρόκειται για διαφημιστικό προϊόν η Yahoo θα πρέπει να εγκρίνει το προϊόν. Αλλιώς, αν πρόκειται για πειραματικούς και μη εμπορικούς σκοπούς μπορούμε να πάρουμε ένα κλειδί API αμέσως. Δημιουργούμε ένα λογαριασμό και στη συνέχεια επιλέγουμε το είδος του κλειδιού που χρειαζόμαστε.

The App Garden

[Create an App](#) | [API Documentation](#) | [Feeds](#) | [What is the App Garden?](#)

First, we need to know whether or not your app is commercial.

Choose **Non-Commercial** if:

- Your app doesn't make money.
- Your app makes money, but you're a family-run, small, or independent business.
- You're developing a product which is not currently commercial, but might be in the future.
- You're building a personal website or blog where you are only using your own images.

[APPLY FOR A NON-COMMERCIAL KEY](#)

or

Choose **Commercial** if:

- You or your agency works for a major brand.

AND one of the following:

- You want to make a profit.
- You charge a fee for your product or services.
- You will bring Flickr content into your product and intend to sell those services.

[APPLY FOR A COMMERCIAL KEY](#)

Στην συνέχεια παραθέτουμε ένα υποθετικό όνομα για την εφαρμογή μας, το οποίο και θα είναι “Flickr-Shots”, και τον λόγο για τον οποίο χρειαζόμαστε το κλειδί, συγκεκριμένα ότι πρόκειται για την εκπόνηση μιας διπλωματικής εργασίας.

The App Garden
Create an App | API Documentation | Feeds | What is the App Garden?

Tell us about your app:

Owner: vonte_204
This app will be associated with your vonte_204 account. You will not be able to change this after you submit your application.

What's the name of your app? Flickr-Shots

What are you building?
(And trust us when we say you can't be detailed enough)
I study for my MSc in Informatics. The subject of my thesis is "Exportation Data from Social Media (Flickr)". I need a key, in order to start working with the API.

I acknowledge that Flickr members own all rights to their content, and that it's my responsibility to make sure that my project does not contravene those rights.

I agree to comply with the [Flickr API Terms of Use](#).

or Cancel

Και πατώντας Submit, μας εμφανίζει ότι χρειαζόμαστε, το key και το secret, ότι χρειαζόμαστε δηλαδή για να διαχειριστούμε τα δεδομένα.

The App Garden
Create an App | API Documentation | Feeds | What is the App Garden?

Done! Here's the API key and secret for your new app:

Flickr-Shots

Key:
518cb877586a7059a8dddf232b7049da

Secret:
03e1d637ac217a7b

[Edit app details](#) - [Edit auth flow for this app](#) - [View all Apps by You](#)

What to do next

If your key is to use in an application that someone else developed, for example to display your own or your group's photos on your website or blog, then you're all done!

If your key is for an app that you're developing, here are things to help you build and promote your app:

- **Build your app**
You're all set to build something cool with the API key and secret! Make sure you know what you can and can't do with the images and data on Flickr before you design your app. Save some time by reading these reference guides first:
 - [API Terms of Use](#)
 - [Community Guidelines](#)
 - [API Documentation](#)
 - [Flickr API Group](#)

Key : **518cb877586a7059a8dddf232b7049da**

Secret : **03e1d637ac217a7b**

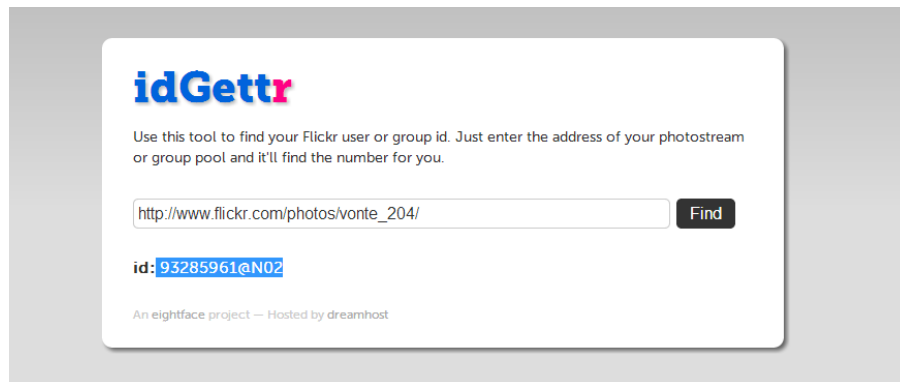
Μετά θα πρέπει να ακολουθήσουμε τρία βήματα.

- Θα πρέπει να στείλουμε στο Flickr το αίτημα των πληροφοριών που θέλουμε. Αυτό γίνεται με την δημιουργία μιας ειδικής διεύθυνσης URL.
- Μόλις το Flickr κατανοήσει το σωστό URL, θα στείλει άμεσα την πληροφορία που θέλουμε.
- Το τελευταίο βήμα είναι να χρησιμοποιήσουμε τα δεδομένα τα οποία το Flickr μας απάντησε.

Για να στείλουμε ένα αίτημα στο Flickr, πρέπει να χρησιμοποιήσουμε τις μεθόδους API, ώστε να δημιουργίσουμε το URL. Όλα τα URLs ξεκινούν με:

<http://api.flickr.com/services/rest/>

Στην συνέχεια προσθέτουμε τις πληροφορίες που χρειαζόμαστε ανάλογα με τα δεδομένα που ζητάμε. Δηλαδή στο τέλος της διεύθυνσης προσθέτουμε την μέθοδο που θέλουμε να χρησιμοποιήσουμε. Και μετά θα πρέπει να μπει το κλειδί και το user id. Για το user id θα πρέπει να μπούμε στο παρακάτω link και να δώσουμε το link του λογαριασμού μας στο flickr. <http://idgettr.com/>



id: 93285961 @N02

Για παράδειγμα με το URL

http://api.flickr.com/services/rest/?method=flickr.places.Avaζήτηση&api_key=10e004cb99ca367a6e63347f14e34a1d&query=paris&format=rest&auth_token=72157634138971163-167b24e97643a2ca&api_sig=3c98967d8ba62fe454119fd63afc1d92

Παίρνουμε πληροφορίες για φωτογραφίες που έχουν σχέση με την λέξη Paris, βέβαια έχουμε χρησιμοποιήσει και την ανάλογη μέθοδο flickr.places.search από το API. Τα αποτελέσματα εμφανίζονται σε xml μορφή, όπως φαίνεται στην παρακάτω εικόνα.

This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
<rsp stat="ok">
  <places query="paris" total="36">
    <place place_id="EsiQVIZXU9_KEA" woeid="615702" latitude="48.856" longitude="2.341" place_url="/France/Ile-de-France/Paris" place_type="locality"
    place_type_id="77" timezone="Europe/Paris" woe_name="Paris">Paris, Ile-de-France, France</place>
    <place place_id="Eis_151TVz2M8JK" woeid="2468279" latitude="33.661" longitude="-95.563" place_url="/United+States/Texas/Paris" place_type="locality"
    place_type_id="77" timezone="America/Chicago" woe_name="Paris">Paris, Texas, United States</place>
    <place place_id="UD665uhTVxz2kgp" woeid="2468280" latitude="36.301" longitude="-88.325" place_url="/United+States/Tennessee/Paris" place_type="locality"
    place_type_id="77" timezone="America/Chicago" woe_name="Paris">Paris, Tennessee, United States</place>
    <place place_id="J0KtUfV0U_2" woeid="4352" latitude="43.190" longitude="-80.381" place_url="/Canada/Ontario/Paris" place_type="locality" place_type_id="77"
    timezone="America/Toronto" woe_name="Paris">Paris, Ontario, Canada</place>
    <place place_id="2sRwIIVVryb.LGL" woeid="2468286" latitude="38.209" longitude="-84.252" place_url="/United+States/Kentucky/Paris" place_type="locality"
    place_type_id="77" timezone="America/Kentucky/Louisville" woe_name="Paris">Paris, Kentucky, United States</place>
    <place place_id="U0r08idTVxkQ3u" woeid="2468283" latitude="39.614" longitude="-87.695" place_url="/United+States/Illinois/Paris" place_type="locality"
    place_type_id="77" timezone="America/Chicago" woe_name="Paris">Paris, Illinois, United States</place>
    <place place_id="00zMSZIVxH0cNu" woeid="2468281" latitude="35.292" longitude="-93.729" place_url="/United+States/Arkansas/Paris" place_type="locality"
    place_type_id="77" timezone="America/Chicago" woe_name="Paris">Paris, Arkansas, United States</place>
    <place place_id="wvus071IVzqzH4I" woeid="2468277" latitude="39.004" longitude="-77.948" place_url="/United+States/Virginia/Paris" place_type="locality"
    place_type_id="77" timezone="America/New York" woe_name="Paris">Paris, Virginia, United States</place>
    <place place_id="WcWQAOVTVzrlx0B" woeid="2468282" latitude="39.481" longitude="-92.000" place_url="/United+States/Missouri/Paris" place_type="locality"
    place_type_id="77" timezone="America/Chicago" woe_name="Paris">Paris, Missouri, United States</place>
    <place place_id="VY4uAhTVzZ5xSn" woeid="2468264" latitude="44.261" longitude="-70.501" place_url="/United+States/Maine/Paris" place_type="locality"
    place_type_id="77" timezone="America/Chicago" woe_name="Paris">Paris, Maine, United States</place>
    <place place_id="ghXG99TVzYtltM" woeid="2468273" latitude="40.797" longitude="-81.163" place_url="/United+States/Ohio/Paris/in-Stark" place_type="locality"
    place_type_id="77" timezone="America/New York" woe_name="Paris">Paris, Ohio, United States</place>
    <place place_id="1tR44NcTVzTQYAE" woeid="2468284" latitude="42.225" longitude="-111.401" place_url="/United+States/Idaho/Paris" place_type="locality"
    place_type_id="77" timezone="America/Boise" woe_name="Paris">Paris, Idaho, United States</place>
    <place place_id="5yXG99TVzXh0cCF" woeid="2468267" latitude="34.178" longitude="-89.455" place_url="/United+States/Mississippi/Paris" place_type="locality"
    place_type_id="77" timezone="America/Chicago" woe_name="Paris">Paris, Mississippi, United States</place>
    <place place_id="UH01FFTVzYtltM" woeid="2468266" latitude="43.776" longitude="-85.502" place_url="/United+States/Michigan/Paris/in-Mecosta"
    place_type="locality" place_type_id="77" timezone="America/Detroit" woe_name="Paris">Paris, Michigan, United States</place>
    <place place_id="sDv0hRIVwQFB2" woeid="2468285" latitude="34.893" longitude="-82.362" place_url="/United+States/SouthCarolina/Paris" place_type="locality"
    place_type_id="77" timezone="America/New York" woe_name="Paris">Paris, South Carolina, United States</place>
    <place place_id="BXDJkz2Uzpy.vNhtg" woeid="20099390" latitude="56.514" longitude="8.489" place_url="/Danmark/Midtyjylland/Paris" place_type="locality"
    place_type_id="77" timezone="Europe/Copenhagen" woe_name="Paris">Paris, Midtyjylland, Danmark</place>
```

Με τον ίδιο τρόπο μπορούμε να κάνουμε άπειρα queries (ερωτήματα), χρησιμοποιώντας πάντα την κατάλληλη μέθοδο και το κατάλληλο URL.

Κεφάλαιο 3

3.1 Βάσεις Δεδομένων

Μία Βάση Δεδομένων (Database – DB) είναι μία οργανωμένη συλλογή δεδομένων. Τα δεδομένα συνήθως οργανώνονται ούτως ώστε να μοντελοποιήσουν πλευρές της πραγματικότητας, με τέτοιον τρόπο που να υποστηρίζει διεργασίες, οι οποίες απαιτούν πληροφορία. Οι βάσεις δεδομένων ορίζονται ως οργανωμένες συλλογές δεδομένων. Παρόλο που με τον όρο “Βάσεις Δεδομένων”, αναφερόμαστε σε ολόκληρο το σύστημα των Βάσεων Δεδομένων, στην πραγματικότητα η αναφορά έχει να κάνει καθαρά με δεδομένα και τις συλλογές αυτών. Το σύστημα που χειρίζεται τα δεδομένα, οι συναλλαγές, τα προβλήματα ή οποιαδήποτε άλλη πτυχή της βάσης δεδομένων είναι το Σύστημα Διαχείρισης Βάσεων Δεδομένων (ΣΔΒΔ).

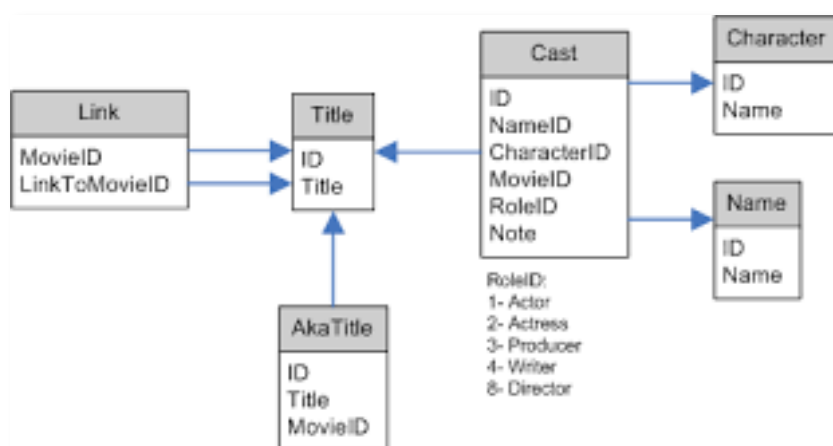
Πολλές εφαρμογές βασίστηκαν στη χρήση των συνδεδεμένων λιστών-πινάκων για τη δημιουργία σχέσεων μεταξύ των δεδομένων και την εύρεση συγκεκριμένων στοιχείων. Τα μοντέλα δεν ήταν τυποποιημένα και απαιτείται εκτεταμένη εκπαίδευση, προκειμένου να καταστεί δυνατή η αποτελεσματική χρήση τους. Αυτά τα μοντέλα και άλλα σημαντικά είδη εξηγούνται παρακάτω.

3.2 Σύστημα Διαχείρισης Βάσεων Δεδομένων (ΣΔΒΔ)

Database management system (DBMS) είναι μια συλλογή από προγράμματα που επιτρέπουν στους χρήστες να δημιουργήσουν και να συντηρήσουν μία βάση δεδομένων. Επομένως, το ΣΔΒΔ είναι ένα γενικής χρήσης σύστημα λογισμικού που διευκολύνει τις διαδικασίες ορισμού, κατασκευής και χειρισμού βάσεων δεδομένων για διάφορες εφαρμογές.

3.3 Σχεσιακή Βάση Δεδομένων

Με τον όρο σχεσιακή βάση δεδομένων εννοείται μία συλλογή δεδομένων οργανωμένη σε συσχετισμένους πίνακες που παρέχει ταυτόχρονα ένα μηχανισμό για ανάγνωση, εγγραφή, τροποποίηση ή και πιο πολύπλοκες διαδικασίες πάνω στα δεδομένα. Οι σχεσιακές βάσεις δεδομένων διαχώρισαν τα δεδομένα από τις εφαρμογές που είχαν πρόσβαση σ' αυτά, επιτρέποντας έτσι τη μεταχείριση της πληροφορίας μέσω της χρήσης μίας γλώσσας ερωτημάτων (query language), όπου η επιλογή συγκεκριμένων δεδομένων μπορούσε να επιτευχθεί αποδοτικά μέσω της δημιουργίας δηλώσεων (statements) που περιείχαν λογικούς τελεστές. Ο σκοπός μιας βάσης δεδομένων είναι η οργανωμένη αποθήκευση πληροφορίας και η δυνατότητα εξαγωγής της πληροφορίας αυτής, ιδίως σε πιο οργανωμένη μορφή, σύμφωνα με ερωτήματα που τίθενται στη σχεσιακή βάση δεδομένων. Το σχεσιακό μοντέλο επέτρεπε την ελεύθερη είσοδο, διαγραφή και τροποποίηση δεδομένων στους πίνακες (tables – δηλαδή σύνολα δεδομένων με τον ίδιο τύπο και με σταθερό μέγεθος). Το μοντέλο αυτό επέτρεπε, επίσης, στα περιεχόμενα της βάσης να εξελιχθούν χωρίς συνεχές επαναγράψιμο των συνδέσμων και των δεικτών. Το σχεσιακό μέρος προέρχεται από το γεγονός ότι οι οντότητες αναφέρονταν σε άλλες οντότητες σε αυτό που αποκαλείται σχέση ενός-προς-πολλούς, όπως τα παραδοσιακά ιεραρχικά μοντέλα, καθώς και σχέσεις πολλών-προς-πολλούς, όπως ένα πλοηγούμενο δικτυακό μοντέλο. Τα δεδομένα είναι δυνατόν να αναδιοργανώνονται με πολλούς διαφορετικούς τρόπους, σε νοητούς πίνακες, χωρίς να είναι απαραίτητη η αναδιοργάνωση των φυσικών πινάκων που τα αποθηκεύουν. Υστερούν μόνο σε ταχύτητα υπολογισμών και σε χώρο αποθήκευσης, αλλά μόνο όταν έχουμε να κάνουμε με πολύ μεγάλες βάσεις δεδομένων, κάτι που θα πρέπει να μας απασχολεί γιατί η συγκεκριμένη εργασία έχει να κάνει με αρκετά μεγάλο όγκο δεδομένων.



Πηγή: Laurent Dupuis 's Home
<http://www.dupuis.me/node/25>

3.4 ACID

Στις Βάσεις Δεδομένων το **ACID** (*ατομικότητα, συνέπεια, απομόνωση, μονιμότητα*) είναι ένα σύνολο ιδιοτήτων το οποίο εγγυάται ότι οι συναλλαγές στην βάση δεδομένων (ΒΔ) λειτουργούν αξιόπιστα. Η ιδέα του ACID είναι να αξιολογεί την δομή των βάσεων δεδομένων και των εφαρμογών. Όσον αφορά τις βάσεις δεδομένων μια λογική ενέργεια που έχει να κάνει με τα δεδομένα της αποκαλείται συναλλαγή. Στην ουσία έχει να κάνει με την ασφάλεια των δεδομένων και κυρίως αναφέρεται σε συναλλαγές και δοσοληψίες. Για παράδειγμα μια μεταφορά κεφαλαίου από ένα λογαριασμό μιας τράπεζας σε έναν άλλο, αν και μπορεί να συμπεριλαμβάνει πολλές ενέργειες που αφορούν την κίνηση του λογαριασμού, είναι μια συναλλαγή.

Χαρακτηριστικά:

Ατομικότητα – Atomicity: Καθιστά τις συνδιαλλαγές ως λειτουργίες που ακολουθούν το όλα-ή-τίποτα σε περίπτωση αποτυχιών. Αυτό σημαίνει, ότι το αποτέλεσμα μία συνδιαλλαγής θα ήταν το «όλα» αν επιτύχει ή το «τίποτα» αν αποτύχει (αν απορριφθεί ή επιστρέψει σε προηγούμενη κατάσταση).

Συνέπεια – Consistency: Παρέχεται από την εφαρμογή. Ο κώδικας της εφαρμογής σε μία συνδιαλλαγή θα πρέπει να εγγυάται πως αν η βάση δεδομένων έχει μια συνεπή κατάσταση, τότε η νέα κατάσταση της βάσης δεδομένων (μετά τη συνδιαλλαγή) θα είναι επίσης συνεπής-σταθερή.

Απομόνωση – Isolation: Παρέχει ατομικότητα συγχρονισμού. Παρέχει δηλαδή την αίσθηση ότι ο χρήστης εκτελεί τη συνδιαλλαγή μόνος του στο σύστημα, ακόμη κι αν εκτελούνται πολλές συνδιαλλαγές ταυτόχρονα.

Ανθεκτικότητα – Durability: Εγγυάται πως οι ενημερώσεις μίας συνδιαλλαγής δεν χάνονται ακόμα και σε περίπτωση σφαλμάτων και αποτυχιών. Οι συνδιαλλαγές είναι ένα πολύ σημαντικό ζήτημα αφαιρετικότητας (abstraction) στο πρόγραμμα καθώς καταργούν δύο πολύ σημαντικά προβλήματα για τον προγραμματισμό εφαρμογών. Το πρώτο είναι η αντιμετώπιση του ταυτοχρονισμού. Οι χρήστες δεν χρειάζεται να νοιάζονται για τον έλεγχό του, όταν προγραμματίζουν εφαρμογές με συνδιαλλαγές για να έχουν πρόσβαση σε διαμοιραζόμενα δεδομένα. Τα πρωτόκολλα που υλοποιούν την ιδιότητα της απομόνωσης θα χειριστούν τις ταυτόχρονες προσβάσεις. Το δεύτερο είναι ότι οι εφαρμογές δε χρειάζεται να αντιμετωπίσουν τις αποτυχίες. Τα πρωτόκολλα ατομικότητας και ανθεκτικότητας παρέχουν αυτόματη επαναφορά σε περίπτωση αποτυχίας. Έτσι οι ιδιότητες ACID, τις οποίες προσφέρουν οι συνδιαλλαγές, απλοποιούν σε σημαντικό βαθμό τη δουλειά των προγραμματιστών.

Πλεονεκτήματα

- Ωριμη τεχνολογία που υποστηρίζει χιλιάδες εφαρμογές σε λειτουργία σήμερα
- Υψηλές αποδόσεις

- Πληθώρα εργαλείων
- Εκπαιδευμένο προσωπικό

Μειονεκτήματα

- Κόστος
- Κλιμάκωση
- Δύσκολη τροποποίηση της βάσης



Πηγή: The Yin-Yang Challenge: Balancing ACID and NoSQL
<https://www.bizcloudnetwork.com/acid-and-nosql/>

3.5 Μη Σχεσιακές Βάσεις Δεδομένων - Solr

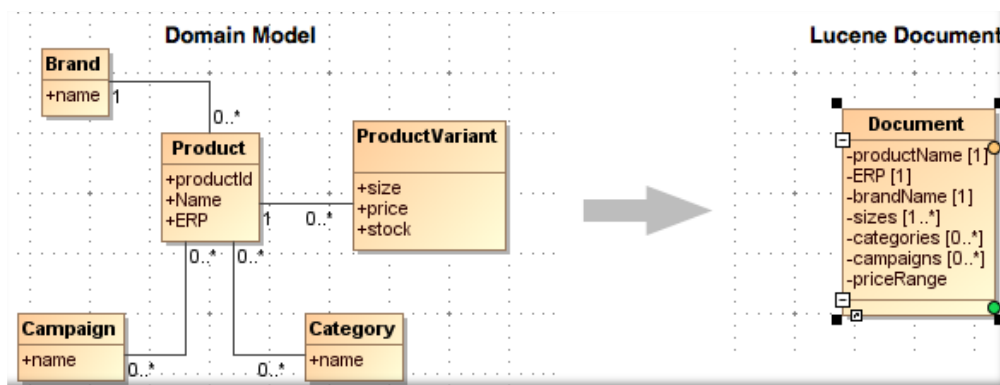
Lucene

Η Lucene/Solr είναι μια open source βιβλιοθήκη υλοποιημένη στην γλώσσα προγραμματισμού Java. Είναι η πιο δημοφιλής και ευρέως χρησιμοποιούμενη μηχανή αναζήτησης. Αποτελεί ένα cross-platform λογισμικό και παρέχει έναν απλό αλλά ισχυρό πυρήνα API που απαιτεί την ελάχιστη κατανόηση των τεχνικών αναζήτησης. Δημιουργήθηκε από τον Doug Cutting και πρόκειται για την αρχική βιβλιοθήκη αναζήτησης και ευρετηρίασης. Μετατρέπει σε index το περιεχόμενο (κατατάσσει σε πίνακα) και παραδίδει το αποτέλεσμα της αναζήτησης γρηγορότερα. Έχει όλες τις δυνατότητες μιας NoSQL βάσης δεδομένων. Η βιβλιοθήκη αναζήτησης Lucene βασίζεται σε έναν ανεστραμμένο ευρετήριο και έχει τρεις πρωταρχικούς στόχους:

- Να διατηρήσει την υπάρχουσα συμπεριφορά από την Java στην C# και να αυτοματοποιήσει πλήρως την διαδικασία, έτσι ώστε το έργο να μπορεί εύκολα να συγχρονιστεί με το χρονοδιάγραμμα απελευθέρωσης της Java Lucene.
- Να διατηρήσει τις υψηλές απαιτήσεις απόδοσης που αναμένονται από την πρώτη κλάση της C# για αναζήτηση στην βιβλιοθήκη
- Να μεγιστοποιήσει τη χρηστικότητα και την δύναμη όταν χρησιμοποιείται στο εσωτερικό της .NET runtime.

Τα χαρακτηριστικά της είναι:

- Δεν χρησιμοποιεί σχήμα.
- Δεν χρησιμοποιεί την SQL ως γλώσσα ερωτημάτων
- Δεν πληρεί τις προϋποθέσεις σύμφωνα με ACID
- Αποθηκεύει αδόμητα δεδομένα
- Ικανότητα να αποθηκεύει και να ανακτά πληροφορία γρηγορότερα
- Δεν υπάρχει συσχέτιση μεταξύ εγγραφών (documents)



Πηγή: Non-Standard Ways of using Lucene

<http://blog.mgm-tp.com/2011/03/non-standard-ways-of-using-lucene/>

3.6 Μη Σχεσιακές Βάσεις Δεδομένων - NoSQL

Η γενιά μετά-σχεσιακών βάσεων δεδομένων τη δεκαετία του 2000 ονομάστηκε NoSQL βάσεις δεδομένων και περιέχει γρήγορες αποθηκεύσεις κλειδιού-τιμής και αρχειοστρεφείς βάσεις. Οι NoSQL βάσεις δεδομένων είναι συχνά πολύ γρήγορες, δεν απαιτούν προκαθορισμένα σχήματα πινάκων, αποφεύγουν λειτουργίες συνδυασμού αποθηκεύοντας από-κανονικοποιημένα δεδομένα και είναι σχεδιασμένες για οριζόντια επεκτασιμότητα (δηλαδή την προσθήκη 23 περισσότερων κόμβων σε ένα σύστημα, όπως την προσθήκη ενός νέου υπολογιστή σε μία κατανεμημένη εφαρμογή λογισμικού). Οι πιο διάσημες βάσεις NoSQL είναι η MongoDB1, η Couchbase2, η Riak, η memcached, η Apache Cassandra και η HBase. Το χαρακτηριστικό που διαχωρίζει τις NoSQL βάσεις δεδομένων είναι η ουσιαστική απόρριψη της «σχεσιακής δόμησης των δεδομένων» που ήταν ενσωματωμένη στα σχεσιακά συστήματα διαχείρισης βάσεων δεδομένων (RDBMS). Η ώθηση των εταιρειών να στραφούν στις μη-σχεσιακές βάσεις είναι η πρόσφατη απότομη αύξηση του όγκου των συναλλαγών που πρέπει να καταγραφούν, καθώς υπάρχει μεγάλος όγκος online κίνησης. Αυτό σε συνδυασμό με την έλευση φθηνού online αποθηκευτικού χώρου έχει καταστήσει τις NoSQL βάσεις δεδομένων αρκετά δημοφιλείς. Παρέχουν πιο εύκολες ad-hoc αλλαγές και το δυναμισμό που χρειάζονται τα παραπάνω συστήματα, σε πολύ μεγαλύτερο βαθμό απ' ό,τι οι σχεσιακές βάσεις δεδομένων. Η δημιουργία μιας σχεσιακής βάσης δεδομένων περιλαμβάνει έρευνα και εξέταση του ποια δεδομένα απαιτείται να καταγραφούν, ούτως ώστε να δημιουργηθεί το «σχήμα» της βάσης δεδομένων. Αντίθετα, μία βάση δεδομένων NoSQL επιτρέπει την αποθήκευση κάθε είδους δεδομένων, ακόμη και αυτών που δεν μπορεί να προβλεφθούν κατά τη δημιουργία της βάσης. Η NoSQL βάση δεδομένων είναι εξαιρετικά επεκτάσιμη, αξιόπιστη και έχει ένα απλό μοντέλο δεδομένων. Δεν διαθέτει μηχανισμό ελέγχου όσον αφορά την συνέπεια και την ακεραιότητα των δεδομένων. Ένα από τα πιο σημαντικά πλεονεκτήματα της NoSQL είναι ότι μπορεί να χειριστεί αδόμητα δεδομένα. Αδόμητα δεδομένα μπορεί να είναι έγγραφα του Word, e-mail, ήχος, βίντεο ή ακόμα δεδομένα από κοινωνικά δίκτυα. Η NoSQL έχει καλύτερη απόδοση σε μεγάλες ποσότητες δεδομένων, το οποίο είναι ζωτικής σημασίας για τις επιχειρήσεις. Κάτι που μας ενδιαφέρει ιδιαίτερα καθώς η συγκεκριμένη εργασία έχει να κάνει με μεγάλο όγκο δεδομένων. Για να επιτρέψει την ταχύτερη απόδοση, η NoSQL συνήθως δεν τηρεί τις ACID προδιαγραφές (ατομικότητα, συνεκτικότητα, απομόνωση, διάρκεια ζωής), τους περιορισμούς που χρησιμοποιούνται σε σχεσιακές βάσεις δεδομένων. Ένα παράδειγμα των επιδόσεων της NoSQL είναι το Facebook, με την εφαρμογή (Κασσάνδρα), που είναι σε θέση να χειρίζεται πάνω από 100 εκατομμύρια χρήστες συνεχώς. Η NoSQL έχει μειονεκτήματα και περιορισμούς. Ένα κοινό πρόβλημα με αυτές τις βάσεις δεδομένων είναι η έλλειψη κρυπτογράφησης για τα αρχεία δεδομένων και η αδυναμία ταυτοποίησης μεταξύ του χρήστη και τους

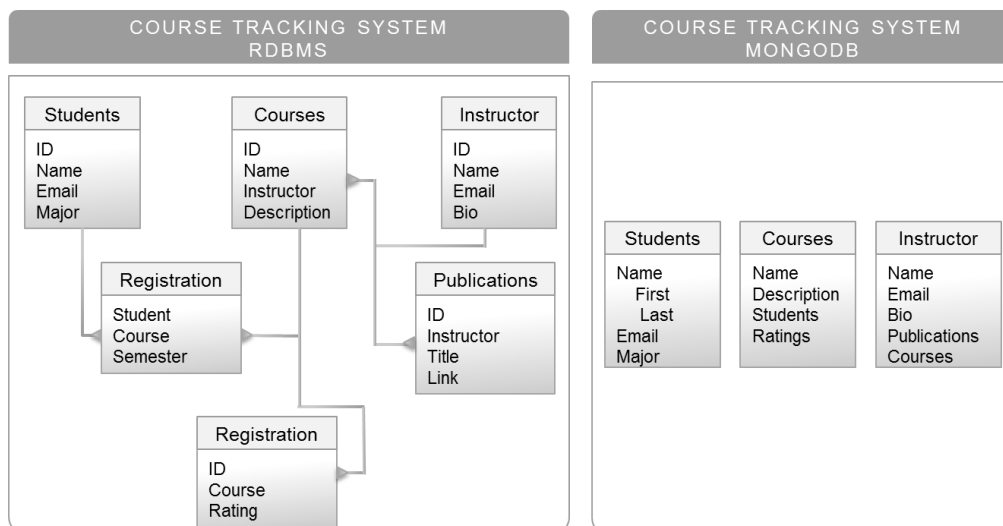
διακομιστές (server). Τέτοιου είδους προβλήματα είναι δυστυχώς ακόμα προς επίλυση.

Πλεονεκτήματα

μη σχεσιακά models
απλουστευμένα query API's
δυναμικό partitioning
αυτόματο replication
load balancing

Μειονεκτήματα

όχι σχεσιακό μοντέλο
όχι SQL
όχι ισχυρή συνέπεια
όχι κεντρικοποίηση



Πηγή: Fusion Power for MongoDB

<http://www.fusionio.com/white-papers/fusion-power-for-mongodb>

3.6.1 MongoDB

Η MongoDB είναι μία βάση δεδομένων εγγράφων που παρέχει μεγάλες επιδόσεις, διαθεσιμότητα και εύκολη επεκτασιμότητα. Παρότι η MongoDB υποστηρίζει λειτουργίες μοναδικού-στιγμιότυπου (singleinstance operations), η εγκατάστασή της είναι συνήθως κατανεμημένη. Τα σύνολα αντιγράφων (replica sets) παρέχουν υψηλών επιδόσεων αντιγραφή με αυτόματο μηχανισμό διόρθωσης σε περίπτωση σφάλματος, ενώ τα διαμοιρασμένα συμπλέγματα καθιστούν δυνατό τον καταμερισμό σε μεγάλα σύνολα δεδομένων πάνω σε πολλά μηχανήματα. Το λογισμικό NoSQL της MongoDB έχει αποκτήσει σημαντικότατο ρόλο στα big data. Ένας από τους λόγους είναι η οριζόντια κλίμακας κατασκευής της και ο παραλληλισμός της. Η MongoDB ανήκει στην οικογένεια των document-oriented NoSQL βάσεων δεδομένων. Πρόκειται για μία βάση ανοιχτού κώδικα, σχεδιασμένη να προσφέρει υψηλή διαθεσιμότητα στις εφαρμογές, επεκτασιμότητα, υψηλή απόδοση και δυνατότητα υποβολής σύνθετων ερωτημάτων. Είναι γραμμένη στην γλώσσα προγραμματισμού C++.

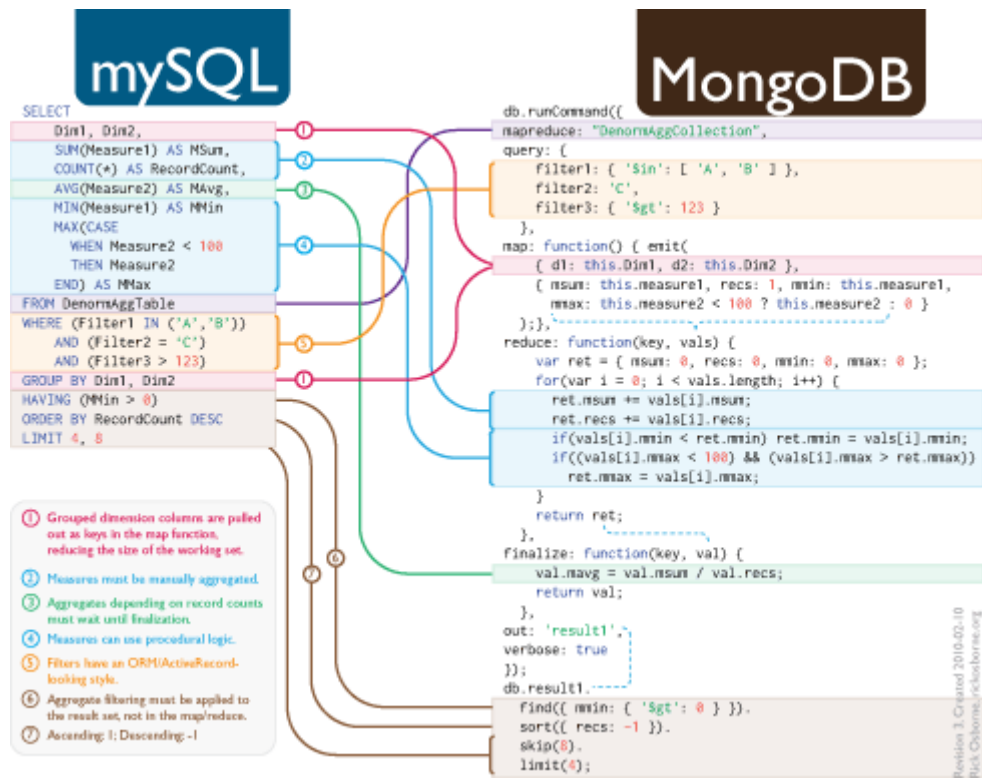
Στην MongoDB τα δεδομένα αποθηκεύονται με schema-free τρόπο σε αρχεία τύπου BSON (Binary JSON), τα οποία ονομάζονται documents. Κάθε document αντιπροσωπεύει μία εγγραφή που περιέχει πεδία σε μορφή key/value pairs. Στο key αποθηκεύεται σε ένα string το όνομα του πεδίου, και στο value μπορούν να αποθηκευτούν τύποι δεδομένων που υποστηρίζουν JSON αρχεία. Τα documents έχουν μέγιστο μέγεθος 16MB, ενώ για αποθήκευση μεγαλύτερων δεδομένων, όπως εικόνες, χρησιμοποιείται το GridFS. Δίνεται ακόμα η δυνατότητα της embedded, ενσωματωμένης αποθήκευσης Objects(αντικειμένων) και Arrays(πινάκων) στο εσωτερικό των documents κάτι που μειώνει την ανάγκη για join από την πλευρά του χρήστη. Ένα σύνολο από collections συνιστά μία βάση δεδομένων. Η MongoDB ως σύστημα διαχείρισης διατηρεί ένα σύνολο από βάσεις δεδομένων.

Η εγκατάσταση της MongoDB φιλοξενεί έναν αριθμό βάσεων δεδομένων. Κάθε βάση δεδομένων από αυτές περιέχει ένα σύνολο από συλλογές (collections). Κάθε συλλογή περιέχει ένα σύνολο από έγγραφα. Η MongoDB αποθηκεύει τα έγγραφα στο δίσκο σε μορφή σειριοποίησης BSON. Το BSON είναι μία δυαδική αναπαράσταση των JSON εγγράφων, αλλά περιέχει πολλούς περισσότερους τύπους δεδομένων από το JSON. Το κέλυφος JavaScript mongo και οι οδηγοί γλώσσας MongoDB (MongoDB language drivers) κάνουν τη μετάφραση μεταξύ BSON και της αναπαράστασης των εγγράφων με βάση τη γλώσσα προγραμματισμού. Όσον αφορά στο πεδίο _id, εξ' ορισμού η MongoDB δημιουργεί ένα μοναδικό ευρετήριο γι' αυτό κατά τη δημιουργία της συλλογής. Είναι πάντα το πρώτο πεδίο στα έγγραφα και μπορεί να περιέχει κάθε τύπο BSON πλην των πινάκων. Για τα ερωτήματα (queries) η MongoDB παρέχει ένα σύνολο από τελεστές για να καθορίσουν πως η μέθοδος find() επιλέγει τα δεδομένα από μία συλλογή με βάση ένα έγγραφο προσδιορισμού

ερωτημάτων που χρησιμοποιεί ένα συνδυασμό ακριβών ταιριασμάτων ισότητας και συνθηκών, με τη χρήση ενός τελεστή ερωτημάτων. Ένα ερώτημα στοχεύει σε μία συγκεκριμένη συλλογή εγγράφων. Τα ερωτήματα καθορίζουν τα κριτήρια ή τις συνθήκες, τα οποία ή οι οποίες ορίζουν τα έγγραφα που η MongoDB θα επιστρέψει στο χρήστη. Μπορούμε να μεταβάλλουμε προαιρετικά τα ερωτήματα, ούτως ώστε να επιβάλλουμε όρια, παραβλέψεις και σειρά ταξινόμησης. Η MongoDB ως βάση δεδομένων NoSQL φέρει πρακτικά όλα τα πλεονεκτήματα αυτού του είδους βάσεων, και πιο συγκεκριμένα τα πλεονεκτήματα των βάσεων με έγγραφα

Χαρακτηριστικά της MongoDB είναι:

- Ερωτήματα ad hoc. Η MongoDB υποστηρίζει την αναζήτηση πεδίου, ερωτήματα εύρους (range queries), αλλά και αναζητήσεις κανονικών εκφράσεων (regular expressions).
- Ευρετήρια. Οποιοδήποτε πεδίο σε ένα έγγραφο MongoDB μπορεί να δεικτοδοτηθεί. Υπάρχει επίσης και δυνατότητα για δευτερεύοντα ευρετήρια.
- Αντίγραφα. Η MongoDB παρέχει μεγάλη διαθεσιμότητα μέσω των συνόλων αντιγράφων της.
- Εξισορρόπηση Φορτίου. Η MongoDB υποστηρίζει οριζόντια επεκτασιμότητα (horizontal scalability), χρησιμοποιώντας θρυμματισμό (sharding). Ο χρήστης 52 επιλέγει ένα κλειδί θρύμματος (shard key) που καθορίζει τον τρόπο με τον οποίο τα δεδομένα θα κατανεμηθούν μέσα σε μία συλλογή. Τα δεδομένα χωρίζονται σε εύρη (βασίζόμενα σε αυτό το κλειδί) και κατανέμονται κατά μήκος πολλών θρυμμάτων. Η MongoDB μπορεί να τρέξει σε πολλούς εξυπηρετητές, εξισορροπώντας το φορτίο και δημιουργώντας αντίγραφα των δεδομένων για τη διατήρηση του συστήματος σε ενεργή κατάσταση σε περίπτωση βλάβης.
- Αποθήκευση Αρχείων. Η MongoDB μπορεί να χρησιμοποιηθεί και ως ένα σύστημα αρχείων, εκμεταλλευόμενη την εξισορρόπηση φορτίου και τα αντίγραφα δεδομένων της.
- Συσσωμάτωση. Η MapReduce μπορεί να χρησιμοποιηθεί για την επεξεργασία παρτίδων των δεδομένων και λειτουργίες συσσωμάτωσης.
- Συλλογές με ανώτατο όριο. Η MongoDB υποστηρίζει και συλλογές σταθερού μεγέθους (fixed-size collections). Αυτού του είδους οι συλλογές λειτουργούν ως μία κυκλική ουρά.

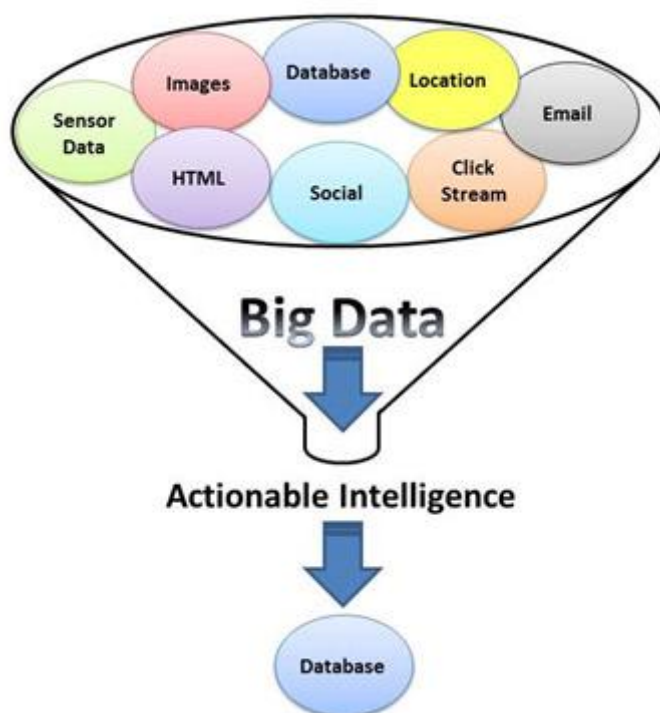


Πηγή: Bi Cortex

<http://bicortex.com/introduction-to-mongodb-nosql-database-for-sql-developers-part-3/>

3.7 Μεγάλα Δεδομένα (Big Data)

Η απότομη αύξηση της χρήσης του διαδικτύου, σε συνδυασμό με τη χρήση εφαρμογών για κινητά και κοινωνικών εφαρμογών, καθώς και οι επικοινωνίες μηχανήμα-προς-μηχάνημα (machine-to-machine communications) οδηγούν στην επανάσταση των «μεγάλων δεδομένων». Η ερευνητική εταιρεία IDC εκτιμά πως το 2013 το συνολικό παγκόσμιο μέγεθος των ψηφιακών δεδομένων ήταν 4.4 zettabytes, δηλαδή 4.4 τρισεκατομμύρια gigabytes. Αυτό το μέγεθος αναμένεται να αυξηθεί σε 44 zettabytes μέχρι το 2020.



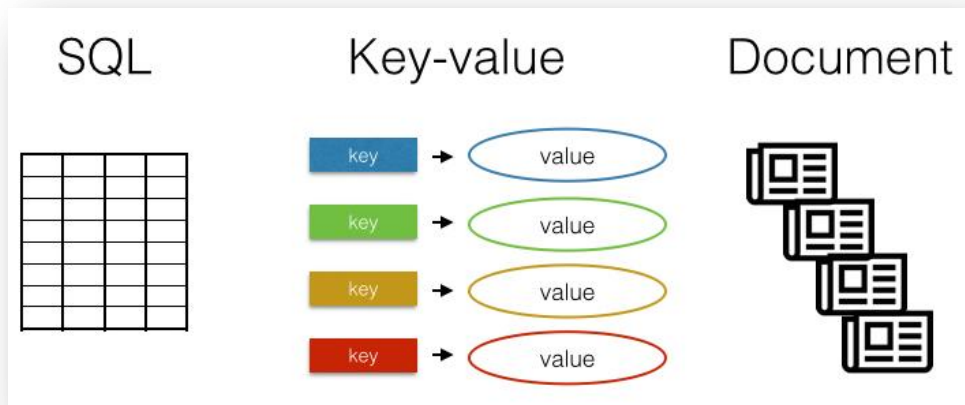
Πηγή: Big Data

<http://www.scaledb.com/big-data.php>

Γιατί MongoDB ?

Επιλέχθηκε για την αποθήκευση η NoSQL βάση δεδομένων, MongoDB ως μία cross-platform document-oriented βάση δεδομένων, η οποία επιτρέπει την ευέλικτη και χωρίς περιορισμούς εξαγωγή και αναζήτηση αρχείων, τα οποία διατηρούν τα δεδομένα τους σε key/value pairs. Ένας ακόμα λόγος που επιλέχθηκε η MongoDB είναι το άμεσο scalability που προσφέρει με προσθήκη κόμβων στο MongoDB cluster, κάτι που σε σχεσιακές βάσεις δεδομένων αποτελεί πολύπλοκη διαδικασία και απαιτεί μεγάλη μεταφορά δεδομένων στο δίκτυο. Η αναζήτηση εκτελείται με (ερωτήματα) MongoDB Queries, προς πλήρη εκμετάλλευση των βελτιστοποιημένων τεχνικών που έχουν αναπτυχθεί ειδικά για το σκοπό αυτό. Στόχος είναι η μελέτη των σύγχρονων τεχνικών εξαγωγής, αποθήκευσης και αναζήτησης μεγάλου όγκου δεδομένων, και της νέας προσέγγισης που προτείνεται από την διπλωματική αυτή, μέσω της MongoDB, μακριά από τις μέχρι τώρα μεθόδους και τεχνικές όπως είναι οι σχεσιακές βάσεις δεδομένων, προς ανάδειξη των καταλληλότερων τεχνικών για κάθε είδους εφαρμογή που υλοποιεί αναζήτηση μεγάλου όγκου δεδομένων με βάση την γεωγραφική τους θέση. Άλλες περιπτώσεις βάσεων, είναι η Lucene, η οποία αν και δεν είναι σχεσιακή βάση αλλά είναι μια Solr βάση, η οποία συγκεκριμένα διαθέτει δυναμική αναζήτηση, πάλι δεν προτείνεται σε περιπτώσεις όπως το Flickr. Ο λόγος είναι γιατί η Lucene είναι κατάλληλη σε περιπτώσεις εξαγωγής εγγράφων και μόνο και όχι φωτογραφιών. Η χρήση της Lucene αντί της MongoDB θα απαιτούσε διπλάσιο χρόνο για να εξάγουμε τα δεδομένα μας. Συνεπώς θα ήταν σοφότερο για τους σκοπούς της εργασίας να χρησιμοποιήσουμε μια NoSQL βάση. Ο βασικότερος λόγος είναι γιατί όπως είδαμε, μπορεί να χειριστεί αδόμητα δεδομένα από κοινωνικά δίκτυα, προτιμάται σε περιπτώσεις μεγάλου όγκου δομένων όπως στην περίπτωση μας, και φυσικά γιατί ενδείκνυται σε API queries. . Απόδειξη της χρησιμότητας και της αποτελεσματικότητας της MongoDB, είναι ότι χρησιμοποιείται από μεγάλα sites, όπως το **eBay**, η ασφαλιστική εταιρία **MetLife**, με 24 terabytes δεδομένων, το **Shutterfly**, με 6.000.000.000 φωτογραφίες, και το **Aadhar**, το οποίο διαθέτει την μεγαλύτερη βάση δεδομένων βιομετρικών στοιχείων. Αυτά είναι μερικά τρανταχτά παραδείγματα χρήσης της MongoDB από μεγάλες εταιρίες με τεράστιο όγκο δεδομένων.

SQL - MongoDB - Lucene



Πηγή: rOpenSci

<https://ropensci.org/blog/2015/05/20/database-interfaces/>

Κεφάλαιο 4

4.1 Αρχιτεκτονική Εφαρμογής

Η εφαρμογή αποτελείται από δυο κομμάτια

A) Το κομμάτι του διαχειριστή, στο οποίο δίνουμε μια **λέξη κλειδί** και κατεβαίνουν δεδομένα από το flickr, που περιέχουν αυτή τη λέξη κλειδί (tag). Τα δεδομένα, αποθηκεύονται στη βάση (MongoDB) με την εξής κωδικοποίηση:

- title (τίτλος φωτογραφίας),
- description (περιγραφή φωτογραφίας),
- latitude, longitude (δίνουμε περιορισμό να κατέβουν μόνο δεδομένα για φωτογραφίες που έχουν γεωγραφικές συντεταγμένες),
- id (ο κωδικός της φωτογραφίας στο flickr),
- imageUrl (για να μπορούμε να δούμε την φωτογραφία σε έναν browser, ή να την ανακαλέσουμε από μία εφαρμογή).
- tags (λέξεις κλειδιά που συνοδεύουν την φωτογραφία).

Η εφαρμογή αυτή μπορεί να τρέξει πολλές φορές από τον διαχειριστή και να εμπλουτίσει την βάση δεδομένων με στοιχεία από διάφορες λέξεις κλειδιά (university, museum κλπ), τις οποίες μπορεί στο β' κομμάτι ο χρήστης να χρησιμοποιήσει

B) Το κομμάτι του χρήστη, στο οποίο ο χρήστης θα δίνει μία πόλη από όλο τον κόσμο (διαλέγοντας από λίστα που θα του προσφέρεται), μία λέξη κλειδί (πχ university) και μια απόσταση και η εφαρμογή θα του εμφανίζει, αφού ψάξει στη βάση δεδομένων, φωτογραφίες που υπάρχουν στη βάση, σε έναν κύκλο με κέντρο την πόλη που ζήτησε και ακτίνα την απόσταση και που περιέχουν στον τίτλο τους την λέξη κλειδί που ζήτησε.

Οι λέξεις κλειδιά που θα μπορεί να δώσει ο χρήστης θα εμπλουτίζονται κάθε φορά που ο διαχειριστής θα τρέχει το δικό του κομμάτι και θα ενημερώνει τη βάση με νέες λέξεις κλειδιά. Πχ αν ο διαχειριστής εμπλουτίσει τη βάση με την λέξη airport, από κει και πέρα, η λέξη airport θα εμφανίζεται σαν δυνατότητα να την επιλέξει ο χρήστης στις δικές του αναζητήσεις.

Secondary Indexes

Στην βάση MondoDb, δημιουργήσαμε τρεις δευτερεύοντες δείκτες:

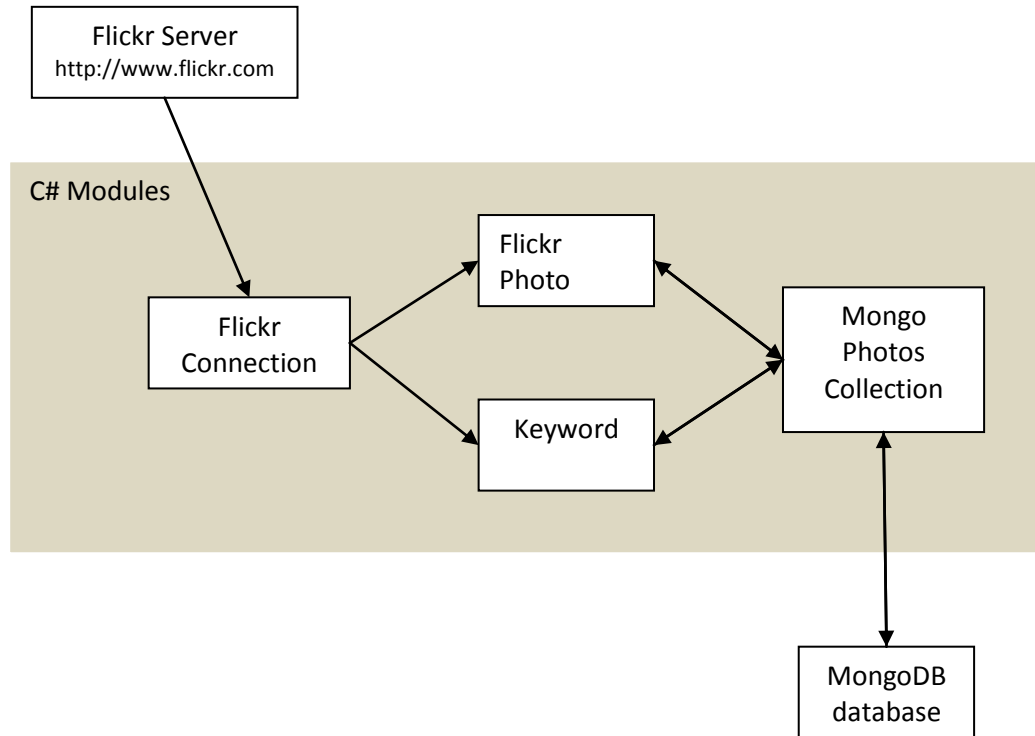
- Έναν Geospatial που διευκολύνει τις αναζητήσεις με longitude και latitude
- Έναν Unique με βάση το PhotoId, ώστε να αποκλείεται να καταχωρηθεί η ίδια φωτογραφία δυο φορές στην βάση
- Έναν με βάση τα tags της κάθε φωτογραφίας για πιο γρήγορες αναζητήσεις με βάση λέξεις κλειδιά.

(Για πρωτεύον κλειδί κάθε καταχώρησης, η MongoDB δημιουργεί έναν μοναδικό κωδικό, που αναφέρεται σαν Objectid).

Ξεχωριστά Modules

- FlickrConnection, η κλάση με την οποία συνδεόμαστε με το Flickr Api και κατεβάζουμε φωτογραφίες με συγκεκριμένα κριτήρια. Χρησιμοποιείται μόνο από την εφαρμογή του Διαχειριστή.
- FlickrPhoto. Στην κλάση αυτή ανήκουν τα αντικείμενα που αποθηκεύονται στην βάση της MondoDb (περιέχει σημαντικά λιγότερα πεδία από την κλάση Photo του Flickr Api. Τα υπόλοιπα δεν τα χρειαζόμαστε). Χρησιμοποιείται και από τις δύο εφαρμογές.
- Keyword. Στην κλάση αυτή ανήκουν τα αντικείμενα που αποθηκεύονται στην βάση της MondoDB σαν λέξεις κλειδιά, τις οποίες έχει κατεβάσει ο διαχειριστής και κάποιες από αυτές μπορεί να χρησιμοποιήσει ο χρήστης.
- MongoPhotosCollection που είναι η κλάση με την οποία συνδεόμαστε με τον τοπικό σέρβερ της MondoDB. Έχει μεθόδους για αποθήκευση φωτογραφιών και λέξεων κλειδιών και μεθόδους για αναζήτηση φωτογραφιών με γεωγραφική θέση και μία, δυο ή τρεις λέξεις κλειδιά.

Τα modules απεικονίζονται στο παρακάτω σχήμα:



4.2 Ιεραρχικό Διάγραμμα

A) Εφαρμογή της πλευράς του διαχειριστή

Καταγραφή των ενεργειών του χρήστη (δηλ. του διαχειριστή του συστήματος)

- 1) Να δώσει μια λέξη κλειδί και να ζητήσει ενημέρωση της τοπικής βάσης με φωτογραφίες που έχουν αυτή την λέξη.
- 2) Να παρακολουθεί την πρόοδο της παραπάνω διαδικασίας.
- 3) Να δει τι λέξεις κλειδιά έχει δώσει μέχρι τώρα.

A) Εφαρμογή της πλευράς του χρήστη

Καταγραφή των ενεργειών του χρήστη (δηλ. του user του συστήματος)

- 1) Να δώσει μια νέα λέξη κλειδί και να κάνει αναζήτηση φωτογραφιών από το Flickr.
- 2) Να κάνει αναζήτηση μιας λέξης από την λίστα, χρησιμοποιώντας το κουμπί Search.
- 3) Να κάνει αναζήτηση για παραπάνω από μία λέξεις κλειδιά.
- 4) Να κάνει αναζήτηση με βάση συγκεκριμένη πόλη που τον ενδιαφέρει
- 5) Να κάνει αναζήτηση μιας πόλης από την λίστα, χρησιμοποιώντας το κουμπί Search.
- 6) Να κάνει αναζήτηση μίας λέξης, όπου θα μπορεί να επιλέξει την επιθυμητή απόσταση (ακτίνα) που θα έχει από το κέντρο της επιλεγμένης πόλης

Κεφάλαιο 5

5.1 Κώδικας

Για αρχή, δημιουργήσαμε ένα μικρό κώδικα σε C# και ζητήσαμε από το Flickr να μας επιστρέψει 100 φωτογραφίες ανά σελίδα, και συγκεκριμένα να μας επιστρέψει την πρώτη σελίδα μόνο, και οι φωτογραφίες έχουν ως geotag την Νέα Υόρκη. Από την μέθοδο χρησιμοποιήσαμε τα Langtitude, Longtitude, hasGeo, perpage και page. Ζητήσαμε να μας επιστρέψει σε ένα Xml αρχείο τα αποτελέσματα. Στην αρχή δεν αποθηκεύεται σε κάποια ΒΔ. Κάθε φορά που τρέχει ο κώδικας δημιουργείται το αρχείο. Οι δυνατότητες είναι πάρα πολλές, και η συγκεκριμένη μέθοδος, μπορεί και τις συνδυάζει σχεδόν όλες. Θα μπορούσαμε για παράδειγμα να κάνουμε αναζήτηση με βάση την ημερομηνία, την παλαιότητα, tagname, το όνομα του ιδιοκτήτη, ή ακόμα θα μπορούσαμε να ζητήσουμε να επιστρέψει αποτελέσματα τα οποία βρίσκονται μέχρι και κάποια χιλιόμετρα από την αρχική μας αναζήτηση. Υπάρχουν και άλλες μέθοδοι, με λιγότερες δυνατότητες, έχουμε πχ την δυνατότητα να δούμε στατιστικά, δηλ πόσα views έχει μια φωτογραφία, η την συγκεκριμένη ημερομηνία ποια φωτογραφία είχε τα περισσότερα comments.κλπ. Παρακάτω το ερώτημα αναφέρεται στην μέθοδο «PhotoSearchOptions» όπου latitude και longitude αντιστοιχούν σε μια συγκεκριμένη περιοχή. Τα αποτελέσματα θα είναι μία (1) σελίδα. Και τα αποτελέσματα της σελίδας θα είναι 100 ανά σελίδα.

```
PhotoSearchOptions options = new PhotoSearchOptions();
options.Lat = 40.714623;
options.Lon = -74.006605;
options.per_page = 100;
options.page = 1 ;
PhotoCollection photos = flickr.PhotosSearch(options);
```

Ακολουθεί εικόνα με τα αποτελέσματα που μας επιστέφει ο κώδικας σε xml μορφή.

```
C:\Users\Maria\Documents\Visual Studio 2012\P...
<?xml version="1.0" encoding="utf-8" ?>
- <ArrayOfPhoto xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:xsd="http://www.w3.org/2001/XMLSchema">
- <Photo>
  <Tags />
  <PhotoId>8694343507</PhotoId>
  <UserId>72107934@N08</UserId>
  <Secret>4b2df18187</Secret>
  <Server>8122</Server>
  <Farm>9</Farm>
  <Title>DSC05337</Title>
  <IsPublic>true</IsPublic>
  <IsFriend>false</IsFriend>
  <IsFamily>false</IsFamily>
  <License>AllRightsReserved</License>
  <OriginalWidth>0</OriginalWidth>
  <OriginalHeight>0</OriginalHeight>
  <DateUploaded>0001-01-01T00:00:00</DateUploaded>
  <LastUpdated>0001-01-01T00:00:00</LastUpdated>
  <DateTaken>0001-01-01T00:00:00</DateTaken>
  <DateAddedToGroup xsi:nil="true" />
  <DateFavorited xsi:nil="true" />
  <SquareThumbnailWidth xsi:nil="true" />
  <SquareThumbnailHeight xsi:nil="true" />
  <LargeSquareThumbnailWidth xsi:nil="true" />
  <LargeSquareThumbnailHeight xsi:nil="true" />
  <ThumbnailWidth xsi:nil="true" />
  <ThumbnailHeight xsi:nil="true" />
  <SmallWidth xsi:nil="true" />
  <SmallHeight xsi:nil="true" />
  <Small320Width xsi:nil="true" />
  <Small320Height xsi:nil="true" />
  <Medium640Width xsi:nil="true" />

```

Η βιβλιοθήκη FlickrNet, μας δίνει μια σειρά από κλάσεις για να επικοινωνήσουμε με το API του Flickr, όπως Photo, PhotoSearchOptions, PhotosCollection κλπ. Εμείς φτιάξαμε μια κλάση **FlickrPhoto** που περιέχει μόνο ορισμένα από τα δεδομένα που περιέχει μια Photo όταν έρχεται από το Flickr (πχ Views και Tags). Η κλάση **FlickrConnection** συνδέεται με το api του flickr και διαθέτει μεθόδους για κατέβασμα φωτογραφιών από το flickr, με βάση λέξεις κλειδιά και ημερομηνία ανεβάσματος. Για να επικοινωνήσουμε με την MongoDB, έχουμε ενσωματώσει δύο βιβλιοθήκες, τις MongoDB.Bson και MongoDB.Driver, που περιέχουν κλάσεις για την επικοινωνία με το api της MongoDB όπως MongoDBDatabase, MongoClient, Query κλπ. Δημιουργήσαμε την κλάση **MongoPhotosCollection** η οποία επικοινωνεί με τον server της MongoDB (που πρέπει να τρέχει στο background). Η κλάση αυτή έχει μεθόδους για να αποθηκεύει αντικείμενα της κλάσης FlickrPhoto, στην βάση MongoDB. Υπάρχει ένας μηχανισμός που απαγορεύει την αποθήκευση της ίδιας φωτογραφίας δυο φορές (με βάση το μοναδικό PhotoId που μας δίνει το Flickr). Επίσης περιέχει μεθόδους για αναζήτηση δεδομένων στη βάση, με βάση λέξεις κλειδιά και γεωγραφικούς περιορισμούς (με βάση ένα σημείο κέντρο και μια ελάχιστη ακτίνα). Τέλος υπάρχει η κλάση **Keyword** για να αποθηκεύονται στην βάση δεδομένων και οι λέξεις κλειδιά τις οποίες έχουμε κατεβάσει.

5.2 Εγκατάσταση

Από το <http://www.mongodb.org/downloads> κατεβάσαμε την έκδοση της MongoDB για 64 bit. Εκτελέσαμε το αρχείο msi που κατέβηκε. Αυτό εγκατέστησε την MongoDB σε έναν φάκελο μέσα στο Program Files. Μέσα στον υποφάκελο bin, υπάρχει το mongod.exe και το mongo.exe. Δημιουργήσαμε τον φάκελο c:\data\db (υποτίθεται ότι είναι ο default φάκελος για αποθήκευση). Σε ένα command prompt, πήγαμε στον φάκελο bin και τρέξαμε mongod -dbpath "c:\data\db. Στην συνέχεια τρέχει μια διεργασία που παγώνει το παράθυρο.

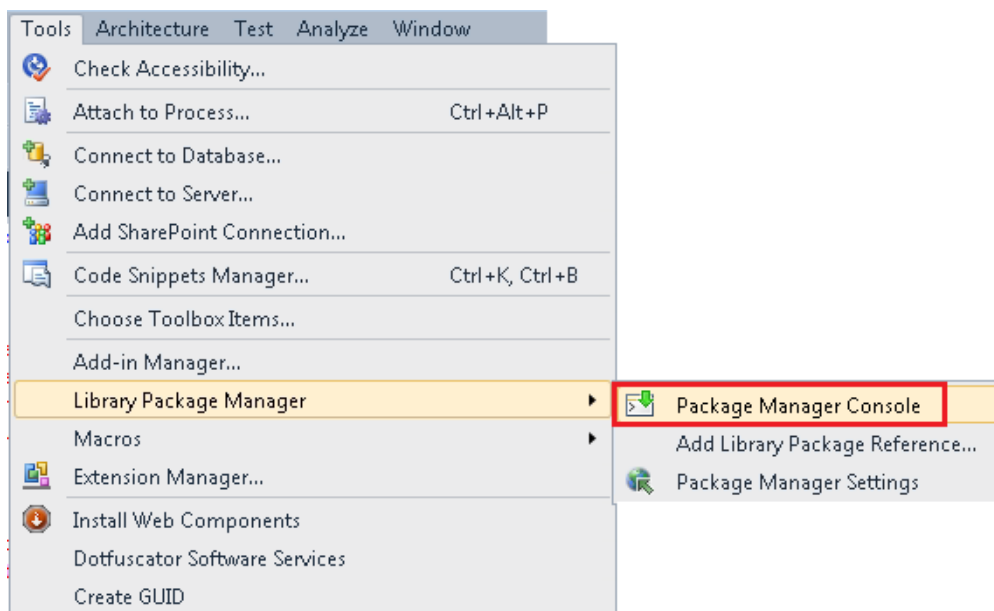
Από ένα άλλο command prompt, τρέχουμε το mongo.exe, στο οποίο μπορούμε να δίνουμε εντολές. Για να τρέχει η MongoDB θα πρέπει κάθε φορά να τρέχουμε το αρχείο mongod.exe μέσω cmd (Command Prompt).

5.2.1 Install C#

Κατεβάσαμε το msi από το <https://github.com/mongodb/mongo-csharp-driver/releases>, τρέχουμε το msi και αυτό εγκαθιστά δυο dll στον φάκελο "Αρχεία εφαρμογών (x86)\MongoDB". Αυτά τα δύο τα κάνουμε Add Reference σε οποιοδήποτε project θέλουμε, όπως και τα :

- using MongoDB.Bson;
- using MongoDB.Driver;

Για να εγκαταστήσουμε την MongoDB στο Visual Studio, ακολουθούμε τα βήματα όπως φαίνεται και στην παρακάτω εικόνα.



Στην συνέχεια γράφουμε την εντολή ***Install-Package Mongocsharp***. Στην εικόνα που ακολουθεί βλέπουμε ότι η βιβλιοθήκη έχει εισαχθεί επιτυχώς στο project μας (getphotos1).

```
PM> Install-Package mongocsharpdriver -Version 1.9.2
```

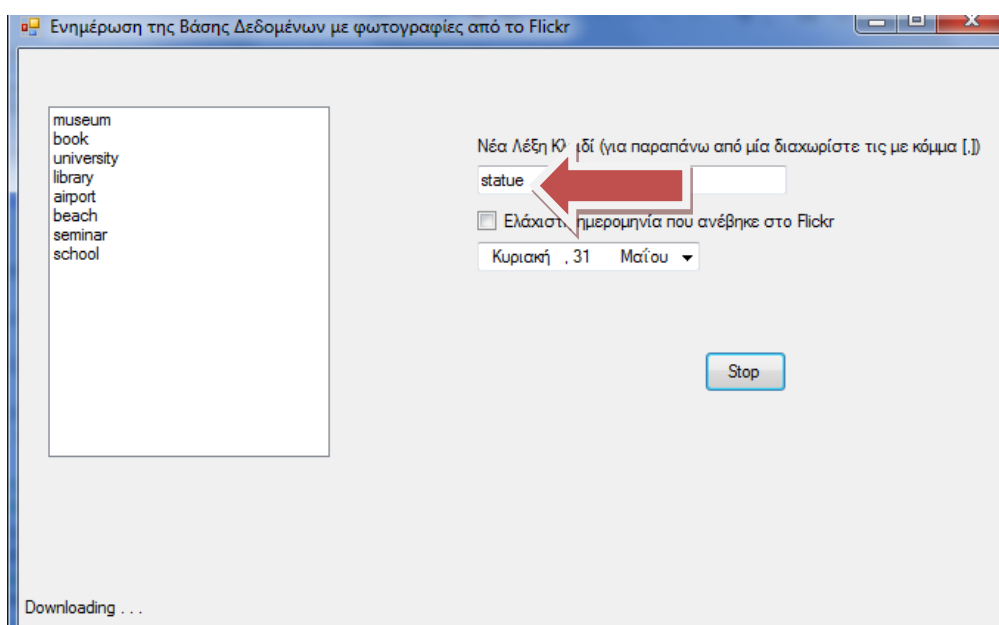
Κεφάλαιο 6

Δοκιμές – Αποτελέσματα

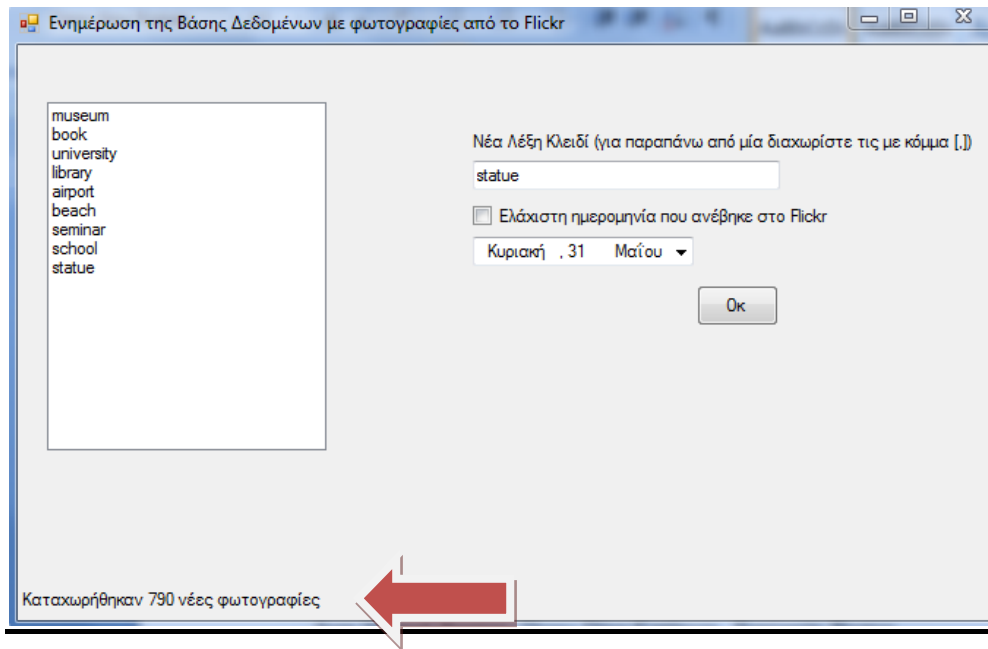
Για τις ανάγκες της εργασίας κατεβάσαμε αρκετές λέξεις (keywords), και παρατηρήσαμε την εξέλιξη του κατεβάσματος (download)

Παράδειγμα 1^ο

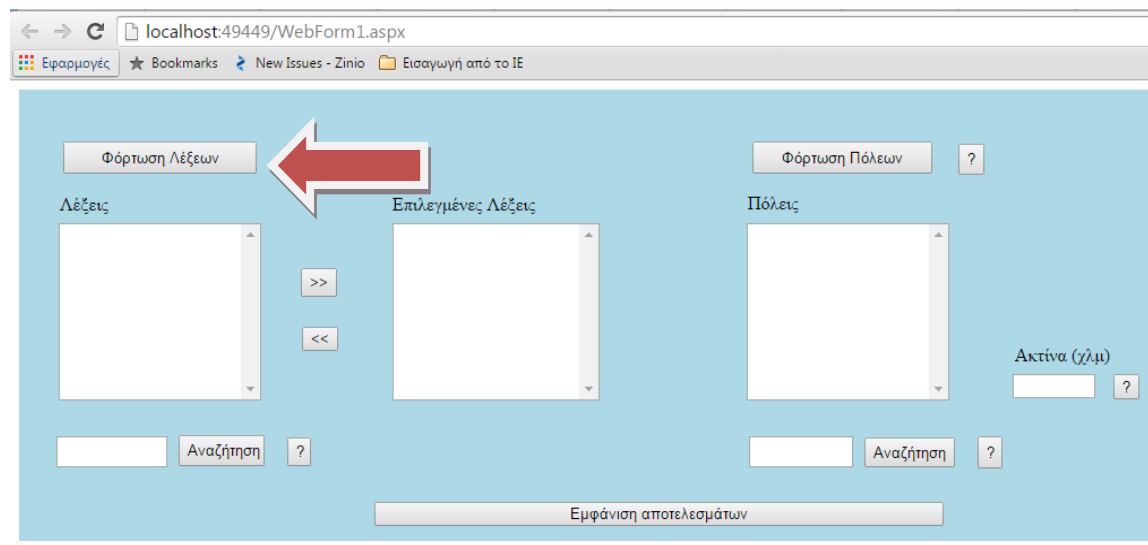
Κατεβάσαμε την λέξη "statue", στις 15:40



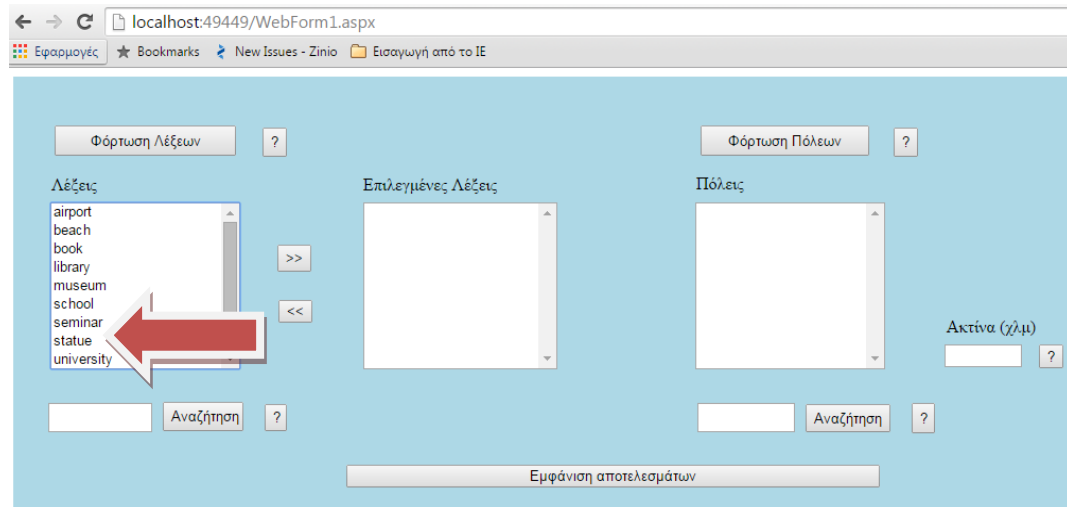
Δύο ώρες μετά 17:40, σταματήσαμε την διαδικασία κατεβάσματος από το κουμπί **OK**. Βλέπουμε τα αποτελέσματα κάτω αριστερά. Για την λέξη "statue" κατέβηκαν 790 φωτογραφίες από το Flickr



Πηγαίνουμε στην εφαρμογή του χρήστη και πατάμε «Φόρτωση λέξεων»

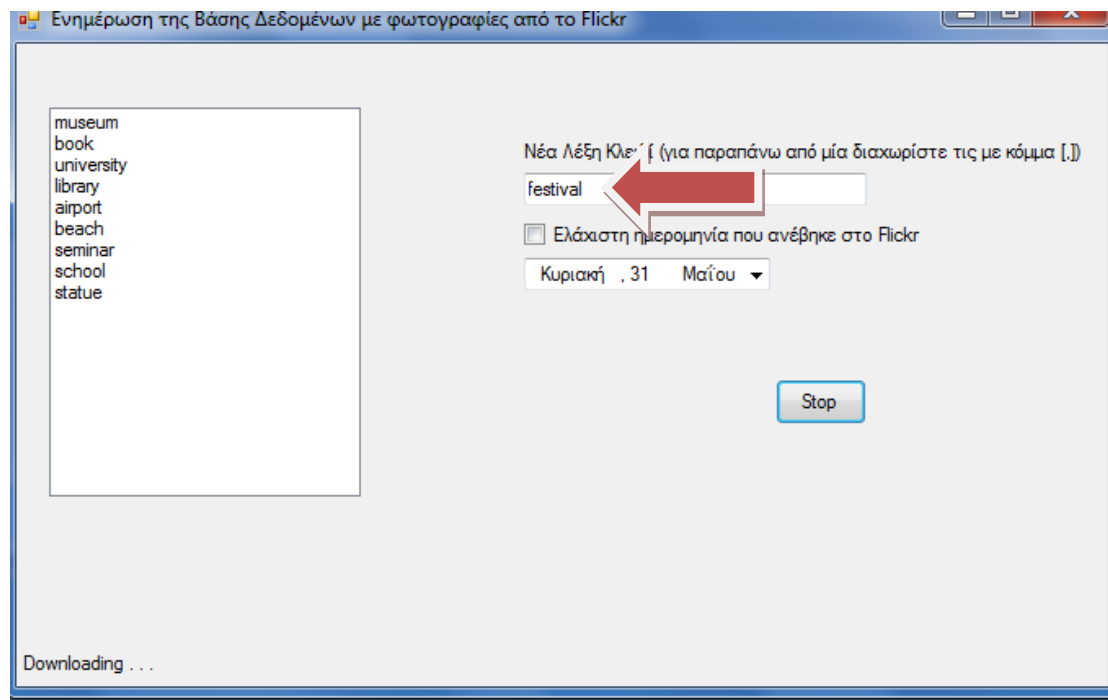


Εκεί βλέπουμε ότι στις διαθέσιμες λέξεις προς αναζήτηση, έχει προστεθεί η λέξη “statue”.

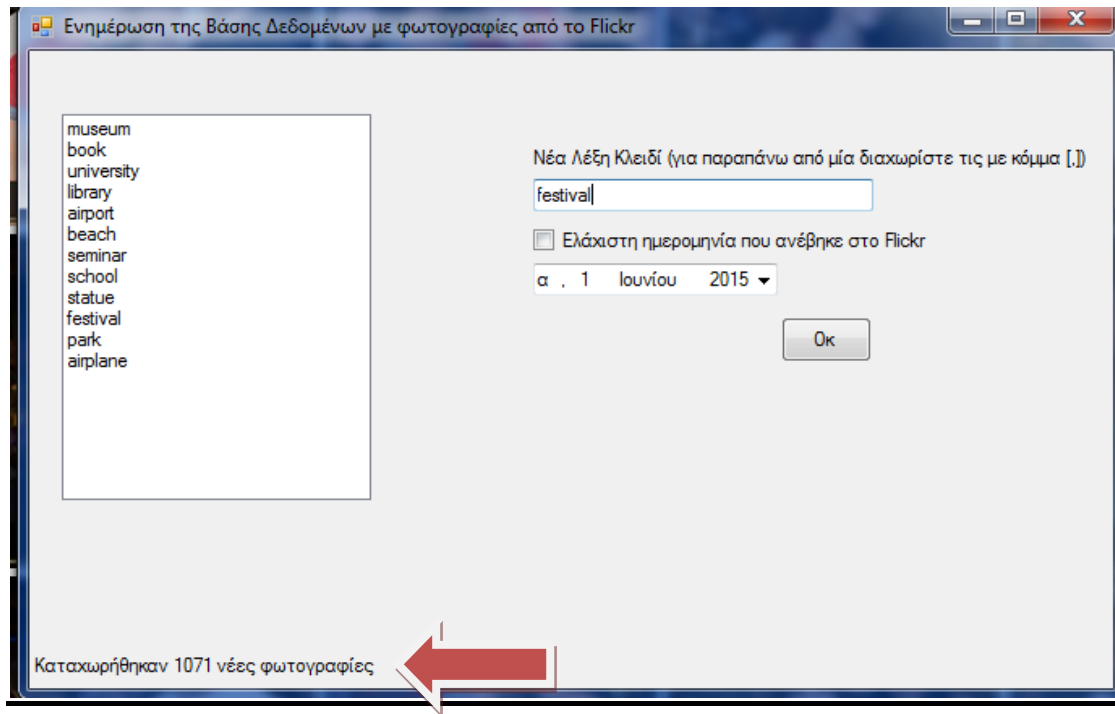


Παράδειγμα 2^ο

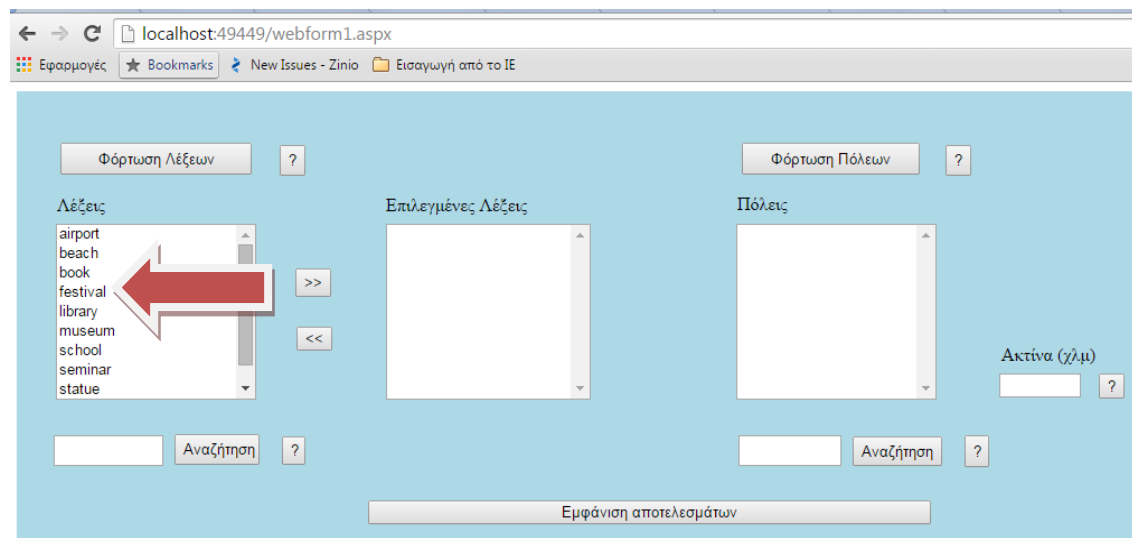
Κατεβάσαμε την λέξη “festival”



Περίπου 4 ώρες μετά βλέπουμε ότι κατέβηκαν 1071 φωτογραφίες

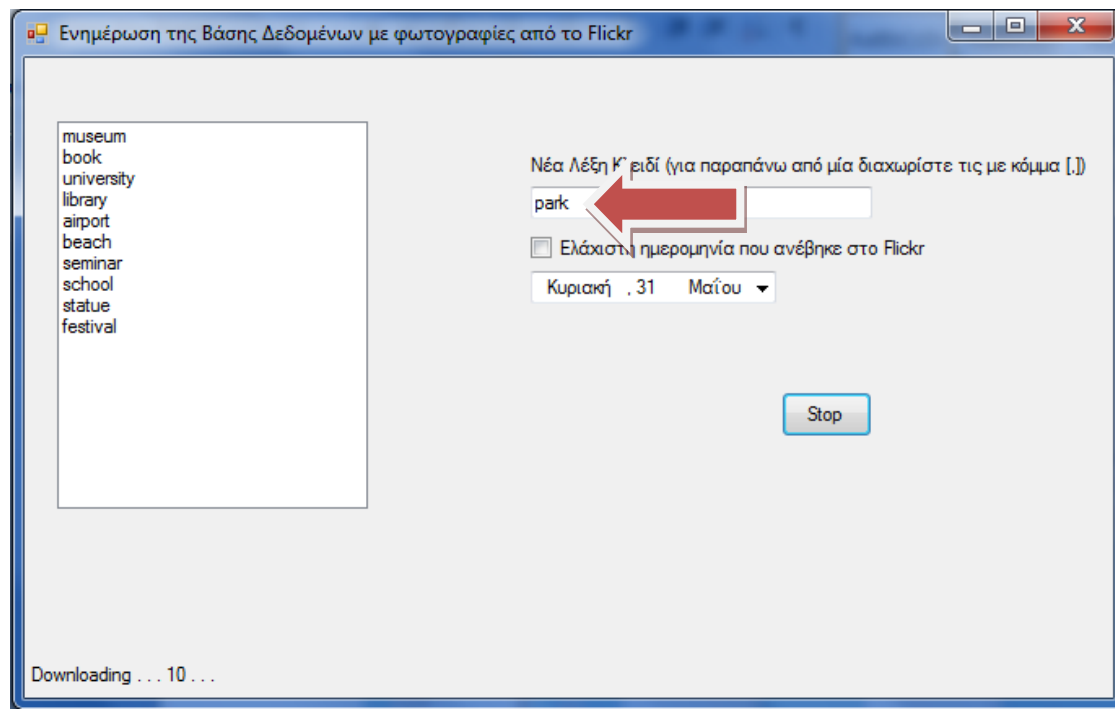


Όπως και πριν βλέπουμε ότι έχει προστεθεί στην εφαρμογή μας η λέξη “festival”

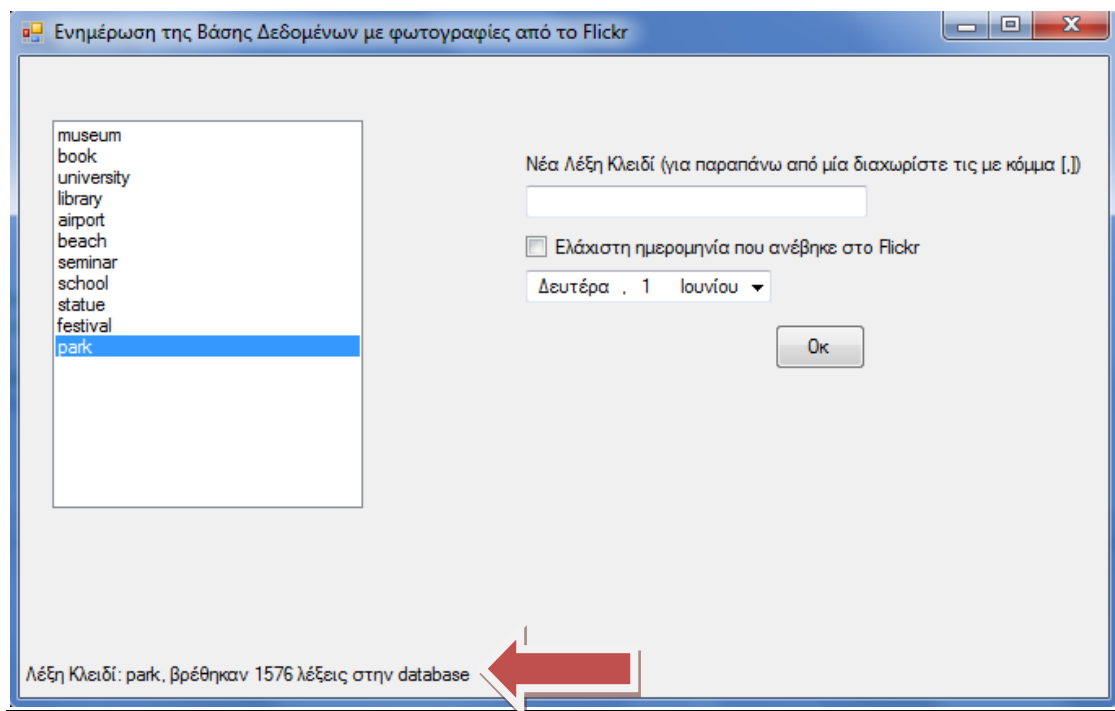


Παράδειγμα 3°

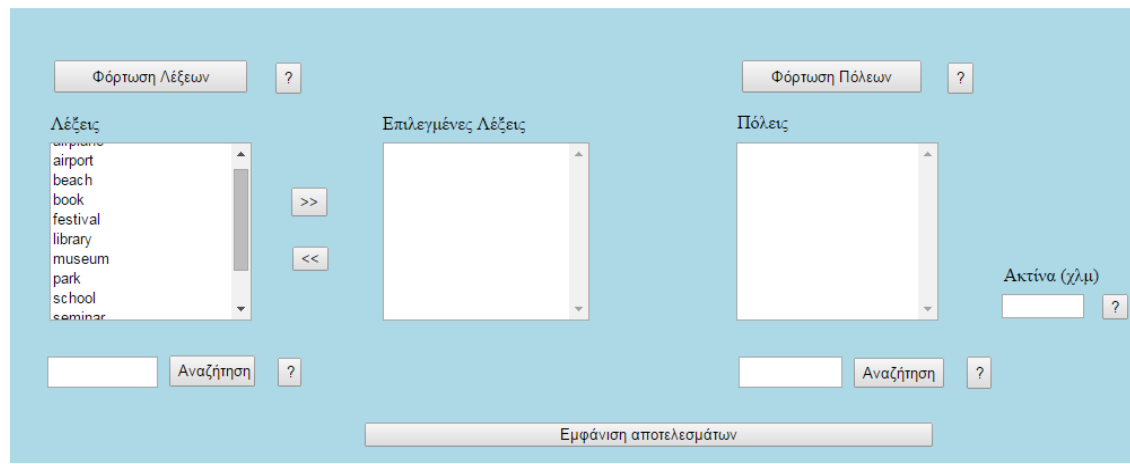
Κατεβάσαμε την λέξη “park” στις 19:10



Τέσσερις ώρες μετά, στις 00:15 σταματήσαμε το κατέβασμα. Βλέπουμε ότι καταχωρήθηκαν 1576 φωτογραφίες στην βάση μας.

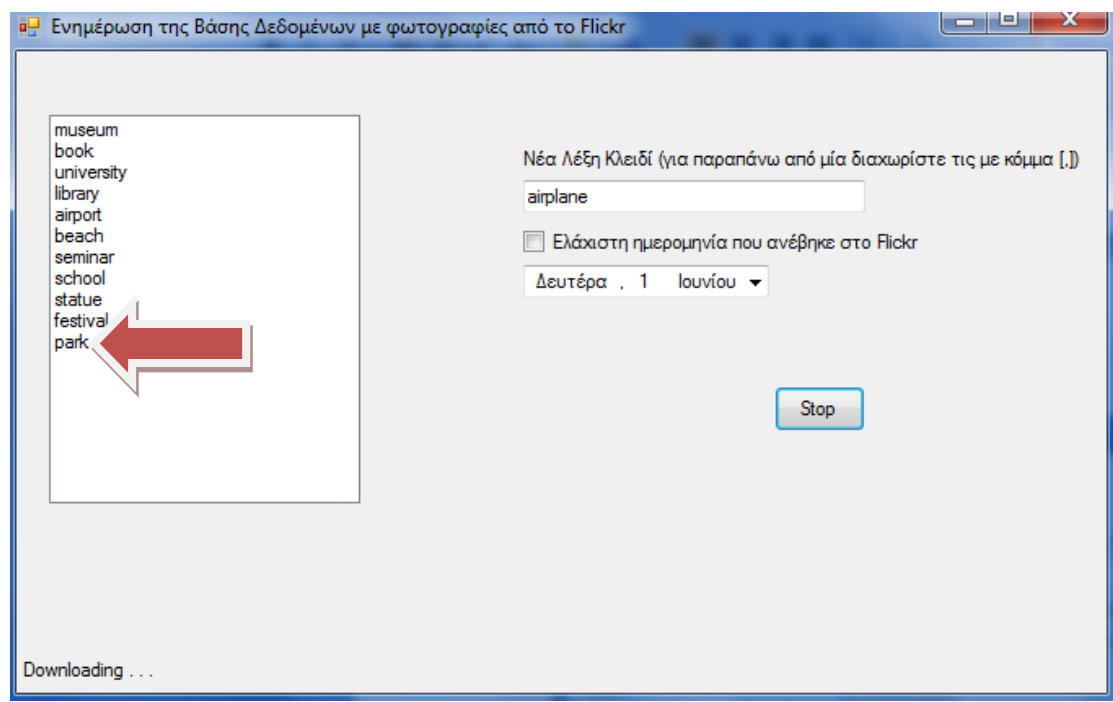


Και πάλι στην εφαρμογή βλέπουμε ότι έχει προστεθεί στην λίστα η νέα λέξη “park”

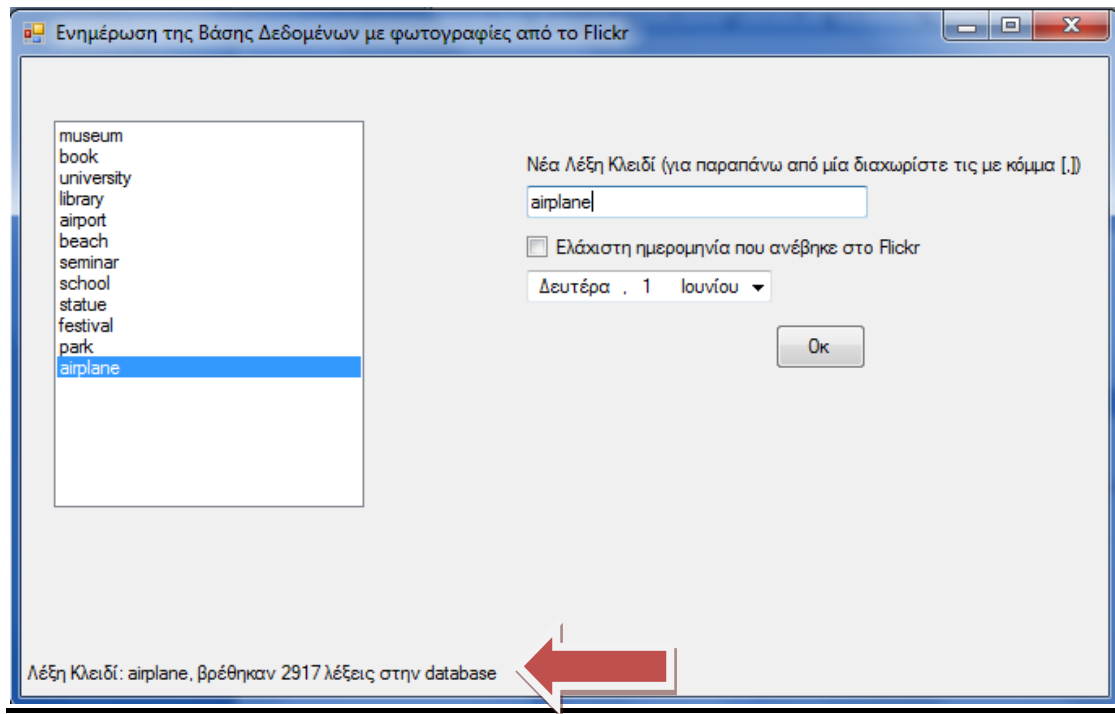


Παράδειγμα 4^ο

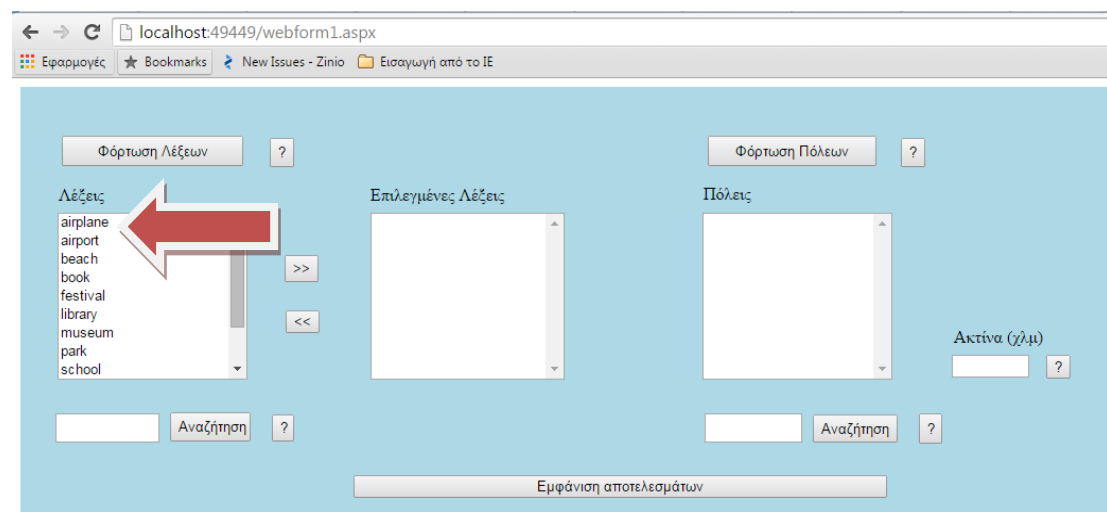
Κατεβάσαμε την λέξη "airplane" στις 17:20



Μετά περίπου 6 ώρες στις 11:10, κατέβηκαν 2917 φωτογραφίες, όπως φαίνεται και παρακάτω.



Ομοίως βλέπουμε στην εφαρμογή, η λέξη “airplane” να έχει προστεθεί στην λίστα μας, έτοιμη προς αναζήτηση.



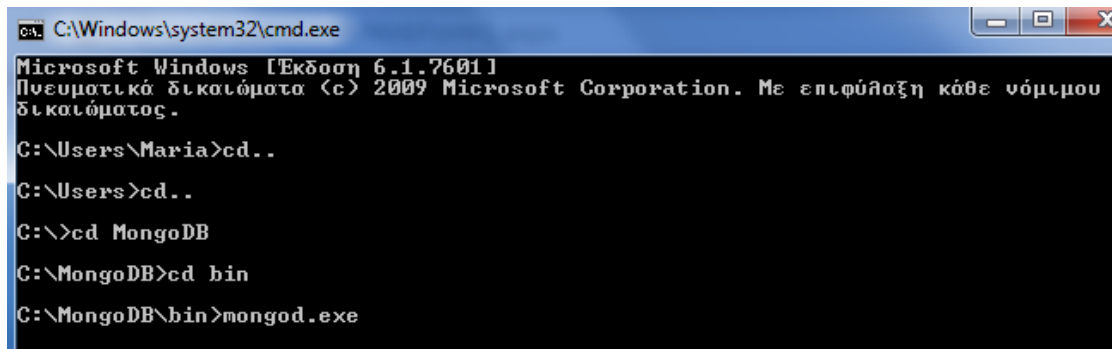
Ακολουθεί πίνακας με τα αναλυτικά αποτελέσματα της εφαρμογής του διαχειριστή

Tag	Σύνολο φωτογραφιών
Museum	2697
Book	937
Beach	295
School	4329
Statue	896
Festival	1071
Park	1625
Airport	1095
Seminar	1570
Library	3155
Airplane	2917

Κεφάλαιο 7

Οδηγίες Εφαρμογής (Manual)

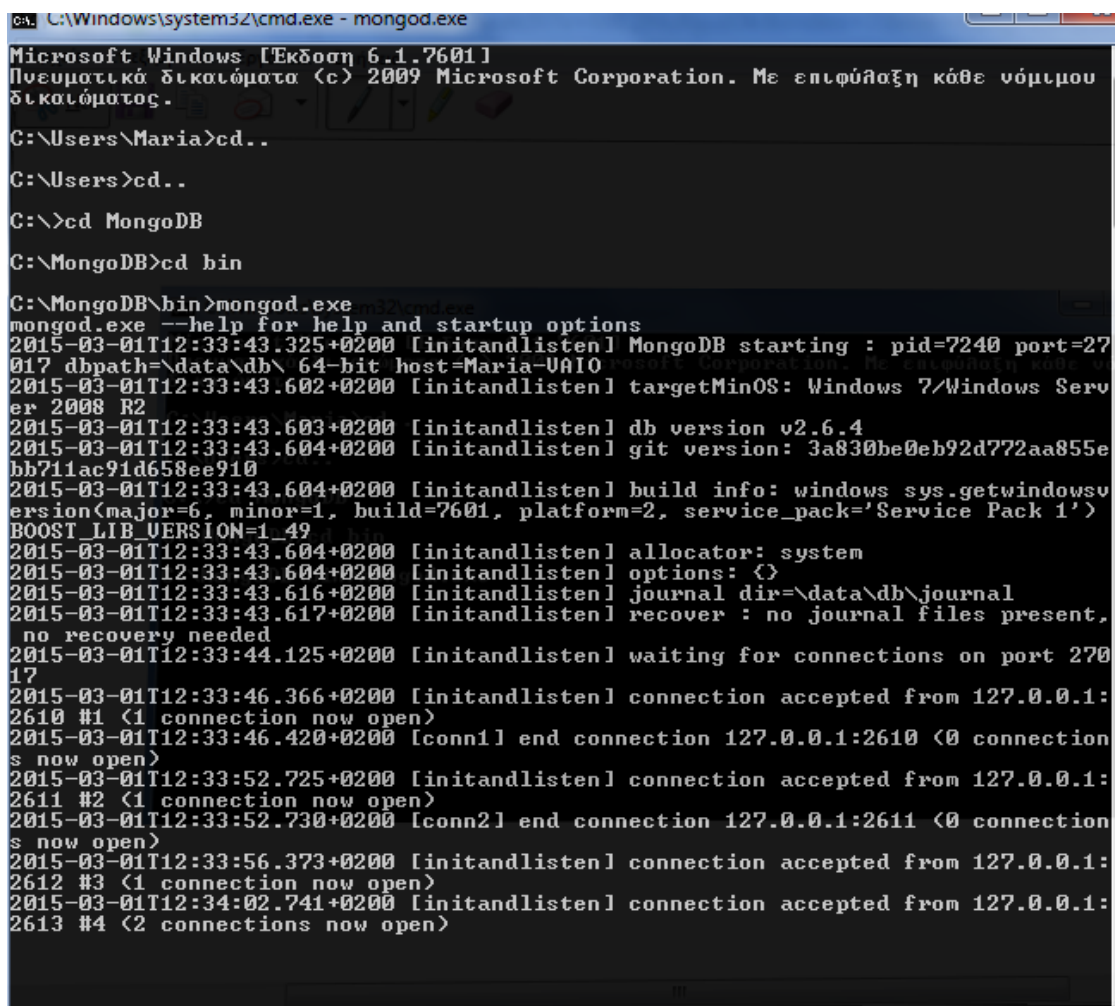
Για να τρέξει η εφαρμογή θα πρέπει πρώτα να τρέξουμε τον server της MongoDB. Για να το κάνουμε αυτό τρέχουμε την εξής εντολή `C:\MongoDB\bin\mongod.exe` μέσω Command Prompt.



```
CA. C:\Windows\system32\cmd.exe
Microsoft Windows [Έκδοση 6.1.7601]
Πνευματικά δικαιώματα (c) 2009 Microsoft Corporation. Με επιφύλαξη κάθε νόμιμου
δικαιώματος.

C:\Users\Maria>cd..
C:\Users>cd..
C:\>cd MongoDB
C:\MongoDB>cd bin
C:\MongoDB\bin>mongod.exe
```

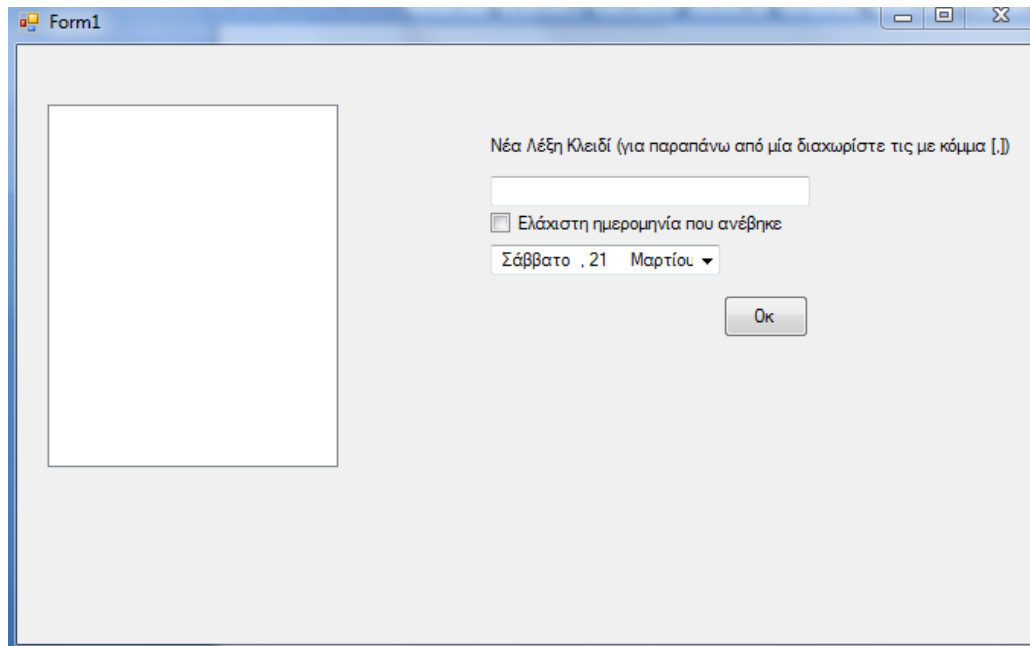
Από την παρακάτω εικόνα βλέπουμε ότι η επικοινωνία είναι ανοιχτή.



```
CA. C:\Windows\system32\cmd.exe - mongod.exe
Microsoft Windows [Έκδοση 6.1.7601]
Πνευματικά δικαιώματα (c) 2009 Microsoft Corporation. Με επιφύλαξη κάθε νόμιμου
δικαιώματος.

C:\Users\Maria>cd..
C:\Users>cd..
C:\>cd MongoDB
C:\MongoDB>cd bin
C:\MongoDB\bin>mongod.exe
mongod.exe --help for help and startup options
2015-03-01T12:33:43.325+0200 [initandlisten] MongoDB starting : pid=7240 port=27
017 dbpath=\data\db\ 64-bit host=Maria-UAIO
2015-03-01T12:33:43.602+0200 [initandlisten] targetMinOS: Windows 7/Windows Serv
er 2008 R2
2015-03-01T12:33:43.603+0200 [initandlisten] db version v2.6.4
2015-03-01T12:33:43.604+0200 [initandlisten] git version: 3a830be0eb92d772aa855e
bb711ac91d658ee910
2015-03-01T12:33:43.604+0200 [initandlisten] build info: windows sys.getwindowsv
ersion(major=6, minor=1, build=7601, platform=2, service_pack='Service Pack 1')
BOOST_LIB_VERSION=1_49
2015-03-01T12:33:43.604+0200 [initandlisten] allocator: system
2015-03-01T12:33:43.604+0200 [initandlisten] options: {}
2015-03-01T12:33:43.616+0200 [initandlisten] journal dir=\data\db\journal
2015-03-01T12:33:43.617+0200 [initandlisten] recover : no journal files present,
no recovery needed
2015-03-01T12:33:44.125+0200 [initandlisten] waiting for connections on port 270
17
2015-03-01T12:33:46.366+0200 [initandlisten] connection accepted from 127.0.0.1:
2610 #1 <1 connection now open>
2015-03-01T12:33:46.420+0200 [conn1] end connection 127.0.0.1:2610 <0 connection
s now open>
2015-03-01T12:33:52.725+0200 [initandlisten] connection accepted from 127.0.0.1:
2611 #2 <1 connection now open>
2015-03-01T12:33:52.730+0200 [conn2] end connection 127.0.0.1:2611 <0 connection
s now open>
2015-03-01T12:33:56.373+0200 [initandlisten] connection accepted from 127.0.0.1:
2612 #3 <1 connection now open>
2015-03-01T12:34:02.741+0200 [initandlisten] connection accepted from 127.0.0.1:
2613 #4 <2 connections now open>
```

Στην συνέχεια πηγαίνουμε στο διαχειριστικό κομμάτι. Εδώ φαίνεται η εφαρμογή του διαχειριστή χωρίς να έχει τρέξει κάποιο ερώτημα.



Form1

Νέα Λέξη Κλειδί (για παραπάνω από μία διαχωρίστε τις με κόμμα [,])

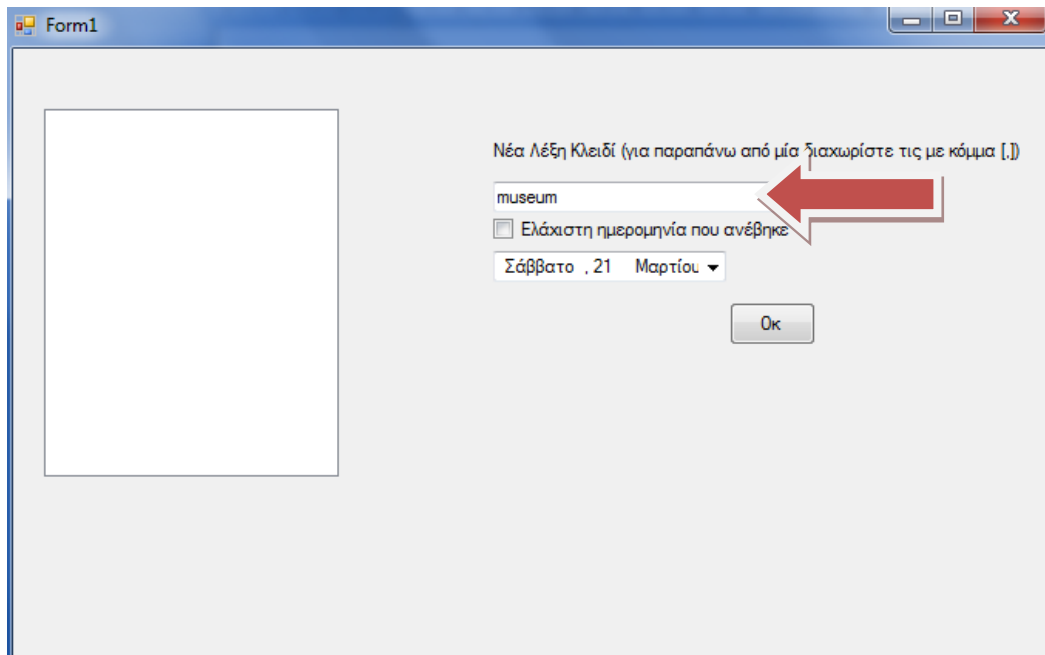
Ελάχιστη ημερομηνία που ανέβηκε

Σάββατο , 21 Μαρτίου ▼

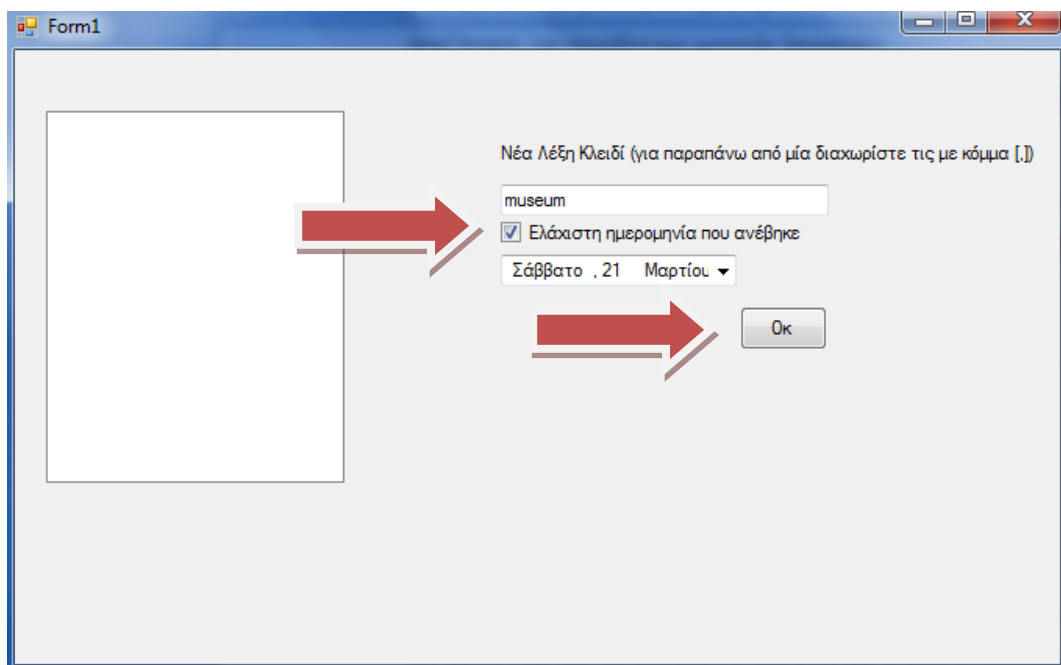
Οκ

Εδώ ο διαχειριστής, τρέχει queries, έτσι ώστε να είναι διαθέσιμα στο χρήστη προς αναζήτηση. Κατεβάζει αποτελέσματα από το Flickr τα οποία αποθηκεύονται στην MongoDB.

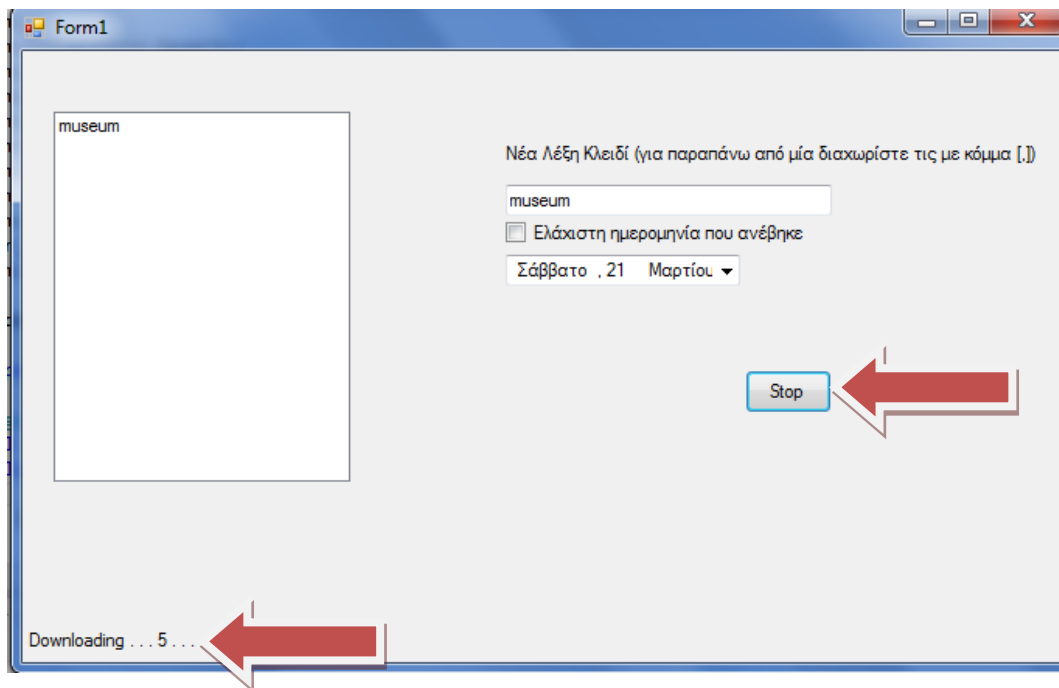
Στο πεδίο **Νέα Λέξη Κλειδί** (keyword), βάζει την λέξη που θέλει να κάνει αναζήτηση, για παράδειγμα μουσεία (museum). Όπως φαίνεται και στην εικόνα για αναζήτηση άνω των 2 λέξεων, θα πρέπει να παρεμβάλλεται κόμμα (,)



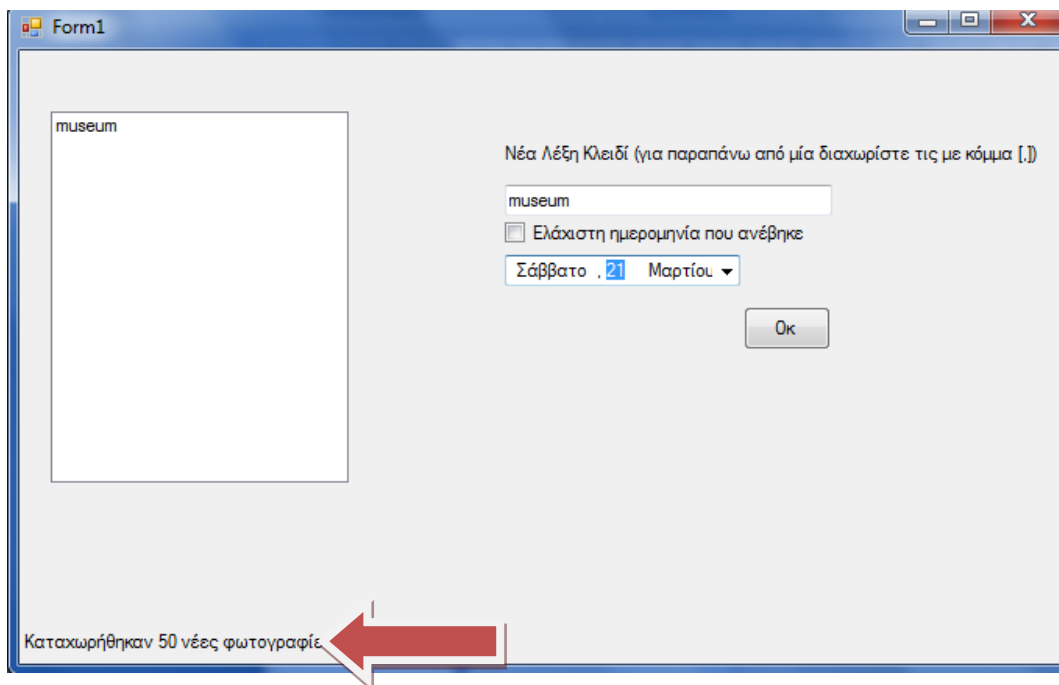
Αν θέλει έχει την δυνατότητα να επιλέξει την ελάχιστη ημερομηνία που ανέβηκε μια φωτογραφία στο Flickr, επιλέγοντας το check box. Στην συνέχεια πατάει **Οκ** για να ξεκινήσει η αναζήτηση.



Κάτω αριστερά μπορεί να παρακολουθεί την διαδικασία κατεβάσματος των αποτελεσμάτων (download). Και πατώντας **Stop**, μπορεί αν θέλει να διακόψει την διαδικασία ανά πάσα στιγμή.



Όταν τελειώσει η διαδικασία βλέπει το τελικό αποτέλεσμα κάτω αριστερά, συγκεκριμένα πόσες φωτογραφίες έχουν κατεβεί συνολικά.



Αν επιλέξει την λέξη museum, βλέπει τις πληροφορίες για την αναζήτηση που έχει κάνει. Δηλαδή ποια ήταν η λέξη που αναζήτησε και πόσα αποτελέσματα καταχωρήθηκαν στην βάση (database).

The screenshot shows a window titled "Form1" with a search interface. On the left, a list box contains the word "museum". A red arrow points from this list box to the right. On the right side, there is a search form with the following elements:

- A text input field containing "museum".
- A checkbox labeled "Ελάχιστη ημερομηνία που ανέβηκε" (Minimum date uploaded) which is currently unchecked.
- A date selection dropdown menu showing "Σάββατο , 21 Μαρτίου" (Saturday, 21 March).
- An "Οκ" (OK) button.

At the bottom left of the window, a status bar displays the text: "Keyword: museum , βρέθηκαν 65 λέξεις στην database". A red arrow points from this status bar to the right.

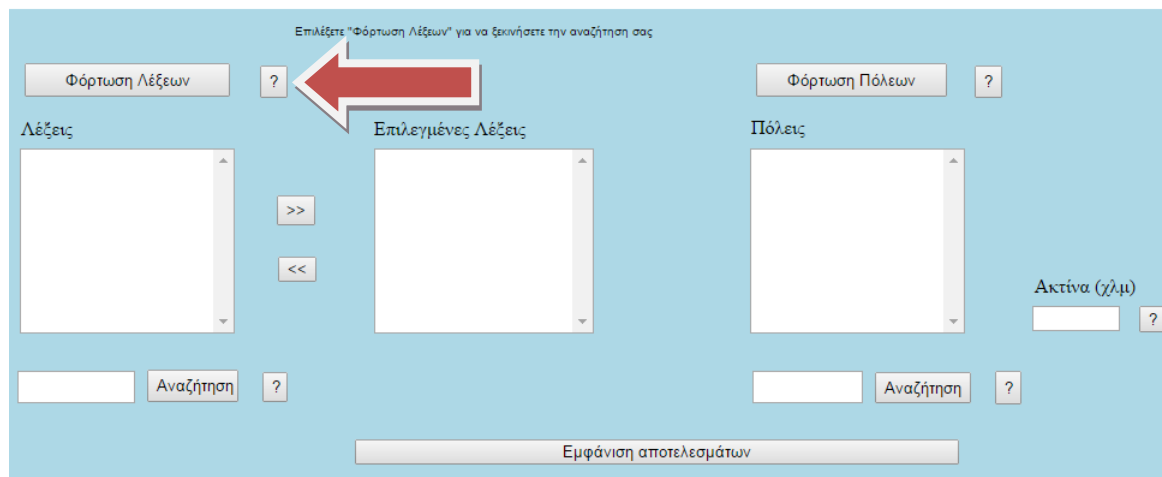
Στην συνέχεια ακολουθεί η εφαρμογή του χρήστη όπως φαίνεται όταν τρέχει για πρώτη φορά.

The screenshot shows a web application interface with a light blue background. At the top, there are two buttons: "Φόρτωση Λέξεων" (Load Words) and "Φόρτωση Πόλεων" (Load Cities), each followed by a question mark icon. Below these are three main sections: "Λέξεις" (Words) on the left, "Επιλεγμένες Λέξεις" (Selected Words) in the middle, and "Πόλεις" (Cities) on the right. Each section has a vertical list box. Between the "Λέξεις" and "Επιλεγμένες Λέξεις" boxes are two buttons: ">>" and "<<". To the right of the "Πόλεις" box is a text input field labeled "Ακτίνα (χλμ)" (Radius in km) with a question mark icon. Below each of the three list boxes is a text input field and a button labeled "Αναζήτηση" (Search) with a question mark icon. At the bottom center, there is a button labeled "Εμφάνιση αποτελεσμάτων" (Show Results).

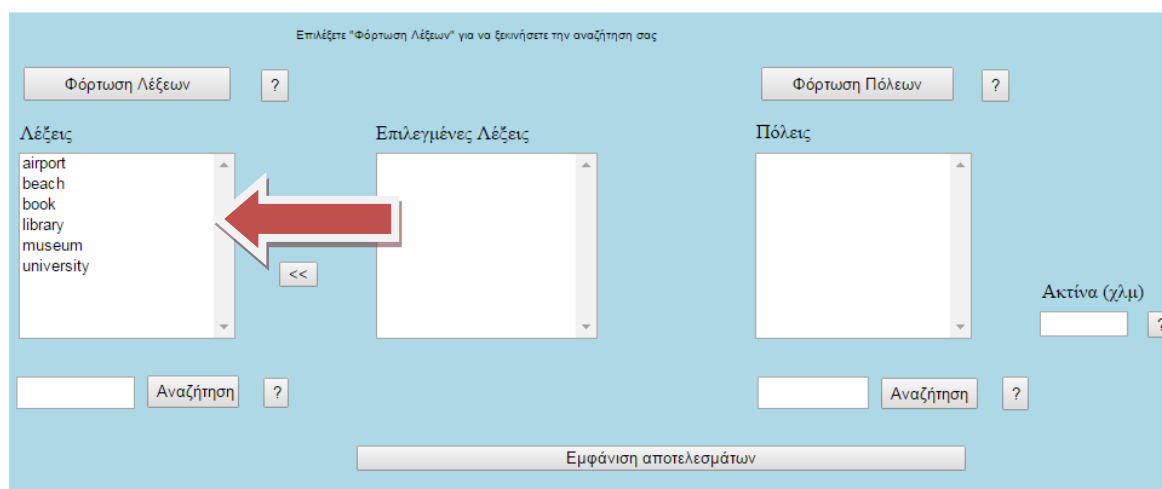
Σε κάποια σημεία υπάρχουν βοηθήματα για τον χρήστη (κουμπιά με ερωτηματικά ?), ώστε να κάνουν την περιήγηση του στην εφαρμογή πιο εύκολη. Για αρχή ο χρήστης πρέπει να πατήσει **Φόρτωση Λέξεων** για να του εμφανιστούν οι λέξεις που υπάρχουν στην βάση της MongoDB, στο πεδίο **Λέξεις**.

This screenshot is identical to the one above, but with five large red arrows pointing to the question mark icons. The arrows point to the question marks next to the "Φόρτωση Λέξεων" and "Φόρτωση Πόλεων" buttons, the question mark next to the "Ακτίνα (χλμ)" input field, and the question marks next to the "Αναζήτηση" buttons under both the "Λέξεις" and "Πόλεις" sections.

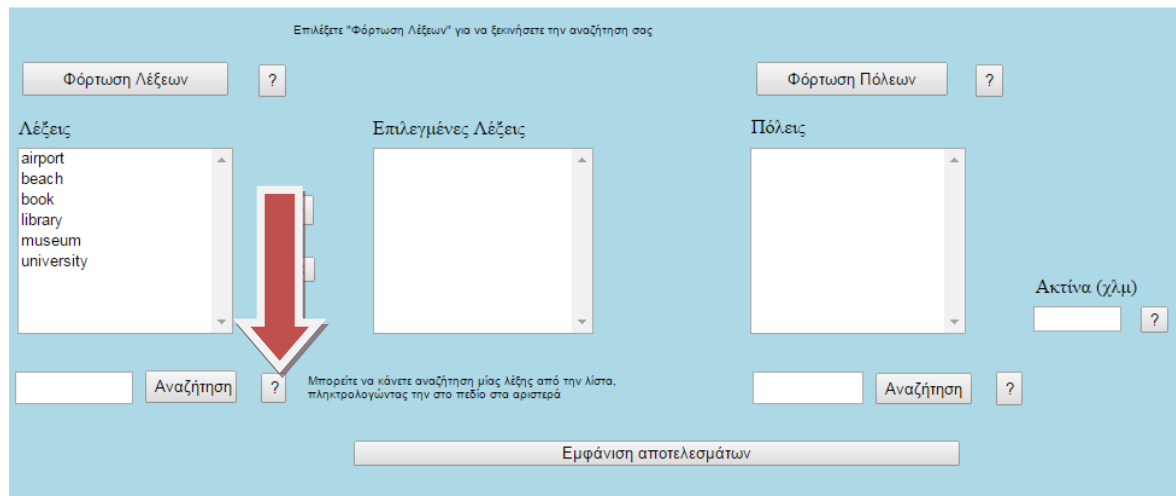
Πατώντας λοιπόν στο πρώτο ερωτηματικό δίπλα από το πεδίο **Φόρτωση Λέξεων**, εμφανίζεται το μήνυμα «**Επιλέξτε Φόρτωση Λέξεων για να ξεκινήσετε την αναζήτηση σας**». Αν ξαναπατήσει πάνω στο ερωτηματικό το μήνυμα εξαφανίζεται. Αυτό ισχύει για όλα τα βοηθητικά μηνύματα της εφαρμογής.



Πατώντας **Φόρτωση Λέξεων**, στο πεδίο **Λέξεις** εμφανίζεται μια λίστα με τις λέξεις, οι οποίες είναι διαθέσιμες προς αναζήτηση από τον χρήστη.



Ένα ακόμα βοηθητικό μήνυμα, ενημερώνει ότι υπάρχει η δυνατότητα αναζήτησης μίας λέξης από την λίστα, πληκτρολογώντας την στο πεδίο στα αριστερά δίπλα από το κουμπί **Αναζήτηση**.



Επιλέξτε "Φόρτωση Λέξεων" για να ξεκινήσετε την αναζήτηση σας

Φόρτωση Λέξεων ? Φόρτωση Πόλεων ?

Λέξεις: airport, beach, book, library, museum, university

Επιλεγμένες Λέξεις

Πόλεις

Ακτίνα (χλμ) ?

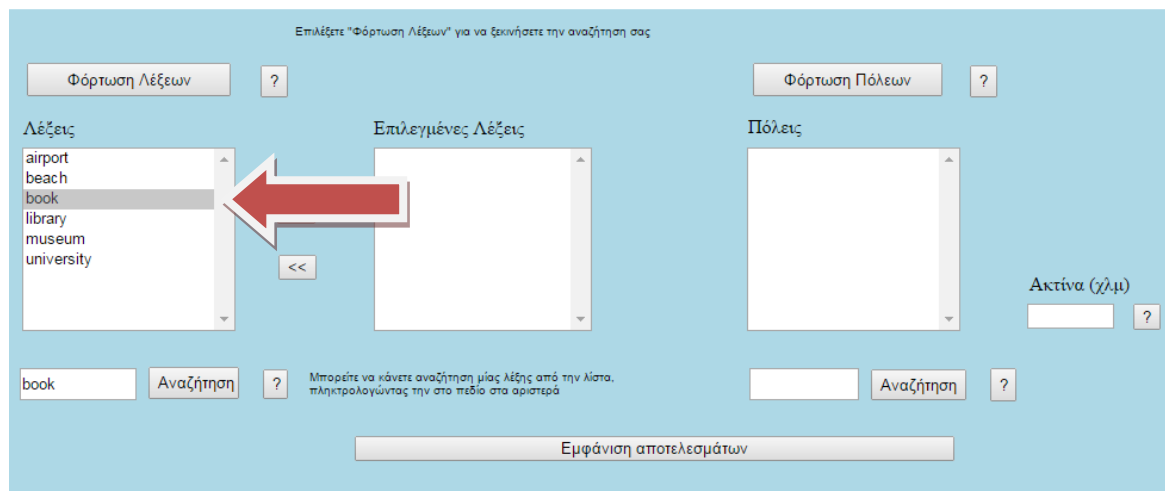
Αναζήτηση ?

Μπορείτε να κάνετε αναζήτηση μίας λέξης από την λίστα, πληκτρολογώντας την στο πεδίο στα αριστερά

Αναζήτηση ?

Εμφάνιση αποτελεσμάτων

Πληκτρολογώντας την λέξη στο πεδίο και πατώντας Αναζήτηση, έχει «γκριζαριστεί» η συγκεκριμένη λέξη, όπως φαίνεται και στην εικόνα.



Επιλέξτε "Φόρτωση Λέξεων" για να ξεκινήσετε την αναζήτηση σας

Φόρτωση Λέξεων ? Φόρτωση Πόλεων ?

Λέξεις: airport, beach, book, library, museum, university

Επιλεγμένες Λέξεις

Πόλεις

Ακτίνα (χλμ) ?

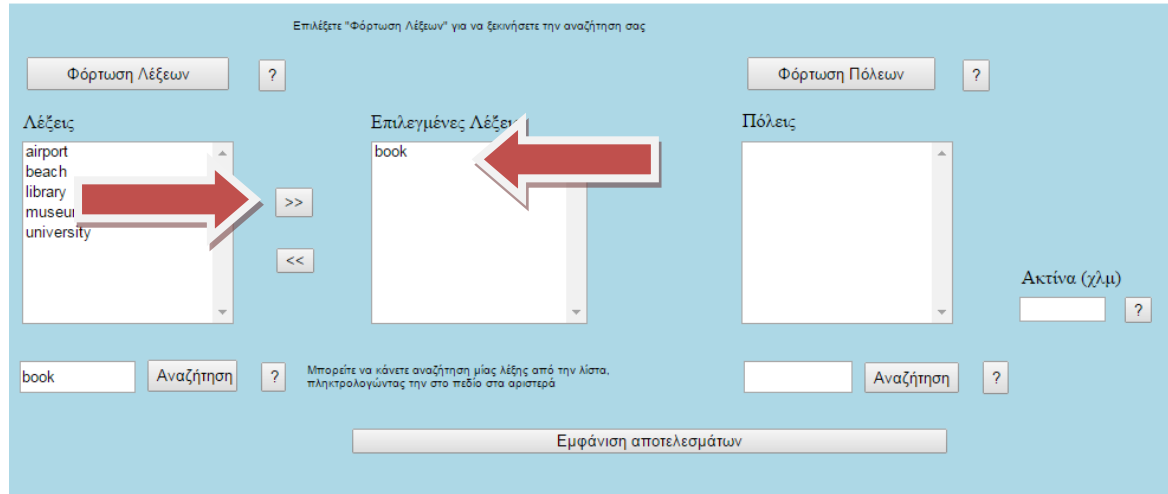
book Αναζήτηση ?

Μπορείτε να κάνετε αναζήτηση μίας λέξης από την λίστα, πληκτρολογώντας την στο πεδίο στα αριστερά

Αναζήτηση ?

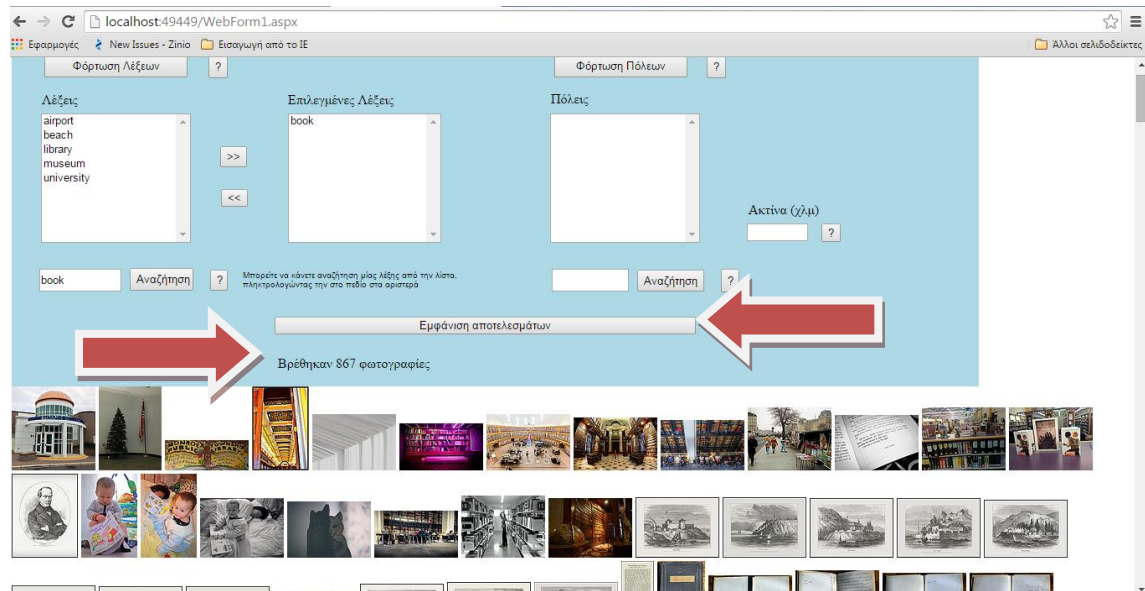
Εμφάνιση αποτελεσμάτων

Σε αυτό το σημείο ο χρήστης, είτε επιλέγοντας την λέξη από την λίστα είτε πληκτρολογώντας την στο πεδίο **Αναζήτηση**, θα πρέπει να μεταφέρει την λέξη που θέλει από το πεδίο **Λέξεις**, στο πεδίο **Επιλεγμένες Λέξεις**, πατώντας το δεξί βέλος όπως φαίνεται και στην εικόνα παρακάτω.



Εδώ ο χρήστης αν θέλει μπορεί να κάνει μια απλή αναζήτηση και να επιλέξει **Εμφάνιση Αποτελεσμάτων**, ώστε να του εμφανιστούν φωτογραφίες που σαν μοναδικό κριτήριο έχουν ως tag την λέξη book.

Πατώντας ο χρήστης λοιπόν στο πεδίο **Εμφάνιση Αποτελεσμάτων**, εμφανίζονται όλες οι φωτογραφίες που υπάρχουν στην βάση της MongoDB, και ακριβώς από κάτω το συνολικό αποτέλεσμα. Πιο συγκεκριμένα για την λέξη “book” αναφέρει «**Βρέθηκαν 867 φωτογραφίες**».

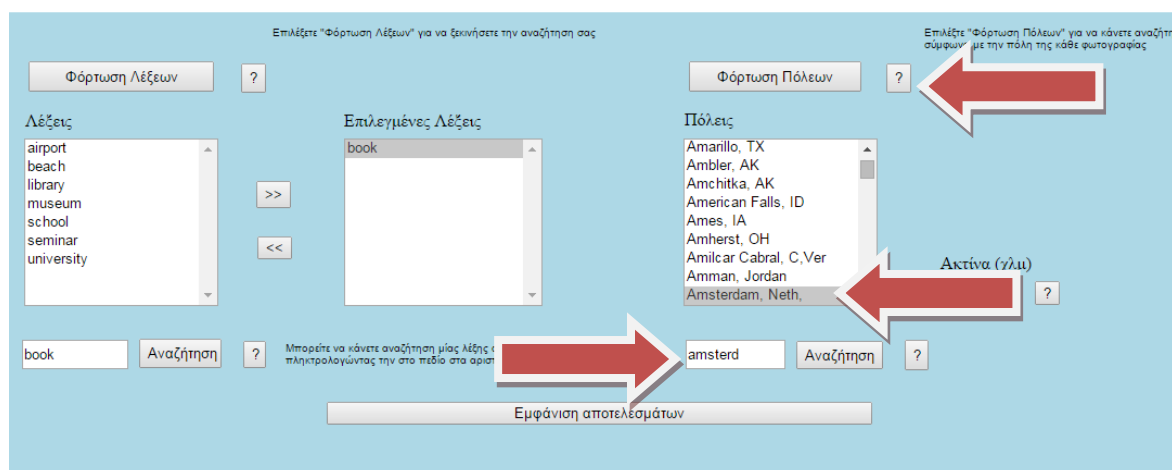


Υποθέτουμε ότι ο χρήστης, θέλει να κάνει μια πιο σύνθετη αναζήτηση, με βάσει την πόλη που έχει τραβηχτεί η κάθε φωτογραφία. Εδώ θα πρέπει να πατήσει το κουμπί **Φόρτωση Πόλεων**, και στο πεδίο **Πόλεις** εμφανίζεται η λίστα με όλες τις πόλεις με αλφαβητική σειρά.

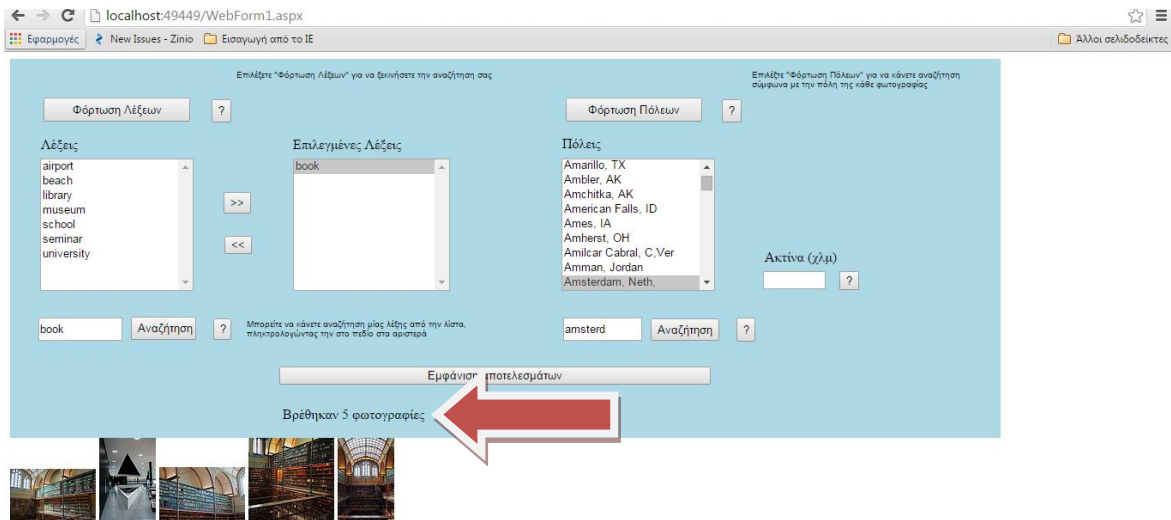


Και εδώ υπάρχει ένα βοηθητικό μήνυμα, δίπλα από το κουμπί **Φόρτωση Πόλεων** που τον ενημερώνει για το τι ακριβώς πρέπει να κάνει.

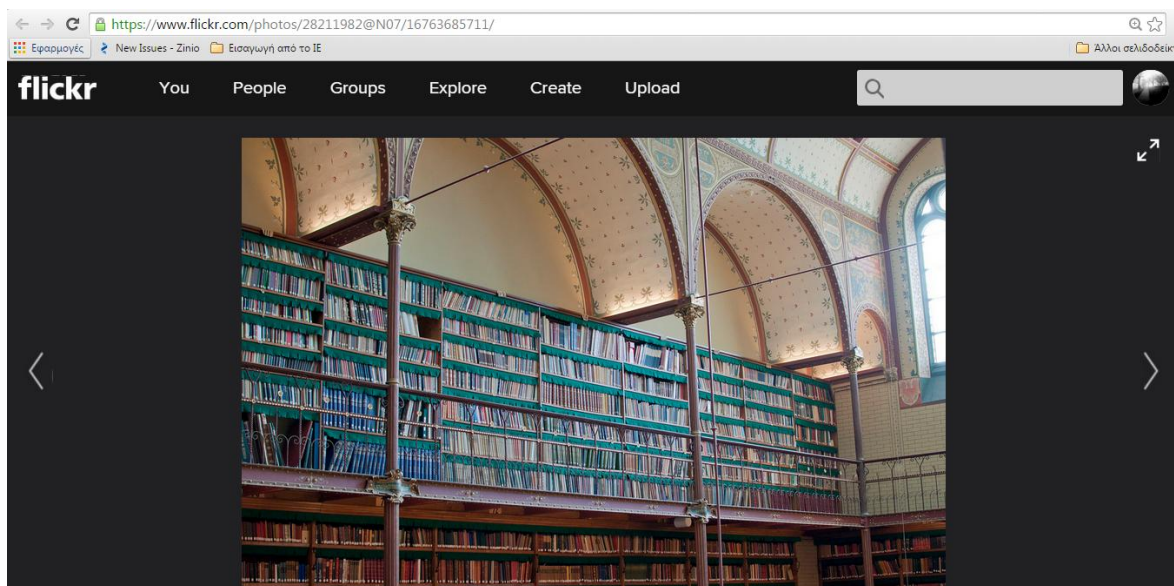
Ομοίως όπως και πριν, για ευχέρεια χρόνου και ευκολία, ο χρήστης μπορεί να πληκτρολογήσει την πόλη που τον ενδιαφέρει, στο πεδίο που βρίσκεται κάτω από την λίστα με τις πόλεις, και να πατήσει το κουμπί **Αναζήτηση**. Και όπως φαίνεται και παρακάτω, η πόλη Άμστερνταμ έχει «γκριζαριστεί».



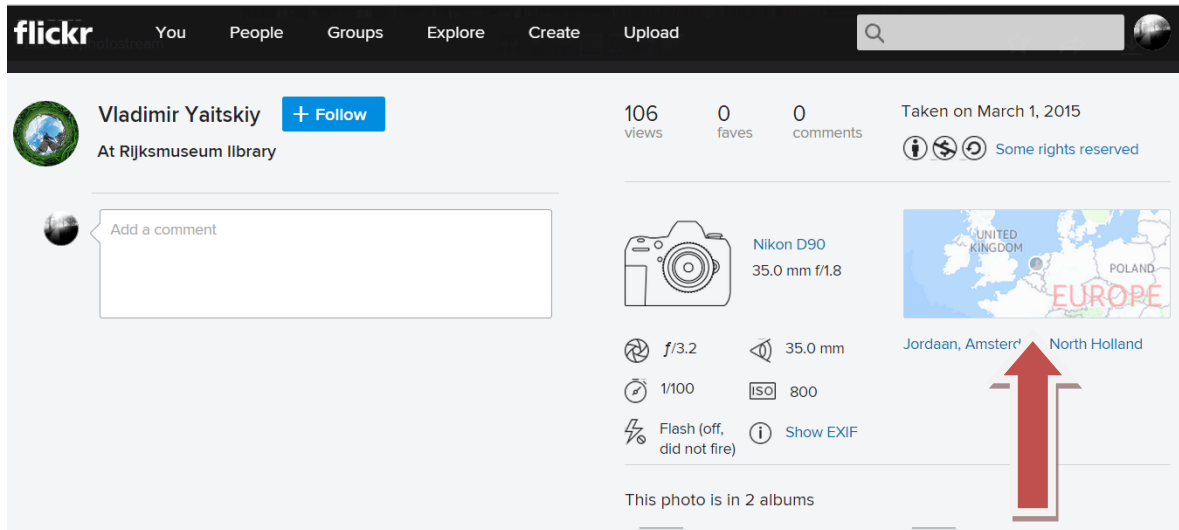
Πατώντας το κουμπί **Εμφάνιση Αποτελεσμάτων**, εμφανίζεται το μήνυμα **Βρέθηκαν 5 φωτογραφίες**. Άρα, στο Flickr, έχουν ανέβει 5 φωτογραφίες που έχουν σαν tag την λέξη book, και έχουν τραβηχτεί στην πόλη του Άμστερνταμ.



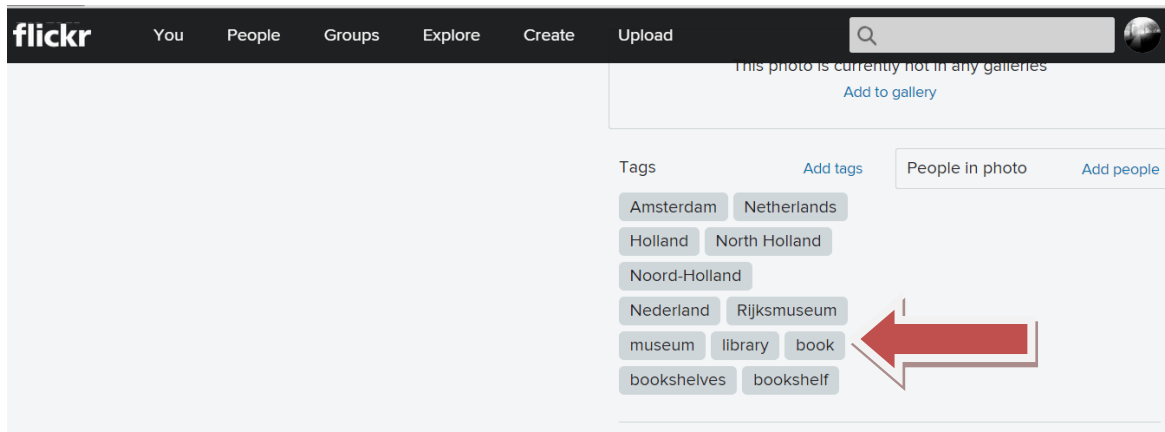
Αν για παράδειγμα ο χρήστης θέλει να δει συγκεκριμένα μία φωτογραφία, θα πατήσει πάνω της και αμέσως θα ανακατευθυνθεί στο url της στο Flickr. Εκεί μπορεί να δει διάφορες πληροφορίες.



Αν περιηγηθεί λίγο στην σελίδα θα δει στον χάρτη, το σημείο που έχει τραβηχτεί η φωτογραφία. Όπως φαίνεται και στην εικόνα, η φωτογραφία σύμφωνα και με την αναζήτηση του χρήστη, βρίσκεται στην Ολλανδία, στο Άμστερνταμ.



Και παρακάτω στην σελίδα του Flickr, βλέπουμε στο πεδίο με τα Tags, την λέξη που αναζήτησε ο χρήστης, την λέξη book.



Ο χρήστης μπορεί αν θέλει να κάνει μια ακόμα πιο σύνθετη αναζήτηση. Πατώντας στο βοηθητικό μήνυμα, βλέπει «Πληκτρολογήστε την απόσταση σε χιλιόμετρα που θέλετε να έχει η φωτογραφία από την επιλεγμένη πόλη». Πιο συγκεκριμένα ο χρήστης μπορεί να αναζητήσει μια φωτογραφία σύμφωνα με την ακτίνα που έχει από το κέντρο της πόλης.

Επιλέξτε "Φόρτωση Λέξεων" για να ξεκινήσετε την αναζήτηση σας

Επιλέξτε "Φόρτωση Πόλεων" για να κάνετε αναζήτηση σύμφωνα με την πόλη της κάθε φωτογραφίας

Φόρτωση Λέξεων ?

Φόρτωση Πόλεων ?

Λέξεις

- airport
- beach
- library
- museum
- school
- seminar
- university

Επιλεγμένες Λέξεις

- book

Πόλεις

- Amarillo, TX
- Ambler, AK
- Amchitka, AK
- American Falls, ID
- Ames, IA
- Amherst, OH
- Amilcar Cabral, C.Ver
- Amman, Jordan
- Amsterdam, Neth.

Πληκτρολογήστε την απόσταση σε χιλμ που θέλετε να έχει η φωτογραφία από την επιλεγμένη πόλη

Ακτίνα (χλμ)

2 ?

book Αναζήτηση ? Μπορείτε να κάνετε αναζήτηση μίας λέξης από την λίστα, πληκτρολογώντας την στο πεδίο στα αριστερά.

amsterd Αναζήτηση ?

Εμφάνιση αποτελεσμάτων

Βρέθηκαν 5 φωτογραφίες

Αν για παράδειγμα θέλει να κάνει την εξής αναζήτηση: «Φωτογραφίες που έχουν σαν tag την λέξη book, σε απόσταση 2 χλμ από το κέντρο του Άμστερνταμ». Τότε στο πεδίο **Ακτίνα**, θα πληκτρολογήσει τον αριθμό 2. Τα τελικά αποτελέσματα μειώθηκαν από 5 σε 3 φωτογραφίες.

Επιλέξτε "Φόρτωση Λέξεων" για να ξεκινήσετε την αναζήτηση σας

Επιλέξτε "Φόρτωση Πόλεων" για να κάνετε αναζήτηση σύμφωνα με την πόλη της κάθε φωτογραφίας

Φόρτωση Λέξεων ?

Φόρτωση Πόλεων ?

Λέξεις

- airport
- beach
- library
- museum
- school
- seminar
- university

Επιλεγμένες Λέξεις

- book

Πόλεις

- Amarillo, TX
- Ambler, AK
- Amchitka, AK
- American Falls, ID
- Ames, IA
- Amherst, OH
- Amilcar Cabral, C.Ver
- Amman, Jordan
- Amsterdam, Neth.

Πληκτρολογήστε την απόσταση σε χιλμ που θέλετε να έχει η φωτογραφία από την επιλεγμένη πόλη

Ακτίνα (χλμ)

2 ?

book Αναζήτηση ? Μπορείτε να κάνετε αναζήτηση μίας λέξης από την λίστα, πληκτρολογώντας την στο πεδίο στα αριστερά.

amsterd Αναζήτηση ?

Εμφάνιση αποτελεσμάτων

Βρέθηκαν 3 φωτογραφίες

Οι πληροφορίες στην βάση για τις πόλεις περιλαμβάνουν τις συντεταγμένες (latitude, longitude). Αυτό σημαίνει ότι κάθε πόλη είναι αντιστοιχισμένη με τις αντίστοιχες συντεταγμένες.

Κεφάλαιο 8

Συμπεράσματα

Τα tags των εικόνων που έχουν δημοσιευτεί στο διαδίκτυο είναι μια καθαρά υποκειμενική πράξη καθώς με αυτόν τον τρόπο οι χρήστες συνεισφέρουν με τον τρόπο τους σε μία τεράστια γνωσιακή βάση δεδομένων. Η κοινωνική σήμανση αντικειμένων ευνοεί την ανάπτυξη κοινοτήτων, οι οποίες αποτελούνται από άτομα με κοινά ενδιαφέροντα και απόψεις, μέσα από τις οποίες παρέχονται πληροφορίες σχετικά με διάφορες περιοχές ενδιαφέροντος. Είναι ένας νέος επαναστατικός τρόπος με τον οποίο σκέφτονται οι χρήστες και δίνει μία νέα οπτική γωνία στον τρόπο με τον οποίο διαμορφώνονται οι ανάγκες πληροφόρησης που έχουν οι χρήστες και στο τρόπο με τον οποίο οι χρήστες επιλέγουν να μοιραστούν την πληροφορία τους με άλλους. Γενικότερα έχουν προκύψει σημαντικά ερωτήματα σχετικά με τη χρησιμότητα και την αντικειμενικότητα των συστημάτων κοινωνικής σήμανσης. Τα ερωτήματα αυτά επικεντρώνονται στη χρήση των συστημάτων αυτών για την ανακάλυψη πόρων καθώς και για την οργάνωση γνώσης που είναι γενικότερα ένας σχετικά αφηρημένος όρος. Οι πληροφορίες οι οποίες προέρχονται από τους χρήστες υπάρχει περίπτωση να μπορούν να ερμηνευθούν με πολλούς διαφορετικούς τρόπους καθώς έχουν ένα καθαρά υποκειμενικό χαρακτήρα. Αυτό συμβαίνει διότι στα περισσότερα συστήματα τέτοιου είδους, δίνεται στους χρήστες η δυνατότητα να επιλέξουν ελεύθερα τις ετικέτες που θα χρησιμοποιήσουν για να χαρακτηρίσουν τα αντικείμενα χωρίς να υπάρχει κάποιου είδους έλεγχος.

Ένα άλλο ζήτημα που προκύπτει στα μέσα κοινωνικής δικτύωσης είναι η γεωγραφική τοποθέτηση των αντικειμένων που χαρακτηρίζονται με ετικέτες. Λόγω της αυξημένης χρήσης των ψηφιακών φωτογραφιών τα τελευταία χρόνια υπάρχουν πολλοί χρήστες οι οποίοι είναι διατεθειμένοι να μοιραστούν τις φωτογραφίες τους και γενικότερα τις εμπειρίες τους μέσα από συστήματα, όπως είναι το Flickr. Το Flickr περιλαμβάνει δισεκατομμύρια φωτογραφίες και γι' αυτό το λόγο η αναζήτηση των χρηστών με βάση κάποια αντικείμενα ενδιαφέροντος τους, δεν είναι πάντα κάτι απλό. Εντούτοις έχει διαπιστωθεί ότι η επιστροφή αποτελεσμάτων από το Flickr δεν επιστρέφει πάντα ικανοποιητικά αποτελέσματα. Είτε πρόκειται για απλές, είτε πρόκειται για σύνθετες αναζητήσεις.

Γενικότερα η εφαρμογή παρόμοιων τεχνικών σε άλλους τομείς εκτός των τοποθεσιών και γενικότερα των χαρακτηριστικών που καθορίζονται από γεωγραφικές συντεταγμένες μπορούν να συμβάλλουν στη βελτίωση της γνώσης που έχουμε για τον κόσμο που μας περιβάλλει.

Όπως αναφέρθηκε και νωρίτερα το Flickr διαθέτει βάση δισεκατομμυρίων φωτογραφιών. Σκοπός λοιπόν αυτής της εργασίας είναι να εξάγει αυτήν την «τεράστια» πληροφορία και μέσω της εφαρμογής που υλοποιήθηκε να μπορέσει να

κάνει αναζήτηση μιας φωτογραφίας με βάση την ετικέτας της (tag), και την γεωγραφική της θέση. Αυτό που μας ενδιέφερε, ως προς την υλοποίηση, είναι πως θα καταφέρουμε να εξάγουμε αυτήν την πληροφορία σε σχετικά γρήγορο χρόνο. Στη συγκεκριμένη εργασία αποφασίσαμε να υλοποιήσουμε μια διαφορετική προσέγγιση, δηλαδή να μην κάνουμε χρήση της παραδοσιακής σχεσιακής βάσης δεδομένων αλλά μιας NoSQL βάσης, της MongoDB. Αν και στο αρχικό μας σχέδιο, υπήρξε και εφαρμόστηκε η Lucene. Στην πορεία όμως της εργασίας και μετά από ενδελεχή έρευνα, αποδείχθηκε ότι η Lucene, αν και αρκετά χρήσιμη στον τομέα της, δηλαδή εξαγωγής μεγάλου όγκου δεδομένων, είναι προτιμότερο να χρησιμοποιείται στην εξαγωγή εγγράφων, πχ pdf, word κλπ. Και όχι στην εξαγωγή φωτογραφιών, γιατί ο χρόνος που απαιτείται είναι περισσότερος από ότι με την χρήση της MongoDB. Στο πρώτο μέρος της εφαρμογής ο διαχειριστής κατεβάζει φωτογραφίες από το Flickr, οι οποίες αποθηκεύονται στην βάση της MongoDB. Υπήρξαν περιπτώσεις όπου στις αναζητήσεις, σε μικρό χρονικό διάστημα κατέβηκε μεγάλος όγκος φωτογραφιών, και άλλες περιπτώσεις όπου για τον ίδιο αριθμό φωτογραφιών χρειάστηκε παραπάνω ή και διπλάσιος χρόνος. Αυτό εξαρτάται καθαρά από το πόσο δημοφιλές είναι ένα tag, και επομένως από το πόσο μεγάλος όγκος φωτογραφιών για αυτήν την λέξη υπάρχει στο Flickr. Πιο συγκεκριμένα λέξεις όπως “university” ή “sun”, είναι λέξεις με τις οποίες κατά την αναζήτηση θα κατέβει μεγάλος όγκος φωτογραφιών. Λέξεις όμως όπως “door”, πόρτα, ή “socks“, κάλτσες, είναι φωτογραφίες που στον ίδιο χρόνο, το συνολικό αποτέλεσμα των φωτογραφιών που θα κατέβει από το Flickr τελικά θα είναι πολύ μικρότερο. Στην συνέχεια από την εφαρμογή του χρήστη, μπορεί κανείς να κάνει αναζήτηση μια φωτογραφίας από αυτές που υπάρχουν στην βάση της MongoDB, και να επιλέξει αν πρόκειται για απλή αναζήτηση ή για πιο σύνθετη, δηλαδή να επιλέξει στην αναζήτηση που θα κάνει, η φωτογραφία να έχει γεωγραφικό περιορισμό. Για παράδειγμα φωτογραφίες που έχουν ως tag, την λέξη “museum”, και βρίσκονται στο Παρίσι, ή φωτογραφίες που έχουν tag την λέξη “university” και βρίσκονται στο Λονδίνο. Επί τω πλείστων τα αποτελέσματα ανταποκρίνονται ορθά στα ερωτήματα μας. Υπάρχουν όμως και περιπτώσεις όπου τα αποτελέσματα είναι παραπλανητικά. Για παράδειγμα κάνουμε αναζήτηση φωτογραφιών με tag την λέξη “museum”. Παρατηρούμε ότι σε κάποιες περιπτώσεις, υπάρχουν χρήστες όπου μπορεί να έχουν ανεβάσει μία φωτογραφία και να της έχουν δώσει το tag “museum”, αλλά στην πραγματικότητα, το περιεχόμενο της φωτογραφίας να μην έχει καμία σχέση με μουσείο, αλλά κάτι εντελώς διαφορετικό. Με αυτό τον τρόπο η ποιότητα των πληροφοριών που παρέχεται από αυτού του είδους τους ιστοτόπους, δεν ανταποκρίνεται πάντα στην πραγματικότητα και μπορεί να είναι ελαφρώς παραπλανητική κάποιες φορές.

Παρ όλα αυτά, το σενάριο το οποίο συνδυάζει τις γεωγραφικούς προορισμούς, όπως μια πόλη με χαρακτηριστικές ετικέτες που την αντιπροσωπεύουν μπορεί να φανεί ιδιαίτερος χρήσιμο για τουριστικούς σκοπούς. Μπορεί να συμβάλλει έτσι ώστε

κάποιος τουρίστας ή γενικότερα κάποιος χρήστη ο οποίος δεν είναι εξοικειωμένος με τη συγκεκριμένη τοποθεσία να ενημερωθεί μέσω φωτογραφιών για το συγκεκριμένο μέρος. Θα ήταν ιδιαίτερα χρήσιμο σε ότι αφορά, σε σημαντικά μνημεία τα οποία κάποιος θεωρεί ως απαραίτητο προορισμό κατά τη διάρκεια ενός ταξιδιού σε κάποια χώρα του εξωτερικού ή της χώρας του. Το σύστημα επιστρέφει σχετικές φωτογραφίες καθώς χρησιμοποιεί και τις γεωγραφικές συντεταγμένες εκτός από τις ετικέτες που χαρακτηρίζουν τις αποθηκευμένες φωτογραφίες. Επίσης μια ακόμα ενδιαφέρουσα πλευρά της εφαρμογής, είναι ότι με τη βοήθεια του γεωγραφικού μήκους και πλάτους επιστρέφονται σαν αποτέλεσμα του ερωτήματος όλες οι τοποθεσίες οι οποίες απέχουν μέχρι μια συγκεκριμένη απόσταση από την αρχική τοποθεσία. Για παράδειγμα, βρες μου φωτογραφίες με tag την λέξη, “library”, οι οποίες απέχουν 10 χιλιόμετρα από το Βερολίνο. Το συγκεκριμένο σενάριο εξυπηρετεί ένα συγκεκριμένο σκοπό. Αν κάποιος χρήστης ενδιαφέρεται για ένα συγκεκριμένο μέρος ή μνημείο μπορεί με αυτό τον τρόπο να ενημερωθεί για τα κοντινά μέρη που υπάρχουν σε αυτόν. Το θετικό σε αυτό είναι ότι ο χρήστης δεν ενημερώνεται μόνο ονομαστικά για τα μέρη αλλά και οπτικά καθώς βλέπει φωτογραφίες που υπάρχουν στο Flickr και αφορούν το συγκεκριμένο μέρος ή το αντικείμενο που τον ενδιαφέρει. Αυτό θα μπορούσε να εφαρμοστεί κάλλιστα σε μια τουριστική εφαρμογή καθώς ο χρήστης θα μπορούσε να δει πόσα ενδιαφέροντα πράγματα υπάρχουν κοντά του, και με κριτήριο τις φωτογραφίες που επιστρέφονται να αποφασίσει ποια από τα κοντινά του μέρη ή αξιοθέατα αξίζει να επισκεφτεί ή πιο γενικά να μπορέσει να δει τις επιλογές του όταν βρίσκεται σε ένα συγκεκριμένο μέρος.

Βιβλιογραφία

- http://news.cnet.com/8301-17939_109-10303576-2.html
- <http://mashable.com/2009/05/02/flickr-search-tools/>
- <http://www.flickr.com/services/api/>
- <http://www.webopedia.com/TERM/A/API.html>
- <http://kylerush.net/blog/flickr-api/>
- <http://www.slideshare.net>
- <http://el.wikipedia.org/>
- <http://sais.aisnet.org/>
- <http://www.epcc.ed.ac.uk/>
- <https://www.nuget.org/packages/mongocsharpdriver>
- <http://docs.mongodb.org/manual/>
- <http://www.edureka.co/blog/real-world-use-cases-of-mongodb/>
- <http://docs.mongodb.org/ecosystem/tutorial/use-csharp-driver/#csharp-driver-tutorial>
- <http://taapps-javasamples.blogspot.gr/2008/01/javaflickrjflickr-java-apiflickrj.html>
- <http://kylerush.net/blog/flickr-api/>
- <http://stevendkay.wordpress.com/tag/flickr-api/>
- <http://stevendkay.wordpress.com/2009/09/25/using-flickr-api-to-get-the-views-faves-and-comments-of-your-most-popular-images/>
- <http://flickrj.sourceforge.net/api/com/aetrion/flickr/photos/PhotosInterface.html>
- <http://net.tutsplus.com/tutorials/javascript-ajax/the-ultimate-guide-to-decoding-the-flickr-api/>
- http://www.winstonprakash.com/articles/netbeans/flickr_enabled_webapp.html
- <http://www.flickr.com/services/api/auth.howto.web.html>
- <https://github.com/callmeal/Flickr4Java>
- <http://www.web-development-blog.com/archives/search-for-photos-using-php-and-the-flickr-api/>
- <http://developer.yahoo.com/java/howto-flickrAuth.html#login>

<http://dbpedialite.org/>

<http://stackoverflow.com/questions/6474427/using-flickr-api-to-search-photos-by-location-or-language>

<http://www.slideshare.net>

<http://hectorcorrea.com/blog/flickr-library-in-c-sharp>

<http://wiki.answers.com>

<http://www.bestopensource.com>

<https://lucenenet.apache.org/>

<https://ropensci.org>

<http://www.scaledb.com/big-data.php>

<http://bicortex.com>

<http://www.dupuis.me/node/25>

<https://www.bizcloudnetwork.com/acid-and-nosql/>

<https://epal-b-sdbd-web.wikispaces.com>

<http://artemis.cslab.ntua.gr/>

<http://blog.mgm-tp.com/2011/03/non-standard-ways-of-using-lucene/>

Library & Information Science Research: Stvilia, B., Jörgensen, C. (2009).

User-generated collection level metadata in an online photo-sharing system. Library & Information Science Research.

How Flickr Helps us Make Sense of the World: Context and Content in Community Contributed Media Collections

Lyndon Kennedy, Mor Naaman, Shane Ahern, Rahul Nair, Tye Rattenbury
Berkeley, CA, USA

Georeferencing Flickr resources based on textual meta-data

Olivier Van Laere, Steven Schockaert, Bart Dhoedt
Department of Information Technology, Ghent University, IBBT, Belgium

Text-Based and Content-Based Image Retrieval on Flickr: DEMO

Juan Manuel Barrios, Diego Díaz-Espinoza, Benjamin Bustos
SISAP, 2009, Similarity Search and Applications, International Workshop, Prague, Czech Republic

Geo-Location Estimation of Flickr Images: Social Web Based Enrichment

Claudia Hauff, Geert-Jan Houben

22. WIS, Delft University of Technology, Delft, The Netherlands