Στα πλαίσια του μαθήματος της πληροφορικής είχαμε μια συζήτηση για τον τρόπο λειτουργίας της μηχανής αναζήτησης Google. Μερικές από τις απορίες μας ήταν οι ακόλουθες:

* Πως βρίσκει το Google τις σελίδες μέσα στον παγκόσμιο ιστό (World Wide Web);
* Που τις αποθηκεύει;
* Πως συσχετίζει τις σελίδες με τις λέξεις κλειδιά;

Διαβάζοντας και ρωτώντας ανακαλύψαμε τις παρακάτω απαντήσεις:

* Το Google έχει ένα πρόγραμμα που ονομάζεται Googlebot το οποίο «σέρνεται» μέσα στον παγκόσμιο ιστό και βρίσκει οποιοδήποτε καινούριο έγγραφο έχει δημοσιευτεί. Ανακαλύψαμε επίσης ότι ο δημιουργός μιας ιστοσελίδας μπορεί αν θέλει να απαγορεύσει την πρόσβαση της Google στην ιστοσελίδα του προσθέτοντας απλά στον κώδικα της σελίδας <meta name="Googlebot" content="nofollow" />. Όταν το Googlebot βρει μια νέα ιστοσελίδα την αντιγράφει και τη στέλνει στην Google
* Η αποθήκευση των σελίδων γίνεται σε μια «φάρμα» υπολογιστών που έχει δημιουργήσει η Google χρησιμοποιώντας εκατομμύρια απλά personal computers. Σύμφωνα με υπολογισμούς η Google στις αρχές του 2013 θα χρησιμοποιεί 2376640 servers, ώστε να μπορεί να αποθηκεύσει 48 δισεκατομμύρια σελίδων.



Η Google «διαβάζει» την κάθε σελίδα αυτόματα και της αναθέτει λέξεις κλειδιά με βάση το περιεχόμενο της. Μελετώντας την ιστορία του διαδικτύου διαπιστώσαμε ότι η Google ήταν από τις πρώτες εταιρίες (αν όχι η πρώτη) που το έκανε αυτό. Προηγουμένως οι μηχανές αναζήτησης είχαν εξειδικευμένο προσωπικό που διάβαζε τη σελίδα και έκανε την καταχώρηση των κλειδιών μηχανικά.

Εμπνευσμένοι από την ιστορία της Google θελήσαμε να κάνουμε τη δική μας «μηχανή αναζήτησης». Σκοπός μας δεν είναι να ξαναανακαλύψουμε την Google. Δεν θα μπορούσαμε άλλωστε. Ανακαλύψαμε όμως ένα τομέα στον οποίο θα μπορούσαμε να εφαρμόσουμε τεχνικές των μηχανών αναζήτησης χρησιμοποιώντας παράλληλα καινοτόμα στοιχεία. Ο τομέας αυτός ήταν οι Ειδήσεις (News). Οι μηχανές αναζήτησης έχουν την επιλογή αναζήτησης ειδήσεων χωρίς όμως να μπορείς να επιλέγεις εσύ τις πηγές. Στη πορεία ανακαλύψαμε ότι αυτό μπορούσε να γίνει με άλλες μεθόδους όπως με το σύστημα RSS. Σε αυτό το σύστημα όμως δεν μπορείς να κάνεις αναζήτηση ειδήσεων. Αποφασίσαμε να «παντρέψουμε» τα δύο συστήματα και να κάνουμε ένα πρόγραμμα στο οποίο θα μπορούσες να κάνεις επιλογή των πηγών σου αλλά και αναζήτηση των ειδήσεων. Ένα επιπλέον χαρακτηριστικό του προγράμματος μας θα είναι ότι ο χρήστης θα έχει τη δυνατότητα να το χρησιμοποιεί ακόμα και offline αφού οι ειδήσεις θα αποθηκεύονται στον υπολογιστή του.

Τέλος αποφασίσαμε ότι ο καταλληλότερος χώρος για την εγκατάσταση του προγράμματος μας θα ήταν η βιβλιοθήκη του σχολείου μας, όπου υπάρχει και η κατάλληλη τεχνική υποδομή.

Όπως συμβαίνει με τη δημιουργία οποιουδήποτε λογισμικού, έτσι και στην περίπτωσή μας ήταν απαραίτητη η ανάλυση και η σχεδίαση του συστήματος πριν από την έναρξη της υλοποίησης.

Προκειμένου να εξακριβώσουμε τις πραγματικές ανάγκες των μελλοντικών χρηστών του λογισμικού μας μιλήσαμε μαζί τους. Πήραμε συνεντεύξεις από 3 καθηγητές και 3 μαθητές και τους ρωτήσαμε τι λειτουργίες θεωρούν ότι θα έπρεπε να έχει ένα πρόγραμμα σαν το δικό μας. Οι απαντήσεις που πήραμε συνοψίζονται στα πιο κάτω σημείο:

* Να μπορεί να γίνεται αναζήτηση με λέξεις κλειδιά, αλλά και με βάση με βάση την ημερομηνία
* Να υπάρχει επιλογή των πηγών
* Να υπάρχει δυνατότητα εκτύπωσης του άρθρου
* Να υπάρχει ταξινόμηση των αποτελεσμάτων με βάση την ημερομηνία

Έχοντας υπόψη τις απαντήσεις που πήραμε μαζί με τα στοιχεία που συλλέξαμε μελετώντας το Google News, το CY Today News και το σύστημα RSS αποφασίσαμε να θέσουμε τις πιο κάτω προδιαγραφές. Πρέπει να αναφέρουμε ότι οι προδιαγραφές που θέσαμε είναι μέσα στα όρια του εφικτού για εμάς.

**Προδιαγραφές**

Το λογισμικό θα πρέπει να:

* Είναι εύκολο στην χρήση του
* Κατεβάζει αυτόματα τις ειδήσεις από επιλεγμένες πηγές
* Αποθηκεύει τα άρθρα στον υπολογιστή για off line χρήση
* Διαβάζει αυτόματα μια είδηση και να της αναθέτει λέξεις κλειδιά
* Μπορεί να κάνει αναζήτηση με βάση λέξεις κλειδιά αλλά και με τη περίοδο έκδοσης των άρθρου
* Να είναι δυνατή και εύκολη μελλοντική επέκταση.

Στη συνέχεια προχωρήσαμε στη σχεδίαση του συστήματος. Στη φάση της σχεδίασης εξετάσαμε θέματα που αφορούν την βάση δεδομένων αλλά και τις οθόνες που θα βλέπει ο χρήστης. Όσο αφορά τη βάση δεδομένων αποφασίσαμε να χρησιμοποιήσουμε την βάση δεδομένων Microsoft Access γιατί είναι διαθέσιμη στο σχολείο μας αλλά και γιατί γνωρίζουμε τη χρήση της. Για κάθε πηγή θα έχουμε 2 πίνακες που θα έχουν την πιο κάτω μορφή.

|  |  |  |
| --- | --- | --- |
| **Όνομα Πεδίου** | **Τύπος Δεδομένων** | **Περιγραφή** |
| Id | Text | Ο κωδικός του άρθρου |
| Title | Text | Ο τίτλος του άρθρου |
| Adate | Date/Time | Η ημερομηνία έκδοσης του άρθρου |
| Category | Text | Η κατηγορία στην οποία ανήκει η είδηση |

Το πρωτεύον κλειδί του πίνακα είναι ο κωδικός του άρθρου(id). Σκοπός του πίνακα είναι να καταχωρεί τον κωδικό, τον τίτλο, την ημερομηνία και την κατηγορία του άρθρου.

|  |  |  |
| --- | --- | --- |
| **Όνομα Πεδίου** | **Τύπος Δεδομένων** | **Περιγραφή** |
| Id | Text | Ο κωδικός του άρθρου |
| Keyword | Text | Λέξη κλειδί για το άρθρο |

Ως πρωτεύον κλειδί του πίνακα είναι ο συνδυασμός και των δύο πεδίων. Σε αυτόν τον πίνακα καταχωρούμε τις λέξεις κλειδιά για κάθε άρθρο.

Τέλος έχουμε έναν πίνακα με λέξεις οι οποίες πρέπει να εξαιρεθούν από τις λέξεις κλειδιά (π.χ. αντωνυμίες)

|  |  |  |
| --- | --- | --- |
| **Όνομα Πεδίου** | **Τύπος Δεδομένων** | **Περιγραφή** |
| Word | Text | Λέξη που πρέπει να εξαιρεθεί |

Όσο αφορά τις φόρμες αποφασίσαμε ότι θα υπάρχει μια κεντρική φόρμα που θα περιλαμβάνει και τα διάφορα μενού. Η πιο κάτω εικόνα δείχνει αυτή τη φόρμα.



Οι άλλες δύο φόρμες που παρουσιάζουμε παρακάτω αφορούν το κατέβασμα των πληροφοριών καθώς και την αναζήτηση.



