**MARMARA UNIVERSITY**

**FACULTY OF ENGINEERING**



**DESIGN AND PROGRAM A HEALTH NEWS RSS AGGREGATOR**

Atilla Gündoğan

**GRADUATION PROJECT REPORT**

Department of Electrical and Electronics Engineering

**Supervisor**

Asst. Prof. Salih Bayar ISTANBUL, 2022

**MARMARA UNIVERSITY**

**FACULTY OF ENGINEERING**

**DESIGN AND PROGRAM A HEALTH NEWS RSS AGGREGATOR**

**by**

**Atilla Gündoğan**

**June 18, 2022, Istanbul**

**SUMBITTED TO THE DEPARTMENT OF ELECTRICAL AND ELECTRONICS ENGINEERING IN**

**PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE**

**OF**

**BACHELOR OF SCIENCE**

**AT**

**MARMARA UNIVERSITY**

Shape

Description automatically generated**The author(s) hereby grant(s) to Marmara University permission to reproduce and to distribute publicly paper and electronic copies of this document in whole or in part and declare that the prepared document does not in any way include copying of previous work on the subject or the use of ideas, concepts, words, or structures regarding the subject without appropriate acknowledgement of the source material.**

Signature of Author(s)

Department of Electrical and Electronics Engineering

Certified By

Project Supervisor, Department of Electrical and Electronics Engineering

Accepted By

Head of the Department of Electrical and Electronics Engineering

# ACKNOWLEDGEMENTS

I would like to thank our beloved teacher Asst. Prof. Salih Bayar for his valuable support and advice on the project, Design and Program a Health News RSS Aggregator, both for guidance and moral support.

**June, 2022**

**Atilla Gündoğan**

# TABLE OF CONTENTS

[ACKNOWLEDGEMENTS ii](#_Toc106657214)

[TABLE OF CONTENTS iii](#_Toc106657215)

[ABSTRACT v](#_Toc106657216)

[1. LIST OF SYMBOLS vi](#_Toc106657217)

[1.1. SQL Special Character Escape Sequences vi](#_Toc106657218)

[1.2. ASCII Characters vii](#_Toc106657219)

[2. ABBREVIATIONS ix](#_Toc106657220)

[3. LIST OF FIGURES x](#_Toc106657221)

[4. LIST OF TABLES xi](#_Toc106657222)

[5. OPERATORS AND RESERVED WORDS xii](#_Toc106657223)

[5.1. PHP xii](#_Toc106657224)

[5.1.1. PHP Arithmetic Operators xii](#_Toc106657225)

[5.1.2. PHP Assignment Operators xii](#_Toc106657226)

[5.1.3. PHP Comparison Operators xiii](#_Toc106657227)

[5.1.4. PHP Increment / Decrement Operators xiii](#_Toc106657228)

[5.1.5. PHP Logical Operators xiv](#_Toc106657229)

[5.1.6. PHP String Operators xiv](#_Toc106657230)

[5.1.7. PHP Array Operators xiv](#_Toc106657231)

[5.2. SQL xv](#_Toc106657232)

[5.2.1. SQL Special Character Escape Sequences xv](#_Toc106657233)

[5.2.2. SQL Keywords xvi](#_Toc106657234)

[6. INTRODUCTION 1](#_Toc106657235)

[6.1. Thesis Content 2](#_Toc106657236)

[7. RESEARCH OBJECTIVE 3](#_Toc106657237)

[8. RELATED WORKS 4](#_Toc106657238)

[9. DESIGN 5](#_Toc106657239)

[9.1. Realistic constraints and conditions 5](#_Toc106657240)

[9.2. Cost of the design 5](#_Toc106657241)

[9.3. Development Standards 5](#_Toc106657242)

[9.4. Details of the design 5](#_Toc106657243)

[10. METHODS 6](#_Toc106657244)

[10.1. TECHNOLOGIES USED 6](#_Toc106657245)

[11. WHAT IS RSS? 7](#_Toc106657246)

[11.1. Why use RSS? 7](#_Toc106657247)

[11.2. Who Should use RSS? 8](#_Toc106657248)

[11.3. Benefits of RSS 8](#_Toc106657249)

[12. WHAT IS XML? 10](#_Toc106657250)

[13. GITHUB 13](#_Toc106657251)

[14. PHP 16](#_Toc106657252)

[15. MYSQL 16](#_Toc106657253)

[16. MARIADB 16](#_Toc106657254)

[17. Hosting 17](#_Toc106657255)

[18. XAMPP 17](#_Toc106657256)

[19. PHPMyAdmin 20](#_Toc106657257)

[19.1. UTF8 vs UTF8MB4 20](#_Toc106657258)

[20. RESULTS AND DISCUSSION 21](#_Toc106657259)

[21. CONCLUSION 22](#_Toc106657260)

[REFERENCES 23](#_Toc106657261)

[APPENDICES 24](#_Toc106657262)

[Appendix 1 24](#_Toc106657263)

# 

# ABSTRACT

RSS Readers was a kind of social media in the 90s before Facebook and its competitors came along. Today RSS finds its place in the hands of avid readers who wants to keep up with the latest news and innovations. Many websites offer this service but not many do enjoy their benefits. For simplification and greater integration, we gather data from RSS XML files provided by publishers and push them to a SQL database. After pushing the data, we can pull them directly or query them to select one or more specific data with requested features. Once we get our requested data, we can create an algorithm and insert them into a Machine Learning application. We can also integrate the data into an existing application creating news panel. Also, an important note, since these are copyrighted work and shares to users consumption free of charge, we can't manipulate the data inside and add our own data. We can share or use it while giving references.

# LIST OF SYMBOLS

## SQL Special Character Escape Sequences

In SQL Databases like MariaDB or MySQL, we may need to query some special characters. Here is a list of SQL special character escape sequences.

**\0 :** An ASCII NUL (0x00)

**\n :** A newline

**\r :** A carriage return

**\t :** A tab

**\Z :** Control-Z

**\' :** A single quote (')

**\" :** A double quote (")

**\b :** A backspace

**\\ :** A backslash (\)

**\% :** A % character

**\\_ :** A \_ character

## ASCII Characters

ASCII is the first character set that ever used between computers over the internet and it was created at 1963 by ANSI (American National Standards Institute). Latest HTML (HTML5) character set (UTF-8) is built on ASCII.

**&#32;** → Space ( )

**&#33;** → Exclamation mark (!)

**&#34;** → Quotation mark (")

**&#35;** → Number sign (#)

**&#36;** → Dollar sign ($)

**&#37;** → Percent sign (%)

**&#38;** → Ampersand (&)

**&#39;** → Apostrophe (')

**&#40;** → Left Parenthesis [(]

**&#41;** → Right Parenthesis [)]

**&#42;** → Asterisk (\*)

**&#43;** → Plus sign (+)

**&#44;** → Comma (,)

**&#45;** → Hyphen (-)

**&#46;** → Period (.)

**&#47;** → Slash (/)

**&#48;** → Digit 0 (0)

**&#49;** → Digit 1 (1)

**&#50;** → Digit 2 (2)

**&#51;** → Digit 3 (3)

**&#52;** → Digit 4 (4)

**&#53;** → Digit 5 (5)

**&#54;** → Digit 6 (6)

**&#55;** → Digit 7 (7)

**&#56;** → Digit 8 (8)

**&#57;** → Digit 9 (9)

**&#58;** → Colon (:)

**&#59;** → Semicolon (;)

**&#60;** → Less-than (<)

**&#61;** → Equals sign (=)

**&#62;** → Greater than (>)

**&#63;** → Question mark (?)

**&#64;** → At sign (@)

**&#91;** → Left square bracket ([)

**&#92;** → Backslash (\)

**&#93;** → Right square bracket (])

**&#94;** → Caret (^)

**&#95;** → Underscore (\_)

**&#96;** → Grave accent (`)

**&#123;** → Left curly brace ({)

**&#124;** → Vertical bar (|)

**&#125;** → Right curly brace (})

**&#126;** → Tilde (~)

# ABBREVIATIONS

**RSS:** Really Simple Syndication

**XML:** eXtensible Markup Language

**VPS:** Virtual Private Server

**VDS:** Virtual Dedicated Server

**VM**: Virtual Machine

**PHP:** Hypertext Pre-processor

**SQL:** Structured Query Language

**NOSQL:** Not Only Structured Query Language

**CPU:** Central Processing Unit

**RAM:** Random Access Memory

**XAMPP:** Cross-platform, Apache, MySQL, PHP, and Perl

**IaaS:** Infrastructure as a Service

**PaaS:** Platform as a Service

**SaaS:** Software as a Service

**ASCII:** American Standard Code for Information Interchange

**CLI:** Command Line Interface

**GB:** Giga Bytes

**GUI:** Graphical User Interface

**MHz:** Mega Hertz

**SEQUEL:** Structured English Query Language

**TCP:** Transport Control Protocol

**WWW:** World Wide Web

# LIST OF FIGURES

[Figure 1 Diagram of an example RSS feed 12](#_Toc106852173)

[Figure 2 Content of an example RSS feed 12](#_Toc106852174)

# LIST OF TABLES

[Table 1 PHP Arithmetic Operators xiii](#_Toc106852185)

[Table 2 PHP Assignment Operators xiii](#_Toc106852186)

[Table 4 PHP Increment and Decrement Operators xiv](#_Toc106852187)

[Table 3 PHP Comparison Operators xiv](#_Toc106852188)

[Table 5 PHP Logical Operators xv](#_Toc106852189)

[Table 6 PHP String Operators xv](#_Toc106852190)

[Table 7 PHP Array Operators xv](#_Toc106852191)

[Table 8 SQL Keywords xvi](#_Toc106852192)

# OPERATORS AND RESERVED WORDS

## PHP

We perform operations on variables and their values with operators. They are grouped as below.

### PHP Arithmetic Operators

Table 1 PHP Arithmetic Operators

|  |  |  |
| --- | --- | --- |
| Operator | Name | Example |
| - | Subtraction | $x - $y |
| + | Addition | $x + $y |
| \* | Multiplication | $x \* $y |
| \*\* | Exponentiation | $x \*\* $y |
| / | Division | $x / $y |
| % | Modulus | $x % $y |

### PHP Assignment Operators

Table 2 PHP Assignment Operators

|  |  |
| --- | --- |
| Assignment | Equal To |
| x = y | x = y |
| x -= y | x = x - y |
| x += y | x = x + y |
| x \*= y | x = x \* y |
| x %= y | x = x % y |
| x /= y | x = x / y |

### PHP Increment / Decrement Operators

Table 4 PHP Increment and Decrement Operators

|  |  |  |
| --- | --- | --- |
| Operator | Name | Description |
| ++$x | Pre-increment | Increments first, then returns $x |
| --$x | Pre-decrement | Decrements first, then returns $x |
| $x++ | Post-increment | Returns $x first, then increments $x |
| $x-- | Post-decrement | Returns $x first, then decrements $x |

### PHP Comparison Operators

Table 3 PHP Comparison Operators

|  |  |  |
| --- | --- | --- |
| Operator | Name | Example |
| == | Equal | $x == $y |
| === | Identical | $x === $y |
| != | Not equal | $x != $y |
| !== | Not identical | $x !== $y |
| < | Less than | $x < $y |
| > | Greater than | $x > $y |
| <= | Less or equal to | $x <= $y |
| >= | Greater or equal to | $x >= $y |

### PHP Logical Operators

Table 5 PHP Logical Operators

|  |  |  |
| --- | --- | --- |
| Operator | Name | Example |
| and | And | $x and $y |
| && | And | $x && $y |
| or | Or | $x or $y |
| || | Or | $x || $y |
| xor | Xor | $x xor $y |
| ! | Not | !$x |

### PHP String Operators

Table 6 PHP String Operators

|  |  |  |  |
| --- | --- | --- | --- |
| Operator | Name | Example | Explanation |
| . | Concatenation | $x . $y | Add $x and $y |
| .= | Concatenation assignment | $x .= $y | Append $y after $tx |

### PHP Array Operators

Table 7 PHP Array Operators

|  |  |  |
| --- | --- | --- |
| Operator | Name | Example |
| + | Union | $x + $y |
| == | Equality | $x == $y |
| != | Inequality | $x != $y |
| <> | Inequality | $x <> $y |
| === | Identity | $x === $y |
| !== | Non-identity | $x !== $y |

## SQL

This part contains SQL reserved words (keywords) which can be used both in MySQL and MariaDB. Reserved keywords have a certain functionality and should not be used to declare variables, tables, or columns.

### SQL Keywords

Table 8 SQL Keywords

|  |  |  |  |
| --- | --- | --- | --- |
| ADD | ADD CONSTRAINT | ALL | ALTER |
| ALTER COLUMN | ALTER TABLE | AND | ANY |
| AS | ASC | WHERE | BETWEEN |
| CASE | CHECK | COLUMN | CONSTRAINT |
| CREATE | CREATE PROCEDURE | CREATE INDEX | VALUES |
| CREATE TABLE | CREATE DATABASE | VIEW | CREATE VIEW |
| DATABASE | DEFAULT | DELETE | DESC |
| DISTINCT | DROP | DROP COLUMN | DROP CONSTRAINT |
| DROP DATABASE | DROP DEFAULT | DROP INDEX | DROP TABLE |
| DROP VIEW | EXEC | EXISTS | FOREIGN KEY |
| FROM | FULL OUTER JOIN | GROUP BY | HAVING |
| IN | INDEX | INNER JOIN | INSERT INTO |
| INSERT INTO SELECT | IS NULL | IS NOT NULL | JOIN |
| LEFT JOIN | LIKE | LIMIT | NOT |
| NOT NULL | OR | ORDER BY | OUTER JOIN |
| PRIMARY KEY | PROCEDURE | RIGHT JOIN | ROWNUM |
| SELECT | SELECT DISTINCT | SELECT INTO | SELECT TOP |
| SET | TABLE | TOP | TRUNCATE TABLE |
| UNION | UNION ALL | UNIQUE | UPDATE |

# INTRODUCTION

Today, readers find themself in a huge pile of information. There is just simply too many publishers, news and researches to keep up with. RSS feeds helped those readers until the 90s but then the technology started to lose its place and now the average person doesn't even know what an RSS feed is. Even though many websites, publishers, blogs and even forum sites offer this service there is not an intuitive way to use this information. In this project we gathered data from RSS XML files from publishers and inserted into a SQL Database so we can use it in other applications and projects. After inserting the data, we can pull them directly or query them to select one or more specific data with requested features.

## Thesis Content

The content should include the following sections.

BU SECTIONDAN ITIBAREN DUZENLE

# RESEARCH OBJECTIVE

Although having too much information doesn't sound bad at all it can have a negative impact. Having too much unorganized information can make the desired info buried under unrelated topics and researches. But even too much organized information can cause problems. One of those problems is information overload. Information overload is probably something you have felt before after hours of researching the web and you realize that there is too much data floating around and you can no longer make decisions about the topic. To prevent this is in this project we only gathered data from Turkish Heath News Publishers. We defined 6 categories including Title, Link, Media, Meta Description, Content and Publication Date. After inserting the data into a database, we can search and filter the data based on containing keywords in the content, name of the publisher, publication date, title, and content itself. Also, only the data in the form of RSS is not helpful. We need an RSS Reader or an RSS Aggregator. The RSS data is in XML Form, and it is hard to read and not intuitive enough for the public. With the help of PHP, we will also display the database content in a web page, that will make it easier to read the data without full permission access to the database itself.

# RELATED WORKS

There is a few closed-source RSS Aggregators like Feedly and Inoreader which you can't run it self-hosted. And then there is NewsBlur, the biggest open-source RSS Reader with a plenty of features. But you must run it a dedicated server alongside Docker. Because you will need a dedicated server which may cost a lot and may add up with the number of publishers and data queries. Also, the code base is very large and thus it is hard to understand without a long read into the documentation. With the scripts in this project, you can run it cost-free alongside your website or on a very cheap website hosting.

There is a few closed-source RSS Aggregators which you can't run it self-hosted. And then there is NewsBlur, the biggest open-source RSS Reader with a plenty of features. But you must run it a dedicated server alongside Docker. Because you will need a dedicated server it may cost a lot. Also, the code base is very large and thus it is hard to understand without a long read into the documentation.

# DESIGN

## Realistic constraints and conditions

Since the RSS data contains copyrighted work we can't edit its content. We can only use it for ourselves or share it while giving credits to the authors and the publisher.

## Cost of the design

Cost of implementing the codes alongside a website installation with an admin panel installation is free. If there is no hosting at hand you can either purchase a cheap web hosting or host the scripts on a personal computer with XAMMP.

## Development Standards

While developing an app or a program we must follow the Programming Language's Syntax, requirements, and standards. Since we only used PHP in this project, we must use it accordingly in an intended way and syntax.

Standards for PHP: <https://www.php-fig.org/psr/>

Also their is industry standards of software development.

Standards for software from American National Standards Institute: <https://webstore.ansi.org/industry/software>

Standards for SQL Database:

<https://webstore.ansi.org/industry/software/technology-languages/SQL>

## Details of the design

The scripts work under 7 main folders. They communicate with require tags. There are 36 PHP files in total, and they all help the project work as intended.

Graphical user interface, application

Description automatically generated

Figure 1 Folder Structure Of The Project

# METHODS

Design (implementation/simulation studies), The experimental setups/the algorithms/the HW designs must be mentioned in detail. The logic behind the study must be explained.

I used GitHub to save and track my progress while I developed new features.

For development purposes I used XAMPP to run Apache, MySQL and phpMyAdmin locally.

Apache is the most used free and open-source webserver software that allows users to deploy their websites on the internet.

I used Visual Studio Code through the entire development process. It is a free and open-source powerful IDE from Microsoft.

PHP stands for Hypertext Preprocessor. PHP can generate dynamic page content, create, open, read, write, and delete files on the server.

MySQL is a relational database which we used to push or insert our data into. We can create databases, create tables into those databases and execute read write operations. We can see our database with the help of phpMyAdmin. Below there is the screenshot of our database configuration.

## TECHNOLOGIES USED

# INTRODUCTION TO RSS

In this project we gather resources through RSS feeds. RSS feeds are used to distribute latest web content from one or more website (source) to the end users (client - receiver). End users can read the data contents with an RSS Reader or an Aggregator.

* Abbreviation of RSS is Really Simple Syndication.
* RSS can be called as a News feed or RSS feed.
* RSS allows fast searching / viewing of latest news and articles without bloat and make it able to gather all contents in on place organized under categories or folders.
* RSS does not follow a guideline and every web publisher can create their non-standard tags.
* It makes it easy to share and display content and data about the content.
* RSS can be configured to update automatically.
* RSS can ben personalized.
* RSS is written/shared is XML format.
* RSS data is small in size and loads fast.
* RSS is read through aggregators.
* RSS aggregator is a website or a program that collects, sorts out and display RSS feeds.

## Why use RSS?

RSS was designed to display specific data. Without RSS, readers will be required to check every website that they want to keep up with. This would consume a lot of time. RSS feeds bring the latest content in a single place so that the consumers (readers) can check the resources faster.

Because RSS data is small in size and loads fast compared to other solutions it can be used with portable gadgets like smart phones and tablets.

## Who Should use RSS?

RSS can be used by anyone especially for the avid readers who constantly want to stay up to date with the latest news. RSS used frequently used in News sites, Calendars, Companies, and even announcing site changes. Every website that updates frequently or posts new content regularly needs an RSS feed.

Generally;

For News sites the XML file contains data like content title, date of publication, description, and content itself.

For Calendars the XML file contains data like upcoming events, holidays, birthdays etc.

For Companies the XML file contains data like news, articles, and products.

For site changes the XML file contains updates and changes of pages.

## Benefits of RSS

There is a lot of benefits of using RSS. Some examples are given below.

1. You choose your news and articles you want to receive

As the benefits of RSS, you can control the publications you receive, check the Informations that interests you and are related to your researches.

1. You can remove undesirable information and unfollow feeds

You are able the separate information that you desire from others that can be classified as spam or ads if you prefer to use RSS.

1. You can increase your website's traffic if you create an RSS feed

If you implement an RSS feed to your website, you can create and manage your own channels and share it to the rest of the world.

## RSS Standards

Unfortunately, there are no official standards for RSS feeds. About half of the feeds use RSS version 0.91. About a quarter of feed use RSS 1.0 and the rest is split between RSS 2.0 and RSS 0.9x versions.

## How RSS Works

RSS is used to share content between distributors and consumers. You can add RSS feed link to services called RSS aggregators. You can even register your feed with them to automate the process.

Below you will find a typical RSS feed in block diagram.

Diagram

Description automatically generated

Figure 2 Diagram of an example RSS feed

Every tag opened must be closed at the end of the part. Inside the tags the content follows. The tag can contain attributes.

Text, letter

Description automatically generated

Figure 3 Content of an example RSS feed

RSS document file must end with .xml extension. Once uploaded to a server you can view it in the browser as plain text or use an RSS aggregator. You can register your feed with services like Feedly so that they index your feed periodically.

If you don't use an aggregator the XML file will the displayed in plain text in a browser. Once displayed in the browser, a typical RSS feed will look like figure below.

Text

Description automatically generated

Figure 4 RSS Feed Example

# WHAT IS XML?

Abbreviation XML stands for eXtensible Markup Language. XML was first used in 1998 and it was designed to store and transfer data across the internet. It is designed to be both human and machine readable. It is widely used in IT and many Internet services across the world. If facilitates the distribution of data over the World Wide Web. It is software and hardware independent so it can be utilized across devices and platform.

Graphical user interface, text, application, email

Description automatically generated

Figure 5 Example XML Code

Some important bullet points include;

* XML is a markup language such as HTML.
* XML was designed to store and transfer data over the WWW.
* XML content is self-descriptive and easy to understand.
* XML is a W3C Recommendation since 1998.
* XML codes does not execute, doesn’t do anything except containing data.
* XML is transferred over TCP protocol.
* XML file may contain sender, receiver information.
* XML file may have a heading and a content body.
* XML is just a form of data that is enclosed by tags.

## Why use XML?

In many IT systems the data is in incompatible format to each other. Exchanging data between different system with incompatible format is a challenge for developers and translation of data to other format is a wasteful in regards of compute power. In conversion of data some data may be lost, or some conversion problems may occur that may disturb the integrity of the data.

Some of the reasons on why we use XML;

* XML helps simplify data sharing
* XML helps transportation of data
* XML is platform independent
* XML is software independent
* XML is hardware independent
* XML improves data availability

XML stores its content in plain text format like the figure below.

Text, timeline

Description automatically generated

Figure 6 XML in Plain Text Format

## The Difference Between XML and HTML

Although both XML and HTML is a markup language, they were designed with different objectives in mind.

* The objective of XML is to carry data while describing data content.
* The objective of HTML is to display data while describing how data looks.
* XML does not have predefined tags.
* HTML only works with predefined tags (ex: <h1>, <b>, <div>).
* XML, as of its name, is extensible, meaning it can be extended or shrunk.

# GITHUB

GitHub is a code hosting platform that offers services like version control, project management and helps team members to work on the project together. Authorized members can pull, push, and merge code to add new features.

A screenshot of a computer

Description automatically generated with medium confidence

Figure 7 Project Repository

Some GitHub essentials are;

* Repositories
* Branches
* Commits
* Pull Requests
* Merge

## Repository

We stored the code in a GitHub repository while in development. If we made any mistake and wanted to revert back to an older version, we could do it in a safely manner through git. We can store any code files, pictures, and documents. We can include a README.md file that we can explain about the project. We may also create a wiki to guide users of our project. GitHub repositories may be public or private.

## Branch

We can use a GitHub branch to develop or share different versions of the repository. We can create a development and stable repository for example. By default, a repository is initiated with a master branch while other branches were a copy of master branch at some point.

## Commits

Any changes to our code base are called commits. Commits has a description on every file that the change occurred.

## Pull Requests

Pull request are a request send to the moderators that the changes you made should be merged to the master branch. Pull request shows code modifications. After a commit, you can create a pull request and start a discussion inside GitHub.

## Merge

Once the development of new features or the bug fixes are finalized in a development repository it is merged into the master branch. Therefore, as soon as you have a pull request, you can merge any changes into the master by clicking a "Merge pull request" button.

# PHP

PHP is a backend scripting language that is mostly used for interactive (dynamic) website development. With PHP, developers can create beautiful dynamic websites and apps. It can be installed and run-on diverse platforms. PHP can interact with files on a server and run queries on many databases.

A screenshot of a computer

Description automatically generated with medium confidence

Figure 8 PHP Code for Reading XML content

## PHP MySQLi vs PDO vs MySQL

While developing for PHP, to interact with a database we can choose from several Database APIs. Developers can choose one of them to access MySQL database, but each have their own advantages and disadvantages.

Text

Description automatically generated

Figure 9 Connecting to MySQL Database Through MySQLi

### MySQL

This was the original extension that eased the communication to MySQL database. However, it is currently deprecated and not recommended to use since PHP 7 and newer versions doesn't support it.

### PHP MySQLi

MySQLi stands for MySQL Improved. It is known to have better integration and features than its predicator. MySQLi has both Object-Oriented and Procedural API. MySQL is strictly dedicated to MySQL and MariaDB and can offer better support for them over PDO.

### PDO

PDO supports most used database systems and provides a uniform method of access to those databases. PDO has support for 11 different database systems. With PDO switching databases are easier but will require to test all SQL queries and responses.

## PHP SimpleXMLElement

With SimpleXML we can get and interact with XML data. From PHP 5, SimpleXMLElement is a core part of PHP. SimpleXML facilitates the process of getting an element's content, name, and attributes in an XML document. SimpleXML allow us to display XML document's data structure.

Text

Description automatically generated

Figure 10 Usage of SimpleXMLElement

## PHP htmlspecialchars\_decode Function

Some XML documents may include HTML special characters. To decode or convert them back to chatacter we use htmlspecialchars\_decode function.

## PHP addslashes Function

The addslashes function returns a string with backslashes added before some characters. We used this function to query single quotes, double quotes and backslahes to the database. Without this the SQL database can't query those characters.

# Apache

Apache is a widely used cross-platform web server to deliver web content. It is a free and open-source webserver that is developed and maintained under the Apache Software Foundation. An Apache server delivers web pages, images, sends documents and files.

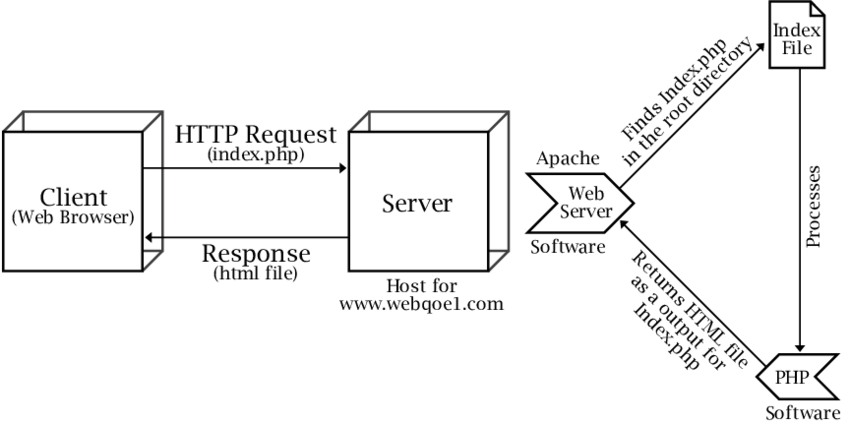


Figure 11 Workflow of Apache Web Server

* Apache receives the request of access.
* Apache runs security checks on HTTP request
* Apache allows the visitor to access the web page.
* Apache can run extra modules.
* Apache serves the documents requested.

# MySQL

MySQL is one of the most popular, widely used database. MySQL is a relational database management system (RDBMS) based on Structured Query Language (SQL). It is developed and maintained by Oracle.

In relational database the data is stored on excel like tables.

--

A database is a structured collection of data. It may be anything from a simple shopping list to a picture gallery or a place to hold the vast amounts of information in a corporate network. In particular, a relational database is a digital store collecting data and organizing it according to the relational model. In this model, tables consist of rows and columns, and relationships between data elements all follow a strict logical structure. An RDBMS is simply the set of software tools used to actually implement, manage, and query such a database.

MySQL is integral to many of the most popular software stacks for building and maintaining everything from customer-facing web applications to powerful, data-driven B2B services. Its open-source nature, stability, and rich feature set, paired with ongoing development and support from Oracle, have meant that internet-critical organizations such as Facebook, Flickr, Twitter, Wikipedia, and YouTube all employ MySQL backends.

# MARIADB

Originally, MySQL DBMS was a part of XAMPP, but now it has been replaced by MariaDB. It is one of the most widely used relational DBMS, developed by MySQL. It offers online services of data storage, manipulation, retrieval, arrangement, and deletion.

MariaDB is an open source relational database management system (DBMS) that is a compatible drop-in replacement for the widely used MySQL database technology. It was created as a software fork of MySQL by developers who played key roles in building the original database; they devised MariaDB in 2009 in response to Oracle Corp.'s acquisition of MySQL.

MariaDB is based on SQL and supports ACID-style data processing with guaranteed atomicity, consistency, isolation and durability for transactions. Among other features, the database also supports JSON APIs, parallel data replication and multiple storage engines, including InnoDB, MyRocks, Spider, Aria, TokuDB, Cassandra and MariaDB ColumnStore.

Much of the development work on the open source database has focused on achieving feature parity between MariaDB and MySQL. MariaDB Corp., the driving force behind MariaDB, says that "for all practical purposes" the database is binary-compatible with MySQL. As a result, most users can switch between the two technologies simply by uninstalling MySQL and installing MariaDB in its place.

# XAMPP

XAMPP is the most popular PHP development environment

XAMPP is a completely free, easy to install Apache distribution containing MariaDB, PHP, and Perl. The XAMPP open source package has been set up to be incredibly easy to install and to use.

XAMPP is one of the widely used cross-platform web servers, which helps developers to create and test their programs on a local webserver. It was developed by the Apache Friends, and its native source code can be revised or modified by the audience. It consists of Apache HTTP Server, MariaDB, and interpreter for the different programming languages like PHP and Perl. It is available in 11 languages and supported by different platforms such as the IA-32 package of Windows & x64 package of macOS and Linux.

What is XAMPP?

XAMPP is an abbreviation where X stands for Cross-Platform, A stands for Apache, M stands for MYSQL, and the Ps stand for PHP and Perl, respectively. It is an open-source package of web solutions that includes Apache distribution for many servers and command-line executables along with modules such as Apache server, MariaDB, PHP, and Perl.

XAMPP helps a local host or server to test its website and clients via computers and laptops before releasing it to the main server. It is a platform that furnishes a suitable environment to test and verify the working of projects based on Apache, Perl, MySQL database, and PHP through the system of the host itself. Among these technologies, Perl is a programming language used for web development, PHP is a backend scripting language, and MariaDB is the most vividly used database developed by MySQL. The detailed description of these components is given below.

## Components of XAMPP

As defined earlier, XAMPP is used to symbolize the classification of solutions for different technologies. It provides a base for testing of projects based on different technologies through a personal server. XAMPP is an abbreviated form of each alphabet representing each of its major components. This collection of software contains a web server named Apache, a database management system named MariaDB and scripting/ programming languages such as PHP and Perl. X denotes Cross-platform, which means that it can work on different platforms such as Windows, Linux, and macOS.

Cross-Platform, Apache, MariaDB, PHP, Perl, phpMyAdmin, OpenSSL, XAMPP Control Panel, Webalizer, Mercury, Tomcat, Filezilla.

## What does Cross-Platform Mean?

Different local systems have different configurations of operating systems installed in it. The component of cross-platform has been included to increase the utility and audience for this package of Apache distributions. It supports various platforms such as packages of Windows, Linus, and MAC OS.

# PHPMyAdmin

It is a tool used for dealing with MariaDB. Its version 4.0.4 is currently being used in XAMPP. Administration of DBMS is its main role.

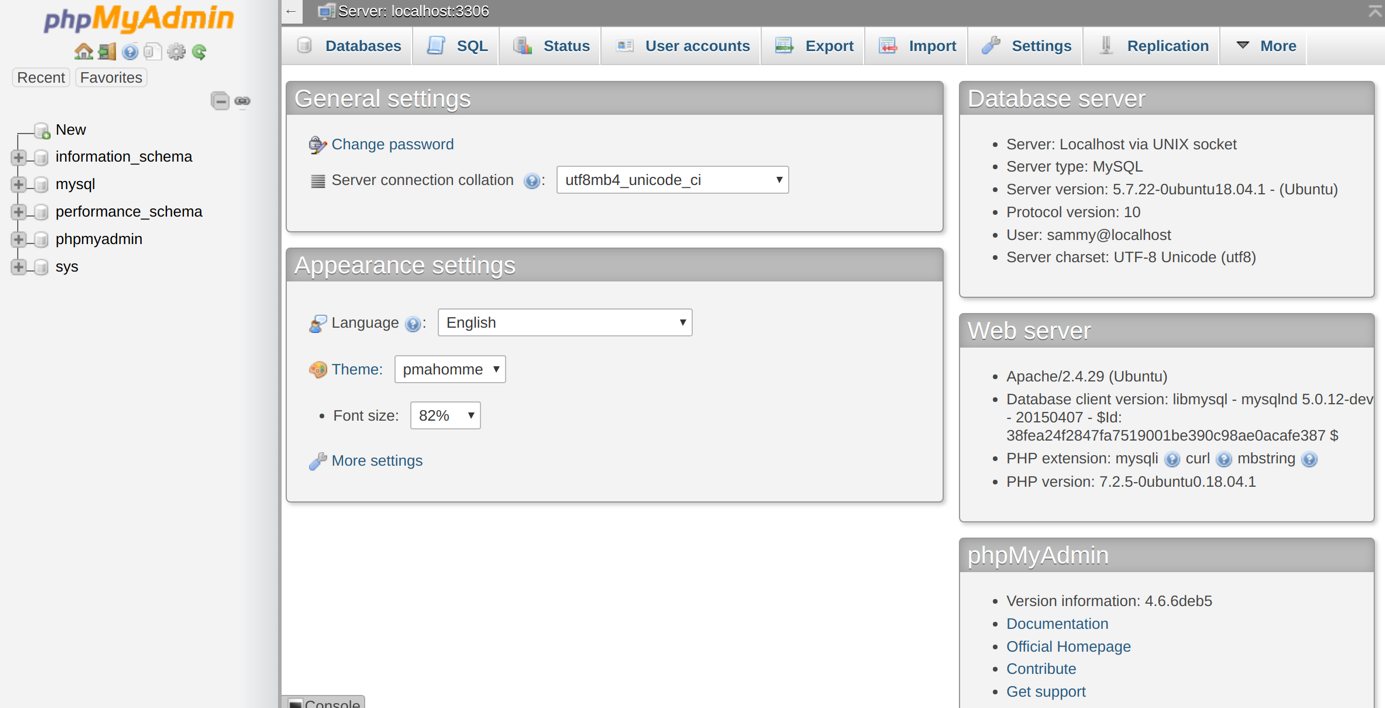


Figure 12 phpMyAdmin Interface

## UTF8 vs UTF8MB4

The utf8mb4 character set is useful because nowadays we need support for storing not only language characters but also symbols, newly introduced emojis, and so on.

According to MySQL 8.0 Reference Manual utf8

Taken from the MySQL 8.0 Reference Manual:

utf8mb4: A UTF-8 encoding of the Unicode character set using one to four bytes per character.

utf8mb3: A UTF-8 encoding of the Unicode character set using one to three bytes per character.

In MySQL utf8 is currently an alias for utf8mb3 which is deprecated and will be removed in a future MySQL release. At that point utf8 will become a reference to utf8mb4.

# RESULTS AND DISCUSSION

Present the results of your study. Comment about the results: Are they satisfactory enough to solve your problem mentioned in chapter 1? Use these results to comment about your study: Which part of your study is not good enough and why? Discuss the satisfactory/unsatisfactory parts.

We have reached the desired result but we can make it work a lot better if we use a NOSQL database like MongoDB or Firebase since it will get rid of some limitations and give us some new features to look deeper into our data. We can also make use of Google’s Machine Learning platform and do more calculations to increase our productivity.

We also did not implement an SSL Certificate and we may need to add it so when we expose our server to the public we can protect our services.

# CONCLUSION

Summary of your work: The important points of the study (from each chapter) should be mentioned, your contribution should be emphasized. The important points of the discussion section should be written and related results should be referred.

Utilizing RSS sources and XML files we have successfully created a Heath News RSS Aggregator. With RSS we can choose to view the news and articles we want, the publications that interest us and are relevant to our work. With RSS we remove unwanted information. RSS services uses XML as a language and XML doesn’t have a standard. RSS Readers doesn’t read and display all the data. Using PHP we can get access to every data and use it across our projects. At the end we made a lightweight program that integrated RSS data to SQL Database. In the later projects we can use this data extensively and create a Machine Learning Application. This project will also serve as a Search Engine as much as it is now serving as an RSS Aggregator. By using user data end queries we can create a simple Google alternative.

# REFERENCES

American National Standards Institute. *SQL Database Standards.* Retrieved from https://webstore.ansi.org/industry/software/technology-languages/SQL

W3Schools. HTML ASCII Reference: The ASCII Character Set. Retrieved from https://www.w3schools.com/charsets/ref\_html\_ascii.asp

MySQL. MySQL 8.0 Reference Manual. What Is MySQL? Retrieved from https://dev.mysql.com/doc/refman/8.0/en/what-is-mysql.html

Apache Friends. About XMAPP. Retrieved from https://www.apachefriends.org/about.html

MariaDB Server Documentation, Using MariaDB Server, SQL Statements and Structures, SQL Language Structures, Reserved Words. Retrieved from https://mariadb.com/kb/en/reserved-words/

PHP Framework Interop Group, PHP Standards Recommendations, Retrieved from https://www.php-fig.org/psr/

Tutorials Point, Basics of Computer Science Abbreviations, Retrieved from https://www.tutorialspoint.com/basics\_of\_computer\_science

ResearchGate, Björn Johansson, A XML sample document including energy consumption from https://www.researchgate.net/figure/A-XML-sample-document-including-energy-c onsumption\_fig3\_221529510

ResearchGate, Mathias Uslar, A sample CIM/XML file modeling idle power (q) and effective power (p) at a corresponding substation, Retrieved from https://www.researchgate.net/figure/A-sample-CIM-XML-file-modeling-idle-power- q-and-effective-power-p-at-a-corresponding\_fig1\_220701370

GeeksForGeeks, What is the difference between MySQL, MySQLi and PDO, Retrieved from https://www.geeksforgeeks.org/what-is-the-difference-between-mysql-mysqli-and- pdo/

Wikipedia, Apache HTTP Server, Feature overview, Retrieved from https://en.wikipedia.org/wiki/Apache\_HTTP\_Server

StackOverflow, What is the difference between utf8mb4 and utf8 charsets in MySQL, Retrieved from, https://stackoverflow.com/questions/30074492/what-is-the- difference-between-utf8mb4-and-utf8-charsets-in-mysql#r

# 

# APPENDICES

## Appendix 1 - SQL\_SCHEMA.md

**Markdown file for SQL Schema, Query Format and Sample Query**

1. ### TABLE SCHEMA
3. id
4. title
5. link = guid = atom:link
6. media = enclosure = media:content = image
7. meta\_description = description
8. content = news = content:encoded
9. pubDate

12. ### QUERY FORMAT
14. $query = "INSERT INTO $tableName
15. (title, link, media, meta\_description, content, pubDate) VALUES
16. ($title, $link, $media, $meta\_description, $content, $pubDate);";

19. ### SAMPLE QUERY
21. ```sql
22. INSERT INTO Entries
23. (title, link, media, meta\_description, content, pubDate) VALUES
24. ('Açık havada oynamak kas iskelet sistemi için çok yararlı', 'https://www.sektorel.com.tr/haber/acik-havada-oynamak-kas-iskelet-sistemi-icin-cok-yararli-33481', 'https://www.sektorel.com.tr/images/haberler/2022/06/acik-havada-oynamak-kas-iskelet-sistemi-icin-cok-yararli-1654067466.jpg', 'Yazın gelmesiyle beraber çocukların açık havada daha fazla zaman geçirmesi gerektiğini belirten Fizik Tedavi ve Rehabilitasyon Uzmanı Doç. Dr. Nihal Özaras, fiziksel olarak hareketli olmanın bedensel ve ruhsal açıdan yararlı olduğunu söyledi. Dünya Sağlık Örgütü’nün çocuk ve ergenlerin günde en az bir saat fiziksel olarak aktif olmalarını önerdiğine dikkat çeken Özaras, “Koşma, zıplama, yürüme, uzanma, tırmanma gibi hareketler içeren sporlar veya ev dışında oynanan oyunların, kas iskelet sistemi, denge ve koordinasyon, kalp damar sistemi açısından sayısız faydaları bulunuyor.” dedi.', 'p style=\"text-align:justify\"><span style=\"font-size:16px\">Çocuklarda fiziksel olarak aktif olmanın hem bedensel ve hem de ruhsal sağlık açısından çok önemli olduğunu vurgulayan </span>Doç. Dr. Nihal Özaras<span style=\"font-size:16px\">, “Koşma, zıplama, yürüme, uzanma, tırmanma gibi hareketler içeren sporlar veya ev dışında oynanan oyunların, kas iskelet sistemi, denge ve koordinasyon, kalp damar sistemi açısından sayısız faydaları bulunuyor. Ayrıca kalori harcanmasını sağladıkları için kilo kontrolünde de yararlıdır.” dedi.</span></p><p style=\"text-align:justify\"><span style=\"font-size:16px\"> </span></p><p style=\"text-align:justify\"><strong><span style=\"font-size:16px\">Hayal gücünü geliştiriyor</span></strong></p><p style=\"text-align:justify\"><span style=\"font-size:16px\"> </span></p><p style=\"text-align:justify\">Doç. Dr. Nihal Özaras<span style=\"font-size:16px\">, araştırmalarda, özellikle planlama olmadan serbestçe oynanan sokak oyunlarının, çocukların hayal güçlerini geliştirdiğini, yükseklik korkusu gibi bazı korkularını azalttığının ortaya çıktığını söyledi. </span></p><p><br>Kaynak: (BHA) - Beyaz Haber Ajansı</p>', 'Wed, 01 Jun 2022 10:11:00 +0300');
25. ```