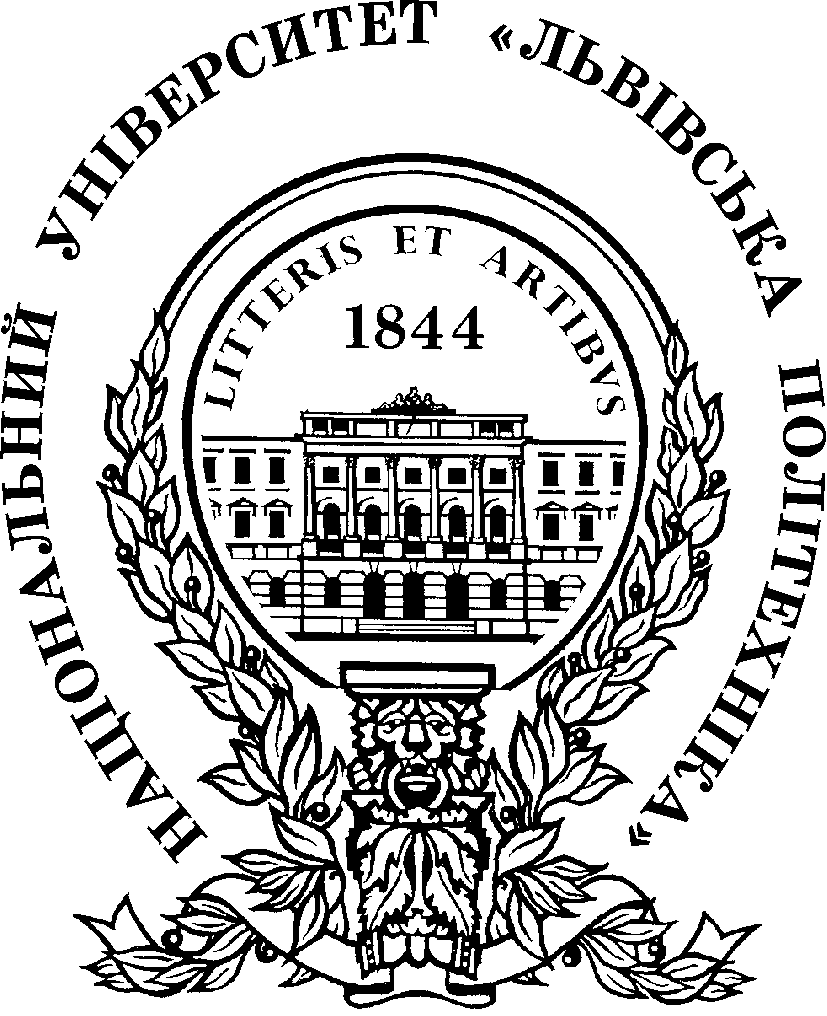
МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ «ЛЬВІСЬКА ПОЛІТЕХНІКА»

Кафедра ІСМ



**Звіт**

До розрахункової роботи

З дисципліни:

«Веб-технології та веб-дизайн»

Виконав:

Студент групи КН-48

Пахолюк С.В.

Прийняв:

Доцент каф. ІСМ,

Василюк А. С.

**Львів-2017**

ЗАВДАННЯ

**Практичне завдання:** За допомогою мови гіпертекстової розмітки HTML написати ієрархічний веб-сайт з трьох сторінок на довільну тему.

**Теоретичне завдання:** Web-mining. Методи та засоби видобування даних з Веб-ресурсів.

**Web Mining** - це використання методів інтелектуального аналізу даних для автоматичного виявлення веб-документів і послуг, вилучення інформації з веб-ресурсів та виявлення загальних законів в Інтернеті.

В Web Mining можна виділити наступні етапи:

* входной етап (англ. input stage) - отримання сировини даних із джерел (логи серверів, тексти електронних документів);
* етап попередньої обробки (англ. preprocessing stage) - дані представляються у формі, необхідної для успішного побудови того чи іншого моделі;
* етап моделювання (англ. stage stage discovery);
* етап аналізу моделі (англ. stage stage analysis) - інтерпретація отриманих результатів.

Це загальні кроки, які необхідно пройти для аналізу даних мережі Інтернет. Конкретні процедури кожної стадії залежать від поставленої задачі. У зв'язку з цим виділяємо різні категорії Web Mining:

* Видобування веб-контенту;
* Веб-структура гірничодобувна;
* Використання веб-гірництва.

**Видобуток веб-контенту** - витяг процесу знань із контенту документів або їх описів, доступних в Інтернеті. Пошуки знань в мережі Інтернет є непростою та трудоемкою задачею. Іменно це напрямок Web Mining вирішує її. Він оснований на об'єднанні можливостей інформаційного пошуку, машинного навчання та інтелектуального аналізу даних.

**Web Structure Mining** - процес виявлення структурної інформації в Інтернеті. Дане напрямок розглядає взаємозв'язок між веб-сторінками, будучи основою на зв'язках між ними. Побудовані моделі можуть бути використані для сортування та пошуку аналогічних веб-ресурсів, а також для розпізнавання авторських сайтів.

**Використання Web Mining**- це автоматичне виявлення шаблонів у русі пересування користувача та пов'язаних з ним даних, зібраних або придбаних внаслідок взаємодії з одним або декількома веб-сайтами. Ця спрямованість базується на вилученні даних з логів веб-серверів. Целью аналізу є виявлення переваг відвідувачів при використанні тих чи інших ресурсів мережі Інтернет.

**Web Mining і інформаційний пошук**

Деякі стверджують, що інформаційний пошук в Інтернеті - це окремий випадок Web Mining, інші асоціюють Web Mining з інтелектуальним інформаційним пошуком. Насправді інформаційний пошук - це автоматичний пошук всіх необхідних документів, однак, в той же час не виключено отримання деяких нерелевантних документів. Основні завдання інформаційного пошуку полягають у пошуку корисних документів, повнотекстовому індексуванні, і в даний час дослідження в області інформаційного пошуку включають в себе моделювання, класифікацію та категоризацію документів, призначених для користувача інтерфейсів, візуалізацію даних, фільтрацію, і т.д. Задача, яку, як вважається, виконує окремий випадок Web Mining - це класифікація або категоризація веб-документів, які можуть бути використані для індексації. У зв'язку з цим, Web Mining є частиною процесу інформаційного пошуку. Проте, слід зазначити, що не всі завдання індексації використовують методи інтелектуального аналізу даних.

**Web Mining і вилучення інформації**

Метою інформаційного вилучення є перетворення колекції документів, зазвичай за допомогою інформаційно-пошукових систем, в легко засвоювану і проаналізовану інформацію. Процес вилучення інформації спрямований на виймання релевантних фактів з документів, в той час як процес інформаційного пошуку спрямований на селекцію релевантних документів. Перший зацікавлений в структурі або поданні документа, тобто працює на рівні тонкої деталізації, а другий розглядає текст документа як колекцію невпорядкованих слів. Проте, відмінності між двома процесами стають несуттєвими, якщо мета інформаційного пошуку - це вилучення інформації.

Завдяки динаміці і різноманітності веб-вмісту, створення ручного режиму систем інформаційного вилучення не представляється можливим. У зв'язку з цим, більшість систем по вилученню даних зосереджують увагу на конкретні веб-сайти. Інші використовують навчальні пристрої або методи інтелектуального аналізу даних і здатні витягувати веб-документи в автоматичному чи напівавтоматичному режимі. З цієї точки зору, Web Mining є частиною процесу вилучення інформації з Інтернету.

**Web Mining і машинне навчання**

Web Mining працює не за тим же принципом, що і застосовуються в Інтернеті методи машинного навчання. З одного боку, існують деякі додатки машинного навчання, які не є окремим випадком Web Mining. Прикладом цього є метод, ефективно використовує веб-павука для конкретної теми, або метод, який акцентував на плануванні кращого шляху, який буде пройдений наступним. З іншого боку, крім методів машинного навчання, існують і інші методи, які застосовуються до Web Mining. Наприклад - деякі запатентовані алгоритми, які використовуються для видобутку хабів і авторитетних сторінок, DataGuides і алгоритми виявлення веб-схем. Проте, між двома областями досліджень існує тісний зв'язок, і методи машинного навчання можуть бути застосовані до процесів Web Mining. Наприклад, недавні дослідження показали, що застосування методів машинного навчання може поліпшити процес класифікації текстів, в порівнянні з результатами роботи традиційних методів інформаційного пошуку.

**Web Content Mining**

Web Content Mining описує автоматичний пошук інформаційних ресурсів в Інтернеті і включає в себе видобуток вмісту з веб-даних. По суті, Web Content Mining є аналогом методу інтелектуального аналізу даних для реляційних баз даних, так як існує можливість знайти схожі типи знань з неструктурованих даних, що знаходяться в веб-документах. Веб-документ може містити кілька типів даних, такі як текст, зображення, аудіо, відео, метадані та гіперпосилання. Деякі з них частково структуровані, такі як HTML-документи, деякі більш структуровані, такі як дані в таблицях або базах даних, але більшість інформації зберігається в неструктурованих текстових даних.

Існують різні методи пошуку інформації в Інтернеті. Найбільш поширеним підходом є пошук на основі ключових слів. Традиційні пошукові системи мають сканери для пошуку і збору корисної інформації в Інтернеті, методи індексування для зберігання інформації і обробки запитів, щоб запропонувати користувачам більш точну інформацію. Web Content Mining виходить за рамки традиційної технології IR (англ. Information Retrieval).

Існує два підходи до Web Content Mining: Агентно і орієнтований на базу даних. У першому випадку, видобуток даних здійснюють програмні агенти, у другому випадку дані розглядаються як належні до бази.

Підхід, заснований на агентах, включає такі системи:

* інтелектуальні пошукові агенти (Intelligent Search Agents);
* фільтрація інформації / класифікація;
* персоніфіковані агенти мережі.

**Web Structure Mining**

Web Structure Mining - це процес виявлення структурної інформації в Інтернеті, який може бути розділений на два види, заснованих на типі структури використовуваної інформації:

* гіперпосилання
* структура документа

Гіперпосилання

Гіперпосилання є структурною одиницею, яка з'єднує локацію в веб-сторінці з іншого, або в межах однієї веб-сторінки або на інший веб-сторінці. Гіперпосилання, яка підключається до іншої частини тієї ж сторінки називається внутрішньо-документної гіперпосиланням, а гіперпосилання, яка з'єднує дві різні сторінки називається меж-документної гіперпосиланням.

Структура документа

Зміст веб-сторінки може бути представлено в деревовидному форматі, заснованому на різних HTML і XML тегах. Завдання полягає в тому, щоб автоматично витягти DOM-структуру (англ. Document object model) з документів.

Web Structure Mining намагається виявити модель, що лежить в основі посилальної структури в Інтернеті. Модель заснована на топології гіперпосилання з або без опису посилання. Ця модель може бути використана для класифікації Веб-сторінки і корисна для отримання інформації, такі як схожість і відносини між веб-сайтами. Посилальна структура містить важливу інформацію, і може допомогти в фільтрації та ранжируванні веб-сторінок. Зокрема, посилання зі сторінки А на сторінку В може вважатися рекомендацією сторінки B автором А.

Були запропоновані деякі нові алгоритми, що використовують кількість посилань структуру не тільки для пошуку за ключовими словами, а й інших завдань, таких як автоматичне створення Yahoo-подібних ієрархій або ідентифікацій спільнот в Інтернеті. Якісне виконання цих алгоритмів, як правило, краще, ніж виконання IR-алгоритмів, оскільки вони використовують більше інформації, ніж просто вміст сторінок.

**Web Usage Mining**

Web Usage Mining - це процес вилучення корисної інформації з користувальницьких журналів доступу, журналів проксі-сервера, браузерних журналів, призначених для користувача сесійних даних. Говорячи простою мовою, Web Usage Mining - це процес з'ясування того, що користувачі шукають в Інтернеті. Деякі користувачі можуть бути зацікавлені тільки в текстових даних, в той час як інші можуть більше приділяти уваги мультимедійних даних.

Аналізується наступна інформація:

* які сторінки переглядав користувач;
* яка послідовність перегляду сторінок.

Аналізується також, які групи користувачів можна виділити серед загального їх числа на основі історії перегляду Web-вузла.

Web Usage Mining включає наступні складові:

* Попередня обробка;
* операційна ідентифікація;
* інструменти виявлення шаблонів;
* інструменти аналізу шаблонів.

На першому етапі відбувається збір даних і попередня обробка даних. Попередня стадія обробки включає в себе очищення clickstream-даних (англ. Clickstream - маршрут переміщення користувача по веб-сайту) і розбиття даних на безліч користувальницьких транзакцій з їх візитами на веб-сайт. На стадії виявлення шаблону, статистичні алгоритми і алгоритми баз даних виконуються на журнали транзакцій, щоб знайти приховані закономірності і поведінку користувачів. На заключному етапі аналізу шаблонів, виявлені зразки з попереднього етапу послідовно обробляються і фільтрують вироблені моделі, які згодом можуть використовуватися як вхідні дані в різні інструменти візуалізації і інструменти генерації звітів.

Статистика фіксує ідентифікаційні дані веб-користувачів разом з їх поведінкою на сайті. Залежно від виду використання даних, результатом роботи Web Usage Mining будуть:

* Дані веб-сервера;
* Дані серверних додатків;
* Дані прикладного рівня.

Дані веб-сервера

Веб-сервером збираються журнали користувача і зазвичай включають в себе IP-адреса, посилання на сторінку і час доступу.

Дані серверних додатків

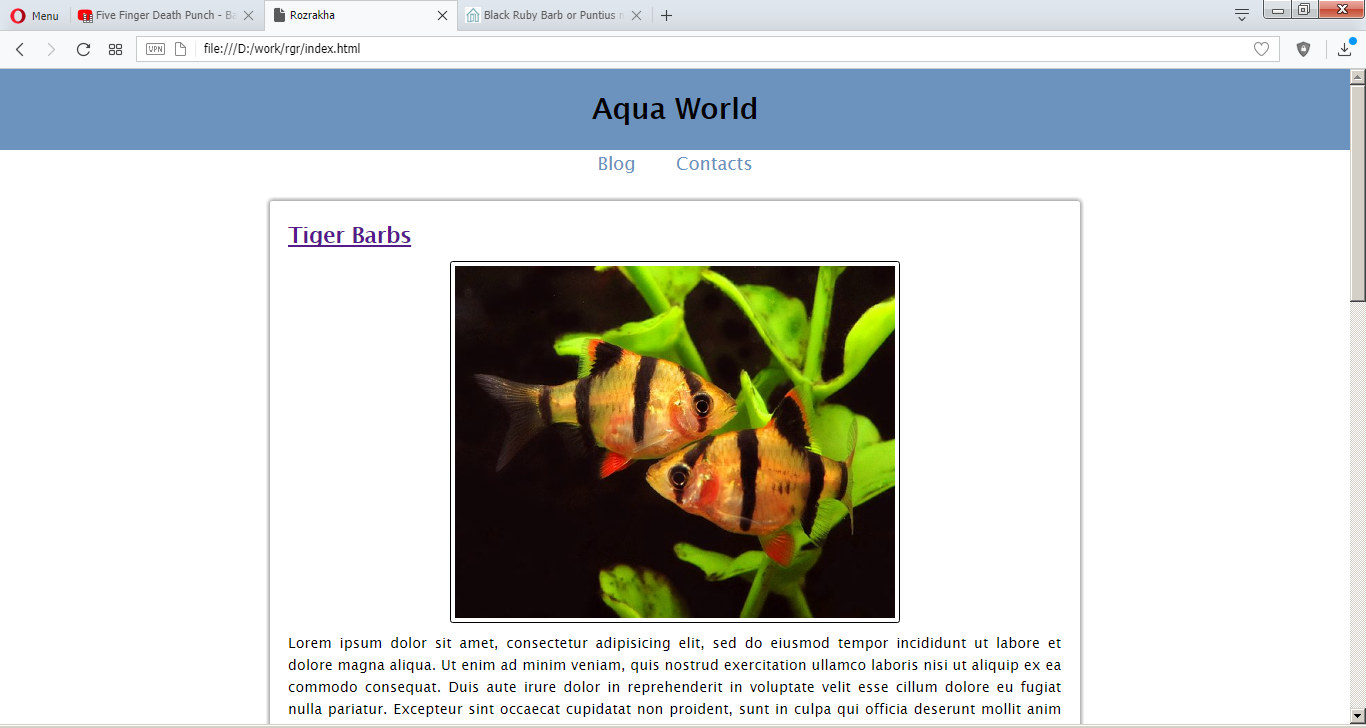
Комерційні сервери додатків, такі як WebLogic, StoryServer, мають значні можливості, що дозволяють влаштуватися на їх вершині додатків для електронної комерції. Ключовою особливістю є можливість відстежувати різні види ділових заходів і реєструвати їх в журналах сервера додатків.

Дані прикладного рівня

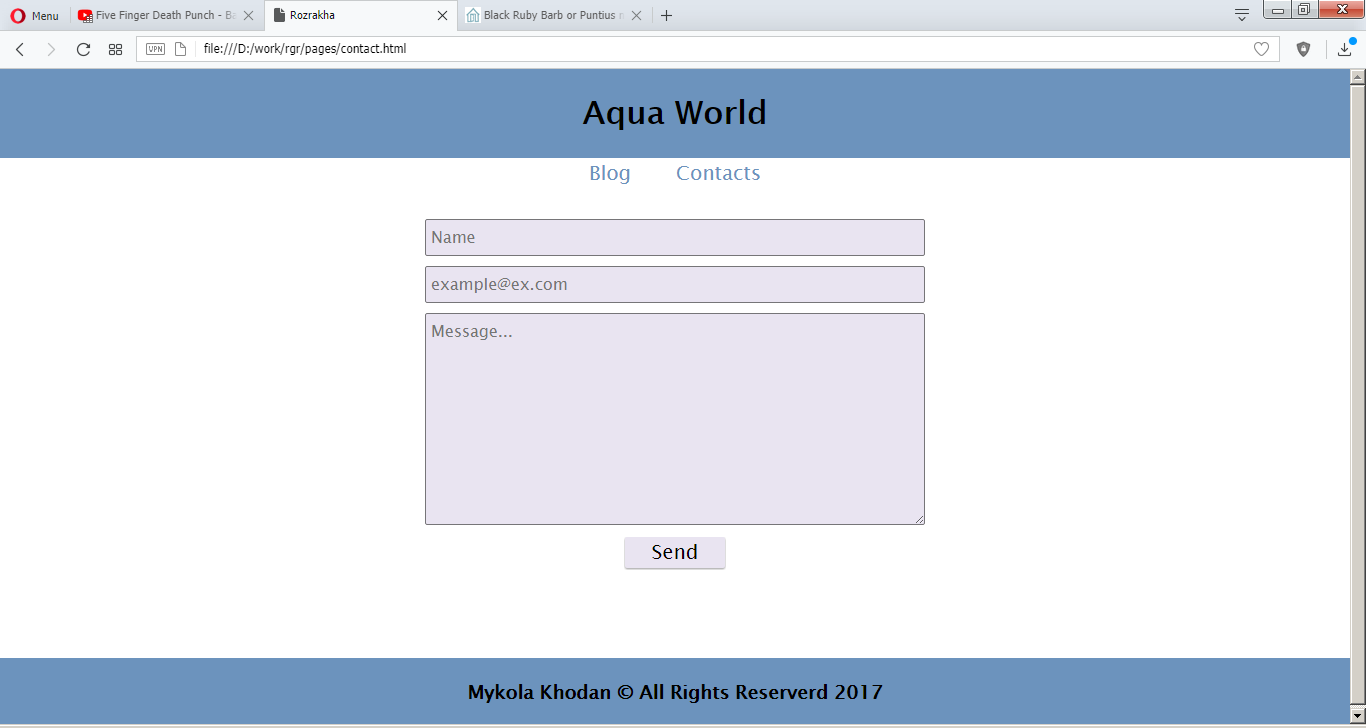
У додатку можуть бути визначені нові види подій, реєстрація яких, може включати історію створення цих подій. Слід зазначити, що багато кінцеві додатки вимагають поєднання одного або декількох методів, що застосовуються в категоріях, зазначених вище.

ПРАКТИЧНА ЧАСТИНА

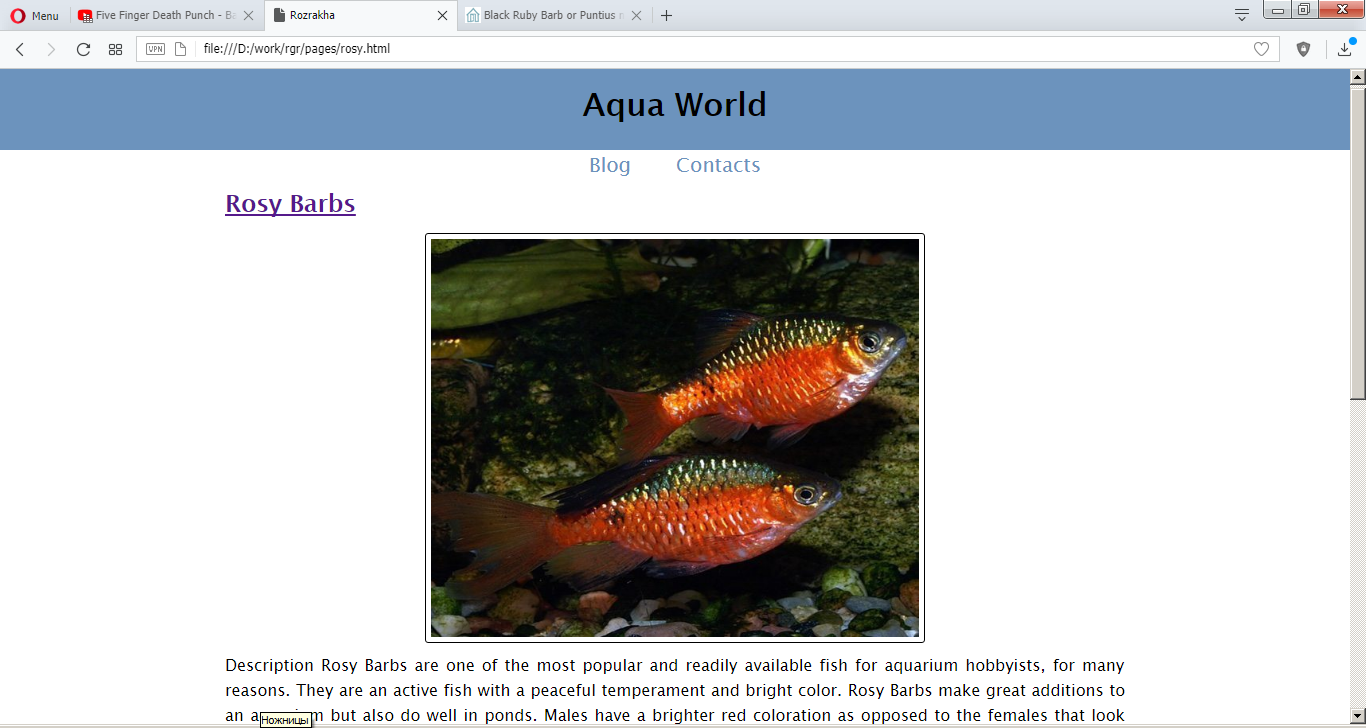
В ході виконання розрахункової роботи було розроблено невеличкий блог на тему акваріумних рибок.



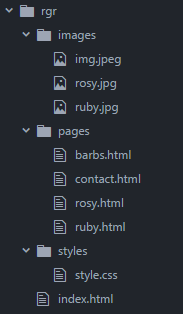
*Рис.1 Головна сторінка сайту.*



*Рис.2 Сторінка з формою зворотного зв’язку.*



*Рис.3 Сторінка однієї з статей блогу.*



*Рис.4 Робочий каталог розробленого сайту.*

Код програми:

index.html:

<!DOCTYPE html>

<html>

<head>

<meta charset="utf-8">

<meta name="viewport" content="width=device-width, initial-scale=1.0">

<title>Rozrakha</title>

<link rel="stylesheet" href="styles/style.css">

</head>

<body>

<header>

<div class="centralizer">

<h1>Aqua World</h1>

</div>

</header>

<nav>

<div class="centralizer">

<a href="index.html" class="nav-link">Blog</a>

<a href="pages/contact.html" class="nav-link">Contacts</a>

</div>

</nav>

<main>

<div class="centralizer">

<div class="blog-item">

<h2>

<a href="pages/barbs.html" class="blog-title">

Tiger Barbs

</a>

</h2>

<img src="images/img.jpeg" alt="title" class="blog-img">

<div class="blog-description">

<p>

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

</p>

</div>

</div>

<div class="blog-item">

<h2>

<a href="pages/rosy.html" class="blog-title">

Rosy Barbs

</a>

</h2>

<img src="images/rosy.jpg" alt="title" class="blog-img">

<div class="blog-description">

<p>

The Rosy Barb is one of the larger members of the Barb species and it can grow up to 6 inches in the wild. This fish enjoys company, and if it's school is large enough, it won't bother others in the aquarium.

</p>

</div>

</div>

<div class="blog-item">

<h2>

<a href="pages/ruby.html" class="blog-title">

Ruby Barbs

</a>

</h2>

<img src="images/ruby.jpg" alt="title" class="blog-img">

<div class="blog-description">

<p>

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

</p>

</div>

</div>

</div>

</main>

<footer>

<div class="centralizer">

<h3>Mykola Khodan &copy; All Rights Reserverd 2017</h3>

</div>

</footer>

</body>

</html>

contact.html:

<!DOCTYPE html>

<html>

<head>

<meta charset="utf-8">

<meta name="viewport" content="width=device-width, initial-scale=1.0">

<title>Rozrakha</title>

<link rel="stylesheet" href="../styles/style.css">

</head>

<body>

<header>

<div class="centralizer">

<h1>Aqua World</h1>

</div>

</header>

<nav>

<div class="centralizer">

<a href="../index.html" class="nav-link">Blog</a>

<a href="#" class="nav-link">Contacts</a>

</div>

</nav>

<main>

<div class="centralizer">

<form class="fb-form" action="index.html" method="post">

<input type="text" name="name" value="" placeholder="Name" class="form-input" required>

<input type="email" name="email" value="" placeholder="example@ex.com" class="form-input" required>

<textarea name="name" rows="8" cols="80" placeholder="Message..." class="form-input" required></textarea>

<input type="submit" name="submit" value="Send" class="form-submit">

</form>

</div>

</main>

<footer>

<div class="centralizer">

<h3>Mykola Khodan &copy; All Rights Reserverd 2017</h3>

</div>

</footer>

</body>

</html>

ВИСНОВКИ

В результаті виконання даної розрахункової роботи було розроблено сайт-блог на тему акваріумних рибок. При розробці даного сайту було використано найновіші засоби та методи розробки веб сайтів, закріплено знання та вміння із HTML та CSS.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1.W3Schools [Електронний ресурс] Режим доступу:

https://www.w3schools.com

2.Mozilla Developers Network [Електронний ресурс] Режим доступу:

https://www.developer.mozilla.org

3. Stackoverflow [Електронний ресурс] Режим доступу:

https://stackoverflow.com