



ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ
ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

**Σχεδίαση και ανάπτυξη ιστοσελίδας για την διαχείριση
ημιδομημένων δεδομένων**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

Χουλιάρη Ιωάννη

Επιβλέπουσα: Βλάχου Ακριβή, Αν. Καθηγήτρια

Μέλη εξεταστικής επιτροπής:

Καρύδα Μαρία, Καθηγήτρια

Κωστούλας Θεόδωρος, Αν. Καθηγητής

Σάμος, Σεπτέμβριος 2023

Πρόλογος και ευχαριστίες

Ευχαριστώ πολύ την επιβλέπουσα καθηγήτρια κυρία Βλάχου Ακριβή για την ευκαιρία που μου έδωσε ώστε να ασχοληθώ με αυτό το θέμα, καθώς και για την καθοδήγηση που μου παρείχε. Επιπλέον θα ήθελα να ευχαριστήσω την οικογένεια μου και τους φίλους μου που με υποστήριξαν καθ' όλη την διάρκεια των σπουδών μου.

© 2023

του

Χουλιάρα Ιωάννη

Τμήμα Μηχανικών Πληροφοριακών και Επικοινωνιακών Συστημάτων

ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ

Πίνακας περιεχομένων

1	Εισαγωγή	1
1.1	Διαχείριση ημιδομημένων δεδομένων	1
1.2	Αντικείμενο διπλωματικής.....	1
1.3	Περιγραφή προβλήματος	2
1.4	Δομή της διπλωματικής	2
2	Ημιδομημένα Δεδομένα	4
2.1	Χαρακτηριστικά ημιδομημένων δεδομένων	5
2.2	Πλεονεκτήματα ημιδομημένων δεδομένων	6
2.3	Μειονεκτήματα ημιδομημένων δεδομένων	7
2.4	Εξαγωγή δεδομένων από ημιδομημένα δεδομένα	9
2.4.1	<i>Εργαλεία και Τεχνικές Εξαγωγής Δεδομένων από Ημιδομημένα Δεδομένα</i>	<i>9</i>
2.5	Ενοποίηση δεδομένων	10
2.5.1	<i>Πλεονεκτήματα ενοποίησης δεδομένων σε ημιδομημένα δεδομένα</i>	<i>10</i>
2.5.2	<i>Τεχνικές και εργαλεία ενοποίησης δεδομένων.....</i>	<i>11</i>
2.6	Ποιότητα δεδομένων.....	12
2.7	Καθαρισμός Δεδομένων	12
2.8	Τυπική Διαδικασία Διαχείρισης Δεδομένων και Διασφάλισης της Ποιότητάς τους	13
3	Ανάπτυξη Ιστοσελίδων	14
3.1	Δυναμικές Ιστοσελίδες.....	14
3.2	Τεχνολογίες για ανάπτυξη δυναμικών ιστοσελίδων	15
3.2.1	<i>HTML (Hypertext Markup Language).....</i>	<i>15</i>
3.2.2	<i>CSS (Cascading Style Sheets).....</i>	<i>15</i>
3.2.3	<i>JavaScript</i>	<i>15</i>
3.2.4	<i>jQuery</i>	<i>16</i>
3.2.5	<i>AJAX (Asynchronous JavaScript and XML).....</i>	<i>16</i>
3.2.6	<i>PHP (Hypertext Preprocessor).....</i>	<i>16</i>
3.3	Σύστημα Διαχείρισης Περιεχομένου (CMS)	16
3.4	Frameworks – ASP .NET MVC	17
4	Αρχιτεκτονική Συστήματος.....	19
4.1	Αρχιτεκτονική λογισμικού Model-View-Controller (MVC)	19
4.1.1	<i>Model.....</i>	<i>20</i>
4.1.2	<i>View</i>	<i>20</i>

4.1.3	Controller	20
4.2	Σχεδιαστικές αποφάσεις.....	21
4.3	Αποθήκευση και ανάκτηση δεδομένων	21
4.4	Ασφάλεια συστήματος.....	22
5	Εργαλεία Ανάπτυξης Συστήματος	23
5.1	Visual Studio 2019.....	23
5.1.1	LINQ	23
5.1.2	XML Serializer.....	24
5.2	Microsoft SQL Server Management Studio (SSMS) 18.....	24
5.3	Internet Information Services (IIS) Manager (v21H2)	24
5.4	GitHub.....	24
5.4.1	Έλεγχος Εκδόσεων	25
6	Παρουσίαση Ιστοσελίδας Εργαστηρίου - Διαχειριστικό	26
6.1	Διαχειριστικό - Dashboard.....	26
6.1.1	Σύνδεση.....	27
6.1.2	Διαχείριση Μελών	28
6.1.3	Ενημέρωση βιβλιογραφίας μέλους.....	29
6.1.4	Διαχείριση Δημοσιεύσεων	32
6.1.5	Διαχείριση Διπλωματικών Εργασιών.....	34
6.1.6	Αποσύνδεση Διαχειριστή.....	35
6.2	Σελίδα Εργαστηρίου	35
6.2.1	Αρχική Σελίδα	36
6.2.2	Σελίδα Μελών	37
6.2.3	Σελίδα Δημοσιεύσεων.....	38
6.2.4	Σελίδα Διπλωματικών Εργασιών.....	39
7	Συμπεράσματα.....	40
	Βιβλιογραφία.....	41

Λίστα Εικόνων

Εικόνα 1: Η ίδια πληροφορία παρουσιάζεται με 3 διαφορετικούς τύπους δομής	5
Εικόνα 2: Απλοποιημένη αναπαράσταση ενοποίησης πολλαπλών ημιδομημένων δεδομένων	10
Εικόνα 3: Απεικόνιση αρχιτεκτονικής MVC	21
Εικόνα 4: Σελίδα σύνδεσης διαχειριστή στο Dashboard	27
Εικόνα 5: Σελίδα διαχείρισης μελών	28
Εικόνα 6: Σελίδα προσθήκης / επεξεργασίας μέλους	29
Εικόνα 7: Παράδειγμα προσπάθειας να ανέβει XML βιβλιογραφίας χωρίς να έχει αποθηκευτεί το μέλος	29
Εικόνα 8: Σελίδα μέλους στο dblp.org - εξαγωγή βιβλιογραφίας σε XML αρχείο	30
Εικόνα 9: Αρχείο XML βιβλιογραφίας από dblp.org	31
Εικόνα 10: Άνοιγμα filepicker για ανέβασμα αρχείου XML βιβλιογραφίας	31
Εικόνα 11: Επεξεργασία μέλους - Διαγραφή Βιβλιογραφίας	32
Εικόνα 12: Σελίδα διαχείρισης δημοσιεύσεων	32
Εικόνα 13: Παράθυρο επιβεβαίωσης διαγραφής δημοσίευσης	33
Εικόνα 14: Σελίδα διαχείρισης διπλωματικών εργασιών	34
Εικόνα 15: Σελίδα προσθήκης / επεξεργασίας διπλωματικής εργασίας	35
Εικόνα 16: Αποσύνδεση διαχειριστή	35
Εικόνα 17: Αρχική σελίδα εργαστηρίου	36
Εικόνα 18: Σελίδα Μελών	37
Εικόνα 19: Σελίδα Δημοσιεύσεων	38
Εικόνα 20: Σελίδα Διπλωματικών Εργασιών	39

Ακρωνύμια

XML	Extensible Markup Language
JSON	JavaScript Object Notation
CSS	Cascading Style Sheets
HTML	HyperText Markup Language
PHP	Hypertext Preprocessor
AJAX	Asynchronous JavaScript and XML
MVC	Model-view-controller
CMS	Content management system
API	Application Programming Interface
LINQ	Language Integrated Query
IDE	Integrated Development Environment
DOM	Document Object Model
ETL	Extract, Transform, Load

Περίληψη

Όλο και περισσότερο αυξάνεται ο όγκος και η πολυπλοκότητα των δεδομένων που παράγονται σε διάφορους τομείς και έτσι τίθενται σημαντικές προκλήσεις για την αποτελεσματική διαχείριση και οργάνωση τους. Τα ημιδομημένα δεδομένα, που χαρακτηρίζονται από την ευελιξία και την εξελισσόμενη φύση τους, έχουν γίνει διαδεδομένα σε πολλούς τομείς, που κυμαίνονται από τα μέσα κοινωνικής δικτύωσης έως την επιστημονική έρευνα. Βάσει των χαρακτηριστικών τους αλλά και επειδή μπορεί να προέρχονται από πολλές πηγές, τα ημιδομημένα δεδομένα, παρουσιάζουν προκλήσεις στην αποτελεσματική διαχείριση τους. Ως αποτέλεσμα, υπάρχει μια αυξανόμενη ανάγκη για αποτελεσματικά συστήματα διαχείρισης, αποθήκευσης και ανάλυσης ημιδομημένων δεδομένων. Στην παρούσα Διπλωματική γίνεται ο σχεδιασμός και η ανάπτυξη ιστοσελίδας για την αποτελεσματική διαχείριση ημιδομημένων δεδομένων. Ο διαχειριστής της ιστοσελίδας απλά εισάγει αρχεία ημιδομημένων δεδομένων και αυτά αυτόματα επεξεργάζονται και μετατρέπονται σε δομημένη πληροφορία που μπορεί εύκολα να διαχειριστεί.

Λέξεις Κλειδιά: ημιδομημένα δεδομένα, ανάπτυξη ιστοσελίδας, δυναμική ιστοσελίδα, ενοποίηση δεδομένων

Abstract

The ever-increasing volume and complexity of the data produced in various fields poses significant challenges for their effective management and organization. Semi-structured data has become prevalent in various fields, from social media to scientific research, and is distinguished by its adaptability and dynamic nature. Dealing with semi-structured data poses inherent challenges in effective management and the additional factor of various sources providing varying forms of semi-structured data further compounds these issues. The demand for effective systems to manage, store, and analyze semi-structured data is therefore increasing. This thesis presents the design and development of a website which aims to efficiently manage semi-structured data. The website administrator simply uploads the semi-structured data files and they are automatically processed into structured data than can be easily managed.

Keywords: *semi-structured data, web development, dynamic webpage, data integration*

1

Εισαγωγή

1.1 Διαχείριση ημιδομημένων δεδομένων

Στο σημερινό τοπίο που βασίζεται στα δεδομένα, η διαχείριση πληροφοριών είναι πρωταρχικής σημασίας. Τα δεδομένα σπάνια προέρχονται από μία μόνο πηγή. Αντίθετα, οι πληροφορίες προέρχονται από διαφορετικές πηγές, με διάφορες μορφές και δομές. Αυτό δημιουργεί προκλήσεις στην ενοποίηση των δεδομένων και μπορεί να έχει επιπτώσεις στην ποιότητα των δεδομένων. Όση λιγότερη δομή έχουν τα δεδομένα τόσο πιο πολύπλοκες γίνονται και οι τεχνικές βελτίωσης της ποιότητας τους. Παράλληλα κάποια χαρακτηριστικά των ημιδομημένων δεδομένων ενδέχεται να δημιουργήσουν περαιτέρω προκλήσεις, όμως με την σωστή διαχείριση τους και εκμεταλλευόμενοι συγκεκριμένα χαρακτηριστικά τους μπορούμε να διευκολύνουμε την διαδικασία ενοποίησης και να βελτιώσουμε την ποιότητα τους με διαδικασίες καθαρισμού δεδομένων. [1]

1.2 Αντικείμενο διπλωματικής

Το αντικείμενο της διπλωματικής είναι η ανάπτυξη ιστοσελίδας για διαχείριση ημιδομημένων δομημένων. Η διαχείριση ημιδομημένων δεδομένων αφορά την επεξεργασία XML αρχείων, ώστε να εξάγουμε την επιθυμητή πληροφορία με την δομή που θέλουμε, να την παρουσιάσουμε στον χρήστη και να του δώσουμε την δυνατότητα να την διαχειριστεί περαιτέρω.

Θα αναπτυχθεί ιστοσελίδα διαχείρισης - Dashboard όπου εκεί θα γίνεται όλη η διαχείριση της ιστοσελίδας εργαστηρίου καθώς και των ημιδομημένων δεδομένων. Η ιστοσελίδα διαχείρισης στοχεύει να αντιμετωπίσει τις προκλήσεις που σχετίζονται με τη διαχείριση ημιδομημένων δεδομένων. Θα παρέχει στους χρήστες μια διαισθητική και φιλική προς το χρήστη διεπαφή για τη

διαχείριση ημιδομημένων δεδομένων, επιτρέποντάς τους να ανεβάζουν, να αποθηκεύουν και να ανακτούν εύκολα δεδομένα. Επιπλέον θα αναπτυχθεί ιστοσελίδα για το εργαστήριο “Βάσεις Δεδομένων” του Τμήματος Μηχανικών Πληροφοριακών και Επικοινωνιακών Συστημάτων του Πανεπιστημίου Αιγαίου, όπου εκεί θα εμφανίζεται η πληροφορία που έχουμε εξάγει από τα ημιδομημένα δεδομένα.

Συγκεκριμένα η διαχείριση ημιδομημένων δομημένων αφορά την επεξεργασία αρχείων XML, από το dblp.com, που περιέχουν την βιβλιογραφία μελών που έχουν προστεθεί στην ιστοσελίδα του εργαστηρίου “Βάσεις Δεδομένων”. Θα εξάγουμε από αυτά την πληροφορία που θέλουμε και θα συμπληρώνεται αυτόματα η βιβλιογραφία του μέλους.

1.3 Περιγραφή προβλήματος

Η διαχείριση δεδομένων έχει γίνει μια βασική εργασία για οργανισμούς όλων των μεγεθών και τύπων. Με τον αυξανόμενο όγκο δεδομένων που παράγονται καθημερινά, η αποτελεσματική διαχείριση δεδομένων έχει γίνει μια πρόκληση. Μία από τις μεγαλύτερες προκλήσεις στη διαχείριση δεδομένων είναι η διαχείριση ημιδομημένων δεδομένων, τα οποία είναι δεδομένα που δεν συμμορφώνονται με ένα σταθερό σχήμα ή μοντέλο δεδομένων.

Τα ημιδομημένα δεδομένα γίνονται όλο και πιο κοινά στο σημερινό ψηφιακό τοπίο. Χρησιμοποιούνται συνήθως σε διαδικτυακές εφαρμογές, ηλεκτρονικά καταστήματα, APIs, και ηλεκτρονικά έγγραφα. Ωστόσο, η διαχείριση ημιδομημένων δεδομένων μπορεί να είναι δύσκολη λόγω της πολύπλοκης και ποικίλης δομής τους. Για την επεξεργασία και την αξιολόγηση ημιδομημένων δεδομένων, απαιτούνται συχνά εξειδικευμένα εργαλεία. Επιπλέον, εάν δεν υπάρχει σαφής δομή, είναι δύσκολο να ενσωματωθούν τα δεδομένα με άλλα συστήματα. [2]

Η ανάλυση και οι πληροφορίες που παρέχουμε για τα ημιδομημένα δεδομένα καθώς και οι τεχνικές και τα εργαλεία διαχείρισης ημιδομημένων δεδομένων που αναφέρουμε μπορούν να βελτιώσουν το πρόβλημα αυτό. Λόγω των ποικίλων μορφών ημιδομημένων δεδομένων δεν μπορούμε να παρουσιάσουμε μια γενική βέλτιστη λύση για το πως μπορούμε να τα διαχειριστούμε. Θα παρουσιάσουμε μια συγκεκριμένη λύση για το Αντικείμενο Διπλωματικής, τις τεχνικές και τις τεχνολογίες που χρησιμοποιήθηκαν.

1.4 Δομή της διπλωματικής

Κεφάλαιο 1: Στο κεφάλαιο 1 γίνεται μια εισαγωγή στο αντικείμενο της διπλωματικής και στο τι θα υλοποιηθεί για αυτήν καθώς και στο πρόβλημα που προσπαθούμε να επιλύσουμε.

Κεφάλαιο 2: Στο κεφάλαιο 2 γίνεται μια ανάλυση για τα ημιδομημένα δεδομένα, παρουσιάζονται τα χαρακτηριστικά τους, τρόποι διαχείρισης τους.

Κεφάλαιο 3: Στο κεφάλαιο 3 γίνεται μια εισαγωγή στην ανάπτυξη δυναμικών ιστοσελίδων και παρουσιάζονται τεχνολογίες που μας βοηθούν σε αυτό.

Κεφάλαιο 4: Στο κεφάλαιο 4 παρουσιάζεται η αρχιτεκτονική του συστήματος που αναπτύξαμε.

Κεφάλαιο 5: Στο κεφάλαιο 5 παρουσιάζονται οι τεχνολογίες και τα εργαλεία που χρησιμοποιήσαμε για να αναπτύξουμε το σύστημα.

Κεφάλαιο 6: Στο κεφάλαιο 6 γίνεται μια παρουσίαση της ιστοσελίδας διαχείρισης εργαστηρίου και της ιστοσελίδας εργαστηρίου.

Κεφάλαιο 7: Στο κεφάλαιο 7 αναφέρουμε τα συμπεράσματά μας από την εκπόνηση της διπλωματική εργασίας.

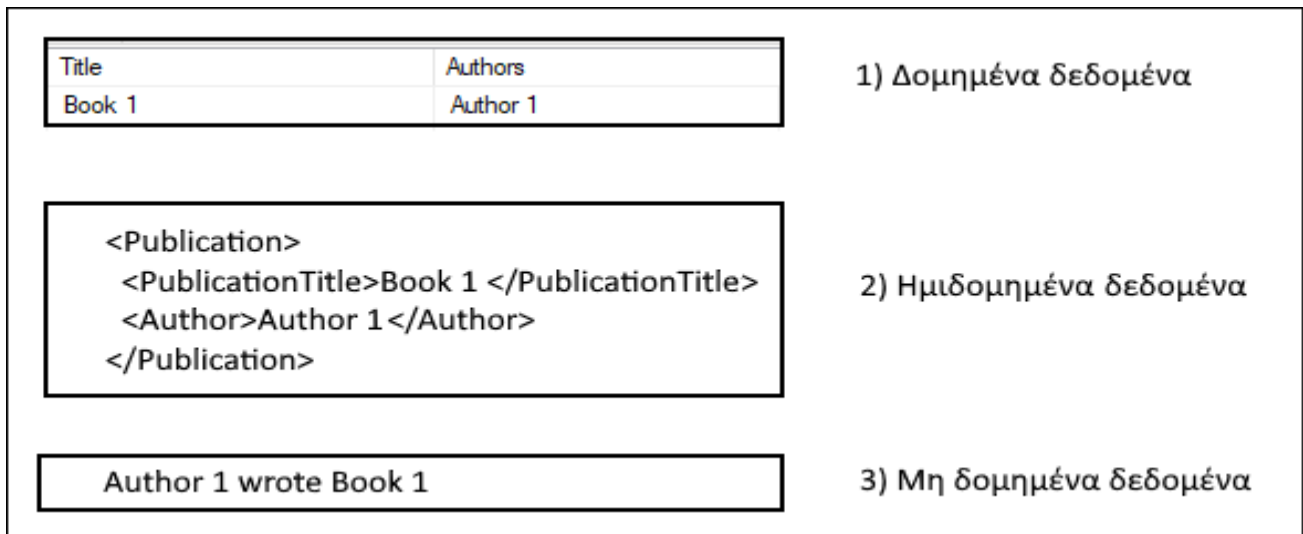
2

Ημιδομημένα Δεδομένα

Τα ημιδομημένα δεδομένα αναφέρονται σε δεδομένα που δεν ακολουθούν μια αυστηρά καθορισμένη μορφή, αλλά εξακολουθούν να έχουν κάποια οργανωτική δομή ή σχέσεις μεταξύ των στοιχείων τους. Χρησιμοποιούνται συχνά για να περιγράψουν δεδομένα που έχουν ένα ευέλικτο σχήμα (schema), πράγμα που σημαίνει ότι η δομή τους δεν είναι σταθερή και μπορεί να αλλάξει. Σε αντίθεση με τα δομημένα δεδομένα που ακολουθούν ένα αυστηρό σχήμα, τα ημιδομημένα δεδομένα επιτρέπουν μεγαλύτερη ευελιξία στην οργάνωση δεδομένων. Η δομή στα ημιδομημένα δεδομένα αναπαρίσταται συνήθως χρησιμοποιώντας ετικέτες, οι οποίες παρέχουν κάποια οργάνωση στα δεδομένα.

Τα πιο γνωστά παραδείγματα ημιδομημένων δεδομένων περιλαμβάνουν τα XML (Extensible Markup Language) και JSON (JavaScript Object Notation). Είναι ευρέως διαδομένη η χρήση τους στην μεταβίβαση πληροφοριών για πολλούς παράγοντες. Η XML είναι ιδιαίτερα δημοφιλής σε πολλούς κλάδους και εύκολα μπορεί κάποιος να αποθηκεύσει και να οργανώσει πολύπλοκες και ιεραρχικές δομές δεδομένων. Το JSON χρησιμοποιείται συνήθως σε εφαρμογές διαδικτύου, καθώς είναι ελαφρύ και εύκολο στην ανάλυση.

Τα ημιδομημένα δεδομένα είναι ένας ευέλικτος και προσαρμόσιμος τρόπος αποθήκευσης και διαχείρισης δεδομένων. Η ιεραρχική τους δομή, το ευέλικτο σχήμα και η έλλειψη κανονικοποίησης τα καθιστούν ιδανικά για τη καταγραφή, διαχείριση και αναπαράσταση δεδομένων από διάφορες πηγές. Συχνά είναι πολύ κοντά στην ανθρώπινη γλώσσα και γίνεται σχετικά εύκολα κατανοητή η δομή τους καθώς και η πληροφορία τους. [2]



Εικόνα 1: Η ίδια πληροφορία παρουσιάζεται με 3 διαφορετικούς τύπους δομής

2.1 Χαρακτηριστικά ημιδομημένων δεδομένων

Μερικά από τα χαρακτηριστικά των ημιδομημένων δεδομένων είναι:

Χαλαρό σχήμα (Loose schema): Τα ημιδομημένα δεδομένα δεν απαιτούν σταθερό σχήμα, επιτρέποντας ευελιξία στην οργάνωση. Αυτό σημαίνει ότι τα δεδομένα μπορούν να οργανωθούν και να αποθηκευτούν με τρόπο που να ταιριάζει καλύτερα στις ανάγκες της εφαρμογής ή του συστήματος. Για παράδειγμα, ένα αντικείμενο JSON μπορεί να περιέχει διαφορετικούς τύπους δεδομένων και η δομή του αντικειμένου μπορεί να αλλάξει ανάλογα με τα δεδομένα που αποθηκεύονται.

Ιεραρχική οργάνωση: Τα ημιδομημένα δεδομένα συχνά ακολουθούν μια ένθετη (nested) ή ιεραρχική οργάνωση. Αυτό επιτρέπει την αναπαράσταση πολύπλοκων σχέσεων και εξαρτήσεων μεταξύ διαφορετικών στοιχείων δεδομένων, κάτι που μπορεί να είναι πιο δύσκολο να επιτευχθεί σε μια δομημένη σχεσιακή βάση δεδομένων.

Αυτοπεριγραφή: Τα ίδια τα δεδομένα συνήθως περιέχουν πληροφορίες σχετικά με τη δομή τους, όπως ονόματα χαρακτηριστικών ή μεταδεδομένα, διευκολύνοντας την κατανόηση του περιεχομένου, της σημασίας και τον τύπο των δεδομένων. Για παράδειγμα τα αρχεία JSON και XML, τα οποία περιέχουν κλειδιά και ετικέτες για την περιγραφή των στοιχείων δεδομένων.

Ετερογένεια: Τα ημιδομημένα δεδομένα μπορούν να αποτελούνται από διαφορετικούς τύπους δεδομένων. Αυτό σημαίνει ότι τα ημιδομημένα δεδομένα μπορούν να προβάλουν ένα ευρύ φάσμα τύπων και δομών, παρέχοντας ευελιξία στην αποθήκευση και επεξεργασία δεδομένων.

Μερικώς δομημένα: Τα ημιδομημένα δεδομένα μπορεί να έχουν κάποια δομή, αλλά αυτή η δομή δεν επιβάλλεται απαραίτητα από ένα προκαθορισμένο σχήμα. Αυτό σημαίνει ότι τα ημιδομημένα δεδομένα μπορούν να είναι μερικώς δομημένα, επιτρέποντας τη συμπερίληψη προαιρετικών ή μεταβλητών στοιχείων, διατηρώντας παράλληλα μια βασική δομή.

Ημιτελή δεδομένα: Τα ημιτελή δεδομένα σε ημιδομημένα δεδομένα αναφέρονται σε δεδομένα από τα οποία λείπουν ένα ή περισσότερα απαιτούμενα στοιχεία, σύμφωνα με το προκαθορισμένο σχήμα ή δομή. Αυτό σημαίνει ότι τα δεδομένα ενδέχεται να μην έχουν όλες τις απαραίτητες

πληροφορίες για χρήση ή επεξεργασία όπως προβλέπεται, γεγονός που μπορεί να περιορίσει τη χρησιμότητά τους.

Διαλειτουργικότητα: Τα ημιδομημένα δεδομένα μπορούν να χρησιμοποιηθούν και να ενσωματωθούν με άλλες μορφές δεδομένων, διευκολύνοντας τη διαλειτουργικότητα μεταξύ διαφορετικών συστημάτων. Για παράδειγμα, ένα αντικείμενο JSON μπορεί εύκολα να μετατραπεί σε άλλες μορφές δεδομένων, όπως XML ή CSV, για να διευκολυνθεί η ανταλλαγή δεδομένων μεταξύ διαφορετικών συστημάτων.

Εξέλιξη σχήματος: Τα ημιδομημένα δεδομένα επιτρέπουν την πραγματοποίηση αλλαγών στο σχήμα με την πάροδο του χρόνου, καθώς εισάγονται νέοι τύποι δεδομένων ή τροποποιούνται οι υπάρχοντες.

2.2 Πλεονεκτήματα ημιδομημένων δεδομένων

Η διαχείριση ημιδομημένων δεδομένων προσφέρει πολλά πλεονεκτήματα έναντι των δομημένων ή μη δομημένων δεδομένων. Παρέχει ευελιξία στην αναπαράσταση δεδομένων, επιτρέποντας την ευκολότερη προσαρμογή σε εξελισσόμενες μορφές και σχήματα δεδομένων. Τα ημιδομημένα δεδομένα επιτρέπουν επίσης την αποτελεσματική αποθήκευση και ανάκτηση πολύπλοκων και ετερογενών πληροφοριών, υποστηρίζοντας την ενοποίηση και ανάλυση δεδομένων.

- **Ευελιξία:** Τα ημιδομημένα δεδομένα επιτρέπουν την αναπαράσταση διάφορων τύπων δεδομένων, μορφών και χαρακτηριστικών σε ένα ενιαίο σύνολο δεδομένων χωρίς αυστηρή τήρηση ενός σταθερού σχήματος. Αυτή η ευελιξία είναι ιδιαίτερα πολύτιμη όταν πρόκειται για δυναμικές πηγές δεδομένων ή για ταχέως μεταβαλλόμενες απαιτήσεις δεδομένων. Για παράδειγμα, στο πλαίσιο των συσκευών IoT ή των δεδομένων μέσω κοινωνικής δικτύωσης, όπου μπορεί να εμφανίζονται συχνά νέοι τύποι πληροφοριών ή χαρακτηριστικών, οι ημιδομημένες μορφές δεδομένων μπορούν να προσαρμοστούν σε αυτές τις αλλαγές χωρίς σημαντική διακοπή.
- **Αναγνωσιμότητα:** Οι ημιδομημένες μορφές δεδομένων, όπως το JSON και το XML, έχουν σχεδιαστεί για να είναι αναγνώσιμες από τον άνθρωπο, με σαφείς, βασισμένες σε κείμενο δομές που μπορούν εύκολα να γίνουν κατανοητές. Αυτή η αναγνωσιμότητα όχι μόνο απλοποιεί την κατανόηση και την ανάλυση δεδομένων, αλλά βοηθά επίσης στον εντοπισμό σφαλμάτων (debugging), την επικύρωση (validation) και την τεκμηρίωση (documentation), καθώς τα δεδομένα μπορούν εύκολα να αναλυθούν και να επεξεργαστούν χρησιμοποιώντας απλούς επεξεργαστές κειμένου ή εργαλεία προγραμματισμού.
- **Αυτοπεριγραφόμενη φύση:** Οι ημιδομημένες μορφές δεδομένων περιλαμβάνουν περιγραφικές ετικέτες, κλειδιά ή μεταδεδομένα που παρέχουν το πλαίσιο και το νόημα για τα στοιχεία δεδομένων. Αυτό το χαρακτηριστικό αυτοπεριγραφής απλοποιεί την ανάλυση, την επεξεργασία και την κατανόηση δεδομένων. Οι προγραμματιστές μπορούν να εντοπίσουν γρήγορα σχετικά στοιχεία δεδομένων, σχέσεις και δομές μέσα στα δεδομένα, ακόμη και όταν διαχειρίζονται ένα σύνολο δεδομένων για πρώτη φορά.
- **Ιεραρχική οργάνωση:** Οι ημιδομημένες μορφές δεδομένων υποστηρίζουν ένθετες ή ιεραρχικές δομές δεδομένων, οι οποίες επιτρέπουν την αναπαράσταση πολύπλοκων

σχέσεων και εξαρτήσεων μεταξύ στοιχείων δεδομένων. Για παράδειγμα, ένα έγγραφο JSON μπορεί να περιέχει ένθετους πίνακες ή αντικείμενα, καθιστώντας δυνατή την αναπαράσταση περίπλοκων σχέσεων γονέα-παιδιού. Αυτή η ιεραρχική οργάνωση είναι επωφελής για την αποθήκευση και την επεξεργασία ιεραρχικών δεδομένων όπως είναι τα συστήματα αρχείων, οργανωτικές δομές ή ταξινομήσεις.

- **Ενοποίηση δεδομένων:** Τα ημιδομημένα δεδομένα διευκολύνουν την ενσωμάτωση δεδομένων από διάφορες πηγές, συμπεριλαμβανομένων δομημένων και μη δομημένων δεδομένων, με την ενσωμάτωση διαφορετικών μορφών και δομών. Αυτή η δυνατότητα απλοποιεί τη διαδικασία ενοποίησης δεδομένων από διάφορες πηγές, όπως βάσεις δεδομένων, API και ιστοσελίδες, σε ένα ενιαίο, ενοποιημένο σύνολο δεδομένων για περαιτέρω ανάλυση και επεξεργασία τους.
- **Επεκτασιμότητα:** Καθώς εμφανίζονται νέα στοιχεία δεδομένων, χαρακτηριστικά ή σχέσεις, τα ημιδομημένα δεδομένα μπορούν εύκολα να επεκταθούν ή να τροποποιηθούν χωρίς σημαντικές αλλαγές στην υπάρχουσα δομή δεδομένων. Αυτή η επεκτασιμότητα είναι πολύ χρήσιμη σε καταστάσεις όπου οι απαιτήσεις δεδομένων εξελίσσονται με την πάροδο του χρόνου, όπως όταν προστίθενται νέες δυνατότητες ή πηγές δεδομένων σε μια εφαρμογή ή όταν τα δεδομένα πρέπει να μετασχηματιστούν ή να εμπλουτιστούν κατά την επεξεργασία.
- **Χειρισμός ελλιπών ή ασυνεπών δεδομένων:** Τα δεδομένα του πραγματικού κόσμου είναι συχνά ακατάστατα, ελλιπή ή ασυνεπή. Οι ημιδομημένες μορφές δεδομένων μπορούν να αντιμετωπίσουν αυτές τις ατέλειες, καθώς δεν επιβάλλουν αυστηρούς περιορισμούς επικύρωσης δεδομένων ή ακεραιότητας. Αυτή η ευελιξία επιτρέπει στους οργανισμούς να επεξεργάζονται και να αναλύουν ατελή δεδομένα, ενώ παράλληλα εξάγουν πολύτιμες πληροφορίες.
- **Προσαρμοστικότητα σε τεχνολογίες μεγάλων δεδομένων:** Η εγγενής ευελιξία και η επεκτασιμότητα των ημιδομημένων δεδομένων τα καθιστούν εξαιρετικά συμβατά με πλαίσια επεξεργασίας μεγάλων δεδομένων, όπως το Apache Hadoop και το Apache Spark. Αυτές οι τεχνολογίες έχουν σχεδιαστεί για την αποτελεσματική αποθήκευση, επεξεργασία και ανάλυση μεγάλου όγκου διαφορετικών δεδομένων, συμπεριλαμβανομένων ημιδομημένων μορφών όπως JSON και XML. Αξιοποιώντας αυτές τις τεχνολογίες μεγάλων δεδομένων, οι οργανισμοί μπορούν να αποκτήσουν πληροφορίες από τεράστια σύνολα δεδομένων, επιτρέποντας την αποτελεσματικότερη λήψη αποφάσεων αφού κατέχουν περισσότερη πληροφορία.

2.3 Μειονεκτήματα ημιδομημένων δεδομένων

Η διαχείριση ημιδομημένων δεδομένων παρουσιάζει ορισμένες προκλήσεις λόγω των χαρακτηριστικών που αναφέραμε πριν. Αυτές οι προκλήσεις περιλαμβάνουν την ασυνέπεια δεδομένων, την έλλειψη τυποποιημένων σχημάτων και την ανάγκη για εξειδικευμένες τεχνικές για την αποτελεσματική εξαγωγή, μετατροπή και φόρτωση των δεδομένων. Επιπλέον, η ετερογένεια και η πολυπλοκότητα των ημιδομημένων πηγών δεδομένων θέτει προκλήσεις για την ενοποίηση και την ανάλυση τους.

- **Έλλειψη επίσημης επιβολής σχήματος:** Σε αντίθεση με τα δομημένα δεδομένα σε σχεσιακές βάσεις δεδομένων, τα ημιδομημένα δεδομένα δεν επιβάλλουν ένα αυστηρό σχήμα ή προκαθορισμένη δομή. Αυτή η έλλειψη επίσημης επιβολής σχήματος μπορεί να οδηγήσει σε προκλήσεις στον ποιοτικό έλεγχο δεδομένων, καθώς δεν υπάρχουν ενσωματωμένοι περιορισμοί ή έλεγχοι για τη διασφάλιση της συνέπειας και της ακρίβειας.
- **Αυξημένη πολυπλοκότητα στη διαχείριση δεδομένων:** Η ευελιξία και οι ποικίλες δομές των ημιδομημένων δεδομένων μπορούν να δημιουργήσουν πολυπλοκότητα στη διαχείριση και την οργάνωση των δεδομένων. Χωρίς ένα προκαθορισμένο σχήμα, η ενοποίηση δεδομένων, ο μετασχηματισμός και η αναζήτηση σε αυτά, μπορεί να απαιτούν πρόσθετη προσπάθεια και να έχουν επιπλέον πολυπλοκότητα, καθώς οι δομές και οι τύποι των δεδομένων μπορεί να διαφέρουν μεταξύ διαφορετικών πηγών.
- **Δυσκολία στην ανάλυση και την αναφορά δεδομένων:** Τα ημιδομημένα δεδομένα ενδέχεται να δημιουργήσουν προκλήσεις κατά την εκτέλεση σύνθετων αναλύσεων δεδομένων και τη δημιουργία αναφορών. Η απουσία ενός συνεπούς σχήματος στα δεδομένα μπορεί να καταστήσει πιο δύσκολο τον ορισμό και την εκτέλεση δομημένων ερωτημάτων. Η εξαγωγή πληροφοριών και η δημιουργία τυποποιημένων αναφορών ενδέχεται να απαιτούν πρόσθετα βήματα επεξεργασίας δεδομένων ή μετασχηματισμούς.
- **Περιορισμένη υποστήριξη εργαλείων:** Ενώ τα δομημένα δεδομένα διαθέτουν καθιερωμένα εργαλεία και τεχνολογίες για αποθήκευση, αναζήτηση και ανάλυση (π.χ. βάσεις δεδομένων SQL), τα εργαλεία που υπάρχουν για τα ημιδομημένα δεδομένα δεν είναι τόσο ανεπτυγμένα. Η ανάπτυξη και χρήση εξειδικευμένων εργαλείων και τεχνολογιών για το χειρισμό ημιδομημένων δεδομένων ενδέχεται να απαιτεί πρόσθετη προσπάθεια και πόρους.
- **Πιθανότητα ασυνέπειας και πλεονασμού δεδομένων:** Η ευέλικτη φύση των ημιδομημένων δεδομένων μπορεί να οδηγήσει σε ασυνέπειες και πλεονασμούς, ειδικά όταν αφορούν την ενοποίηση δεδομένων. Πολλαπλές πηγές μπορεί να έχουν διαφορετικές αναπαραστάσεις ή ερμηνείες παρόμοιων στοιχείων δεδομένων, γεγονός που μπορεί να μας οδηγήσει στο να έχουμε πολλαπλές εγγραφές για την ίδια πληροφορία ή να έχουμε αντικρουόμενες πληροφορίες.
- **Θέματα απόδοσης:** Η ευέλικτη και ένθετη δομή των ημιδομημένων δεδομένων μπορεί να επηρεάσει την απόδοση, ειδικά όταν πρόκειται για μεγάλο όγκο δεδομένων. Πολύπλοκα ερωτήματα ή μετασχηματισμοί σε ημιδομημένα δεδομένα ενδέχεται να απαιτούν περισσότερους υπολογιστικούς πόρους και ενδεχομένως να οδηγήσουν σε πιο αργούς χρόνους επεξεργασίας σε σύγκριση με τα δομημένα δεδομένα.

Είναι σημαντικό να σημειωθεί ότι αυτά τα μειονεκτήματα δεν είναι εγγενή ελαττώματα των ημιδομημένων δεδομένων, αλλά προκλήσεις που πρέπει να αντιμετωπιστούν αποτελεσματικά για να αξιοποιηθούν τα οφέλη τους. Πρέπει να λαμβάνουμε υπόψη αυτές τις πτυχές των ημιδομημένων δεδομένων όταν εργαζόμαστε με αυτά ώστε να εξασφαλίσουμε τη σωστή διαχείριση δεδομένων, τον ποιοτικό έλεγχο και την αποτελεσματική επεξεργασία τους. [3]

2.4 Εξαγωγή δεδομένων από ημιδομημένα δεδομένα

Η εξαγωγή δεδομένων από ημιδομημένα δομημένα είναι απαραίτητη ώστε να κατανοήσουμε και να αποδώσουμε την πληροφορία που περιέχουν. Η διαδικασία αυτή μπορεί να είναι δύσκολη, καθώς τα δεδομένα μπορεί να περιέχουν ασυνέπειες, πλεονασμούς ή ασάφειες. Σε αυτό το κεφάλαιο, θα διερευνήσουμε ορισμένες τεχνικές και εργαλεία για την εξαγωγή δεδομένων από ημιδομημένα δεδομένα. [4]

Πριν εμβαθύνουμε σε τεχνικές εξαγωγής δεδομένων από ημιδομημένα δεδομένα, είναι απαραίτητο να κατανοήσουμε τα μοναδικά χαρακτηριστικά των ημιδομημένων δεδομένων. Όπως αναφέρθηκε πριν, τα ημιδομημένα δεδομένα είναι ευέλικτα, είναι μερικώς δομημένα και μπορούν να αποθηκευτούν σε ένα ευρύ φάσμα μορφών. Αυτά τα χαρακτηριστικά καθιστούν δύσκολη την εξαγωγή σημαντικών δεδομένων από ημιδομημένα δεδομένα, καθώς τα δεδομένα ενδέχεται να περιέχουν ασυνέπειες ή πλεονασμούς που πρέπει να επιλυθούν.

Για να προετοιμαστούμε για την εξαγωγή δεδομένων, πρέπει πρώτα να κατανοήσουμε τη δομή και τη μορφή των δεδομένων. Αυτό περιλαμβάνει τον προσδιορισμό των πηγών δεδομένων, των τύπων των διαθέσιμων δεδομένων και να αποφασίσουμε ποια είναι η πληροφορία που θέλουμε να εξάγουμε. Αφού κατανοήσουμε ξεκάθαρα τα δεδομένα, μπορούμε να επιλέξουμε την καταλληλότερη μέθοδο εξαγωγής δεδομένων, τα εργαλεία και τις τεχνικές εξαγωγής δεδομένων από ημιδομημένα δεδομένα. Κάποια από αυτά παρουσιάζονται στην Ενότητα 2.4.1.

2.4.1 Εργαλεία και Τεχνικές Εξαγωγής Δεδομένων από Ημιδομημένα Δεδομένα

Τα παρακάτω είναι μερικά από τα διαθέσιμα εργαλεία και τεχνικές για την εξαγωγή δεδομένων από ημιδομημένα δεδομένα:

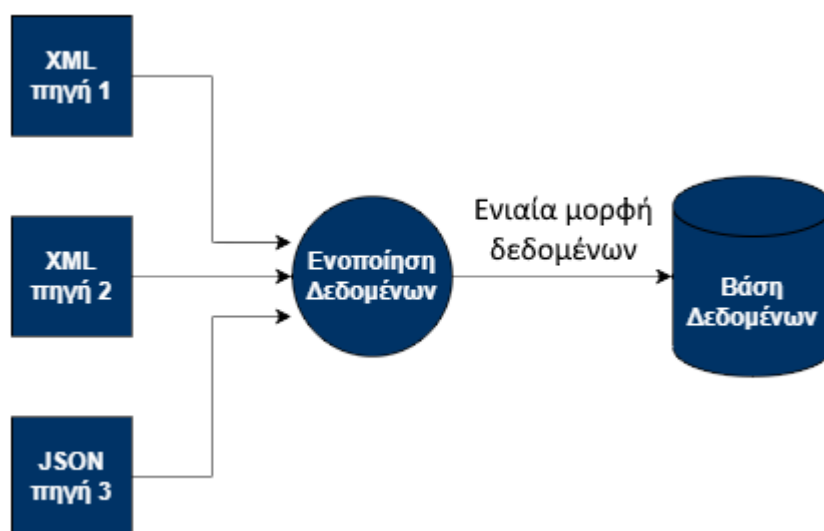
- **XMLSerializer:** Η κλάση XMLSerializer είναι μέρος του framework.NET και μπορεί να χρησιμοποιηθεί για την ανάγνωση και εγγραφή δεδομένων XML. Παρέχει έναν απλό τρόπο εξαγωγής δεδομένων από έγγραφα XML.
- **XPath:** Η XPath είναι μια γλώσσα ερωτημάτων που χρησιμοποιείται για την προσπέλαση εγγράφων XML και την επιλογή στοιχείων ή χαρακτηριστικών που ταιριάζουν με ένα συγκεκριμένο μοτίβο.
- **XQuery:** Η XQuery είναι μια γλώσσα ερωτημάτων που χρησιμοποιείται για την εξαγωγή δεδομένων από έγγραφα XML. Είναι παρόμοια με την SQL, αλλά έχει σχεδιαστεί ειδικά για XML.
- **Κανονικές εκφράσεις (Regular Expressions):** Οι κανονικές εκφράσεις μπορούν να χρησιμοποιηθούν για την εξαγωγή δεδομένων από ημιδομημένα δεδομένα. Είναι ιδιαίτερα χρήσιμα όταν ασχολούμαστε με μη ιεραρχικά δεδομένα, όπως αρχεία καταγραφής.
- **Apache Tika:** Το Apache Tika είναι μια εργαλειοθήκη για τον εντοπισμό και την εξαγωγή μεταδεδομένων και κειμένου από διάφορες μορφές αρχείων, συμπεριλαμβανομένης της XML. Παρέχει μια ενοποιημένη διεπαφή για την εξαγωγή δεδομένων από διαφορετικές μορφές αρχείων.
- **Hadoop MapReduce:** Το Hadoop MapReduce είναι ένα κατανεμημένο πλαίσιο επεξεργασίας που μπορεί να χρησιμοποιηθεί για την παράλληλη επεξεργασία μεγάλου

όγκου δεδομένων. Είναι ιδιαίτερα χρήσιμο για την επεξεργασία μεγάλων όγκων ημιδομημένων δεδομένων.

2.5 Ενοποίηση δεδομένων

Η ενοποίηση δεδομένων είναι η διαδικασία συνδυασμού δεδομένων από πολλαπλές πηγές και η επεξεργασία τους προκειμένου να καταλήξουν να έχουν μια ενιαία μορφή. Είναι μια απαραίτητη διαδικασία για έναν οργανισμό στην σημερινή εποχή αφού θα έχει δεδομένα που προέρχονται από πολλαπλές πηγές και θα έχουν διαφορετικές μορφές όπως δεδομένα από, ιστοσελίδες, APIs, βάσεις δεδομένων, αρχεία ημιδομημένων δεδομένων.

Η διαδικασία αυτή έχει επιπλέον δυσκολίες όταν αναφερόμαστε σε ημιδομημένα δεδομένα, καθώς τα δεδομένα αποθηκεύονται συχνά σε διαφορετικές μορφές και με διαφορετικές δομές. Παρόλα αυτά είναι μια πολύ σημαντική διαδικασία αφού όταν επιτευχθεί αυτή, θα βελτιωθεί η ποιότητα των δεδομένων, θα αυξηθεί η ποσότητα των δεδομένων και η ανάλυση τους και η επεξεργασία τους θα γίνει πολύ πιο απλά και εύκολα αφού πλέον θα έχουμε τα δεδομένα σε μια ενιαία μορφή. Προκειμένου να γίνει αποτελεσματικά η ενοποίηση δεδομένων στα ημιδομημένα δεδομένα, μπορούν να χρησιμοποιηθούν διάφορες τεχνικές και εργαλεία. [4]–[6]



Εικόνα 2: Απλοποιημένη αναπαράσταση ενοποίησης πολλαπλών ημιδομημένων δεδομένων

2.5.1 Πλεονεκτήματα ενοποίησης δεδομένων σε ημιδομημένα δεδομένα

- **Ενιαία μορφή:** Τα ημιδομημένα δεδομένα αποθηκεύονται συχνά σε διαφορετικές μορφές και δομές, γεγονός που καθιστά δύσκολη την ανάλυση τους και την επεξεργασία τους ώστε να εξάγουμε την πληροφορία που χρειαζόμαστε. Με την ενοποίηση δεδομένων θα είναι όλα τα δεδομένα σε μια ενιαία μορφή απλουστεύοντας έτσι την επεξεργασία τους.
- **Ποιότητα δεδομένων:** Τα ημιδομημένα δεδομένα είναι επιρρεπή σε λάθη και ασυνέπειες λόγω της ευέλικτης δομής τους. Η ενοποίηση δεδομένων μπορεί να συμβάλει στη βελτίωση της ποιότητας των δεδομένων εντοπίζοντας και διορθώνοντας λάθη και ασυνέπειες.

- **Εμπλουτισμός δεδομένων:** Τα ημιδομημένα δεδομένα, ξεχωριστά από κάθε πηγή, ενδέχεται να μην μας παρέχουν όλες τις πληροφορίες που χρειαζόμαστε. Η ενοποίηση δεδομένων μπορεί να βοηθήσει στον εμπλουτισμό των δεδομένων προσθέτοντας δεδομένα που λείπουν ή ενσωματώνοντάς τα με άλλες πηγές δεδομένων.
- **Ανάλυση δεδομένων:** Η ενοποίηση δεδομένων μπορεί να βοηθήσει να γίνουν τα δεδομένα πιο εύκολα και πιο χρήσιμα στην ανάλυση τους και επιπλέον να καταλάβουμε σχέσεις και μοτίβα μεταξύ των διαφόρων πηγών δεδομένων. Η ανάλυση γίνεται πιο απλή και λιγότερο χρονοβόρα αφού θα πρέπει να γίνει για την ενιαία μορφή που θα έχουμε μετά την ενοποίηση των δεδομένων.

2.5.2 Τεχνικές και εργαλεία ενοποίησης δεδομένων

Μια τεχνική για την ενοποίηση ημιδομημένων δεδομένων είναι η χρήση ενός κοινού μοντέλου δεδομένων που μπορεί να χρησιμοποιηθεί για την αναπαράσταση των δεδομένων από όλες τις πηγές. Αυτό μπορεί να βοηθήσει στην απλοποίηση της διαδικασίας ενοποίησης και να διασφαλίσει ότι τα δεδομένα είναι συνεπή σε όλες τις πηγές. Επιπλέον, εργαλεία όπως οι αναλυτές XML και οι μετατροπείς μπορούν να χρησιμοποιηθούν για τη μετατροπή των δεδομένων από την αρχική τους μορφή, στο κοινό μοντέλο δεδομένων. Παρακάτω αναφέρονται κάποια δημοφιλή εργαλεία ενοποίησης δεδομένων που χρησιμοποιούνται για ημιδομημένα δεδομένα. Αυτά τα εργαλεία παρέχουν δυνατότητες εξαγωγής, μετατροπής και φόρτωσης δεδομένων (Extract, Transform, Load - ETL) από διάφορες πηγές, συμπεριλαμβανομένων ημιδομημένων μορφών όπως JSON και XML.

Apache Nifi: Το Apache Nifi είναι μια πλατφόρμα ενοποίησης δεδομένων ανοιχτού κώδικα που υποστηρίζει την επεξεργασία διαφορετικών μορφών δεδομένων, συμπεριλαμβανομένων των ημιδομημένων δεδομένων. Προσφέρει μια γραφική διεπαφή για το σχεδιασμό ροών δεδομένων, επιτρέποντας στους χρήστες να εξάγουν, να μετασχηματίζουν και να φορτώνουν εύκολα δεδομένα από διαφορετικές πηγές και να πραγματοποιούν μετασχηματισμούς σε ημιδομημένα δεδομένα.

Talend: Το Talend είναι ένα ολοκληρωμένο εργαλείο ενοποίησης δεδομένων και ETL που υποστηρίζει την διαχείριση ημιδομημένων δεδομένων. Παρέχει μια γραφική διεπαφή για το σχεδιασμό ροών εργασίας ενοποίησης δεδομένων, επιτρέποντας στους χρήστες να εξάγουν δεδομένα από διάφορες πηγές, να εφαρμόζουν μετασχηματισμούς και να φορτώνουν τα δεδομένα σε άλλα συστήματα. Το Talend υποστηρίζει το χειρισμό ημιδομημένων μορφών δεδομένων όπως JSON και XML.

Pentaho Data Integration: Το Pentaho Data Integration, γνωστό και ως Kettle, είναι ένα εργαλείο ανοιχτού κώδικα ETL που προσφέρει ισχυρή υποστήριξη για την διαχείριση ημιδομημένων δεδομένων. Παρέχει δυνατότητες οπτικού σχεδιασμού ροών εργασίας ενοποίησης δεδομένων, επιτρέποντας στους χρήστες να εξάγουν δεδομένα από πολλαπλές πηγές, να πραγματοποιούν μετασχηματισμούς σε ημιδομημένα δεδομένα και να τα φορτώνουν σε διαφορετικά συστήματα.

Informatica PowerCenter: Το Informatica PowerCenter είναι ένα δημοφιλές εργαλείο ενοποίησης δεδομένων που παρέχει πολλές λειτουργίες για την διαχείριση δομημένων, ημιδομημένων και μη δομημένων δεδομένων. Υποστηρίζει εξαγωγή δεδομένων από διάφορες

πηγές, μετασχηματισμό δεδομένων και φόρτωση σε συστήματα. Το Informatica PowerCenter προσφέρει ειδικά σχεδιασμένα εργαλεία και πρόσθετα την διαχείριση ημιδομημένων μορφών δεδομένων.

Η επιλογή του εργαλείου εξαρτάται από παράγοντες όπως οι απαιτήσεις του έργου ενοποίησης δεδομένων, οι περιορισμοί προϋπολογισμού, οι ανάγκες επεκτασιμότητας και το επίπεδο πολυπλοκότητας που συνεπάγεται με την διαχείριση των ημιδομημένων δεδομένων.

2.6 Ποιότητα δεδομένων

Η ποιότητα των δεδομένων είναι πολλή σημαντική σε κάθε σύστημα που βασίζεται σε δεδομένα, αφού επηρεάζει άμεσα την αξιοπιστία και την εγκυρότητα των πληροφοριών, των αποφάσεων και των ενεργειών που βασίζονται στα δεδομένα αυτά. Η ποιότητα των δεδομένων αναφέρεται στο κατά πόσο τα δεδομένα είναι αξιόπιστα, ακριβή, έγκυρα, συνεπή, πλήρη και κατάλληλα για το σύστημα που θα χρησιμοποιηθούν.

Η εισαγωγή δεδομένων από πολλαπλές πηγές σε ένα σύστημα, μπορεί να έχει αρνητικές συνέπειες για την ποιότητα των δεδομένων του αλλά και μπορεί να του προσφέρει την δυνατότητα να βελτιώσει την ποιότητα τους. Η ποιότητα δεδομένων κινδυνεύει να υποβαθμιστεί κυρίως όταν δεν γίνονται σωστές διαδικασίες και έλεγχοι για την εισαγωγή δεδομένων από εξωτερικές πηγές. Από την άλλη έχοντας πολλαπλές πηγές μπορεί να γίνει πιο εύκολα ο εντοπισμός και η διόρθωση τυχών σφαλμάτων, έτσι βελτιώνεται η ποιότητα δεδομένων του συστήματος. [1]

2.7 Καθαρισμός Δεδομένων

Τα δεδομένα, ειδικά όταν αυτά μπορεί να προέρχονται από πολλαπλές πηγές, συχνά έχουν σφάλματα και ασυνέπειες που έχουν επιπτώσεις στην συνολική ποιότητα δεδομένων του συστήματος. Τα δεδομένα με αυτά τα προβλήματα πρέπει είτε να διορθωθούν είτε να αφαιρεθούν. Ο καθαρισμός δεδομένων είναι η διαδικασία εντοπισμού και διόρθωσης ή αφαίρεσης σφαλμάτων ασυνεπειών, ανακρίβειών σε ένα σύνολο δεδομένων για τη βελτίωση της ποιότητας, της αξιοπιστίας και της ακρίβειας του. Τα σφάλματα αυτά μπορεί να προέρχονται από λάθη κατά την εισαγωγή δεδομένων (data entry) ή μπορεί να προκύπτουν από την ίδια την εφαρμογή που αποθηκεύει τα δεδομένα. Επιπλέον τα XML αρχεία είναι πιο επιρρεπή σε τέτοια σφάλματα και ασυνέπειες. [7]

Ο καθαρισμός δεδομένων και η ποιότητα των δεδομένων είναι ένα πολύ σημαντικό θέμα, ειδικά αν λάβουμε υπόψιν πως έχει προβλεφθεί ότι πάνω από το 25% των κρίσιμων δεδομένων των εταιρειών Fortune 1000 είναι ελαττωματικά, και υπάρχουν αρκετές αναφορές στην βιβλιογραφία για μεθοδολογίες και εργαλεία που βοηθούν σε αυτή την πτυχή των δεδομένων ενός συστήματος. [8], [9]

2.8 Τυπική Διαδικασία Διαχείρισης Δεδομένων και Διασφάλισης της Ποιότητάς τους

1. **Εισαγωγή Δεδομένων (Data Ingestion):** Συλλέγονται τα δεδομένα από όλες τις διαθέσιμες πηγές και αποθηκεύονται σε ένα κοινό σύστημα ώστε να αναλυθούν και να επεξεργαστούν.
2. **Μετατροπή Δεδομένων (Data Transformation):** Τα δεδομένα μετατρέπονται σε κοινές μονάδες, όπως τα μίλια σε χιλιόμετρα, χρησιμοποιώντας κανόνες και εργαλεία μετατροπής. Έχοντας κοινές μονάδες στα δεδομένα διευκολύνεται η διαδικασία ανάλυσης και σύγκρισής τους.
3. **Βελτίωση Ποιότητας Δεδομένων (Data Quality):** Μέσω διαδικασιών καθαρισμού δεδομένων, γίνεται εντοπισμός και διόρθωση σε σφάλματα που βρίσκονται στα δεδομένα. Αυτά τα σφάλματα μπορεί να είναι ελλιπής ή εσφαλμένες τιμές. Με αυτό τον τρόπο διασφαλίζεται η ποιότητά και η ακρίβεια τους.
4. **Ευθυγράμμιση Σχημάτων (Schema Matching):** Ο στόχος αυτού του βήματος είναι να δημιουργηθεί μια κοινή δομή για τα δεδομένα, προκειμένου να εξασφαλιστεί η ομοιομορφία και η συμβατότητά τους. Είναι απαραίτητο βήμα για την σύγκριση των δεδομένων μεταξύ των πηγών. [10]
5. **Αφαίρεση Διπλοτύπων και Σύνδεση Εγγραφών (Data Deduplication and Record Linkage):** Μέσω αυτής της διαδικασίας εντοπίζονται οι διπλότυπες εγγραφές και αφαιρούνται, καθώς και ομαδοποιούνται εγγραφές που αναφέρονται στην ίδια οντότητα.
6. **Επιλογή Κανονικών Τιμών (Golden Values):** Γίνεται επιλογή των πιο αξιόπιστων τιμών για τα δεδομένα ώστε να δημιουργηθεί το τελικό ενοποιημένο σύνολο δεδομένων.

Μετά από όλα αυτά τα βήματα, έχει δημιουργηθεί ένα εκτενές, ποιοτικό, οργανωμένο και συνεπές σύνολο δεδομένων που συγκεντρώνει πληροφορίες από όλες τις πηγές. Αυτό το τελικό σύνολο δεδομένων μπορεί να χρησιμοποιηθεί στο σύστημα που το απαιτεί. [9],[11]

3

Ανάπτυξη Ιστοσελίδων

3.1 Δυναμικές Ιστοσελίδες

Οι δυναμικές ιστοσελίδες, όπως αναφέρει και το όνομα τους είναι ιστοσελίδες που το περιεχόμενό τους δημιουργείται και μεταβάλλεται δυναμικά. Αυτό έρχεται σε αντίθεση με τις στατικές ιστοσελίδες, οι οποίες παρουσιάζουν ένα προκατασκευασμένο περιεχόμενο που παραμένει το ίδιο για κάθε χρήστη. Οι δυναμικές ιστοσελίδες κατασκευάζονται χρησιμοποιώντας έναν συνδυασμό client-side και server-side γλωσσών επιτρέποντας στους διαχειριστές να τροποποιούν εύκολα το περιεχόμενο και το σχεδιασμό χωρίς να χρειάζεται να κάνουν αλλαγές σε κάθε σελίδα του ιστότοπου. Επιπλέον μπορούν να παρέχουν μια πιο διαδραστική και εξατομικευμένη εμπειρία στους χρήστες, κάτι που είναι απαραίτητο για πολλές σύγχρονες εφαρμογές Ιστού.

Πρακτικά:

1. Ο χρήστης ανοίγει την ιστοσελίδα και στέλνεται αίτημα (request) στον διακομιστή.
2. Ο διακομιστής λαμβάνει και επεξεργάζεται το αίτημα εκτελώντας τον κώδικα που έχει γραφτεί για το συγκεκριμένο αίτημα. Αλληλεπιδρά με την βάση ή με άλλες πηγές δεδομένων για την ανάκτηση των απαιτούμενων δεδομένων και δημιουργεί τον κώδικα HTML με τα τελικά δεδομένα.
3. Το πρόγραμμα περιήγησης εμφανίζει την HTML και ο χρήστης βλέπει το αποτέλεσμα στην οθόνη του.
4. Ο χρήστης αλληλεπιδρά με την ιστοσελίδα κάνοντας ενέργειες που προκαλούν αλλαγές στο περιεχόμενο της ιστοσελίδας είτε κατευθείαν client-side μέσω JavaScript είτε υπάρχει πάλι επικοινωνία με τον διακομιστή και απάντηση αυτού στο πρόγραμμα περιήγησης ώστε να ενημερωθεί το περιεχόμενο.

Από την στιγμή που το περιεχόμενο της ιστοσελίδας μπορεί να δημιουργηθεί δυναμικά από την βάση δεδομένων, μπορεί να δημιουργηθεί μια διαχείριση του δυναμικού περιεχομένου ώστε οι διαχειριστές της ιστοσελίδας να μπορούν να το μεταβάλουν χωρίς να έχουν γνώσεις προγραμματισμού ούτε να χρειάζεται να έχουν πρόσβαση στα αρχεία της ιστοσελίδας. [12]

3.2 Τεχνολογίες για ανάπτυξη δυναμικών ιστοσελίδων

Οι βασικές τεχνολογίες για την ανάπτυξη μιας σύγχρονης δυναμικής ιστοσελίδας είναι HTML, CSS, JavaScript, μια server-side γλώσσα (PHP, C#, Python, Ruby), ένα σύστημα διαχείρισης βάσης (SQL Server, MySQL) και η χρήση AJAX. Επιπλέον με χρήση της τεχνικής AJAX μπορεί η ιστοσελίδα να είναι πιο δυναμική και διαδραστική. Μπορούν να χρησιμοποιηθούν και frameworks όπως ASP.NET MVC, React, Angular που βοηθάνε την διαδικασία ανάπτυξης δυναμικών ιστοσελίδων. Σημαντική είναι και η χρήση ενός συστήματος διαχείρισης βάσεων δεδομένων (SQL Server, MySQL) ώστε να αποθηκεύονται τα δεδομένα που χρειάζεται η δυναμική ιστοσελίδα.

3.2.1 HTML (Hypertext Markup Language)

Η HTML είναι η βασική γλώσσα σήμανσης που χρησιμοποιείται για τη δημιουργία ιστοσελίδων και αποτελείται από ένα σύνολο ετικετών (tags) και χαρακτηριστικών (attributes). Εκτός από τον καθορισμό της δομής και του περιεχομένου μιας ιστοσελίδας, η HTML μπορεί επίσης να χρησιμοποιηθεί σε συνδυασμό με άλλες τεχνολογίες Ιστού, όπως CSS και JavaScript, για τη δημιουργία δυναμικών και διαδραστικών ιστοσελίδων. Η HTML5 έχει αλλάξει αρκετές εκδόσεις, η τωρινή έκδοση είναι η HTML5 που έγινε δημόσια διαθέσιμη το 2008.

3.2.2 CSS (Cascading Style Sheets)

Η γλώσσα CSS χρησιμοποιείται για τον έλεγχο της εμφάνισης και της μορφοποίησης των ιστοσελίδων, συμπεριλαμβανομένων των χρωμάτων, των γραμματοσειρών, του φόντου και της διάταξης. Το CSS χωρίζεται σε τρεις κύριες κατηγορίες: ενσωματωμένο (inline), εσωτερικό και εξωτερικό. Το ενσωματωμένο CSS εφαρμόζεται απευθείας σε ένα μεμονωμένο στοιχείο HTML χρησιμοποιώντας το χαρακτηριστικό "style". Το εσωτερικό CSS συνήθως ορίζεται στην ενότητα κεφαλίδας (header) ενός εγγράφου HTML χρησιμοποιώντας την ετικέτα "<style>". Το εξωτερικό CSS ορίζεται σε ένα ξεχωριστό αρχείο CSS, το οποίο στη συνέχεια συνδέεται με το έγγραφο HTML χρησιμοποιώντας την ετικέτα "link". Το εξωτερικό CSS είναι η πιο κοινή προσέγγιση, καθώς επιτρέπει τον διαχωρισμό περιεχομένου και παρουσίασης και επιτρέπει την επαναχρησιμοποίηση στυλ σε πολλές ιστοσελίδες.

3.2.3 JavaScript

Η JavaScript είναι μια υψηλού επιπέδου γλώσσα προγραμματισμού που χρησιμοποιείται κυρίως για τη δημιουργία διαδραστικού και δυναμικού περιεχομένου Ιστού. Χρησιμοποιείται για client-side λειτουργίες και παρέχει διαδραστικότητα, χειρισμό στοιχείων της ιστοσελίδας και χειρισμό συμβάντων από τις ενέργειες του χρήστη στην ιστοσελίδα. Η JavaScript μπορεί επίσης να χρησιμοποιηθεί server-side με τεχνολογίες όπως το Node.js. Ορισμένες δημοφιλείς βιβλιοθήκες (libraries) και frameworks της JavaScript περιλαμβάνουν τα jQuery, React, Angular και Vue.js.

3.2.4 *jQuery*

Η jQuery είναι μια δημοφιλής βιβλιοθήκη JavaScript που απλοποιεί τη διαδικασία δημιουργίας διαδραστικού και δυναμικού περιεχομένου Ιστού. Παρέχει πολλές λειτουργίες και μεθόδους για την ανάπτυξη ιστοσελίδων, όπως η προσπέλαση και η διαμόρφωση στοιχείων του DOM, ο χειρισμός HTML και CSS, ο χειρισμός συμβάντων, προσθήκη animations και η διαχείριση των αιτημάτων AJAX για την επικοινωνία με τον διακομιστή. Χρησιμοποιείται ευρέως για την ευκολία χρήσης και τη συμβατότητα μεταξύ προγραμμάτων περιήγησης.

3.2.5 *AJAX (Asynchronous JavaScript and XML)*

Είναι μια τεχνική για την ανταλλαγή δεδομένων με έναν διακομιστή στο παρασκήνιο, δηλαδή η επικοινωνία μπορεί να γίνεται ασύγχρονα. Το βασικό πλεονέκτημα είναι ότι επιτρέπει στις εφαρμογές Ιστού να ενημερώνουν το περιεχόμενό τους χωρίς να απαιτείται από τον χρήστη να πραγματοποιήσει πλήρη ανανέωση της σελίδας. Αυτό σημαίνει ότι οι ιστοσελίδες μπορούν να ενημερώνουν το περιεχόμενό τους σε πραγματικό χρόνο, χωρίς να διακόπτουν την εμπειρία του χρήστη. Αυτό καθιστά δυνατή τη δημιουργία πιο δυναμικών, ελκυστικών και διαδραστικών εφαρμογών Ιστού που παρέχουν καλύτερη εμπειρία στον χρήστη. Επιτρέπει στις εφαρμογές Ιστού να ενημερώνουν δυναμικά το περιεχόμενό τους, καθιστώντας τις πιο ανταποκρινόμενες και διαδραστικές για τους χρήστες. Οι σύγχρονες υλοποιήσεις συνήθως χρησιμοποιούν JSON αντί για XML για την ανταλλαγή δεδομένων μεταξύ του διακομιστή.

3.2.6 *PHP (Hypertext Preprocessor)*

Η PHP είναι μια server-side γλώσσα που χρησιμοποιείται για την δημιουργία εφαρμογών Ιστού. Είναι μια γλώσσα ανοιχτού κώδικα (open-source) που μπορεί να ενσωματωθεί σε κώδικα HTML για τη δημιουργία δυναμικών ιστοσελίδων που μπορούν να αλληλεπιδρούν με βάσεις δεδομένων, να επεξεργάζονται δεδομένα φόρμας και να εκτελούν άλλες server-side εργασίες. Η PHP μπορεί επίσης χρησιμοποιείται σε δημοφιλή συστήματα διαχείρισης περιεχομένου (CMS) όπως το WordPress, το Drupal και το Joomla, για τη δημιουργία προσαρμοσμένων θεμάτων και plugins που επεκτείνουν τη λειτουργικότητα αυτών των πλατφορμών.

3.3 *Σύστημα Διαχείρισης Περιεχομένου (CMS)*

Τα συστήματα διαχείρισης περιεχομένου είναι ένας πολύ δημοφιλής τρόπος για την ανάπτυξη δυναμικών ιστοσελίδων και χρησιμοποιούνται συχνά για την κατασκευή ιστοσελίδων, ιστολογίων και ηλεκτρονικών καταστημάτων. Παρέχουν μια φιλική προς τον χρήστη διεπαφή για τη διαχείριση περιεχομένου και συχνά έχουν ενσωματωμένες δυνατότητες για την προσαρμογή του σχεδιασμού και της λειτουργικότητας ενός ιστότοπου. Επιτρέπουν στους χρήστες να δημιουργούν, να δημοσιεύουν και να διαχειρίζονται περιεχόμενο, όπως άρθρα, εικόνες, βίντεο και άλλα μέσα, χωρίς να απαιτείται γνώση στην ανάπτυξη ιστού. Τα πιο δημοφιλή συστήματα διαχείρισης περιεχομένου είναι το WordPress, το Drupal και το Joomla. Είναι μια καλή επιλογή για ιστότοπους που απαιτούν συχνές ενημερώσεις, καθώς παρέχουν μια εύχρηστη διεπαφή για τη διαχείριση περιεχομένου χωρίς να απαιτούν ιδιαίτερες γνώσεις στην ανάπτυξης ιστού από τους διαχειριστές. [13]

Τα Συστήματα Διαχείρισης Περιεχομένου παρέχουν μια φιλική προς το χρήστη διεπαφή για τη δημιουργία και την επεξεργασία περιεχομένου. Ένας χρήστης μπορείτε να φτιάξει μια ιστοσελίδα χωρίς χρειάζονται τεχνικές γνώσεις και γνώσεις κώδικα. Οι χρήστες μπορούν να εισάγουν κείμενο, εικόνες, βίντεο και άλλους τύπους μέσω σε μια δομημένη μορφή. Τα δεδομένα αυτά αποθηκεύονται σε μια βάση δεδομένων που την παρέχει το CMS και έτσι γίνεται εύκολα η ανάκτηση και η παρουσίαση του περιεχομένου.

Τα Συστήματα Διαχείρισης Περιεχομένου διαθέτουν συστήματα διαχείρισης χρηστών που μπορεί να οριστεί το επίπεδο πρόσβασης του κάθε χρήστη. Προσφέρουν έτοιμα πρότυπα και θέματα που καθορίζουν την εμφάνιση του ιστότοπου. Οι χρήστες μπορούν να διαλέξουν το θέμα που θέλουν και στην συνέχεια έχουν την δυνατότητα να το προσαρμόσουν ώστε να αποδώσουν στην ιστοσελίδα την εμφάνιση που θέλουν.

Επιπλέον τα Συστήματα Διαχείρισης Περιεχομένου μέσω των προσθέτων (plugins) μπορούν να επεκτείνουν την λειτουργικότητα τους σε διάφορους τομείς, από το να κάνουν ενσωμάτωση μέσω κοινωνικής δικτύωσης στην ιστοσελίδα μέχρι και να προσθέσουν δυνατότητες ηλεκτρονικού εμπορίου. Συγκεκριμένα για τον σκοπό της διπλωματικής μας, υπάρχουν πρόσθετα που επιτρέπουν το ανέβασμα XML αρχείων και επιτρέπουν την αντιστοίχιση XML πεδίων με πεδίων στην βάση δεδομένων.

Τα Συστήματα Διαχείρισης Περιεχομένου είναι τα ισχυρά εργαλεία που διαθέτουν και απλοποιούν και διευκολύνουν την ανάπτυξη και διατήρηση δυναμικών ιστοσελίδων ακόμα και σε χρήστες που δεν έχουν τεχνικές γνώσεις στην ανάπτυξη ιστού.

3.4 Frameworks – ASP.NET MVC

Η χρήση ενός ισχυρού και αποτελεσματικού Framework μπορεί να βοηθήσει σημαντικά στη δημιουργία δυναμικών και επεκτάσιμων ιστοσελίδων. Ένα τέτοιο Framework είναι το ASP .NET MVC, το οποίο προσφέρει πολλά πλεονεκτήματα. Το ASP .NET MVC ακολουθεί το αρχιτεκτονικό μοτίβο Model-View-Controller, παρέχοντας μια δομημένη προσέγγιση στην ανάπτυξη εφαρμογών Ιστού.

Ένας από τους κύριους λόγους για να επιλεγεί το ASP .NET MVC είναι το ολοκληρωμένο σύνολο εργαλείων και βιβλιοθηκών που διευκολύνουν τη διαδικασία ανάπτυξης σε όλες τις πτυχές της ιστοσελίδας που θέλουμε να δημιουργήσουμε, από την δημιουργία της ιστοσελίδας που αλληλεπιδρά ο χρήστης, τις λειτουργίες που γίνονται στον διακομιστή μέχρι και την επεξεργασία και διαχείριση των ημιδομημένων δεδομένων.

Ένα άλλο σημαντικό πλεονέκτημα του ASP .NET MVC είναι η ισχυρή ενσωμάτωσή του με τη στοίβα τεχνολογίας της Microsoft. Η συγγραφή του κώδικα έγινε στο Microsoft Visual Studio, η βάση δεδομένων είναι δημιουργήθηκε με Microsoft SQL Server Management Studio και η ιστοσελίδα έχει γίνει host σε Windows Server μέσω του Internet Information Services (IIS) Manager. Αυτή η ενοποίηση εξασφαλίζει συμβατότητα, βελτιστοποίηση απόδοσης και πρόσβαση σε ένα ευρύ φάσμα πρόσθετων πόρων και λειτουργιών.

Επιπλέον, το ASP .NET MVC προσφέρει ισχυρά χαρακτηριστικά ασφαλείας, συμπεριλαμβανομένης της ενσωματωμένης προστασίας από κοινά τρωτά σημεία ιστού. Το πλαίσιο παρέχει μηχανισμούς για την εφαρμογή ελέγχου ταυτότητας, εξουσιοδότησης και

επικύρωσης εισόδου, διασφαλίζοντας την ακεραιότητα και την ασφάλεια του ιστότοπου και των δεδομένων του.

Δεδομένων αυτών των λόγων, το ASP .NET MVC αναδείχθηκε ως η προτιμώμενη επιλογή για την ανάπτυξη του ιστότοπου, παρέχοντας τα απαραίτητα εργαλεία, ευελιξία και ασφάλεια για την αποτελεσματική δημιουργία μιας δυναμικής ιστοσελίδας για την διαχείριση ημιδομημένων δομημένων.

4

Αρχιτεκτονική Συστήματος

Στο κεφάλαιο αυτό θα αναλύσουμε την αρχιτεκτονική του συστήματος. Θα γίνει μια επισκόπηση της αρχιτεκτονικής συστήματος για τον ιστότοπο που αναπτύχθηκε για τη διαχείριση ημιδομημένων δεδομένων. Το κεφάλαιο περιγράφει τη δομή του συστήματος, τα διάφορα στοιχεία που απαρτίζουν το σύστημα και πως αυτά αλληλεπιδρούν μεταξύ τους.

Ο ιστότοπος που αναπτύχθηκε για τη διαχείριση ημιδομημένων δεδομένων είναι μια διαδικτυακή εφαρμογή που επιτρέπει στους χρήστες να ανεβάζουν και να διαχειρίζονται αρχεία ημιδομημένων δεδομένων. Το σύστημα χωρίζεται σε τρία κύρια στοιχεία: τη διεπαφή χρήστη, τον διακομιστή και τη βάση δεδομένων. Η διεπαφή χρήστη είναι υπεύθυνη για την αλληλεπίδραση με τον χρήστη, ενώ ο διακομιστής είναι υπεύθυνος για την επεξεργασία αιτημάτων από τη διεπαφή χρήστη και την ανάκτηση δεδομένων από τη βάση δεδομένων. Η βάση δεδομένων αποθηκεύει όλα τα δεδομένα που ανεβάζει ο χρήστης.

4.1 Αρχιτεκτονική λογισμικού Model-View-Controller (MVC)

Η αρχιτεκτονική του συστήματος βασίζεται στο μοντέλο αρχιτεκτονικής λογισμικού Model-View-Controller (MVC). Το μοντέλο αρχιτεκτονικής λογισμικού MVC διαχωρίζει τη διεπαφή χρήστη, τον διακομιστή και τη βάση δεδομένων σε τρία ξεχωριστά στοιχεία, επιτρέποντας ευκολότερη συντήρηση και επεκτασιμότητα.

Η διεπαφή χρήστη επικοινωνεί με τον διακομιστή χρησιμοποιώντας αιτήματα HTTP. Ο διακομιστής έχει κατασκευαστεί χρησιμοποιώντας το πλαίσιο MVC .NET, το οποίο παρέχει ένα σύνολο εργαλείων και βιβλιοθηκών για τη δημιουργία εφαρμογών Ιστού. Ο διακομιστής χειρίζεται αιτήματα από τη διεπαφή χρήστη, ανακτά δεδομένα από τη βάση δεδομένων και στέλνει απαντήσεις πίσω στη διεπαφή χρήστη.

Η βάση δεδομένων υλοποιείται χρησιμοποιώντας τον Microsoft SQL Server, ένα δημοφιλές σύστημα διαχείρισης σχεσιακών βάσεων δεδομένων. Η βάση δεδομένων αποθηκεύει όλα τα δεδομένα που εμφανίζονται στις δυναμικές σελίδες του ιστότοπου. Το σχήμα της βάσης δεδομένων έχει σχεδιαστεί για να είναι ευέλικτο, επιτρέποντας την εύκολη τροποποίηση της δομής δεδομένων, όπως απαιτείται. [14], [15]

4.1.1 Model

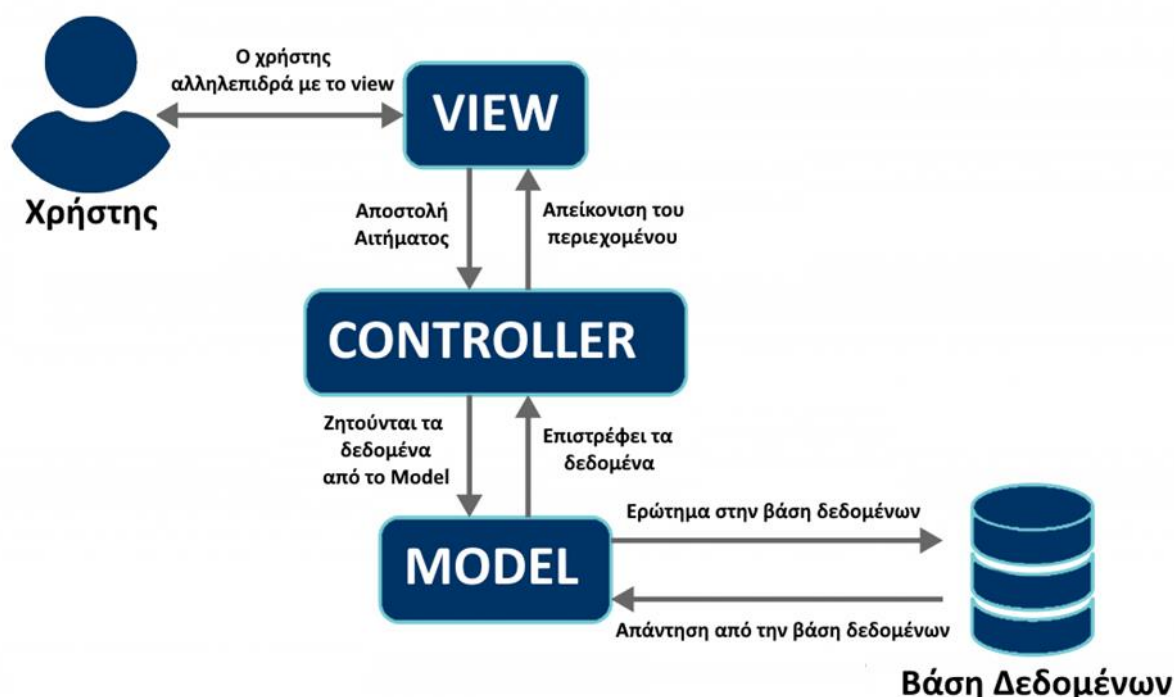
Το στοιχείο Model είναι το βασικό κομμάτι της αρχιτεκτονικής συστήματος MVC καθώς αναπαριστά τα δεδομένα και την λογική του συστήματος. Περιλαμβάνει επίσης ένα σύνολο κλάσεων C# που ενσωματώνουν την λογική του συστήματος, συμπεριλαμβανομένης της επικύρωσης δεδομένων, των ερωτημάτων και των διάφορων διαδικασιών που εκτελεί ο χρήστης.

4.1.2 View

Το στοιχείο View της αρχιτεκτονικής του συστήματος, που είναι στην ουσία η διεπαφή χρήστη, περιλαμβάνει δύο διακριτά μέρη: την ιστοσελίδα του εργαστηρίου και την ιστοσελίδα διαχείρισης - Dashboard. Η ιστοσελίδα του εργαστηρίου έχει σχεδιαστεί για να παρουσιάζει τις πληροφορίες που ορίζει ο διαχειριστής από την ιστοσελίδα διαχείρισης, που περιλαμβάνει και τα εξαγόμενα δεδομένα από τα ημιδομημένα δομημένα, στους επισκέπτες με φιλικό προς τον χρήστη και οπτικά ελκυστικό τρόπο. Η ιστοσελίδα διαχείρισης, από την άλλη πλευρά, έχει σχεδιαστεί για να γίνεται η διαχείριση των μελών, των δημοσιεύσεων και γενικά των πληροφοριών που εμφανίζονται στην ιστοσελίδα εργαστηρίου. Η διεπαφή χρήστη αναπτύσσεται χρησιμοποιώντας HTML, CSS, JavaScript, Razor για τη δημιουργία δυναμικών ιστοσελίδων.

4.1.3 Controller

Το στοιχείο Controller της αρχιτεκτονικής συστήματος περιλαμβάνει ένα σύνολο κλάσεων C# που χειρίζονται τα δεδομένα εισόδου και τις ενέργειες του χρήστη και τα μετατρέπει σε εντολές για το Model ή το View. Εφαρμόζει την λογική του συστήματος, συμπεριλαμβανομένου του χειρισμού του ελέγχου ταυτότητας χρήστη, της διαχείρισης περιόδων σύνδεσης και της δρομολόγησης αιτημάτων στα κατάλληλα Views.



Εικόνα 3: Απεικόνιση αρχιτεκτονικής MVC

Η Εικόνα 3 δείχνει ένα μπλοκ διάγραμμα υψηλού επιπέδου της αρχιτεκτονικής MVC. Το διάγραμμα απεικονίζει τα διάφορα στοιχεία του συστήματος και τον τρόπο αλληλεπίδρασης μεταξύ τους.

4.2 Σχεδιαστικές αποφάσεις

Κατά την ανάπτυξη του συστήματος λήφθηκαν αρκετές σχεδιαστικές αποφάσεις. Αυτές οι αποφάσεις βασίστηκαν σε παράγοντες όπως η απόδοση, η επεκτασιμότητα και η χρηστικότητα. Για παράδειγμα, η απόφαση να χρησιμοποιηθεί το MVC .NET ελήφθη λόγω της ικανότητάς του να χειρίζεται μεγάλες ποσότητες κίνησης και να κλιμακώνεται εύκολα. Επιπλέον το .NET Framework παρέχει πολλές βιβλιοθήκες και εργαλεία για την δημιουργία εφαρμογών Ιστού, διαχείρισης ημιδομημένων δεδομένων και απλοποιεί τα ερωτήματα στην βάση με την χρήση της LINQ (Language Integrated Query).

Η απόφαση να χρησιμοποιηθεί ο Microsoft SQL Server ως βάση δεδομένων ελήφθη λόγω της αξιοπιστίας, της επεκτασιμότητας, της υποστήριξης για τη διαχείριση της σχεσιακής βάσης δεδομένων καθώς και της απλής διαδικασίας σύνδεσης του με το MVC .NET.

4.3 Αποθήκευση και ανάκτηση δεδομένων

Τα δεδομένα αποθηκεύονται στη βάση δεδομένων του SQL Server σε δομημένη μορφή. Στις διαδικασίες όπου χρειάζεται αποθήκευση ή η επεξεργασία μεγάλο όγκου δεδομένων, όπως η αποθήκευση της βιβλιογραφίας, γίνεται μαζική εισαγωγή των εγγράφων ανά 300 εγγραφές ώστε

να μειωθούν τα ερωτήματα στην βάση και να είναι πιο γρήγορη η διαδικασία. Τα δεδομένα ανακτώνται από τη βάση δεδομένων υποβάλλοντας ερωτήματα SQL με την χρήση της LINQ.

4.4 Ασφάλεια συστήματος

Έχουν εφαρμοστεί διάφορα μέτρα ασφαλείας για να διασφαλιστεί ότι το σύστημα είναι ασφαλές. Αυτά τα μέτρα περιλαμβάνουν έλεγχο ταυτότητας με όνομα χρήστη και κωδικό για να συνδεθεί κάποιος στην ιστοσελίδα του διαχειριστικού.

5

Εργαλεία Ανάπτυξης Συστήματος

Σε αυτό το κεφάλαιο θα αναφερθούν τα εργαλεία που χρησιμοποιήθηκαν για την ανάπτυξη της ιστοσελίδας για την διαχείριση των ημιδομημένων δεδομένων.

5.1 Visual Studio 2019

Η ανάπτυξη της ιστοσελίδας του εργαστήριου και του διαχειριστικού, όλες οι λειτουργίες που έχουν αυτά, καθώς και η διαχείριση των ημιδομημένων δομημένων έγιναν στο Visual Studio με χρήση κώδικα C#, HTML, JavaScript, jQuery κ.α.

Είναι ένα ολοκληρωμένο περιβάλλον ανάπτυξης (IDE – Integrated Development Environment) της Microsoft και χρησιμοποιείται για την ανάπτυξη εφαρμογών, προγραμμάτων, ιστοσελίδων, υπηρεσιών ιστού. Το Visual Studio χρησιμοποιεί πλατφόρμες της Microsoft όπως Windows Forms, Windows API, Microsoft Silverlight και Window Store. Παρέχει μια σειρά από δυνατότητες, όπως συμπλήρωση κώδικα (IntelliSense), εργαλεία εντοπισμού σφαλμάτων και φιλική προς το χρήστη διεπαφή που επιτρέπει στους προγραμματιστές να δημιουργούν, να δοκιμάζουν και να αναπτύσσουν εφαρμογές αποτελεσματικά.

5.1.1 LINQ

Η LINQ (Language-Integrated Query) είναι ένα ισχυρό εργαλείο για την αναζήτηση και τον χειρισμό δεδομένων σε εφαρμογές .NET. Επιτρέπει στους προγραμματιστές να χρησιμοποιούν σύνταξη τύπου SQL για πρόσβαση σε δεδομένα που είναι αποθηκευμένα σε διάφορες πηγές, όπως βάσεις δεδομένων, λίστες δεδομένων και έγγραφα XML. Τα ερωτήματα στην βάση δεδομένων καθώς και διάφοροι απαραίτητοι έλεγχοι και διαδικασίες έγιναν με την χρήση LINQ.

5.1.2 XML Serializer

Η XML Serializer είναι μια κλάση που παρέχεται από το Visual Studio και χρησιμοποιείται για τη μετατροπή ενός αντικειμένου σε έγγραφο XML και αντίστροφα. Επιτρέπει στους προγραμματιστές να σειριοποιήσουν εύκολα και να αποσειροποιήσουν ένα αντικείμενο προς και από ένα έγγραφο XML.

Η XML (eXtensible Markup Language) είναι μια δημοφιλής μορφή για την αποθήκευση και την ανταλλαγή δεδομένων, ιδιαίτερα σε εφαρμογές Ιστού. Για να εργαστούν με δεδομένα XML σε .NET, οι προγραμματιστές μπορούν να χρησιμοποιήσουν τις δυνατότητες που παρέχει η κλάση XML Serializer. Επιτρέπει στους προγραμματιστές να μετατρέπουν XML αρχεία σε μοντέλα κλάσεων που έχουν φτιάξει στον κώδικα τους.

Χρησιμοποιήσαμε μεθόδους που παρέχει η κλάση XML Serializer για να μετατρέψουμε τα ημιδομημένα δεδομένα XML σε μια μορφή που μπορούμε εύκολα να διαχειριστούμε και να εξάγουμε την πληροφορία που χρειαζόμαστε. Συγκεκριμένα, με την χρήση της μεθόδου αποσειροποίησης της XML Serializer μπορέσαμε να εξάγουμε την πληροφορία που θέλουμε από τα XML που ανεβάζει ο διαχειριστής στο διαχειριστικό του εργαστηρίου και να συμπληρώσουμε έτσι τις δημοσιεύσεις των μελών.

5.2 Microsoft SQL Server Management Studio (SSMS) 18

Η βάση δεδομένων που χρησιμοποιήθηκε στην ιστοσελίδα εργαστηρίου είναι μια βάση σχεσιακή βάση δεδομένων SQL Server που φτιάχτηκε στο Microsoft SQL Server Management Studio. Μέσα από το περιβάλλον διαχείρισης φτιάχτηκε η βάση και οι πίνακές της.

Το Microsoft SQL Server Management Studio παρέχει μια φιλική προς το χρήστη διεπαφή για προγραμματιστές για τη δημιουργία και την τροποποίηση αντικειμένων βάσης δεδομένων, όπως πίνακες, προβολές (views) και αποθηκευμένες διαδικασίες (stored procedures), και για την εκτέλεση εργασιών χειρισμού δεδομένων.

5.3 Internet Information Services (IIS) Manager (v21H2)

Η υπηρεσία IIS χρησιμοποιήθηκε ώστε να ρυθμιστεί η ιστοσελίδα σε εξυπηρετητή και να μπορούν οι χρήστες να πλοηγηθούν, μέσω του προγράμματος περιηγητή ιστού τους, στην ιστοσελίδα του εργαστηρίου και στην ιστοσελίδα του διαχειριστικού του εργαστηρίου.

Το εργαλείο παρέχει μια φιλική προς το χρήστη διεπαφή για προγραμματιστές για τη διαχείριση των ρυθμίσεων του διακομιστή ιστού και την παρακολούθηση της απόδοσης του διακομιστή.

5.4 GitHub

Το GitHub χρησιμοποιήθηκε για να αποθηκεύσει τον πηγαίο κώδικα της διαδικτυακής εφαρμογής και για να έχουμε έλεγχο εκδόσεων. Η διαδικασία έγινε μέσω του Visual Studio 2019 αφού παρέχει εύκολα και άμεσα την δυνατότητα διασύνδεσης του με το GitHub.

Το GitHub είναι μια διαδικτυακή πλατφόρμα, με χρήση του Git, για έλεγχο εκδόσεων και συνεργασία μεταξύ χρηστών που επιτρέπει στους προγραμματιστές να αποθηκεύουν και να

διαχειρίζονται κώδικα, καθώς και να συνεργάζονται με άλλους προγραμματιστές σε έργα λογισμικού. Το εργαλείο παρέχει μια σειρά λειτουργιών, όπως επισκόπηση κώδικα, παρακολούθηση ζητημάτων και εργαλεία διαχείρισης έργου που ενισχύουν τη διαδικασία ανάπτυξης και διευκολύνουν τη συνεργασία μεταξύ των μελών της ομάδας.

5.4.1 Έλεγχος Εκδόσεων

Ο έλεγχος εκδόσεων είναι η διαχείριση αλλαγών σε έγγραφα, αρχεία ή κώδικα με την πάροδο του χρόνου. Είναι μια βασική διαδικασία στη ανάπτυξη λογισμικού και χρησιμοποιείται για την παρακολούθηση διαφορετικών εκδόσεων του πηγαίου κώδικα, διευκολύνοντας τους προγραμματιστές να συνεργάζονται και να διατηρούν μια βάση κώδικα με την πάροδο του χρόνου.

Τα συστήματα ελέγχου εκδόσεων επιτρέπουν στους προγραμματιστές να παρακολουθούν αλλαγές σε μια βάση κώδικα, να προβάλλουν ιστορικές εκδόσεις κώδικα, να συγχωνεύουν αλλαγές που έγιναν από διαφορετικούς προγραμματιστές και να επιστρέφουν σε προηγούμενες εκδόσεις του κώδικα όταν χρειάζεται. Παρέχουν επίσης δυνατότητες όπως η διακλάδωση (branching), η οποία επιτρέπει στους προγραμματιστές να δημιουργούν διαφορετικές εκδόσεις μιας βάσης κωδικών για διαφορετικούς σκοπούς, όπως δοκιμή νέων λειτουργιών ή διόρθωση σφαλμάτων.

6

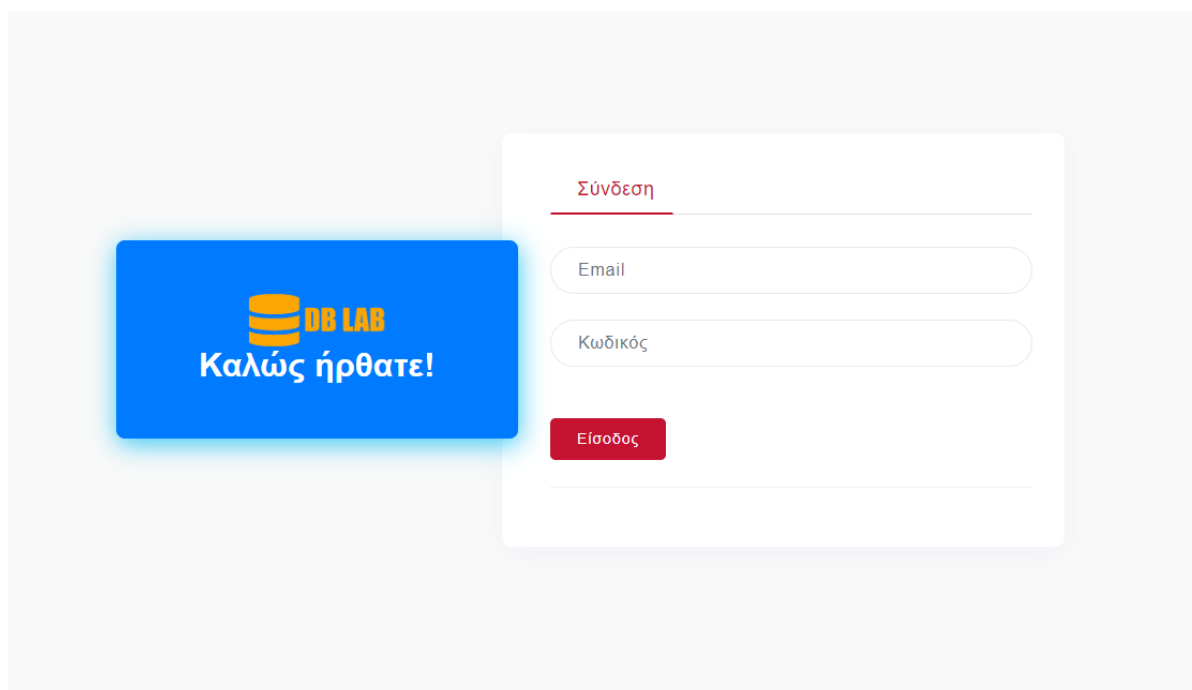
Παρουσίαση Ιστοσελίδας Εργαστηρίου - Διαχειριστικό

Σε αυτό το κεφάλαιο θα γίνει μια περιγραφή για το περιβάλλον που αναπτύχθηκε για τις ανάγκες της διπλωματικής εργασίας. Το σύστημα που υλοποιήθηκε έχει δύο διακριτά μέρη, το Διαχειριστικό (Dashboard) όπου βλέπει ο διαχειριστής και την Ιστοσελίδα Εργαστηρίου όπου βλέπουν οι απλοί χρήστες.

6.1 Διαχειριστικό - Dashboard

Το Dashboard είναι η ιστοσελίδα διαχείρισης των πληροφοριών που εμφανίζονται στην ιστοσελίδα του εργαστηρίου. Θα ακολουθήσει ανάλυση του Dashboard καθώς και ένας σύντομος οδηγός για την σωστή λειτουργία του.

6.1.1 Σύνδεση

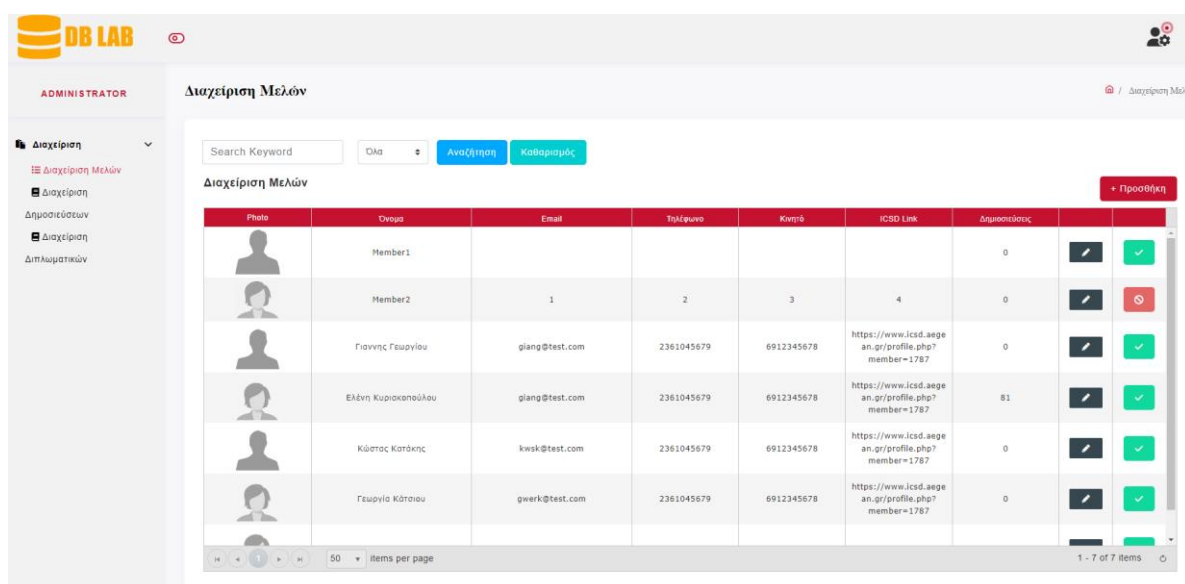


Εικόνα 4: Σελίδα σύνδεσης διαχειριστή στο Dashboard

Όταν ο διαχειριστής ανοίγει την σελίδα του Dashboard και δεν είναι συνδεδεμένος, μεταφέρεται στην σελίδα σύνδεσης διαχειριστή (Εικόνα 4) όπου συμπληρώνει στην φόρμα τα στοιχεία email και κωδικό ώστε να συνδεθεί στο διαχειριστικό.

Όταν συμπληρωθούν τα στοιχεία και πατηθεί το κουμπί είσοδος γίνεται έλεγχος για το αν είναι έγκυρα τα στοιχεία του. Συγκεκριμένα στέλνεται στον διακομιστή με AJAX αίτημα η φόρμα με τα στοιχεία και ελέγχεται ο πίνακας tblusers για το αν περιέχει ενεργό χρήστη με το email και τον κωδικό που συμπληρώθηκε. Εάν δεν βρεθεί ενεργός διαχειριστής με αυτά τα στοιχεία εμφανίζεται σχετικό μήνυμα στον χρήστη. Εάν υπάρχει ενεργός διαχειριστής με αυτά τα στοιχεία, ο διακομιστής ανακατευθύνει τον χρήστη στην σελίδα Διαχείρισης Μελών. Αποθηκεύεται σχετικό HTTP Cookie στο πρόγραμμα περιήγησης ώστε να μην χρειάζεται ο διαχειριστής να κάνει συνέχεια την διαδικασία σύνδεσης. Ο διαχειριστής πηγαίνει στην σελίδα που θέλει από το μενού που βρίσκεται στα αριστερά της οθόνης.

6.1.2 Διαχείριση Μελών



Εικόνα 5: Σελίδα διαχείρισης μελών

Η Εικόνα 5 απεικονίζει τη σελίδα διαχείρισης μελών. Στη συγκεκριμένη σελίδα γίνεται η προσθήκη και η επεξεργασία των μελών που εμφανίζονται στην ιστοσελίδα του εργαστήριου. Άμεσα εμφανίζονται βασικές πληροφορίες για τα ήδη αποθηκευμένα μέλη καθώς και ένα κουμπί επεξεργασίας και ένα κουμπί αλλαγής κατάστασης μέλους. Επιπλέον υπάρχει αναζήτηση και φίλτρο κατάστασης για να βρίσκει ο διαχειριστής εύκολα τα μέλη που θέλει.

Οι πληροφορίες που εμφανίζονται στις 9 στήλες του πίνακα μελών αναλύονται παρακάτω.

1. **Εικόνα μέλους:** Σε αυτή την στήλη εμφανίζεται η εικόνα που έχει ανεβάσει ο διαχειριστής για αυτό το μέλος.
2. **Όνομα:** Στην δεύτερη στήλη εμφανίζεται το όνομα του μέλους.
3. **Email:** Στην τρίτη στήλη εμφανίζεται το email του μέλους.
4. **Τηλέφωνο:** Σε αυτή την στήλη εμφανίζεται το τηλέφωνο του μέλους.
5. **Κινητό:** Στην τέταρτη στήλη εμφανίζεται το κινητό του μέλους.
6. **ICSD Link:** Εμφανίζεται ο σύνδεσμος που έχει το μέλος στην ιστοσελίδα της σχολής.
7. **Αριθμός Δημοσιεύσεων:** Σε αυτή την στήλη εμφανίζεται ο αριθμός δημοσιεύσεων του μέλους που υπολογίζεται δυναμικά, από τις δημοσιεύσεις του.
8. **Κουμπί Επεξεργασίας:** Με κλικ σε αυτό το κουμπί ο διαχειριστής μεταφέρεται στην σελίδα επεξεργασίας του συγκεκριμένου μέλους.
9. **Κουμπί Κατάστασης:** Αυτό το κουμπί επιτρέπει στον διαχειριστή να αλλάζει εύκολα και άμεσα την κατάσταση του μέλους από Δημοσιευμένη σε Μη Δημοσιευμένη και το αντίστροφο. Τα μέλη με κατάσταση Δημοσιευμένη φαίνονται με ένα πράσινο κουμπί κατάστασης, ενώ τα μέλη με κατάσταση Μη Δημοσιευμένη φαίνονται με ένα κόκκινο κουμπί κατάστασης.

Εικόνα 6: Σελίδα προσθήκης / επεξεργασίας μέλους

Στη συνέχεια, η Εικόνα 6 αφορά την σελίδα προσθήκης / επεξεργασίας των μελών. Πιο συγκεκριμένα, ο διαχειριστής προσθέτει τις πληροφορίες για το μέλος. Τα υποχρεωτικά πεδία εμφανίζονται με έναν αστερίσκο δίπλα (*), στην συγκεκριμένη περίπτωση υποχρεωτικό πεδίο είναι μόνο το όνομα του μέλους. Για να αποθηκευτεί εικόνα ή να γίνει ανέβασμα XML βιβλιογραφίας, πρέπει πρώτα να έχει αποθηκευτεί το μέλος στην βάση. Βγαίνουν σχετικές ειδοποιήσεις πάνω δεξιά στην οθόνη για να ενημερώνουν τον διαχειριστή. Στην Εικόνα 7 βλέπουμε την ειδοποίηση που εμφανίζεται όταν ο διαχειριστής προσπαθεί να ανεβάσει XML βιβλιογραφίας, χωρίς να έχει αποθηκεύσει πρώτα το μέλος.

Εικόνα 7: Παράδειγμα προσπάθειας να ανέβει XML βιβλιογραφίας χωρίς να έχει αποθηκευτεί το μέλος

6.1.3 Ενημέρωση βιβλιογραφίας μέλους

Όταν το μέλος έχει αποθηκευτεί τότε μπορεί να γίνει ενημέρωση της βιβλιογραφίας του. Παρακάτω φαίνεται το περιβάλλον της ιστοσελίδας dblp.org (Εικόνα 8) καθώς και το πως μπορεί

να γίνει η εξαγωγή της βιβλιογραφίας σε XML αρχείο. Αφού έχουμε βρει στο dblp.org τον χρήστη του οποίου θέλουμε να εξαγάμε την βιβλιογραφία του, πηγαίνουμε τον κέρσορα στο κουμπί “export bibliography” και επιλέγουμε το “XML”. Τότε ανάλογα με το πρόγραμμα περιηγητή ιστού που χρησιμοποιούμε, το αρχείο θα κατέβει ή θα ανοίξει νέο παράθυρο στο πρόγραμμα περιήγησης. Αν ανοίξει σε νέο παράθυρο το αποθηκεύουμε ώστε να ξεκινήσουμε την διαδικασία ενημέρωσης βιβλιογραφίας μέλους. Το αρχείο XML που θα έχουμε αποθηκεύσει φαίνεται στην Εικόνα 9.

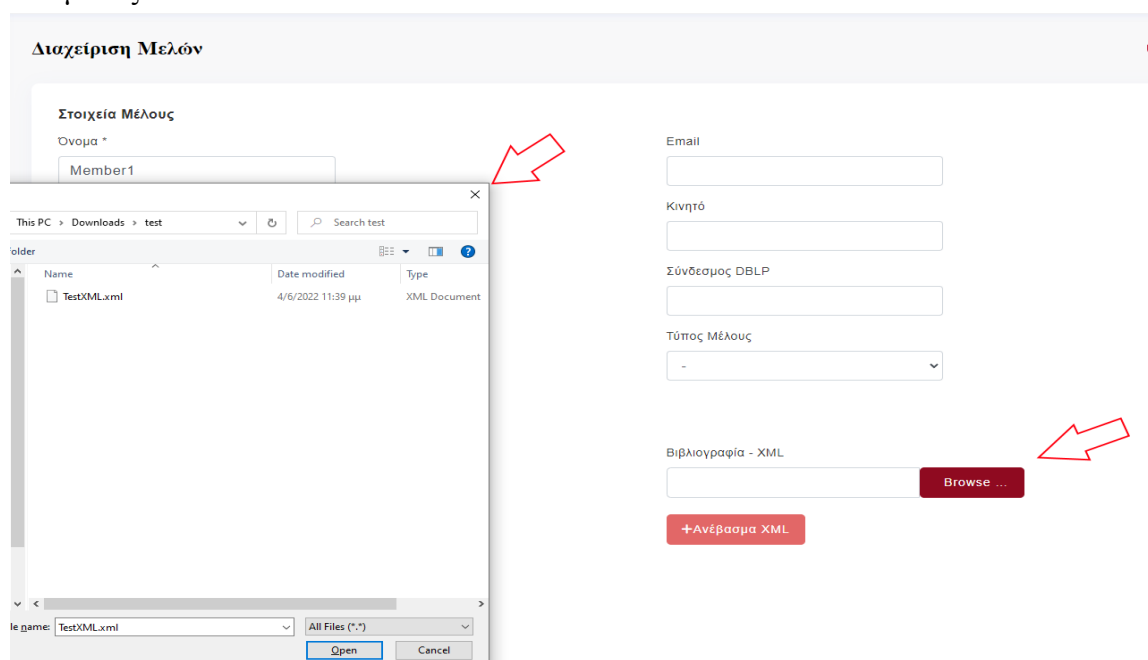
The screenshot shows the dblp.org website interface. At the top, there is a navigation bar with links: home, browse, search, about. Below this, the user profile for Akrivi Vlachou is displayed. A red arrow points to the 'export bibliography' button, which has opened a dropdown menu. In this menu, the 'XML' option is highlighted with another red arrow. The main content area shows a list of publications, with the first two from 2022 and one from 2021. On the right side, there is a 'Refine list' section with a bar chart and various filters for refining the search results.

Εικόνα 8: Σελίδα μέλους στο dblp.org - εξαγωγή βιβλιογραφίας σε XML αρχείο

```
<dblpperson name="Akrivi Vlachou" pid="15/6276" n="84">
  <person key="homepages/15/6276" mdate="2016-02-03">
    <author pid="15/6276">Akrivi Vlachou</author>
    <note type="affiliation">Norwegian University of Science and Technology, Norway</note>
    <url>http://www.idi.ntnu.no/~vlachou</url>
  </person>
</dblpperson>
<article key="journals/fgcs/CuzzocreaKV22" mdate="2021-09-28">
  <author pid="c/AlfredoCuzzocrea">Alfredo Cuzzocrea</author>
  <author pid="08/5342">Panagiotis Karras</author>
  <author pid="15/6276">Akrivi Vlachou</author>
  <title>Effective and efficient skyline query processing over attribute-order-preserving-free encrypted data in cloud-enabled databases.</title>
  <pages>237-251</pages>
  <year>2022</year>
  <volume>126</volume>
  <journal>Future Gener. Comput. Syst.</journal>
  <ee>https://doi.org/10.1016/j.future.2021.08.008</ee>
  <url>db/journals/fgcs/fgcs126.html#CuzzocreaKV22</url>
</article>
<article key="journals/kais/SantipantakisKG22" mdate="2022-09-22">
  <author pid="48/7871">Georgios M. Santipantakis</author>
  <author pid="49/5229">Konstantinos I. Kotis</author>
  <author pid="123/5458">Apostolos Glenis</author>
  <author pid="v/GeorgeAVouros">George A. Vouros</author>
  <author pid="14/3086">Christos Doukeridis</author>
  <author pid="15/6276">Akrivi Vlachou</author>
  <title>RDF-Gen: generating RDF triples from big data sources.</title>
  <pages>2985-3015</pages>
  <year>2022</year>
  <volume>64</volume>
  <journal>Knowl. Inf. Syst.</journal>
  <number>11</number>
  <ee>https://doi.org/10.1007/s10115-022-01729-x</ee>
  <url>db/journals/kais/kais64.html#SantipantakisKG22</url>
</article>
<inproceedings key="conf/dawak/VlachouDRN22" mdate="2022-07-28">
  <author pid="15/6276">Akrivi Vlachou</author>
  <author pid="14/3086">Christos Doukeridis</author>
  <author pid="41/7270">João B. Rocha-Junior</author>
  <author pid="n/KjetilNorvag">Kjetil Norvag</author>
  <title>On Decisive Skyline Queries.</title>
  <pages>61-73</pages>
  <year>2022</year>
```

Εικόνα 9: Αρχείο XML βιβλιογραφίας από dblp.org

Εφόσον αποθηκεύσουμε το αρχείο XML με την βιβλιογραφία στον υπολογιστή μας, κάνουμε κλικ στο κουμπί “Browse” στο πεδίο “Βιβλιογραφία - XML”, όπως φαίνεται στην Εικόνα 10, και σηκώνεται ένα παράθυρο filepicker. Στο filepicker επιλέγουμε το XML αρχείο βιβλιογραφίας που έχουμε κατεβάσει από το dblp.org και πατάμε Άνοιγμα. Το όνομα του αρχείου που επιλέξαμε φαίνεται πλέον στο πεδίο “Βιβλιογραφία - XML” και πατάμε το κουμπί “+ Ανέβασμα XML”. Όταν ολοκληρωθεί η διαδικασία εμφανίζεται μήνυμα επιτυχίας και έχει περαστεί η βιβλιογραφία στο μέλος.



Εικόνα 10: Άνοιγμα filepicker για ανέβασμα αρχείου XML βιβλιογραφίας

Για να είναι εύκολη η διαχείριση της πληροφορίας από το αρχείο XML έχει δημιουργηθεί μοντέλο βάσει της δομής του XML. Αυτό επιτρέπει να γίνει εύκολα αποσειροποίηση του ημιδομημένου XML αρχείου σε μοντέλο, να επεξεργαστεί και εν τέλει να αποθηκευτεί στην βάση μας η πληροφορία με την δομή που θέλουμε.

Στοιχεία Μέλους

Όνομα *

Τηλέφωνο

Σύνδεσμος ICSD

Σύνδεσμος Scholar

Is Published
☒

Φωτογραφία
 Browse ...

Email

Κινητό

Σύνδεσμος DBLP

Τύπος Μέλους

Βιβλιογραφία - XML
 Browse ...

+Ανέβασμα XML **+Διαγραφή Βιβλιογραφίας**

+Ανέβασμα Φωτογραφίας ***Αποθήκευση** **Διαχει**

Εικόνα 11: Επεξεργασία μέλους - Διαγραφή Βιβλιογραφίας

Εάν ένα μέλος έχει ανεβάσει ήδη βιβλιογραφία, εμφανίζεται το κουμπί “Διαγραφή Βιβλιογραφίας” (Εικόνα 11) ώστε εύκολα να μπορεί να διαγράψει την βιβλιογραφία του. Με κλικ στο κουμπί εμφανίζεται μήνυμα για επιβεβαίωση της ενεργείας διαγραφής.

6.1.4 Διαχείριση Δημοσιεύσεων

Διαχείριση Δημοσιεύσεων

Search Keyword Όλα

Τίτλος	Member	Ημ. έκδοσης		
Effective and efficient skyline query processing over attribute-order-preserving-free encrypted data in cloud-enabled databases.	Ελένη Κυριακοπούλου	01/01/2022	✓	✖
Pruning techniques for parallel processing of reverse top-k queries.	Ελένη Κυριακοπούλου	01/01/2021	✓	✖
Parallel and scalable processing of spatio-temporal RDF queries using Spark.	Ελένη Κυριακοπούλου	01/01/2021	✓	✖
A Demonstration of NoDA: Unified Access to NoSQL Stores.	Ελένη Κυριακοπούλου	01/01/2021	✓	✖
A Survey on Big Data Processing Frameworks for Mobility Analytics.	Ελένη Κυριακοπούλου	01/01/2021	✓	✖
A Novel Indexing Method for Spatial-Keyword Range Queries.	Ελένη Κυριακοπούλου	01/01/2021	✓	✖
Declarative Querying of Heterogeneous NoSQL Stores.	Ελένη Κυριακοπούλου	01/01/2021	✓	✖
SPARTAN: Semantic integration of big spatio-temporal data from streaming and archival sources.	Ελένη Κυριακοπούλου	01/01/2020	✓	✖
The SPADES Framework for Scalable Management of Spatio-temporal Data	Ελένη Κυριακοπούλου	01/01/2020	✓	✖

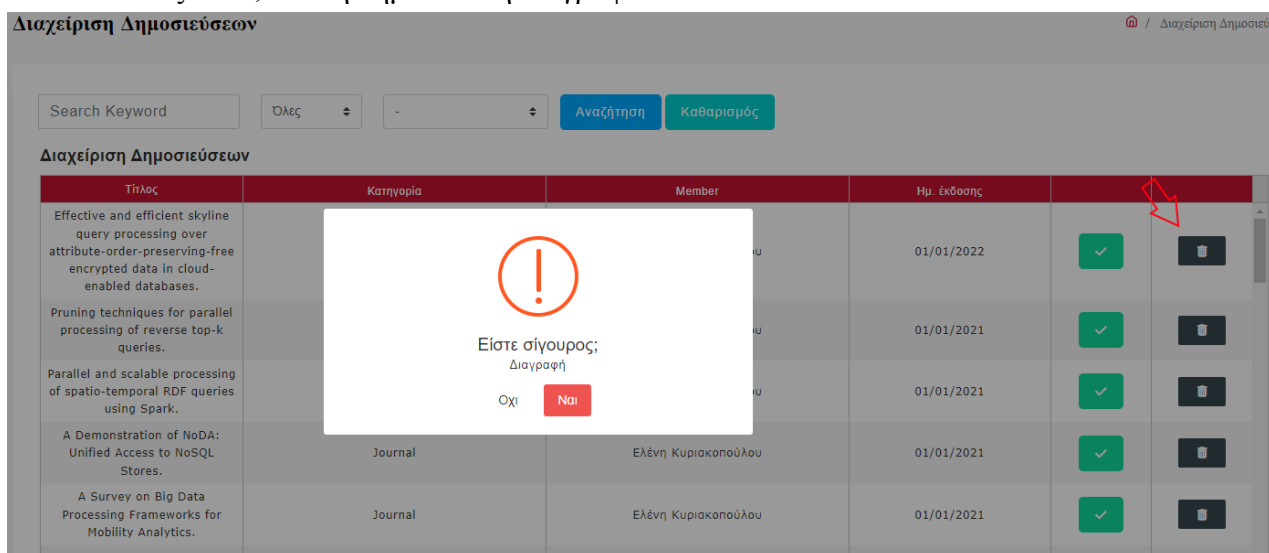
1 - 50 of 84 items

Εικόνα 12: Σελίδα διαχείρισης δημοσιεύσεων

Σε αυτή την σελίδα (Εικόνα 12) εμφανίζονται βασικές πληροφορίες των δημοσιεύσεων. Μπορεί να γίνει αναζήτηση βάσει του τίτλου, να φιλτραριστούν βάσει κατάστασης (Ενεργή - Μη Ενεργή) καθώς και να εμφανιστούν οι δημοσιεύσεις για συγκεκριμένο μέλος. Ο διαχειριστής μπορεί εύκολα να αλλάξει την κατάσταση των δημοσιεύσεων ή να διαγράψει δημοσιεύσεις.

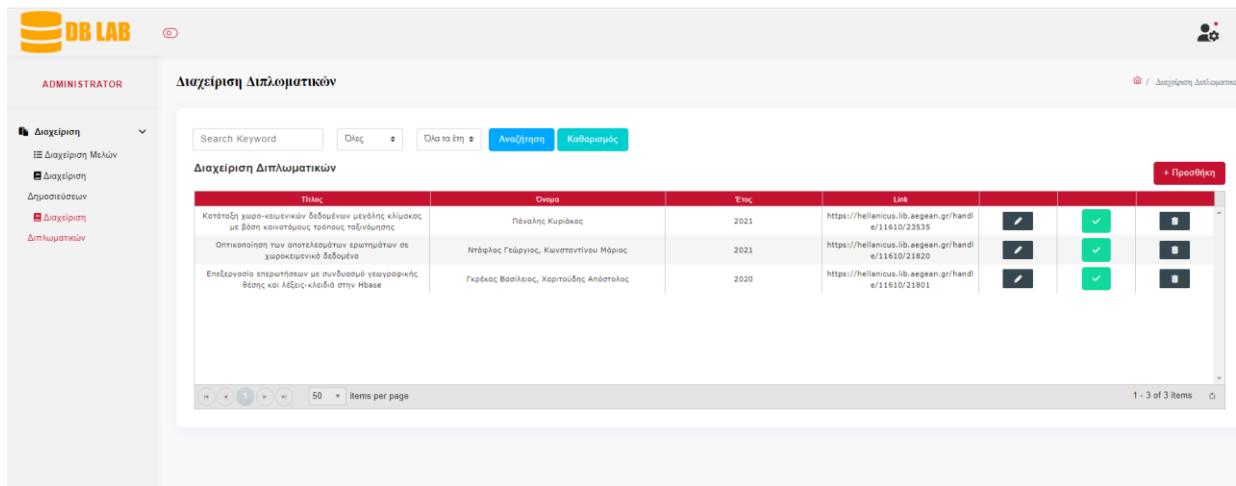
Οι πληροφορίες που εμφανίζονται στις 6 στήλες του πίνακα δημοσιεύσεων αναλύονται παρακάτω.

1. **Τίτλος:** Σε αυτή την στήλη εμφανίζεται ο τίτλος της δημοσίευσης.
2. **Κατηγορία:** Στην δεύτερη στήλη εμφανίζεται η κατηγορία της δημοσίευσης. Οι κατηγορίες είναι Journal, Conference, Book Chapters, Book.
3. **Μέλος:** Στην τρίτη στήλη εμφανίζεται το όνομα του μέλους που του έχει αποδοθεί η δημοσίευση.
4. **Ημερομηνία Έκδοσης:** Σε αυτή την στήλη φαίνεται η ημερομηνία της δημοσίευσης.
5. **Κουμπί Κατάστασης:** Αυτό το κουμπί επιτρέπει στον διαχειριστή να αλλάζει εύκολα και άμεσα την κατάσταση της δημοσίευσης από Δημοσιευμένη σε Μη Δημοσιευμένη και το αντίστροφο. Οι δημοσιεύσεις με κατάσταση Δημοσιευμένη φαίνονται με ένα πράσινο κουμπί κατάστασης, ενώ οι δημοσιεύσεις με κατάσταση Μη Δημοσιευμένη φαίνονται με ένα κόκκινο κουμπί κατάστασης.
6. **Κουμπί Διαγραφής:** Με κλικ σε αυτό το κουμπί εμφανίζεται παράθυρο επιβεβαίωσης, όπως φαίνεται στην εικόνα 13, για την διαγραφή της δημοσίευσης. Εάν ο διαχειριστής επιλέξει ναι, τότε η δημοσίευση διαγράφεται.



Εικόνα 13: Παράθυρο επιβεβαίωσης διαγραφής δημοσίευσης

6.1.5 Διαχείριση Διπλωματικών Εργασιών



Εικόνα 14: Σελίδα διαχείρισης διπλωματικών εργασιών

Η Εικόνα 14 αφορά την σελίδα “Διαχείριση Διπλωματικών”, όπου γίνεται η προσθήκη και η επεξεργασία των διπλωματικών εργασιών που εμφανίζονται στην ιστοσελίδα του εργαστήριου. Άμεσα εμφανίζονται βασικές πληροφορίες για τις ήδη αποθηκευμένες διπλωματικές καθώς και κουμπιά για επεξεργασία, αλλαγή κατάστασης και διαγραφής διπλωματικής. Επιπλέον υπάρχει αναζήτηση, φίλτρο κατάστασης και φίλτρο έτους για να βρίσκει ο διαχειριστής εύκολα τις διπλωματικές εργασίες που θέλει.

Οι πληροφορίες που εμφανίζονται στις 6 στήλες του πίνακα δημοσιεύσεων αναλύονται παρακάτω.

1. **Τίτλος:** Σε αυτή την στήλη εμφανίζεται ο τίτλος της διπλωματικής.
2. **Όνομα:** Στην δεύτερη στήλη εμφανίζονται τα ονόματα των φοιτητών που εκπόνησαν την διπλωματική.
3. **Έτος Παρουσίασης:** Σε αυτή την στήλη φαίνεται το έτος που παρουσιάστηκε η διπλωματική.
4. **Σύνδεσμος:** Στην τέταρτη στήλη φαίνεται ο σύνδεσμος που υπάρχει η διπλωματική.
5. **Κουμπί Επεξεργασίας:** Με κλικ σε αυτό το κουμπί ο διαχειριστής μεταφέρεται στην σελίδα επεξεργασίας της συγκεκριμένης διπλωματικής.
6. **Κουμπί Κατάστασης:** Αυτό το κουμπί επιτρέπει στον διαχειριστή να αλλάζει εύκολα και άμεσα την κατάσταση της διπλωματικής από Δημοσιευμένη σε Μη Δημοσιευμένη και το αντίστροφο. Οι διπλωματικές με κατάσταση Δημοσιευμένη φαίνονται με ένα πράσινο κουμπί κατάστασης, ενώ οι διπλωματικές με κατάσταση Μη Δημοσιευμένη φαίνονται με ένα κόκκινο κουμπί κατάστασης.
7. **Κουμπί Διαγραφής:** Με κλικ σε αυτό το κουμπί εμφανίζεται παράθυρο επιβεβαίωσης για την διαγραφή της διπλωματικής. Εάν ο διαχειριστής επιλέξει ναι, τότε η διπλωματική διαγράφεται.

Διαχείριση Διπλωματικών

Στοιχεία Διπλωματικής

Τίτλος *

Συγγραφέας *

Έτος *

Σύνδεσμος

Is Published

Αποθήκευση Διαχείριση Διπλωματικών

Εικόνα 15: Σελίδα προσθήκης / επεξεργασίας διπλωματικής εργασίας

Η Εικόνα 15 απεικονίζει την σελίδα προσθήκης / επεξεργασίας διπλωματικής εργασίας. Στην συγκεκριμένη σελίδα ο διαχειριστής προσθέτει και επεξεργάζεται τις πληροφορίες για την διπλωματική εργασία. Συμπληρώνει τα πεδία, προαιρετικά το πεδίο σύνδεσμος και αποθηκεύει την διπλωματική.

6.1.6 Αποσύνδεση Διαχειριστή

Εάν ο διαχειριστής θέλει να αποσυνδεθεί από την σελίδα, αρκεί να πάει τον κέρσορα του πάνω δεξιά στο εικονίδιο χρήστη με το γρανάτζι και να πατήσει το “Εξοδος” (Εικόνα 16). Σβήνεται το HTTP Cookie που έχει αποθηκευτεί και ο διαχειριστής ανακατευθύνεται στην σελίδα σύνδεσης διαχειριστή - Dashboard.

ADMINISTRATOR

Διαχείριση Μελών

Search Keyword [] [OK] [Αναζήτηση] [Καθαρισμός]

Διαχείριση Μελών

Photo	Όνομα	Email	Τηλέφωνο	Χώρα	ICSD Link	Δημοσιεύσεις		
	Member1					0		✓
	Member2	1	2	3	4	0		✗
	Γιάννης Γεωργίου	giang@test.com	2361045679	6912345678	https://www.icsd.aegean.gr/profile.php?member=1787	0		✓
	Ελένη Κυριακοπούλου	giang@test.com	2361045679	6912345678	https://www.icsd.aegean.gr/profile.php?member=1787	84		✓
	Κώστας Κατάκης	kwsk@test.com	2361045679	6912345678	https://www.icsd.aegean.gr/profile.php?member=1787	0		✓
	Γεωργία Κάτσιου	gwerk@test.com	2361045679	6912345678	https://www.icsd.aegean.gr/profile.php?member=1787	0		✓

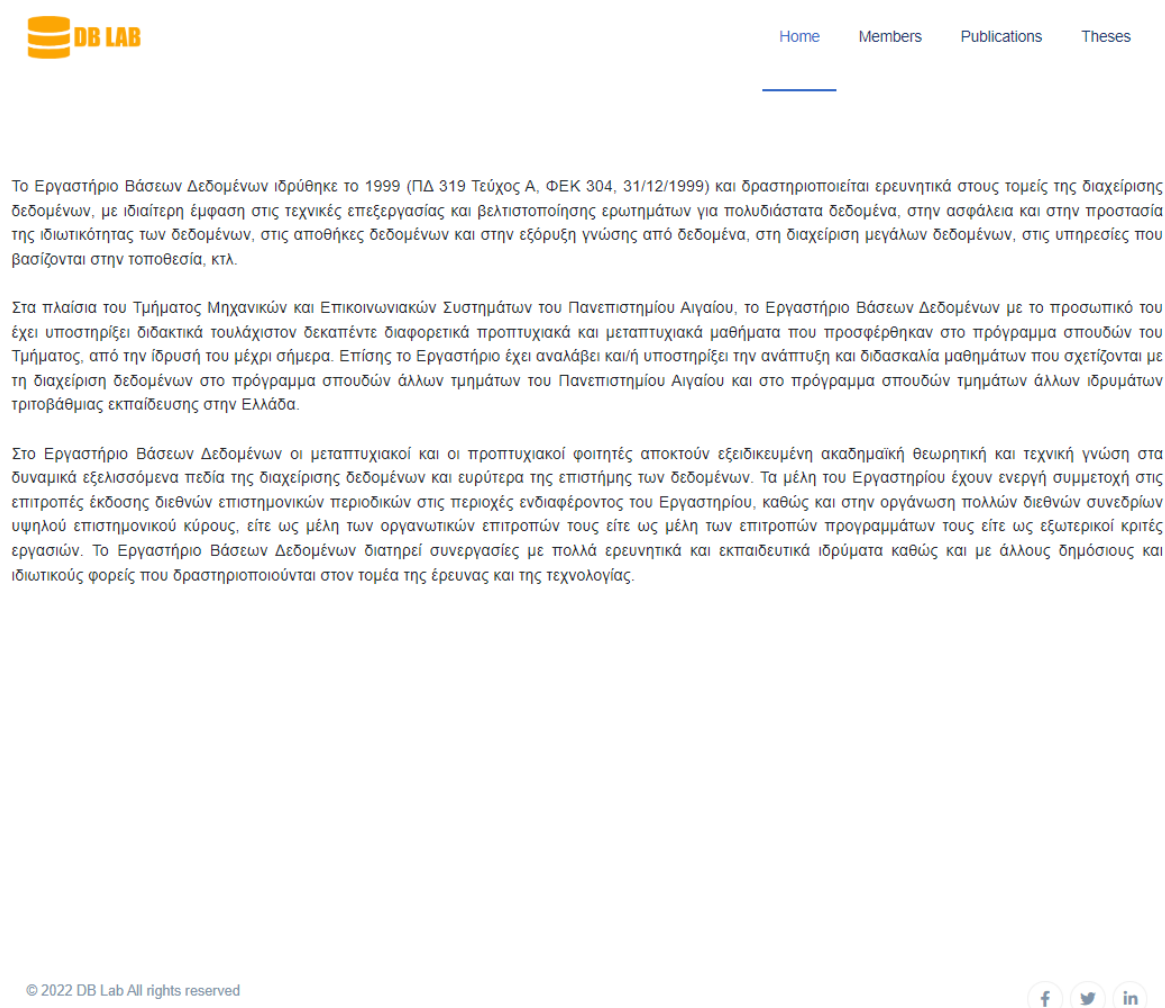
Εξοδος

Εικόνα 16: Αποσύνδεση διαχειριστή

6.2 Σελίδα Εργαστηρίου

Η σελίδα του εργαστηρίου είναι η ιστοσελίδα που θα επισκέπτονται οι χρήστες και εκεί εμφανίζονται όλες οι πληροφορίες που έχει επιλέξει και επεξεργαστεί ο διαχειριστής. Περιέχει μια αρχική σελίδα με μια εισαγωγή για το εργαστήριο και ένα μενού πλοήγησης για να μπορεί να μεταφερθεί και στις υπόλοιπες ενότητες της σελίδας εργαστηρίου.

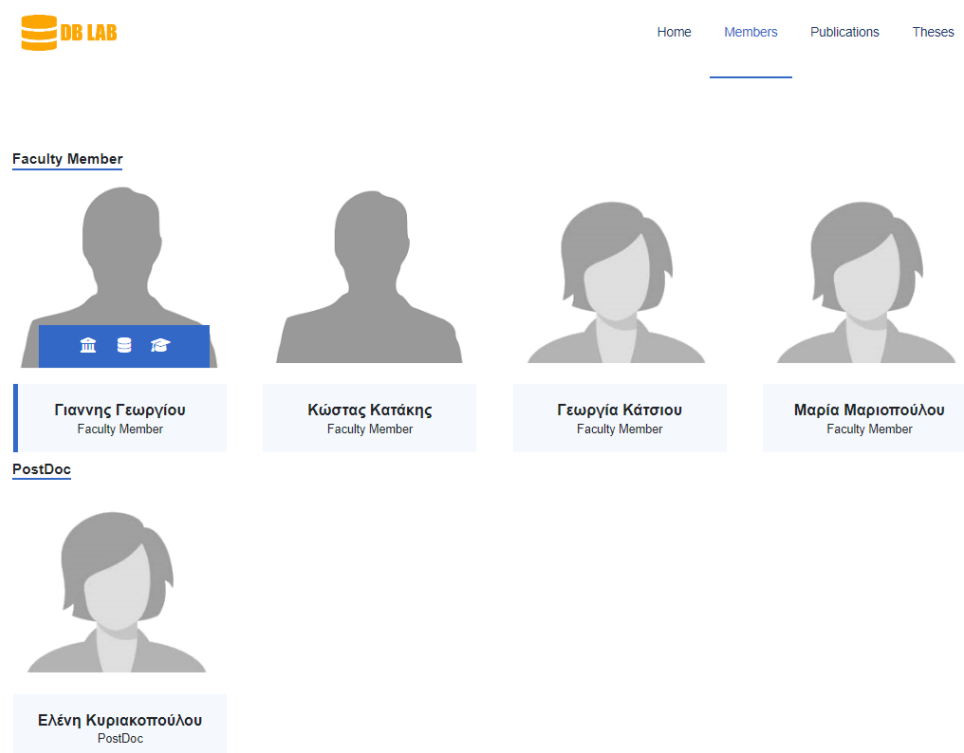
6.2.1 Αρχική Σελίδα



Εικόνα 17: Αρχική σελίδα εργαστηρίου

Στην αρχική σελίδα εμφανίζονται πληροφορίες σχετικά με το εργαστήριο (Εικόνα 17). Πάνω δεξιά βρίσκεται το μενού όπου μπορεί να πλοηγηθεί ο χρήστης στις υπόλοιπες σελίδες του εργαστηρίου. Το μενού αφορά τα μέλη του εργαστηρίου “Members”, τις σχετικές δημοσιεύσεις “Publications” καθώς και τις διπλωματικές εργασίες που έχουν παρουσιαστεί μέχρι στιγμής “Theses”.

6.2.2 Σελίδα Μελών

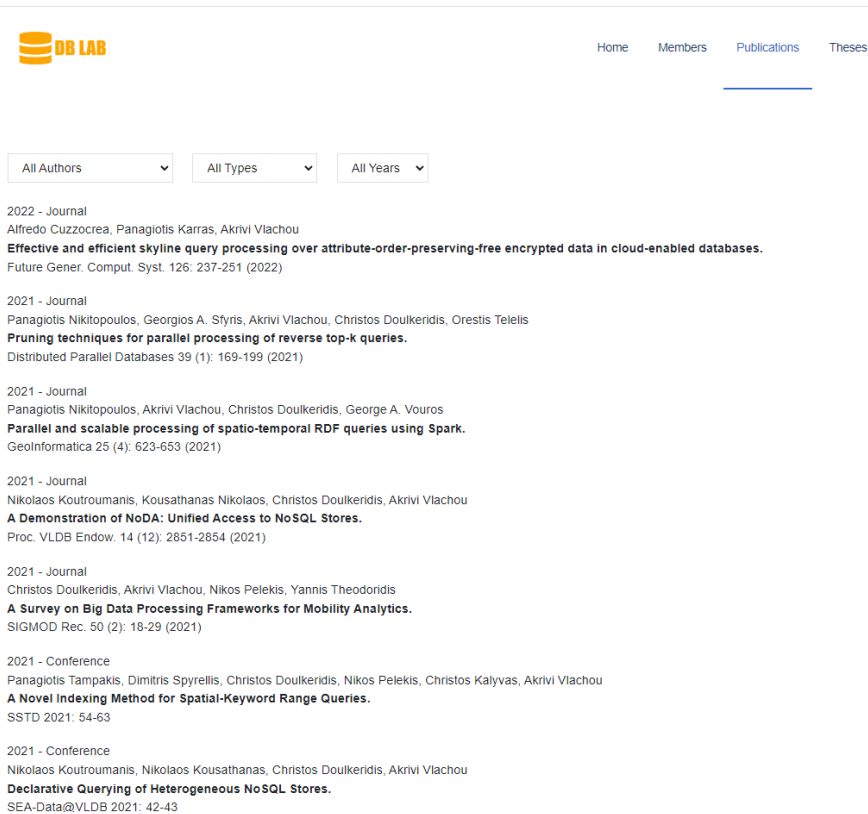


Εικόνα 18: Σελίδα Μελών

Στην Εικόνα 18 φαίνεται η σελίδα μελών, που περιέχει τα μέλη χωρισμένα ανά κατηγορία. Όταν μετακινηθεί ο κέρσορας πάνω από ένα μέλος εμφανίζονται εικονίδια με εξωτερικούς συνδέσμους, εάν έχουν συμπληρωθεί. Από το διαχειριστικό μπορεί να προστεθούν σύνδεσμοι για την σελίδα τους στο πανεπιστήμιο, για την προσωπική τους σελίδα στο dblp.org καθώς και για το Google Scholar. Με κλικ πάνω στο όνομα μέλους, μεταφέρεται ο χρήστης στις δημοσιεύσεις του συγκεκριμένου μέλους.

Εμφανίζονται μόνο τα μέλη που έχουν κατάσταση Δημοσιευμένη. Επιπλέον τα μέλη χωρίζονται σε κατηγορίες Faculty Member, PostDoc, PhD, PostGrad και εμφανίζονται με την σειρά που αναφέρθηκαν.

6.2.3 Σελίδα Δημοσιεύσεων



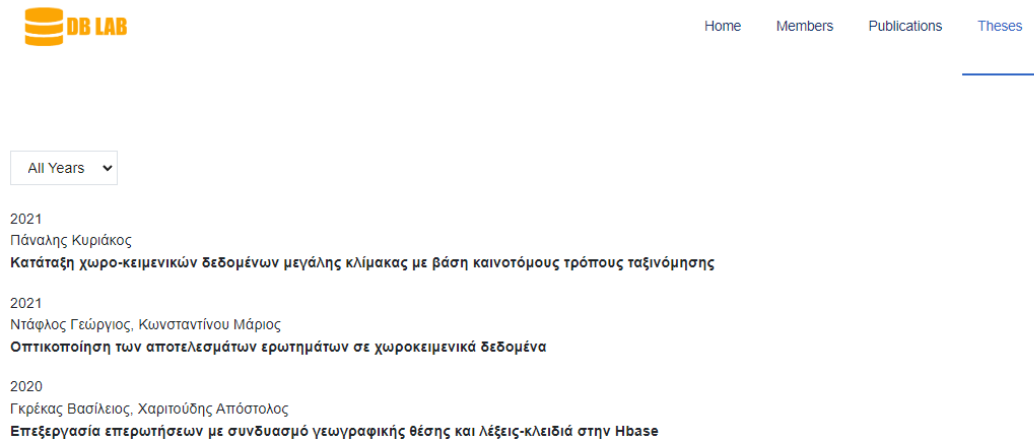
The screenshot shows the DB LAB website's publication page. At the top, there is a navigation bar with links for Home, Members, Publications (which is highlighted), and Theses. Below the navigation bar, there are three dropdown filters: 'All Authors', 'All Types', and 'All Years'. The main content area displays a list of publications, each with its year, journal/conference name, authors, title, and publication details. The publications listed are:

- 2022 - Journal: Alfredo Cuzzocrea, Panagiotis Karras, Akrivi Vlachou. **Effective and efficient skyline query processing over attribute-order-preserving-free encrypted data in cloud-enabled databases.** Future Gener. Comput. Syst. 126: 237-251 (2022)
- 2021 - Journal: Panagiotis Nikitopoulos, Georgios A. Sfyris, Akrivi Vlachou, Christos Doukeridis, Orestis Telelis. **Pruning techniques for parallel processing of reverse top-k queries.** Distributed Parallel Databases 39 (1): 169-199 (2021)
- 2021 - Journal: Panagiotis Nikitopoulos, Akrivi Vlachou, Christos Doukeridis, George A. Vouros. **Parallel and scalable processing of spatio-temporal RDF queries using Spark.** Geoinformatica 25 (4): 623-653 (2021)
- 2021 - Journal: Nikolaos Koutroumanis, Kousathanas Nikolaos, Christos Doukeridis, Akrivi Vlachou. **A Demonstration of NoDA: Unified Access to NoSQL Stores.** Proc. VLDB Endow. 14 (12): 2851-2854 (2021)
- 2021 - Journal: Christos Doukeridis, Akrivi Vlachou, Nikos Pelekis, Yannis Theodoridis. **A Survey on Big Data Processing Frameworks for Mobility Analytics.** SIGMOD Rec. 50 (2): 18-29 (2021)
- 2021 - Conference: Panagiotis Tampakis, Dimitris Spyrellis, Christos Doukeridis, Nikos Pelekis, Christos Kalyvas, Akrivi Vlachou. **A Novel Indexing Method for Spatial-Keyword Range Queries.** SSTD 2021: 54-63
- 2021 - Conference: Nikolaos Koutroumanis, Nikolaos Kousathanas, Christos Doukeridis, Akrivi Vlachou. **Declarative Querying of Heterogeneous NoSQL Stores.** SEA-Data@VLDB 2021: 42-43

Εικόνα 19: Σελίδα Δημοσιεύσεων

Σε αυτή την σελίδα (Εικόνα 19) εμφανίζονται όλες οι δημοσιεύσεις των μελών. Αρχικά φαίνεται το έτος και ο τύπος της δημοσίευσης, στην συνέχεια φαίνονται οι συγγραφείς και μετά με έντονη γραμματοσειρά ο τίτλος. Τέλος εμφανίζεται η υπόλοιπη πληροφορία που περιέχει που μπορεί να διαφέρει ανάλογα τον τύπο δημοσίευσης, όπως κωδικός βιβλίου, κεφάλαιο, σελίδες, συνέδριο κ.α.. Ο χρήστης έχει την δυνατότητα να φιλτράρει τις δημοσιεύσεις, ανά συγγραφέα-μέλος, ανά τύπο και ανά έτος. Οι επιλογές των φίλτρων έρχονται δυναμικά. Στο φίλτρο “Συγγραφείς” εμφανίζονται μόνο τα μέλη που έχουν δημοσιεύσεις και στο φίλτρο “Έτος” εμφανίζονται μόνο τα έτη με δημοσιεύσεις.

6.2.4 Σελίδα Διπλωματικών Εργασιών



Εικόνα 20: Σελίδα Διπλωματικών Εργασιών

Στην σελίδα διπλωματικών εργασιών (Εικόνα 20) ο χρήστης μπορεί να δει όλες τις διπλωματικές εργασίες που έχουν προστεθεί από το διαχειριστικό και έχουν κατάσταση Δημοσιευμένη. Εμφανίζονται με φθίνουσα ταξινόμηση βάσει του έτους. Αρχικά φαίνεται το έτος, μετά οι συγγραφείς και τέλος ο τίτλος της διπλωματικής. Οι διπλωματικές εργασίες μπορούν να φιλτραριστούν βάσει έτους. Οι τιμές που περιέχει το φίλτρο του έτους δημιουργούνται δυναμικά βάσει του πεδίου “έτος” που έχουν οι διπλωματικές.

7

Συμπεράσματα

Η διαχείριση ημιδομημένων δεδομένων είναι μια πολύ σημαντική διαδικασία στις σύγχρονες εφαρμογές που βασίζονται σε δεδομένα από εξωτερικές πηγές. Ο σχεδιασμός και η ανάπτυξη του ιστότοπου του εργαστηρίου “Βάσεις δεδομένων” του Τμήματος Μηχανικών Πληροφοριακών και Επικοινωνιακών Συστημάτων του Πανεπιστημίου Αιγαίου, αναδεικνύει τη δυνατότητα αξιοποίησης ημιδομημένων δεδομένων από εξωτερικές πηγές για την απόκτηση χρήσιμης πληροφορίας. Στην συγκεκριμένη περίπτωση η βιβλιογραφία των μελών του εργαστηρίου, εισάχθηκε απλά με την μεταφόρτωση των αρχείων XML από το dblp.org στην σελίδα επεξεργασίας του εκάστοτε μέλους του εργαστηρίου. Μέσω των διαδικασιών που εκτελέστηκαν η πληροφορία μετατράπηκε σε δομημένη, αποθηκεύτηκε στην βάση και από εκεί και πέρα μπορούμε να την αξιοποιήσουμε.

Λόγω της δομής και των εργαλείων που έχουν χρησιμοποιηθεί στο σύστημα, είναι σχετικά εύκολη η επέκταση των λειτουργιών του. Η πιο σημαντική επέκταση που μπορεί να γίνει είναι η προσθήκη επιπλέον πηγών ημιδομημένων δεδομένων, ώστε να έχουμε περισσότερες πληροφορίες στην διάθεσή μας. Αυτό φυσικά συνεπάγεται και την συνεχή ανάπτυξη και βελτιστοποίηση διαδικασίας ενοποίησης δεδομένων με κάθε προσθήκη νέας πηγής, ώστε να έχουμε πάντα την πιο σωστή και ολοκληρωμένη πληροφορία που μπορούμε.

Βιβλιογραφία

- [1] C. Batini, C. Cappiello, C. Francalanci, and A. Maurino, “Methodologies for data quality assessment and improvement,” *ACM Comput Surv*, vol. 41, no. 3, Jul. 2009, doi: 10.1145/1541880.1541883.
- [2] S. Abiteboul, Serge, Buneman, Peter, D. Suciu, and Dan, “Data on the Web: From Relations to Semistructured Data and XML,” 2000.
- [3] S. Davidson and D. Suciu, “Adding Structure to Unstructured Data,” 1997. [Online]. Available: <http://repository.upenn.edu/cis>
- [4] F. Cauteruccio, P. Lo Giudice, L. Musarella, G. Terracina, D. Ursino, and L. Virgili, “A Lightweight Approach to Extract Interschema Properties from Structured, Semi-Structured and Unstructured Sources in a Big Data Scenario,” *Int J Inf Technol Decis Mak*, vol. 19, no. 3, pp. 849–889, May 2020, doi: 10.1142/S0219622020500182.
- [5] Y. Papakonstantinou, H. Garcia-Molina, and J. Widom, “Object Exchange Across Heterogeneous Information Sources.”
- [6] D. Skoutas and A. Simitsis, “Ontology-Based Conceptual Design of ETL Processes for Both Structured and Semi-Structured Data,” *Int. J. Semantic Web Inf. Syst.*, vol. 3, pp. 1–24, Jul. 2007, doi: 10.4018/jswis.2007100101.
- [7] D. Milano, M. Scannapieco, and T. Catarci, “Using ontologies for XML data cleaning,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2005, pp. 562–571. doi: 10.1007/11575863_75.
- [8] N. Swartz, “Gartner warns firms of ‘dirty data,’” *Information Management Journal*, vol. 41, no. 3, 2007.
- [9] I. F. Ilyas and X. Chu, “Trends in cleaning relational data: Consistency and deduplication,” *Foundations and Trends in Databases*, vol. 5, no. 4, pp. 281–393, 2015, doi: 10.1561/19000000045.
- [10] M. Labreche, S. Weill, S. Jean-Pierre Adi, and T. Sébastien, “A general approach for of schema matching problem: case of databases.”
- [11] M. Stonebraker and I. F. Ilyas, “Data Integration: The Current Status and the Way Forward,” *IEEE Data Eng. Bull.*, vol. 41, pp. 3–9, 2018, [Online]. Available: <https://api.semanticscholar.org/CorpusID:49407081>
- [12] A. E. Quora, S. M. Drucker, A. Edmonds, R. W. White, and D. Morris, “Instrumenting the Dynamic Web. Machine Teaching View project Creativity Support Measurement View project

INSTRUMENTING THE DYNAMIC WEB,” 2007. [Online]. Available: <https://www.researchgate.net/publication/220538168>

- [13] S. K. Patel, A. Prof Acharya, M. Patel, V. R. Rathod, and J. B. Prajapati, “Performance Analysis of Content Management Systems-Joomla, Drupal and WordPress General Terms Open source Content Management System,” 2011.
- [14] S. I. Ahmad, T. Rana, and A. Maqbool, “A Model-Driven Framework for the Development of MVC-Based (Web) Application,” *Arab J Sci Eng*, vol. 47, no. 2, pp. 1733–1747, Feb. 2022, doi: 10.1007/s13369-021-06087-4.
- [15] G. E. Krasner and S. T. Pope, “A Description of the Model-View-Controller User Interface Paradigm in the Smalltalk-80 System,” 1988.