DATABASE PROJECT REPORT

2020 - SPRING

Team Members: Ayşe Ceren Çiçek - 170709009
Gizem Kurnaz - 170709059
Gizem Pesen - 170709050

Description of the Project

Our data gives general information about top 50 ranked sites from each 191 countries along with their traffic. The purpose of this project is to compare websites and give information to users about them (e.g. daily traffics, security status).

The Changes Made in the Project

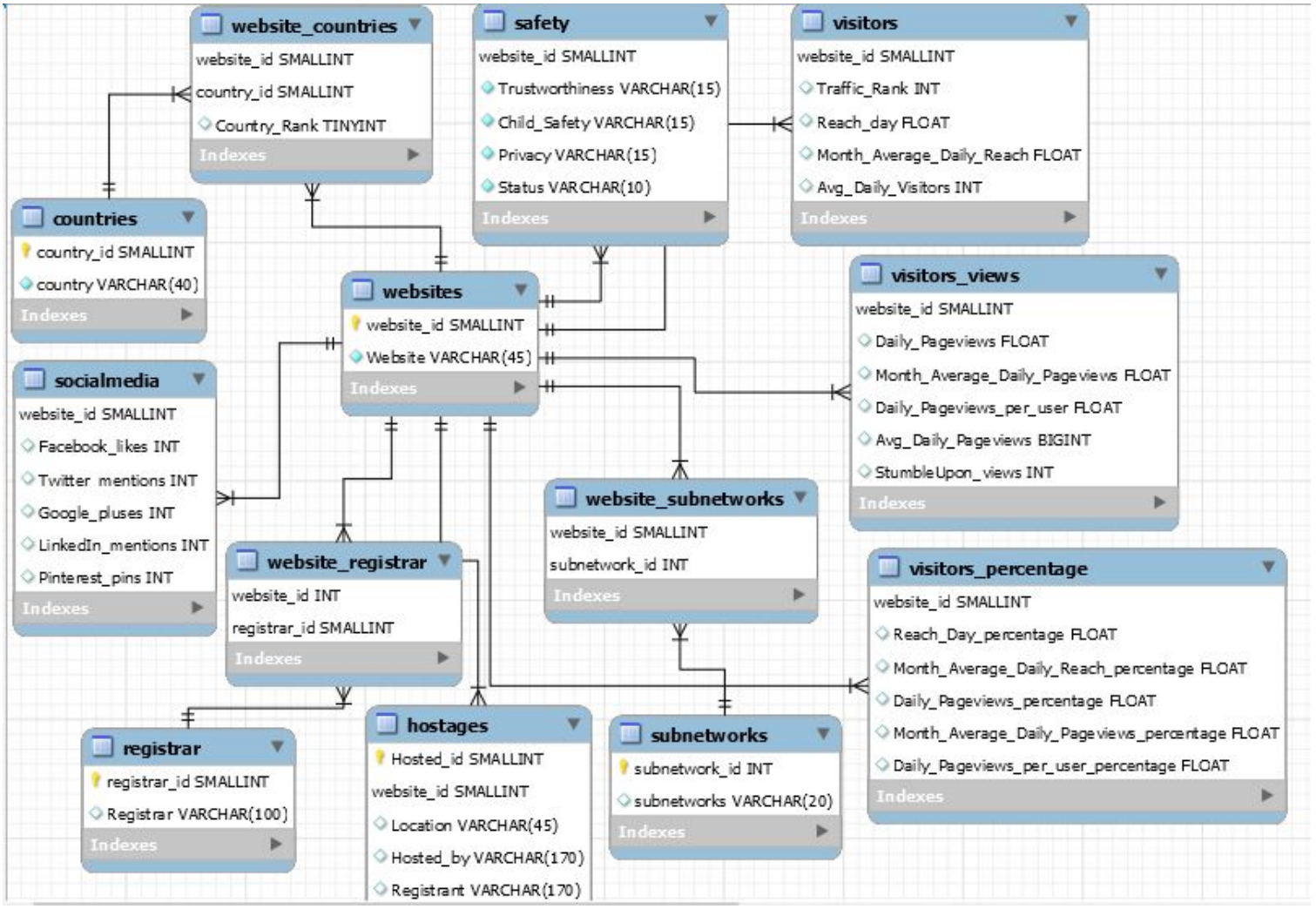
While working on the dataset, we decided to add three columns subnetwork_id, registrar_id and Hosted_id to avoid partial dependencies and to normalize our data. We made several changes on the design of our database. First of all, we realized that country_rank depends on both website_id and country_id, we added it to website_countries and deleted from countries table. Also, we find out that the registrar and subnetwork columns depends on the website and we split them into new tables ‘registrar’ and ‘subnetworks’ to avoid too much repetition in the data.

Loading the Data

Firstly, we cleaned the dataset (null values) and added IDs with Open Refine software. To upload the data to MySql, we added a new table called ‘denormalized’ which includes all attributes of our data. Then we loaded our data via csv file using ‘load data infile’. With insert statements, we parsed the data into tables. We uploaded all the data.

The source of our data: https://www.kaggle.com/bpali26/popular-websites-across-the-globe

ER DIAGRAM



Stored Procedures/ Views

1) This stored procedure gets the country name as input and returns the top 50 websites of this country.

```
create procedure getWebsitesByCountry(in countryName varchar(40))
select websites.website_id, Website, Country_Rank
from websites join website_countries on websites.website_id = website_countries.website_id
where country_id in (select country_id
                     from countries
                     where country = countryName)
order by Country_Rank;
```

2) The stored procedure shown below, gets the name of the country as input parameter and returns the average traffic rank value as output parameter.

```
create procedure avgTrafficRankByCountry(in countryName varchar(40), out avgTrafficRank float)
select avg(Traffic_Rank)
into avgTrafficRank
from visitors join websites on visitors.website_id = websites.website_id
      join website_countries on websites.website_id = website_countries.website_id
      join countries on website_countries.country_id = countries.country_id
where country = countryName;
```

3) This stored procedure deletes the given website from websites table.

```
create procedure deleteWebsite(inout website_name varchar(45))
delete from websites
where Website = website_name;
```

4)The stored procedure shown below, returns the social media information by website ID.

```
create procedure getSocialMedia(in websiteID SMALLINT)
select websites.website_id, Facebook_likes, Twitter_mentions, Google_pluses, LinkedIn_mentions, Pinterest_pins
from websites join socialmedia on websites.website_id = socialmedia.website_id
where websites.website_id = websiteID;
```

5)This stored procedure returns the websites which are safe in a given country

```
create procedure getSafeWebsitesByCountry(in countryName varchar(40))
select Website
from websites join safety on websites.website_id = safety.website_id
where Trustworthiness = "Excellent" and
      Child_Safety = "Excellent" and Privacy = "Excellent"
      and websites.website_id in (select website_countries.website_id
                  from website_countries join countries on website_countries.country_id = countries.country_id
                  where country = countryName);
```

6) The view shown below creates a virtual table which includes view and social media information of websites.

```
create view website_Info_View as
select Website, Traffic_Rank, Facebook_likes, Twitter_mentions, Google_pluses, LinkedIn_mentions, Pinterest_pins,
        Reach_day, Daily_Pageviews, StumbleUpon_views
from ((websites join visitors on websites.website_id = visitors.website_id)
        join socialmedia on websites.website_id = socialmedia.website_id)
        join visitors_views on socialmedia.website_id = visitors_views.website_id;
```

7) This view includes some network informations about websites.

```
create view website_Network_Info as
select Website, Hosted_by, Location, Registrant, Registrar, Subnetworks
from ((((websites join website_registrar on websites.website_id = website_registrar.website_id)
      join registrar on website_registrar.registrar_id = registrar.registrar_id)
      join hostages on websites.website_id = hostages.website_id)
      join website_subnetworks on websites.website_id = website_subnetworks.website_id)
      join subnetworks on website_subnetworks.subnetwork_id = subnetworks.subnetwork_id;
```

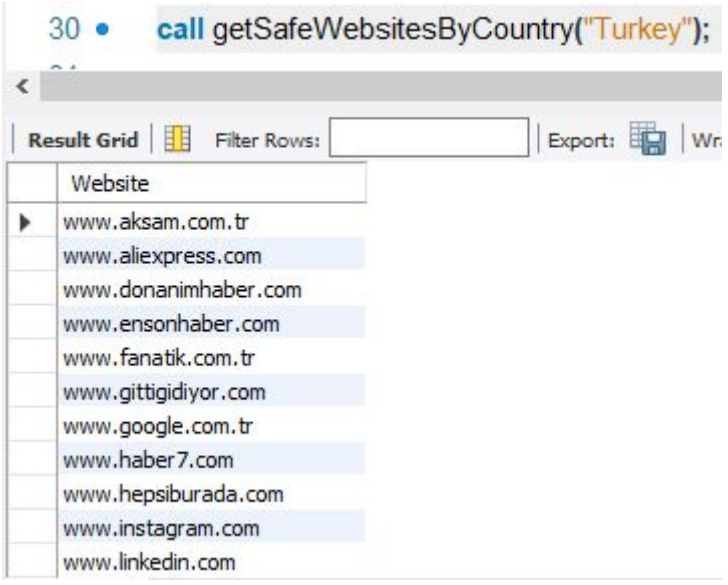
Output of the first stored procedure:

```
10 •    call getWebsitesByCountry("Turkey");
```

Result Grid | Filter Rows: | Export:

website_id	Website	Country_Rank
1275	www.google.com.tr	1
3348	www.youtube.com	2
4038	www.google.com	3
1018	www.facebook.com	4
2581	www.sabah.com.tr	5
2274	www.onedio.com	6
1397	www.haber7.com	7
2586	www.sahibinden.com	8
972	www.ensonhaber.com	9
1981	www.milliyet.com.tr	10
1829	www.live.com	11

Output of the fifth stored procedure:



System's Limitations and Suggestions

MySQL Workbench 8.0, Intel(R) Core(™) i5-4210U CPU @ 2.40 GHz, 4 GB Ram, 64 bit processor, Windows 10

MySQL Workbench 8.0, Intel(R) Core(TM) i5-8250U CPU @ 1.80 GHz, 8 GB Ram, 64 bit processor, Windows 10

MySQL Workbench 8.0, Intel(R) Core(™) i5-7200U CPU @ 2.70 GHz, 8 GB Ram, 64 bit processor, Windows 10

MySQL Workbench 8.0 supported systems:

	Minimum	Recommended
CPU	64 bit x86 CPU	Multi Core 64 bit x86 CPU, 8 GB RAM
RAM	4 GB	8 GB or higher

To improve our database performance, we can upgrade our hardware with stronger CPU and allocate more memory. But according to the table above, all devices that we use are convenient to download and use MySQL Workbench platform. Also we can optimize queries to improve the performance.

The Database in SQL Database Definition Language

```
CREATE TABLE IF NOT EXISTS `websites` (
  `website_id` SMALLINT NOT NULL AUTO_INCREMENT,        #website_id = 4038
  `Website` VARCHAR(45) NOT NULL,                       #Website = “www.google.com”
  PRIMARY KEY (`website_id`));

CREATE TABLE IF NOT EXISTS `countries` (
  `country_id` SMALLINT NOT NULL AUTO_INCREMENT,        #country_id = 176
  `country` VARCHAR(40) NOT NULL,                       #country = “Turkey”
  PRIMARY KEY (`country_id`));

CREATE TABLE IF NOT EXISTS `website_countries` (
  `website_id` SMALLINT NOT NULL,                       #website_id = 4038
  `country_id` SMALLINT NOT NULL,                       #country_id = 176
  `Country_Rank` TINYINT NULL,                          #Country_Rank = 3
  PRIMARY KEY (`website_id`, `country_id`));

CREATE TABLE IF NOT EXISTS `socialmedia` (
  `website_id` SMALLINT NOT NULL,                       #website_id = 4038
  `Facebook_likes` INT NULL,                            #Facebook_likes = 942000
  `Twitter_mentions` INT NULL,                          #Twitter_mentions = 112000
  `Google_pluses` INT NULL,                             #Google_pluses = 117000000
  `LinkedIn_mentions` INT NULL,                         #LinkedIn_mentions = 247000
  `Pinterest_pins` INT NULL,                            #Pinterest_pins = 108000
  PRIMARY KEY (`website_id`));


CREATE TABLE IF NOT EXISTS `safety` (
  `website_id` SMALLINT NOT NULL,                       #website_id = 4038
  `Trustworthiness` VARCHAR(15) NOT NULL,               #Trustworthiness = “Excellent”
  `Child_Safety` VARCHAR(15) NOT NULL,                  #Child_Safety = “Excellent”
  `Privacy` VARCHAR(15) NOT NULL,                       #Privacy = “Excellent”
  `Status` VARCHAR(10) NOT NULL,                        #Status = “ok”
  PRIMARY KEY (`website_id`));


CREATE TABLE IF NOT EXISTS `hostages` (
  `Hosted_id` SMALLINT NOT NULL AUTO_INCREMENT,         #Hosted_id = 43
  `website_id` SMALLINT NOT NULL,                       #website_id = 4038
  `Location` VARCHAR(45) NULL,                          #Location = “United States”
  `Hosted_by` VARCHAR(170) NULL,                        #Hosted_by = “Google Inc.”
  `Registrant` VARCHAR(170) NULL,                               #Registrant = “DNS Admin
PRIMARY KEY (`Hosted_id`, `website_id`));               #                     (Google Inc.)”
```

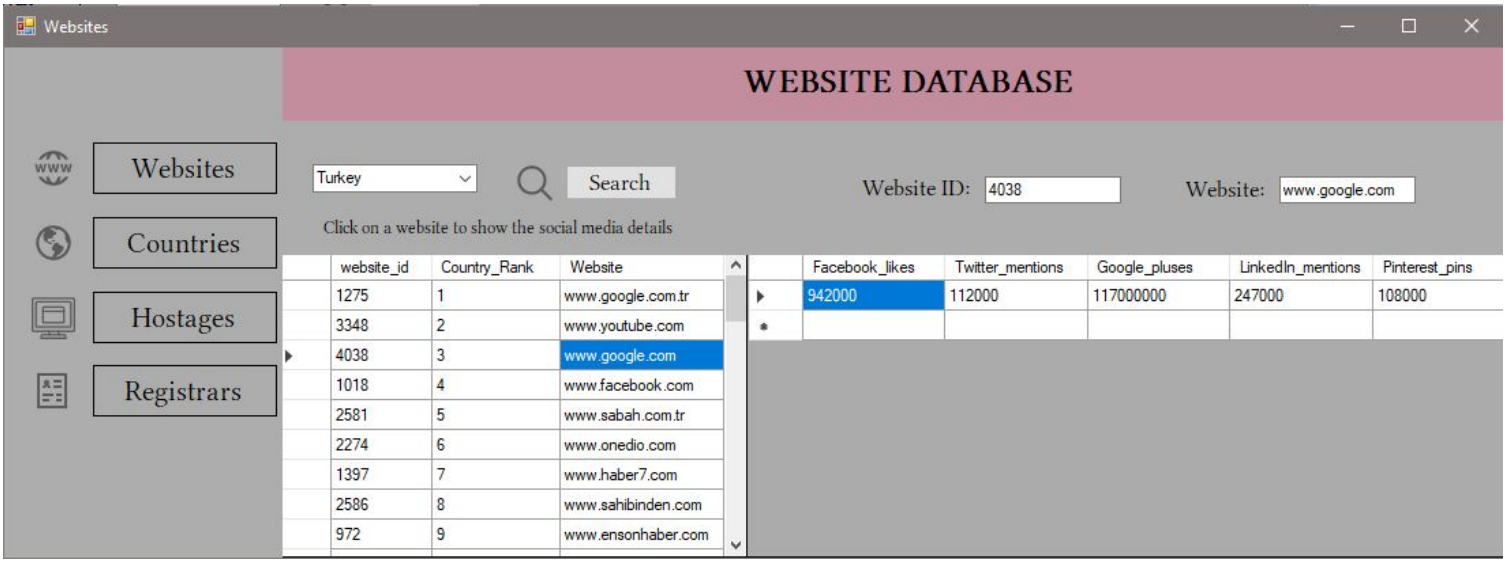
```
CREATE TABLE IF NOT EXISTS `visitors_percentage` (
  `website_id` SMALLINT NOT NULL,                          #website_id = 4038
  `Reach_Day_percentage` FLOAT NULL,                       #8.7
  `Month_Average_Daily_Reach_percentage` FLOAT NULL,       #1.1
  `Daily_Pageviews_percentage` FLOAT NULL,                 #45.8
  `Month_Average_Daily_Pageviews_percentage` FLOAT NULL,   #3.5
  `Daily_Pageviews_per_user_percentage` FLOAT NULL,        #33.9
  PRIMARY KEY (`website_id`));

CREATE TABLE IF NOT EXISTS `visitors_views` (
  `website_id` SMALLINT NOT NULL,                          #website_id = 4038
  `Daily_Pageviews` FLOAT NULL,                            #8.44896
  `Month_Average_Daily_Pageviews` FLOAT NULL,              #7.80634
  `Daily_Pageviews_per_user` FLOAT NULL,                   #8496
  `Avg_Daily_Pageviews` BIGINT NULL,                       #4192159833
  `StumbleUpon_views` INT NULL,                            #246000
  PRIMARY KEY (`website_id`));

CREATE TABLE IF NOT EXISTS `visitors` (
  `website_id` SMALLINT NOT NULL,                          #website_id = 4038
  `Traffic_Rank` INT NULL,                                 #Traffic_Rank = 1
  `Reach_day` FLOAT NULL,                                  #Reach_day = 42.3744
  `Month_Average_Daily_Reach` FLOAT NULL,                  #41.1494
  `Avg_Daily_Visitors` INT NULL,                           #Avg_Daily_Visitors = 518108189
  PRIMARY KEY (`website_id`));

CREATE TABLE IF NOT EXISTS `subnetworks` (
  `subnetwork_id` INT NOT NULL AUTO_INCREMENT,             #subnetwork_id = 3185
  `subnetworks` VARCHAR(20) NULL,                          #subnetworks = 216.58.194.132
  PRIMARY KEY (`subnetwork_id`));

CREATE TABLE IF NOT EXISTS `website_subnetworks` (
  `website_id` SMALLINT NOT NULL,                          #website_id = 4038
  `subnetwork_id` INT NOT NULL,                            #subnetwork_id = 3185
  PRIMARY KEY (`website_id`, `subnetwork_id`));

CREATE TABLE IF NOT EXISTS `registrar` (
  `registrar_id` SMALLINT NOT NULL,                        #registrar_id = 146
  `Registrar` VARCHAR(100) NULL,                           #registrar = “MARKMONITOR INC”
  PRIMARY KEY (`registrar_id`));

CREATE TABLE IF NOT EXISTS `website_registrar` (
  `website_id` INT NOT NULL,                               #website_id = 4038
  `registrar_id` SMALLINT NOT NULL,                        #registrar_id = 146
  PRIMARY KEY (`website_id`, `registrar_id`));
```

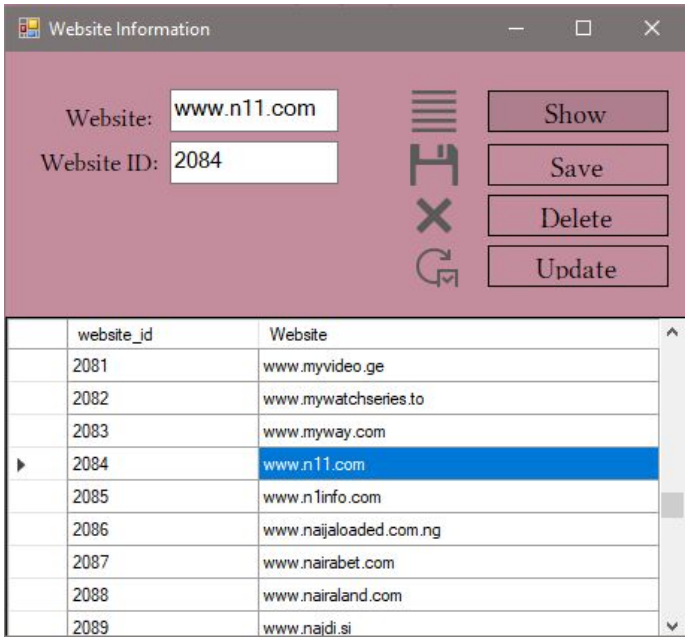
GUI

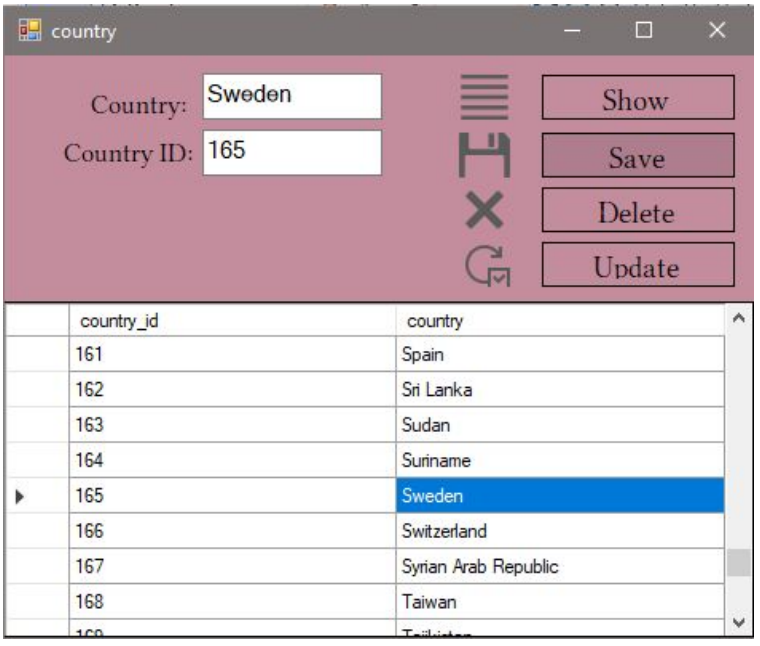
We created a simple GUI with Visual Studio 2015. We coded in C# using Windows Forms which is a graphical class library.

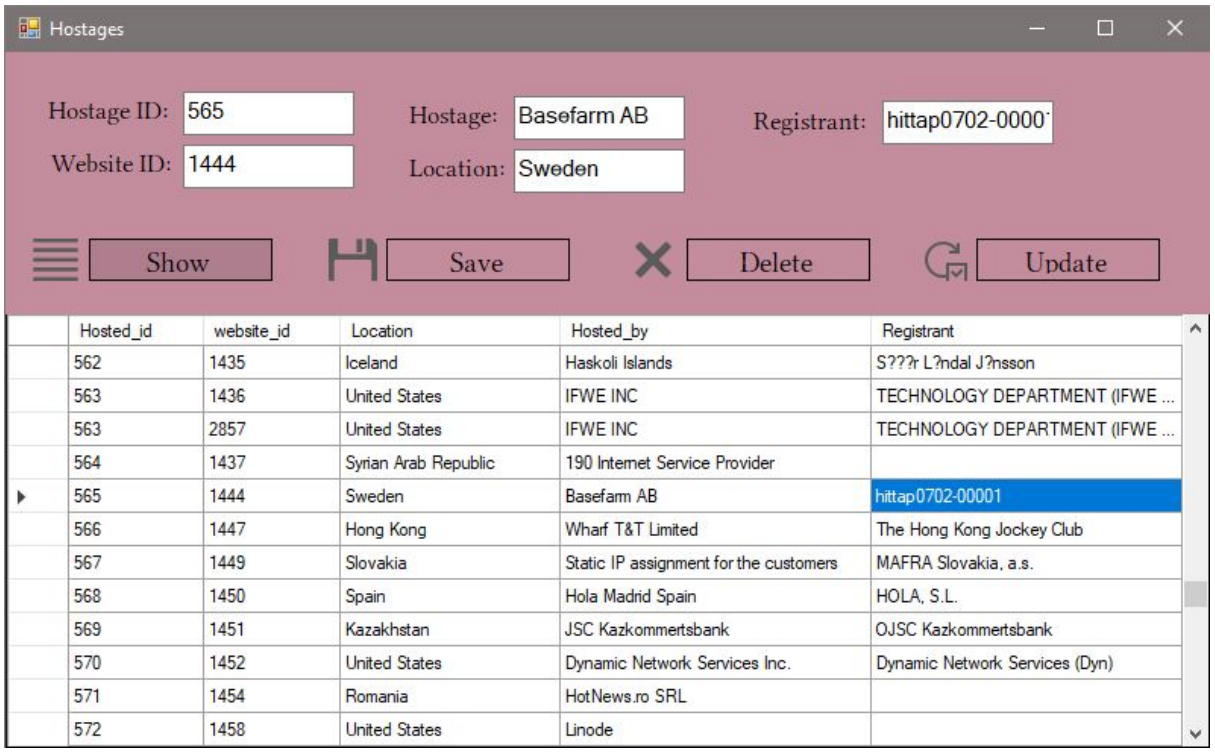
The codes of the GUI: https://github.com/crncck/WebsitesDB_GUI/tree/master/Website

ScreenShots:

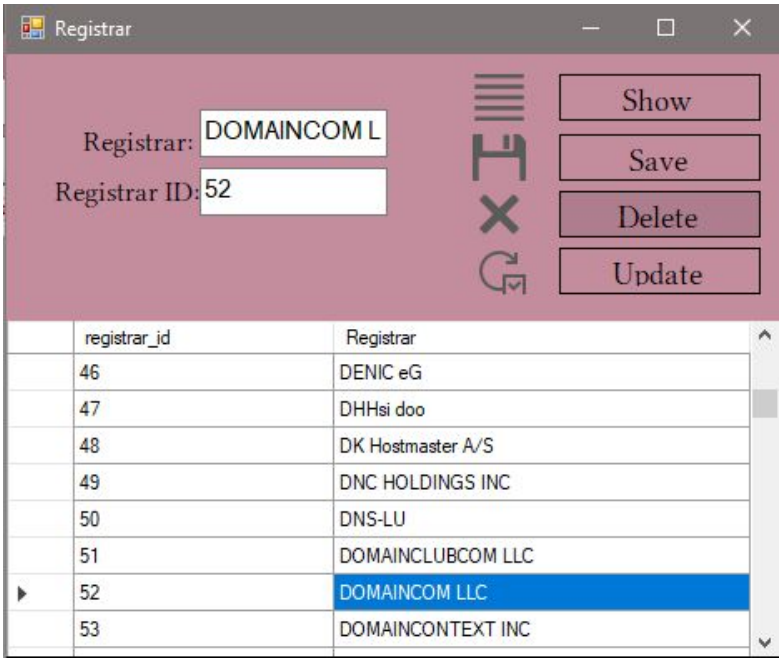








Hosted_id	website_id	Location	Hosted_by	Registrant
562	1435	Iceland	Haskoli Islands	S???r L?ndal J?nsson
563	1436	United States	IFWE INC	TECHNOLOGY DEPARTMENT (IFWE ...
563	2857	United States	IFWE INC	TECHNOLOGY DEPARTMENT (IFWE ...
564	1437	Syrian Arab Republic	190 Internet Service Provider	
565	1444	Sweden	Basefarm AB	hittap0702-00001
566	1447	Hong Kong	Wharf T&T Limited	The Hong Kong Jockey Club
567	1449	Slovakia	Static IP assignment for the customers	MAFRA Slovakia, a.s.
568	1450	Spain	Hola Madrid Spain	HOLA, S.L.
569	1451	Kazakhstan	JSC Kazkommertsbank	OJSC Kazkommertsbank
570	1452	United States	Dynamic Network Services Inc.	Dynamic Network Services (Dyn)
571	1454	Romania	HotNews.ro SRL	
572	1458	United States	Linode	



registrar_id	Registrar
46	DENIC eG
47	DHHsi doo
48	DK Hostmaster A/S
49	DNC HOLDINGS INC
50	DNS-LU
51	DOMAINCLUBCOM LLC
52	DOMAINCOM LLC
53	DOMAINCONTEXT INC

All SQL Code Used in Our System

```
show variables like "secure_file_priv";

load data
infile "C:\\ProgramData\\MySQL\\MySQL Server 8.0\\Uploads\\website3.csv"
into table denormalized
columns terminated by ';'
ESCAPED BY '\\';

load data
infile "C:\\ProgramData\\MySQL\\MySQL Server 8.0\\Uploads\\subnetworks.csv"
into table subnetworks
columns terminated by ';'
ESCAPED BY '\\';

load data
infile "C:\\ProgramData\\MySQL\\MySQL Server 8.0\\Uploads\\registrar.csv"
into table registrar
columns terminated by ';'
ESCAPED BY '\\';

insert into websites (website_id, Website)
select distinct website_id, Website
from denormalized;

insert into website_subnetworks (website_id, subnetwork_id)
select distinct website_id, subnetwork_id
from denormalized;

insert into website_registrar (website_id, registrar_id)
select distinct website_id, registrar_id
from denormalized;

insert into visitors_percentage (website_id, Reach_Day_percentage,
Month_Average_Daily_Reach_percentage, Daily_Pageviews_percentage,
Month_Average_Daily_Pageviews_percentage, Daily_Pageviews_per_user_percentage)
```

```
select distinct website_id, Reach_Day_percentage, Month_Average_Daily_Reach_percentage,
Daily_Pageviews_percentage, Month_Average_Daily_Pageviews_percentage,
Daily_Pageviews_per_user_percentage
from denormalized;

insert into socialmedia (website_id, Facebook_likes, Twitter_mentions, Google_pluses,
LinkedIn_mentions, Pinterest_pins)
select distinct website_id, Facebook_likes, Twitter_mentions, Google_pluses, LinkedIn_mentions,
Pinterest_pins
from denormalized;

insert into safety (website_id, Trustworthiness, Child_Safety, Privacy, Status)
select distinct website_id, Trustworthiness, Child_Safety, Privacy, Status
from denormalized;

insert into website_countries (website_id, country_id, Country_Rank)
select distinct website_id, country_id, Country_Rank
from denormalized;

insert into countries (country_id, country)
select distinct country_id, country
from denormalized;

insert into hostages (Hosted_id, website_id, Location, Hosted_by, Registrant)
select distinct Hosted_id, website_id, Location, Hosted_by, Registrant
from denormalized;

insert into visitors_views (website_id, Daily_Pageviews, Month_Average_Daily_Pageviews,
Daily_Pageviews_per_user, Avg_Daily_Pageviews, StumbleUpon_views)
select distinct website_id, Daily_Pageviews, Month_Average_Daily_Pageviews,
Daily_Pageviews_per_user, Avg_Daily_Pageviews, StumbleUpon_views
from denormalized;

insert into visitors (website_id, Traffic_Rank, Reach_day, Month_Average_Daily_Reach,
Avg_Daily_Visitors)
```

```
select distinct website_id, Traffic_Rank, Reach_day, Month_Average_Daily_Reach,
Avg_Daily_Visitors
from denormalized;
```