ПРИЛОЖЕНИЕ ЗА ОБРАБОТВАНЕ НА ХИДРОМЕТРИЧНИ ДАННИ

Благовест Пижев, Национален Институт по Метеорология и Хидрология, pizhevsoft@gmail.com

Емил Хаджиколев, Пловдивски университет „Паисий Хилендарски“, emil.hadjikolev@gmail.com

Резюме: В статията е представена работата по изграждане на софтуерното приложение HydroAnalyzer за съхранение и анализ на хидрометрични данни. Описан е моделът на софтуерната архитектура, съобразен със специфичните начини за събиране и съхранение на хидрометрични данни. Разгледани са компонентите на приложението и основните последователности на употребата им. Реализираните функционалности включват обработка и сравнение на хидрометрични данни, графични и таблични представяния на данните, анализ и известяване на потребителя за грешни данни. Бъдещото усъвършенстване на HydroAnalyzer е в посока едновременен анализ на множество разнородни източници на данни и предсказване на стойности на различни хидроложки величини. Приложението се използва експериментално от служители на Националния институт по метеорология и хидрология.

Ключови думи: съхранение и анализ на хидрометрични данни, обработване на хидрометрични данни, Hydro Analyzer.

APPLICATION FOR HYDROMETRIC DATA PROCESSING

Blagovest Pizhev, National Institute of Meteorology and Hydrology, pizhevsoft@gmail.com

Emil Hadzhikolev, Plovdiv University „Paisii Hilendarski“, emil.hadjikolev@gmail.com

Abstract: The paper presents the work on building the HydroAnalyzer software application for storage and analysis of hydrometric data. It describes the software architecture model tailored to the specific ways of collecting and storing hydrometric data. The paper examines the application components and the main sequences of their use. The implemented functionalities include processing and comparison of hydrometric data, graphical and tabular data presentation, analysis and notifications of incorrect data. Future development of HydroAnalyzer is directed towards simultaneous analysis of multiple heterogeneous data sources and predicting values of different hydrological measurements. The application is used experimentally by employees of the National Institute of Meteorology and Hydrology.

Keywords: storage and analysis of hydrometric data, processing of hydrometric data, Hydro Analyzer application.

1. Въведение

Националният институт по метеорология и хидрология (НИМХ) е държавна институция с предмет на дейност наблюдение, анализ и прогноза на процесите в атмосферата и хидросферата, и изследване на пространствено-времеви характеристики на климатичните и водните ресурси [1]. Институтът предоставя хидрометрични данни на различни клиенти – частни лица, фирми, полиция, прокуратура, застрахователни компании и др. Базовата наблюдателна хидрологична мрежа обхваща множество реки, кладенци и извори [2]. НИМХ разполага с десетки хидроложки, валежомерни, климатични и синоптични станции. Последните предоставят през три часа данни за валеж, вятър и температура. Събирането на коректни данни, съхранението, анализирането и визуализирането им е важна задача на НИМХ за предоставяне на качествени услуги. Изработваните прогнози се получават на база числови модели (метеорологични, хидроложки, модели на ветровото вълнение и др.) [3].

Събирането на значителна част от хидрологичните данни се извършва автоматично, чрез автоматични измервателни станции и/или от хидронаблюдатели и се съхранява в релационни бази данни. За осигуряване на качеството на данните, получени по различни способи, е необходимо те да бъдат сравнявани, и при значителни различия, да се предприемат съответни коригиращи операции. ***Обработването на хидрометрични данни*** и последващата им визуализация в различен формат за различните видове потребители – хидролози, ръководители на звена в НИМХ, администратори и др., е друга трудоемка задача. Необходим е механизъм за автоматизирането на тези процеси.

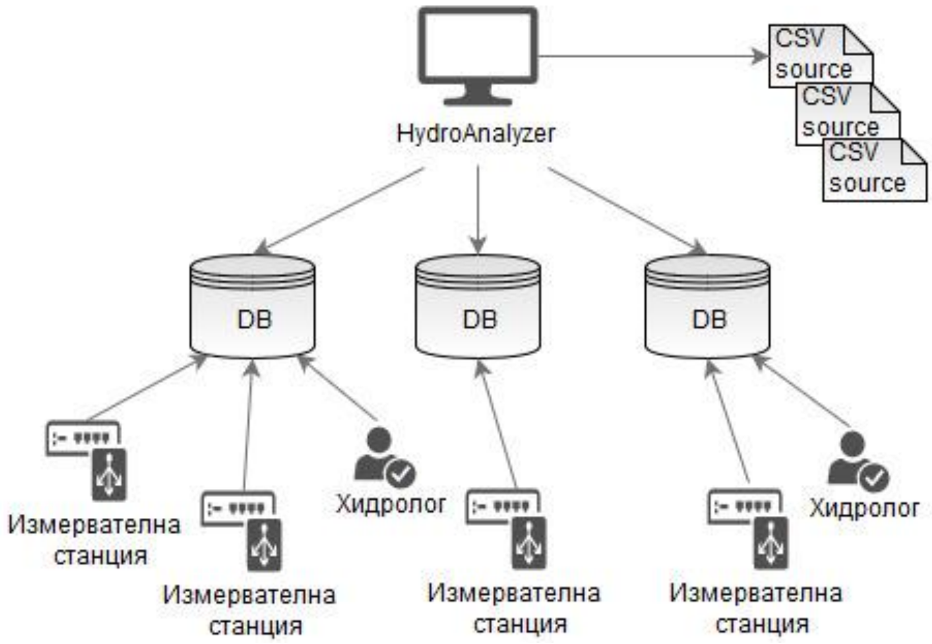
В статията е представено ***приложението HydroAnalyzer***. То поддържа инструменти за обработка, сравнение и известяване на потребителя за грешни хидроложки данни. Приложението е с клиент-сървър архитектура, но има и десктоп модул, който работи с данни, сериализирани в CSV формат.

2. Софтуерно приложение HydroAnalyzer за съхранение и анализ на хидрометрични данни

Отчитайки спецификата на предметната област и нуждите на различните типове потребители, пред разработваната система са поставени следните изисквания:

- възможност за търсене и извличане на данни по определени критерии – номер на станция, времеви период и др.;
- визуализиране на данните в табличен и графичен формат,
- търсене на несъответствия в данни, касаещи един и същи времеви период и измервателна станция;
- контрол на достъпа до приложението, съобразен с длъжностните характеристики и професионалните задължения на потребителите на базата данни;
- възможност за работа с приложението офлайн, и др.

Част от хидроложките данни се отчитат на всеки час – от самопишещи уреди или автоматизирани измервателни станции. В допълнение на това, ежедневно данните се отчитат и от хидронаблюдатели. Информацията се съхранява в множество бази данни с еднотипна структура. HydroAnalyzer работи както с данните, съхранени в базата данни, така и с CSV файлове, съдържащи сходна структурирана информация (фиг. 1). Тъй като историческата информация първоначално се е съхранявала в CSV файлов формат, а в последствие се въвеждат релационните бази данни и автоматизираната им обработка, възниква необходимостта от проверка на данните чрез сравнение на годишните изводи (месечни минимални, максимални и средни) в двата типа структури от данни, получени по различни способи.



Фиг. 1. Основна схема за събиране, съхранение и работа с данните

Данните за дневни минимални, максимални и средни водни количества се изчисляват от ежедневната информация в релационната БД и се сравняват с тези в архивите от CSV файлове – за минимални, максимални и средни месечни и годишни водни количества.

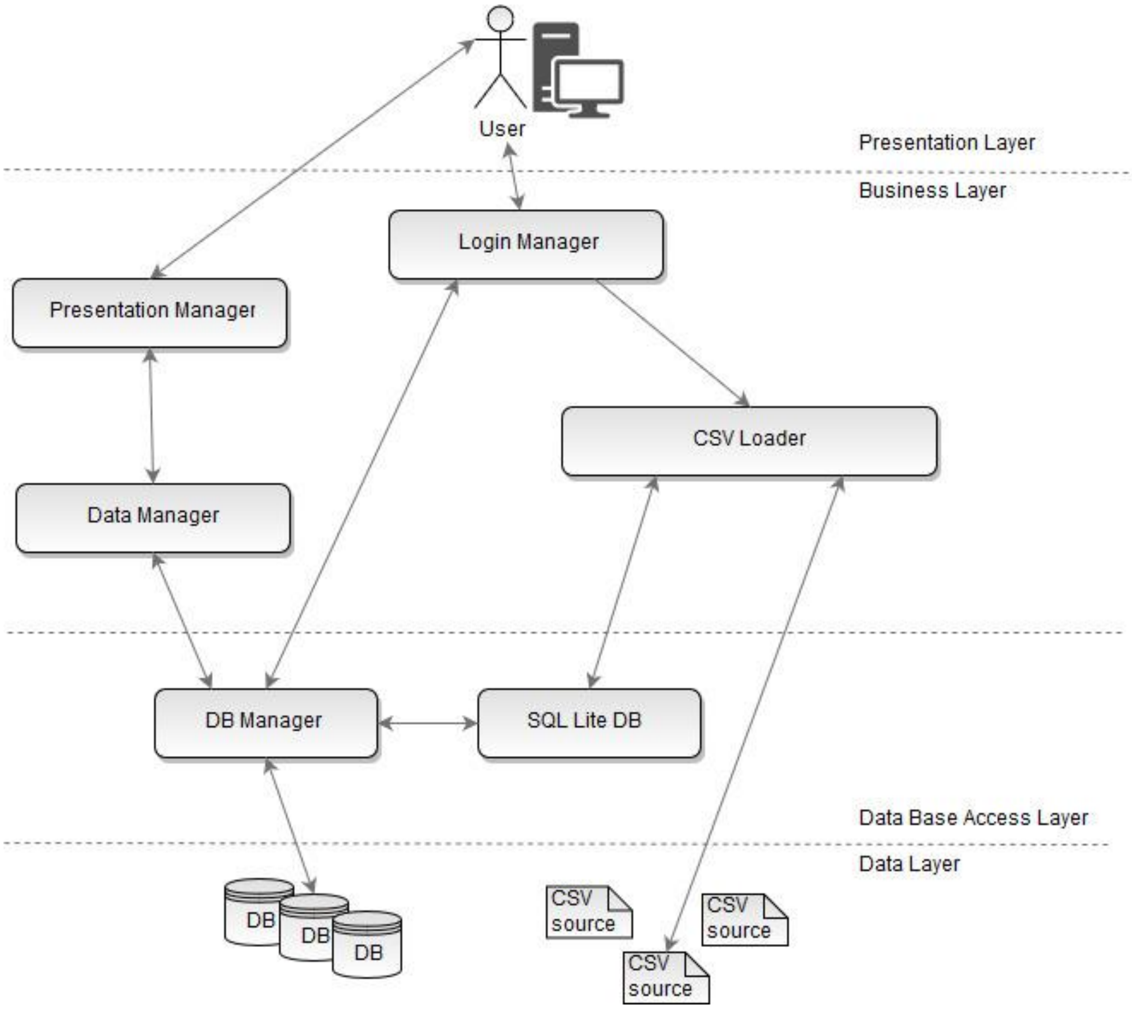
Методите за сравнение на данните се основават на преобразуването на автоматично получените дневни минимални, максимални и средни стойности до месечни или годишни, и последващото им сравнение със записаните от хидролозите архивни данни. Поради спецификата на съхранение на дневните данни като редове от таблица от базата данни, са необходими специализирани методи за автоматичното им преобразуване до месечни или годишни. Единият възможен подход е – данните да се извлекат от базата данни в списък в приложението, където да се преобразуват в месечни или годишни. В случаи, в които се обработват големи обеми данни, този подход изисква голям обем оперативна памет и процесорна мощ, което може да направи приложението практически неизползваемо. При реализирания от нас вариант, изчисленията се извършват от системата за управление на базата данни, която е оптимизирана за бързо изпълнение на агрегатни функции. Успешно изпълнено предизвикателство беше създаването на една единствена SQL заявка, предоставяща цялата необходима информация за минимални, максимални и средни стойности по месеци и/или години.

Приложението може да се използва в два основни режима от различни потребители. Докато ръководителите на звена имат пълен достъп до данните в различните MySQL бази данни [4], то подчинените им хидролози работят само с ограничен брой локално съхранени CSV-файлове с дневни и месечни данни. При това, в първия случай приложението използва клиент-сървър архитектура: основната реализация се явява клиент на сървъра за управление на бази от данни. Във втория случай, при работа с CSV-файлове, хидролозите могат да използват приложението офлайн, без достъп до Интернет и сървър за бази данни.

Друго интересно предизвикателство при работата със CSV-файловете е определяне на логиката за обработката им. Вместо реализиране на нови алгоритми при зареждането на данните от CSV-файловете в списъци с автоматично изчислени и ръчно въведени данни по месеци, задачата беше сведена до вече решена такава. За целта данните се зареждат във вградена в приложението SQLite база данни [5], имаща структура на оригиналните бази от данни. По този начин, логиката за достъп до отдалечените MySQL и вградената SQLite бази данни може да се изолира в отделен слой, който се ползва от работната логика на приложението.

Приложението е реализирано на езика C#, с помощта на средата за разработка Visual Studio [6]. Архитектурата му е показана на фиг. 2. То се състои от 4 слоя, всеки от които съдържа специфични компоненти:

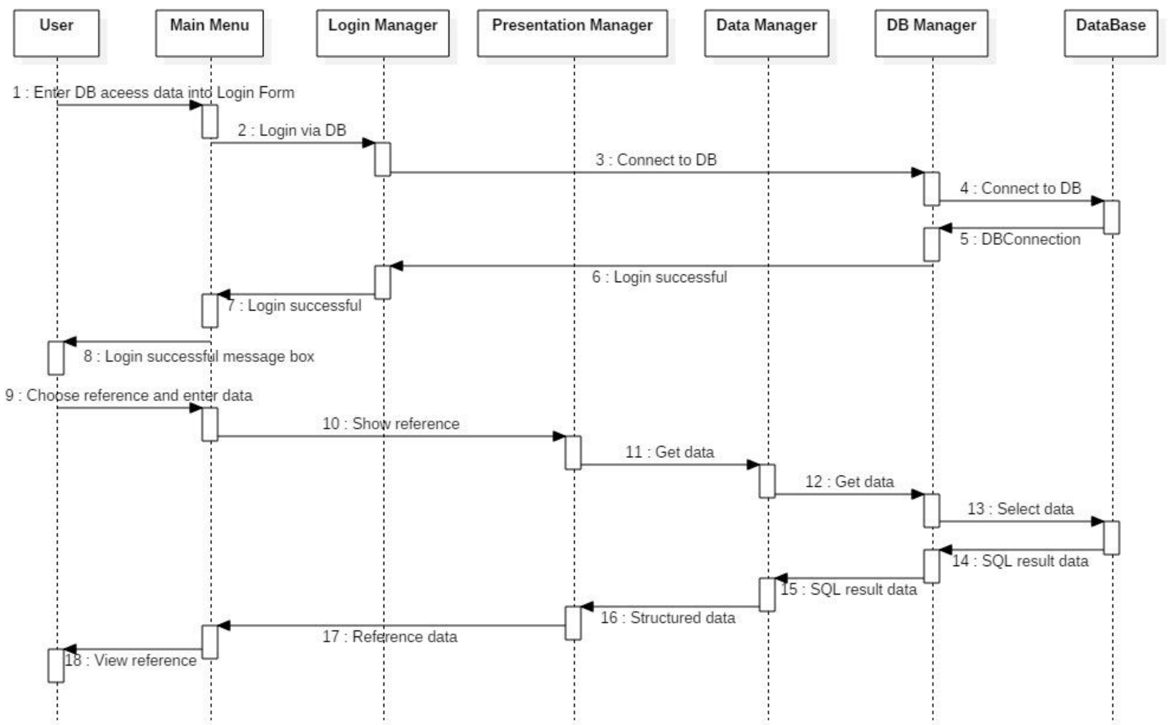
- ***Presentation layer*** – слоят е отговорен за представителната логика, обработване на клиентските заявки за достъп до приложението и визуализиране на резултатите;
- ***Business layer*** – в този слой се извършва фактическата обработка на данните, вкл. избор/зареждане на източник на данни, сравняване на данни, анализ и обработване на данни, в подходящ за визуализиране вид;
- ***Data Base access layer*** – слоят съдържа компоненти за извличане на данни от база данни, а също така и за зареждане на данни в „олекотена“ SQLite база данни в оперативната памет;
- ***Data layer*** – слой за данните, вкл. релационни бази данни и CSV данни.



Фиг. 2. Архитектура на HydroAnalyzer

Основните софтуерни компоненти и съответните отговорности са следните:

- ***Login Manager*** обработва входните данни за достъп и определя дали приложението ще се използва в онлайн или в офлайн режим. Ако потребителят е въвел коректни данни за идентификация, име и IP адрес на база данни, Login Manager предава управлението на компонента DB Manager. В противен случай извиква компонента CSV Loader за работа с CSV-файлове.
- ***CSV Loader*** е отговорен за извличането на данни от локален файл със CSV формат, преформатирането им чрез подходящи SQL заявки и предаването им към SQL Lite DB.
- ***SQL Lite DB*** е база данни, която работи локално, в оперативната памет на компютъра. Тя функционира без да стартира отделен сървърен процес и без допълнителна конфигурация, което я прави лесна за употреба. Използва се като алтернативна олекотена система за бази данни при зареждане на данните от CSV-файлове.
- ***DB Manager*** е отговорен за извличане на конкретни данни от базата данни и предаването им към Data Manager. Имплементирани са специфични особености в SQL синтаксиса при работа с двата вида бази данни – MySQL и SQLite.
- ***Data Manager*** извършва обработката на данните, получени от DB Manager – създава вътрешни за приложението представяния на данните, сравнява ги и открива грешни стойности.
- ***Presentation Manager*** е отговорен за това, данните получени от Data Manager да се преобразуват до вид, подходящ за различни таблични или графични компоненти, използвани при визуализацията.



Фиг. 3. Диаграма на последователностите за работа с HydroAnalyzer

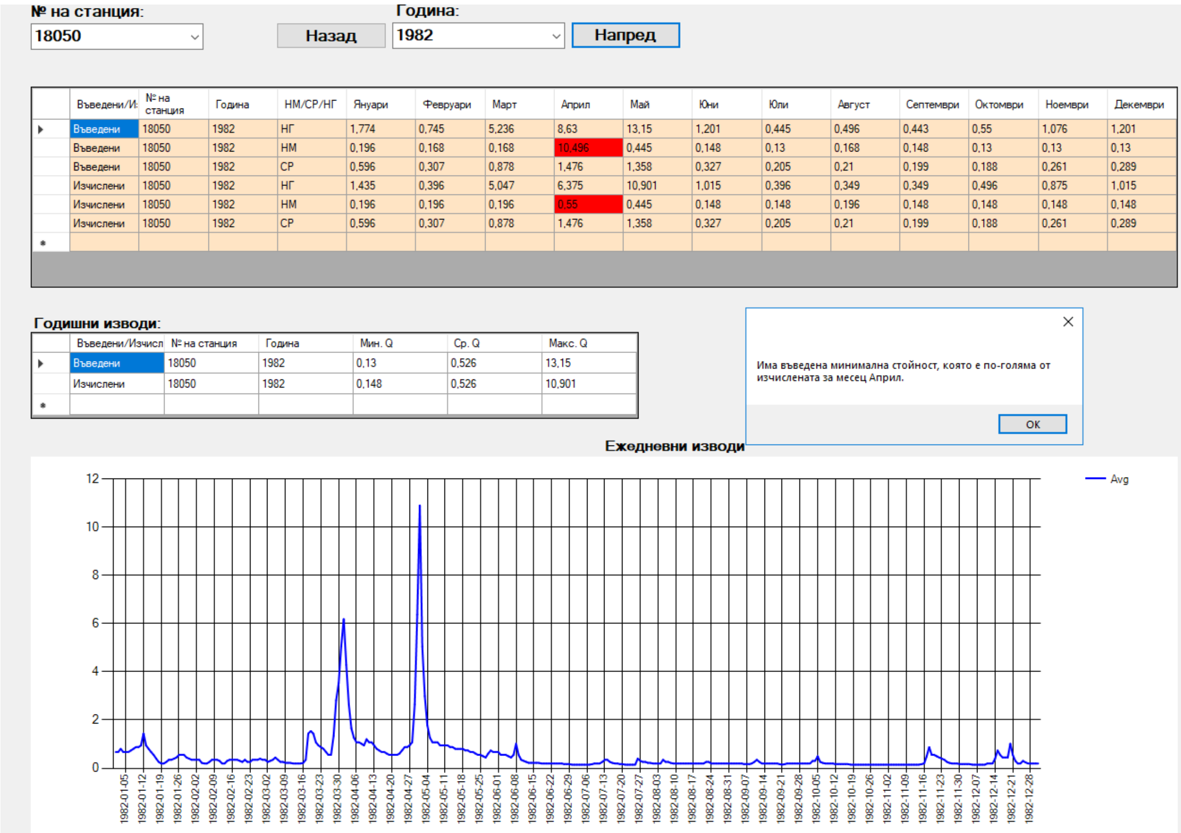
Преформатирането на данните се извършва по следния начин: данните се извличат от CSV файла, и посредством специфични SQL заявки се записват в SQL Lite DB. След това, чрез Data Manager се зареждат в подходящи структури на C#. Presentation Manager ги визуализира в табличен и/или графичен вид.

На фиг. 3 е показан целия процес на работа на приложението в онлайн режим. Първоначално потребителят въвежда данни за идентификация в началния екран на приложението, вкл. потребителско име, парола, име и IP адрес за връзка с база данни. Login Manager проверява данните, и предава информация за конкретната база данни към DB Manager. Компонентът се свързва с базата данни и взема „обект за връзка с базата данни“, след което връща съобщение към Login Manager за успешно извършване на операцията. На потребителят се визуализира форма, в която може да избере някоя от функционалностите на приложението – извличане на данни, сравняване или обработка на данни. Чрез последователното извикване на компонентите Presentation Manager, Data Manager и DB Manager, данните се извличат от базата данни. След това, DB Manager предава данните в SQL формат към Data Manager, който ги преобразува в подходящи структури чрез възможностите на езика за програмиране C# и ги предава на Presentation Manager. Този компонент, от своя страна, ги организира във вид подходящ за използваните в приложението стандартни таблични и графични компоненти, в които се визуализират за потребителя.

Фиг. 4. Жизнен цикъл на данните в HydroAnalyzer

Обработката на данните преминава през различни етапи, които са обобщени на фиг. 4. Данните в отдалечената MySQL база данни се въвеждат по независими от приложението начини или в локална SQLite база данни чрез използването на CSV-файл. В последствие те се преобразуват в подходящи структури за обработка в приложението, и на по късен етап – в данни, пригодни за табличните и графичните компоненти.

На фиг. 5 е показан екран от приложението. За конкретна година и станция, в табличен вид се представят по редове месечна информация за въведените от хидролога и автоматично изчислените минимални, максимални и средни стойности за водните количества. При отклонения между въведените и изчислените стойности за един месец, и един и същи вид стойности, клетките на таблицата се оцветяват в различен цвят. Работещият с приложението може да отчете грешките, като в офлайн версията на приложението има възможност за записване на коментари в локална директория на компютъра. От ежедневните автоматизирани отчитания се извежда графика. Годишните изводи за въведените и изчислени стойности се визуализират в отделна таблица.



Фиг. 5. Екран от работата на приложение HydroAnalyzer

3. Заключение

Оптимизирането на дейностите по съхранение, анализ, обработване и визуализиране на коректни хидрометрични данни е от съществено значение за изработване на добри анализи и прогнози. Откриването на некоректни данни и предприемането на мерки за отстраняването им води до повишаване качеството на данните.

В статията е представено приложението HydroAnalyzer, имащо за основна цел да облекчи работата на специалистите от НИМХ. То работи в два режима – онлайн и офлайн. Онлайн режимът позволява в реално време достъп до хидроложките данни, тяхната обработка и визуализация. Служителите, които имат ограничен достъп до данни, могат да ползват приложението офлайн – чрез импортиране на данни от наличен CSV файл. Те зареждат в приложението данните от файла и могат да ги обработват по същия начин, по който и в онлайн режим.

Идеите за бъдещо развитие на системата включват създаването на алгоритми, обобщаващи и анализиращи данните от различни станции, както и на предсказания за стойности на различни хидроложки величини с използването на методи на статистически анализ и невронни мрежи [7-9].

Благодарности: Работата е подкрепена от проект СП17-ФМИ-005 „Студентска школа за ИКТ иновации в бизнеса и обучението" към Фонд „Научни изследвания" при Пловдивския университет „П. Хилендарски".

Литература

1. *Национален институт по метеорология и хидрология*, онлайн: http://www.meteo.bg, последно посетен на 16.07.2018 г.
2. *Базова наблюдателна мрежа на НИМХ*, онлайн: http://hydro.bg/gwMap/htdocs/gwmap.phtml, последно посетен на 16.07.2018 г.
3. *Прогнози на НИМХ*, онлайн: http://www.meteo.bg/bg/prognozi, последно посетен на 16.07.2018 г.
4. *MySQL*, онлайн: https://www.mysql.com, последно посетен на 16.07.2018 г.
5. *SQLite*, онлайн: https://www.sqlite.org, последно посетен на 16.07.2018 г.
6. *Microsoft Visual Studio*, онлайн: https://visualstudio.microsoft.com, последно посетен на 16.07.2018 г.
7. Taylor, M., M. Koning, *The Math of Neural Networks*, Publ. Blue Windmill Media, 2017.
8. Hassoun, M., *Fundamentals of Artificial Neural Networks*, Publ. A Bradford Book, 2003, ISBN: 0262514672.
9. Hagan, M., H. Demuth, M. Beale, *Orlando De Jesús, Neural Network Design*, Publ. Martin Hagan, 2014, ISBN: 0971732116.