МИНИСТЕРСТВО науки и высшего ОБРАЗОВАНИЯ РОссИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ

УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«МОСКОВСКИЙ АВИАЦИОННЫЙ ИНСТИТУТ

(национальный исследовательский университет)»

Кафедра 319 «Интеллектуальные системы мониторинга»

**КУРСОВАЯ РАБОТА**

по дисциплине

«Основы построения промышленных программных систем»

**«Проектирование и разработка веб-приложения классификации текстов с применением методов машинного обучения»**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Студент | | |  | Родионов В. С. |
| Группа | | | М3О-221М-20 |  |
| Руководитель | | |  | Полицына Е. В. |
| Оценка |  | Дата защиты «25» декабря 2021 г. | | |

**Москва 2021**

МИНИСТЕРСТВО науки и высшего ОБРАЗОВАНИЯ РОссИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ

УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«МОСКОВСКИЙ АВИАЦИОННЫЙ ИНСТИТУТ

(национальный исследовательский университет)»

Кафедра 319 «Интеллектуальные системы мониторинга»

УТВЕРЖДАЮ

Заведующий кафедрой

д.т.н., профессор Нагибин С.Я.

(И.О.Фамилия)

«\_\_\_\_»\_\_\_\_\_\_\_ 2021 г.

**З А Д А Н И Е**

на курсовую работу по дисциплине

Основы построения промышленных программных систем

Студент М3О-221М-20, Родионов Вадим Станиславович

(№ группы, Ф. И. О.)

Тема «Проектирование и разработка веб-приложения классификации текстов с применением методов машинного обучения»

Перечень вопросов, подлежащих разработке в курсовой работе

Реализовать клиент-серверную систему классификации для свой предметной области, которая представляет собой веб-приложение на Java/Python с пользовательским веб- интерфейсом.

Рекомендуемая литература

Дронов В. А. Django 3.0. Практика создания веб-сайтов на Python

Лутц М. Изучаем Python

Задание выдано «12» сентября 2021 г.

Руководитель Полицына Е.В., к.т.н. доцент кафедры 319\_\_\_\_\_\_

(Ф. И. О., должность, подпись)

Студент\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(подпись)

**СОДЕРЖАНИЕ**

[1 ОПИСАНИЕ ВОЗМОЖНОСТЕЙ ПРИЛОЖЕНИЯ 4](#_Toc90223318)

[1.1 Общие сведения 4](#_Toc90223319)

[1.1.1 Полное наименование системы и её условное обозначение 4](#_Toc90223320)

[1.1.2 Наименование предприятий (объединений) разработчика и заказчика (пользователя) системы и их реквизиты 4](#_Toc90223321)

[1.1.3 Основания для разработки 4](#_Toc90223322)

[1.1.4 Плановые сроки начала и окончания работы по созданию системы 4](#_Toc90223323)

[1.1.5 Порядок оформления и предъявления заказчику результатов работ по созданию системы (её частей), по изготовлению и наладке отдельных средств (технических, программных, информационных) и программно-технических (программно-методических) комплексов системы. 4](#_Toc90223324)

[1.2 Назначения и цели 5](#_Toc90223325)

[1.2.1 Назначение системы 5](#_Toc90223326)

[1.2.2 Цели создания системы 5](#_Toc90223327)

[2 ТРЕБОВАНИЯ 6](#_Toc90223328)

[2.1 Требования к системе в соответствии с ТЗ 6](#_Toc90223329)

[2.2 Функциональные требования 6](#_Toc90223330)

[2.3 Нефункциональные требования 7](#_Toc90223331)

[3 АРХИТЕКТУРА ПРИЛОЖЕНИЯ 8](#_Toc90223332)

[4 ОПИСАНИЕ КЛАССИФИКАТОРА 10](#_Toc90223333)

[4.1 Классы 10](#_Toc90223334)

[4.2 Вектор признаков 10](#_Toc90223335)

[4.3 Обучающие и тестовые коллекции 10](#_Toc90223336)

[5 ОПИСАНИЕ ИНФРАСТРУКТУРЫ РАЗРАБОТКИ 14](#_Toc90223337)

[6 РЕЗУЛЬТАТЫ ТЕСТИРОВАНИЯ 16](#_Toc90223338)

[6.1 Сценарии тестирования 16](#_Toc90223339)

[6.1.1 Сценарий ручного тестирования клиентской части. 16](#_Toc90223340)

[6.2 Результаты тестирования 27](#_Toc90223341)

[7 ИНТЕРФЕЙС И ВОЗМОЖНОСТИ СИСТЕМЫ 29](#_Toc90223342)

[8 АНАЛИЗ РЕЗУЛЬТАТОВ 33](#_Toc90223343)

[СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ 34](#_Toc90223344)

# 1 ОПИСАНИЕ ВОЗМОЖНОСТЕЙ ПРИЛОЖЕНИЯ

## 1.1 Общие сведения

### 1.1.1 Полное наименование системы и её условное обозначение

Веб-приложение «Django Unchained для коллекций статей с сайта newsru.com. Просмотр, классификация, редактирование».

*Условное обозначение:* Классификатор категорий по текстам.

### 1.1.2 Наименование предприятий (объединений) разработчика и заказчика (пользователя) системы и их реквизиты

Заказчик – кафедра 319, МАИ (национальный исследовательский университет), Полицына Е. В.

Разработчик – студент группы М3О-221М-20 Родионов В. С.

### 1.1.3 Основания для разработки

Курсовая работа по предмету «Основы построения промышленных программных систем», целью которой является изучение методов машинного обучения, проектирование, моделирование, разработка и тестирование веб-приложения с использованием изученных методов классификации.

### 1.1.4 Плановые сроки начала и окончания работы по созданию системы

Начало работ по проектированию и реализации системы: сентябрь 2021 г.

Окончание работ по созданию системы: декабрь 2021 г

### 1.1.5 Порядок оформления и предъявления заказчику результатов работ по созданию системы (её частей), по изготовлению и наладке отдельных средств (технических, программных, информационных) и программно-технических (программно-методических) комплексов системы.

К результатам труда разработчика относится:

* оригинальное программное обеспечение;
* оригинальные модели классификаторов;
* оригинальные алгоритмы работы программы;
* проектная и рабочая документация.

Заказчику передаются:

* ссылка на репозиторий с исходным кодом проекта;
* исполняемый файл проекта и необходимые модели классификатора;
* устная просьба поставить оценку «Отлично»;
* сопроводительная документация проекта.

Проектная документация должна быть разработана в соответствии с ГОСТ 34.602-89. Процедуры приемки – передачи результатов работ оформляются актами приемки-передачи.

## 1.2 Назначения и цели

### 1.2.1 Назначение системы

Классификатор категорий из введённого текста предназначен для автоматизации процесса классифицирования объектов по 10 следующим классам:

* Автоновости;
* В мире;
* В России;
* Инопресса;
* Культура;
* Медицина;
* Недвижимость;
* Спорт;
* Технологии;
* Экономика;

А также из текста, введённым пользователем.

### 1.2.2 Цели создания системы

Целью создания системы является:

* автоматизация процесса хранения текстов;
* автоматизация классификации текста;
* анализ результатов практического применения машинного обучения.

# 2 ТРЕБОВАНИЯ

## 2.1 Требования к системе в соответствии с ТЗ

Система должна позволять:

1. Загрузить текст и как минимум ответить на вопрос, соответствует ли загруженный текст предметной области системы (бинарная классификация на основе методов машинного обучения);

2. При построении модели классификатора в качестве способа векторизации можно использовать любой, минимум “bag-of-words”;

3. Предоставлять возможность пользователю через веб-интерфейс просмотреть объем коллекции и примеры документов в системе с возможностью задания фильтров по каким-либо признакам, которые есть в корпусе текстов системы (например, “автор”).

Требования к реализации:

1. Бэкенд должен быть реализован на Java/Python;

2. Оформление кода должно соответствовать Java Style Guide/PEP8;

3. Для реализации фронтенда могут быть использованы любые технологии.

4. Для реализации классификатора могут быть использованы любые библиотеки, реализующие методы машинного обучения;

5. Для реализации векторизации документов или других функций могут быть использованы любые библиотеки, фреймворки, внешние API и т.д.;

6. Обязательно использование системы контроля версий для хранения всех артефактов проекта (предпочтительно GitHub);

7. Обязательно поддержание проекта в работоспособном состоянии с настроенной системой сборки и развертывания системы;

## 2.2 Функциональные требования

В соответствии с требованиями, изложенными выше, и с последующими согласованиями, были предъявлены следующие требования:

* Создание статей;
* Просмотр статей;
* Удаление статей;
* Обновление статей;
* Фильтрация и сортировка статей по признакам, который укажет пользователь;
* Определение отношения статьи к интересующему пользователю классу;

Дополнительные требования:

* Возможность записи статей с локального хранилища в формате .xml в БД веб-сервиса, полученные с помощью приложения «newsGrabb3rForCoursework».

## 2.3 Нефункциональные требования

В соответствии с требованиями, изложенными выше, и с последующими согласованиями, были предъявлены следующие требования:

* Реализуемая система должна быть масштабируемой для обеспечения дальнейшей поддержки и развития проекта
* Реализуемая система должна поддерживать разные режимы работы растеризации страницы браузером (минимальное разрешение – 800х600);
* Реализуемая система должна быть устойчива к некорректным пользовательским запросам;
* Клиентская часть должна быть понятна пользователю.

# 3 АРХИТЕКТУРА ПРИЛОЖЕНИЯ

В качестве основной технологии разработки было отдано предпочтение Django [3] на языке Python, ввиду простоты написания кода и развёртывания программы [4].

Основным паттерном проектирования является Model-View или «Модель-представление». В соответствии с этим паттерном приложение представляет собой отдельные друг от друга компоненты, тем самым позволяя реализовывать логику, не меняя других элементов системы.

Таким образом, разработанное веб-приложение представляет собой клиентскую часть (веб-страницы передаются на браузер клиента) и серверную часть, которая выполняет функции, которые захочет пользователь, но не более тех, которые описаны документацией.

Выбранные технологии для Frontend:

1. HTML – язык разметки для веб-сайтов. Просто, удобно, надёжно.

2. CSS – язык стилей для веб-сайтов. Предназначен для оформления страниц и стилей, создания кнопок и других форм без использования Js.

3. Js – популярный скриптовый язык для сайтов. Для сценариев и асинхронных запросов.

Backend:

1. Python 3.6 и Django 3.2 – язык программирования и Фреймворк.

2. Библиотека scikit-learn и ntlk для реализации машинного обучения.

3. Встроенная БД SQLite.

Для разработки непосредственно самого приложения выбрана IDE PyCharm Professional 2021, ввиду большого функционала и удобства работы.

Пример архитектуры проекта представлен на рисунке 1.

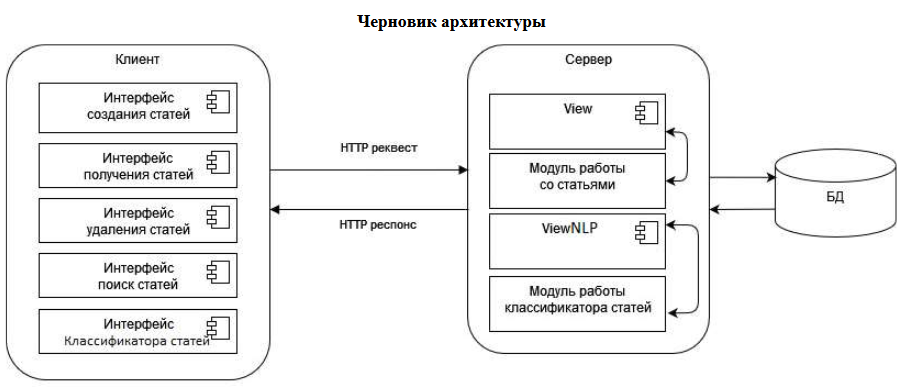


Рисунок 1 – Архитектура проекта

Точкой входа в систему является адрес «127.0.0.1» по порту 8000. После перехода по указанному адресу отображается главная страница с возможностью выбора того, что хочется пользователю.

# 4 ОПИСАНИЕ КЛАССИФИКАТОРА

## 4.1 Классы

Были выбраны следующие классы:

* Автоновости – новости про автомобили и связанные с ними изобретениями;
* В мире – новости, происходящие в мире;
* В России – новости, происходящие на территории РФ (Крым так же включён в эту выборку);
* Инопресса – мнения и новости иностранных агентств о РФ на русском;
* Культура – новости со всего мира из мира культуры;
* Медицина – новости про медицину с соответствующей терминологией;
* Недвижимость – новости про недвижимость со всего мира;
* Спорт – новости про спорт со всего мира;
* Технологии – новости про технологии со всего мира;
* Экономика – новости про экономику со всего мира;

Так же, на сайте newwsru.com присутствует раздел «мнения». Его было решено не включать, так как он сильно схож с классами «В мире» и «В России».

## 4.2 Вектор признаков

Вектор признаков представляет собой набор чисел с плавающей запятой, представленных в виде сумки слов, связанных с темой текстов обучающей выборки, где TF-IDF – надстройка над способом векторизации.

## 4.3 Обучающие и тестовые коллекции

Составлена из 50 статей для каждой категории, которая так же является и тестовой выборкой. Для выбора статей был написан запрос к БД следующего вида и представлен в листинге 1.

SELECT id FROM XMLCollection\_article where category='category' limit 50;

Листинг 1 – выборка статей из БД

Где параметр «category» соответствует одному из классов, указанных в пункте 1.

Обучение модели классификатора производилось через POST запрос «studyMLFromNLP», а для его оценки использовались закомментированные участки кода в методе «studyClassificator» файла «metodsNLP.py». Так, в представленных таблицах 1-4 был проанализирован результат и сделан выбор в пользу классификатора LogisticRegression ввиду большей точности по сравнению с другими.

Таблица 1 – Классификатор SVC

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Точность: 0,76 | | | | |
|  | precision | recall | f1-score | support | |
| 0 | 1.00 | 0.89 | 0.94 | 9 | |
| 1 | 0.60 | 0.43 | 0.50 | 8 | |
| 2 | 0.40 | 0.80 | 0.53 | 8 | |
| 3 | 0.67 | 0.67 | 0.67 | 12 | |
| 4 | 0.86 | 0.86 | 0.86 | 1 | |
| 5 | 0.80 | 0.89 | 0.84 | 10 | |
| 6 | 1.00 | 0.65 | 0.79 | 11 | |
| 7 | 1.00 | 1.00 | 1.00 | 13 | |
| 8 | 1.00 | 0.69 | 0.82 | 10 | |
| 9 | 0.69 | 0.90 | 0.78 | 9 | |
|  |  |  |  |  | |
| accuracy |  |  | 0.76 | 100 | |
| macro avg | 0.80 | 0.78 | 0.77 | 100 | |
| weighted avg | 0.82 | 0.76 | 0.77 | 100 | |

Таблица 2 – Классификатор LogisticRegression

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Точность: 0.8 | | | | |
|  | precision | recall | f1-score | support |
| 0 | 1.00 | 0.89 | 0.94 | 9 |
| 1 | 0.67 | 0.57 | 0.62 | 7 |
| 2 | 0.64 | 0.70 | 0.67 | 10 |
| 3 | 0.75 | 0.67 | 0.71 | 9 |
| 4 | 0.67 | 0.86 | 0.75 | 7 |
| 5 | 0.80 | 0.89 | 0.84 | 9 |
| 6 | 0.88 | 0.75 | 0.81 | 20 |
| 7 | 0.86 | 1.00 | 0.92 | 6 |
| 8 | 0.91 | 0.77 | 0.83 | 13 |
| 9 | 0.77 | 1.00 | 0.87 | 10 |
|  |  |  |  |  |
| accuracy |  |  | 0.80 | 100 |
| macro avg | 0.79 | 0.78 | 0.80 | 100 |
| weighted avg | 0.81 | 0.76 | 0.80 | 100 |

Таблица 3 – Классификатор KNeighborsClassifier

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Точность: 0.73 | | | | |
|  | precision | recall | f1-score | support |
| 0 | 0.89 | 0.89 | 0.89 | 9 |
| 1 | 0.44 | 0.57 | 0.50 | 7 |
| 2 | 0.50 | 0.70 | 0.58 | 10 |
| 3 | 0.50 | 0.33 | 0.40 | 9 |
| 4 | 0.71 | 0.71 | 0.71 | 7 |
| 5 | 0.75 | 0.67 | 0.71 | 9 |
| 6 | 0.84 | 0.80 | 0.82 | 20 |
| 7 | 1.00 | 1.00 | 1.00 | 6 |
| 8 | 0.92 | 0.85 | 0.88 | 13 |
| 9 | 0.70 | 0.70 | 0.70 | 10 |
|  |  |  |  |  |
| accuracy |  |  | 0.73 | 100 |
| macro avg | 0.73 | 0.72 | 0.72 | 100 |
| weighted avg | 0.74 | 0.73 | 0.73 | 100 |

Таблица 4 – Классификатор DecisionTreeClassifier

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Точность: 0.53 | | | | |
|  | precision | recall | f1-score | support |
| 0 | 0.78 | 0.78 | 0.78 | 9 |
| 1 | 0.25 | 0.57 | 0.35 | 7 |
| 2 | 0.40 | 0.20 | 0.27 | 10 |
| 3 | 0.55 | 0.67 | 0.60 | 9 |
| 4 | 0.67 | 0.57 | 0.62 | 7 |
| 5 | 0.50 | 0.44 | 0.47 | 9 |
| 6 | 0.69 | 0.55 | 0.61 | 20 |
| 7 | 0.50 | 1.00 | 0.67 | 6 |
| 8 | 0.60 | 0.46 | 0.52 | 13 |
| 9 | 0.43 | 0.30 | 0.35 | 10 |
|  |  |  |  |  |
| accuracy |  |  | 0.53 | 100 |
| macro avg | 0.54 | 0.55 | 0.52 | 100 |
| weighted avg | 0.56 | 0.53 | 0.53 | 100 |

# 5 ОПИСАНИЕ ИНФРАСТРУКТУРЫ РАЗРАБОТКИ

В проекте при разработке применялись методы доступные в документации и методы, доступ к которым ныне закрыт, однако доступ к ним может быть предоставлен владельцем или обслуживающим персоналом, который знаком с документацией.

Таким образом, для сборки нужных статей для обучающей и тестовой выборки применяется метод «preparedLearnNLP», а уже для обучения – метод «studyMLFromNLP».

Преимуществом такого подхода (разделения методов) является удобство сбора нужных статей для нужных категорий (в случае неполадок появляется возможность легко отследить в каком месте именно), а недостатком – при желании обновить классификатор, необходимо снова делать выборку статей с указанием в POST методе «prepare=» так же количество, на котором будет сделана выборка.

В процессе разработки применялись следующие пакеты:

* Spacy – для леммизации;
* Nltk – для процесса векторизации;
* Pandas – для создания таблиц;
* Numpy – для математических массивов;
* Sklearn – для классификатора;
* Json – для обработки JSON;
* Lxml – для обработки XML файлов;
* Mimetypes – для выделения прав на static.

В процессе разработки так же было уделено внимание выходу программы в релиз, в частности:

* Отключение настройки DEBUG;
* Указанием портов доступа;
* Отладкой всех узлов системы и их тестировании в «боевых» условиях.

Так же были описаны страницы ошибок, которые в случае возникновения вернут пользователя на страницы веб-приложения.

Для того, чтобы запустить приложение [1] на нужной машине необходимо прописать «py manage.py runserver». В случае отсутствия библиотек NTLK модуль классификации работать не будет.

# 6 РЕЗУЛЬТАТЫ ТЕСТИРОВАНИЯ

## 6.1 Сценарии тестирования

### 6.1.1 Сценарий ручного тестирования клиентской части.

Для тестирования API применялся инструмент Postman, с помощью которого делались обращения напрямую к серверу как с корректными данными, так и с некорректными, для получения результатов обращений в рамках протокола HTTP.

Таблица 5 – Тестирование запросов с кодами ответов и нужными ответами

|  |  |  |
| --- | --- | --- |
| Ожидаемый код ответа | Полученный код ответа | Запрос |
| 200 | 200 | GET http://127.0.0.1:8000 |
| 200 | 200 | GET http://127.0.0.1:8000/id=20 |
| 500 | 500 | POST http://127.0.0.1:8000/add (без тела запроса) |
| 200 | 200 | POST http://127.0.0.1:8000/add (с телом запроса) |
| 500 | 500 | POST http://127.0.0.1:8000/save (без тела запроса) |
| 200 | 200 | POST http://127.0.0.1:8000/ save (с телом запроса) |
| 200 | 200 | POST http://127.0.0.1:8000/delete |
| 200 | 200 | POST http://127.0.0.1:8000/search |
| 200 | 200 | GET http://127.0.0.1:8000/search |
| 200 | 200 | POST http://127.0.0.1:8000/check |
| 200 | 200 | POST http://127.0.0.1:8000/prepare=50 |
| 200 | 200 | POST http://127.0.0.1:8000/study |
| 200 | 200 | POST http://127.0.0.1:8000/exam=40 |

Для ручного тестирования применялись следующие браузеры: Chrome и Firefox. Результаты ручного тестирования представлены в формате тест-кейсов в таблицах 6-12.

Таблица 6 – Тестирование интерфейса веб-страницы «newsru | Коллекция»

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| № сценария | | Т-1 | | | | Приоритет | Высокий | | |
| Что проверяется | | Интерфейс веб-страницы «newsru | Коллекция» | | | | | | | |
| Предусловия | | | | | | | | | |
| Использование определённого браузера, Переход на адрес 127.0.0.1:8000 | | | | | | | | | |
| История изменений | | | | | | | | | |
| Создан | 08.12.2021 | | кем | Родионов Вадим | | | | Тестирование интерфейса веб-приложения | |
| Изменен | 08.12.2021 | | кем | Родионов Вадим | | | | Повторное тестирование интерфейса веб-приложения. | |
| Закрыт | 08.12.2021 | | кем | Родионов Вадим | | | |  | |
| Тестовый сценарий | | | | | | | | | |
| Действие | | | | | Ожидаемый результат | | | | Фактический результат (выполнено, не выполнено, выполнено с ошибкой). |
| 1. Кликнуть левой кнопкой мыши на пункт «классифицировать текст?». | | | | | 1. Выделенная кнопка открыла окно для ввода данных с клавиатуры. | | | | 1. Выполнено. |
| 2. Нажать левой кнопкой мыши на крестик. | | | | | 2. Окно ввода текста закрывается, появляется страница «newsru | Коллекция» веб-приложения.. | | | | 2. Выполнено. |
| 3. Выбрать пункт «Создать статью». | | | | | 3. Открывается страница «newsru | Создание статьи» веб-приложения с полями ввода для создания статьи. | | | | 3. Выполнено. |
| 4. Нажать комбинацию клавиш «alt + стрелка влево» | | | | | 4. Возврат на страницу «newsru | Коллекция» веб-приложения. | | | | 4. Выполнено. |
| 5. Нажать левой кнопкой мыши на пункт «классифицировать?». | | | | | 5. Окно с предупреждением о том, что «Режим классификации включён» и кнопки ОК. | | | | 5. Выполнено. |
| Продолжение таблицы 6 | | | | | | | | | |
| 6. Нажать левой кнопкой мыши на кнопку ОК. | | | | | 6. Окно с предупреждением закрывается. | | | | 6. Выполнено. |
| 7. Повторить действие 5 | | | | | 7. Окно с предупреждением о том, что «Режим классификации выключен» и кнопки ОК. | | | | 7. Выполнено. |
| 8. Повторить действие 6. | | | | | 8. Результат действия 6. | | | | 8. Выполнено. |
| 9. Нажать левой кнопкой мыши на пункт «найти статью» | | | | | 9. Появляется выпадающее меню с выбором из трёх полей и кнопкой «поиск». | | | | 9. Выполнено. |
| 10. Нажать левой кнопкой мыши на пункт «скрыть поиск» | | | | | 10. Выпадающее меню «скрыть поиск» исчезает и становится «найти статью» | | | | 10. Выполнено. |
| 11. Нажать левой кнопкой мыши на «любую» категорию в левом меню. | | | | | 11. Открывается страница «newsru | Коллекция» с коллекцией выбранных соответствующих статей и их количеством. | | | | 11. Выполнено. |
| 12. Повторить действие 4. | | | | | 12. Результат действия 4. | | | | 12. Выполнено. |
| 13. Нажать левой кнопкой мыши на «Следующая страница». | | | | | 13. Открывается страница «newsru | Коллекция» с новыми статьями, следующими за предыдущими, внизу страницы появляется так же кнопка «предыдущая страница», а количество страниц увеличивается в большую сторону на 1. | | | | 13. Выполнено. |
| 14. Повторить действие 4. | | | | | 14. Открывается страница «newsru | Коллекция» со статьями, предшествующие пункту 13, внизу страницы исчезает кнопка «предыдущая страница», а количество страниц уменьшается в меньшую сторону на 1. | | | | 14. Выполнено. |
|  | | | | | Результат | | | | Выполнено |

Таблица 7 - Тестирование интерфейса веб-страницы «newsru | Создание статьи»

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| № сценария | | Т-2 | | | | Приоритет | Высокий | | |
| Что проверяется | | Создание статьи «newsru | Создание статьи» | | | | | | | |
| Предусловия | | | | | | | | | |
| Использование определённого браузера, Переход на адрес 127.0.0.1:8000/add | | | | | | | | | |
| История изменений | | | | | | | | | |
| Создан | 08.12.2021 | | кем | Родионов Вадим | | | | Тестирование интерфейса веб-приложения | |
| Изменен | 08.12.2021 | | кем | Родионов Вадим | | | | Повторное тестирование интерфейса веб-приложения | |
| Закрыт | 08.12.2021 | | кем | Родионов Вадим | | | |  | |
| Тестовый сценарий | | | | | | | | | |
| Действие | | | | | Ожидаемый результат | | | | Фактический результат (выполнено, не выполнено, выполнено с ошибкой). |
| 1. Нажать левой кнопкой мыши на выпадающее меню «Категория» и выбрать произвольную «категорию» | | | | | 1. Открылось выпадающее меню с возможностью выбора «категории статьи», а после нажатия на выбранную «категорию», меню закрылось, и категория получила отображение выбранной категории. | | | | 1. Выполнено. |
| 2. Нажать левой кнопкой мыши на поле ввода «Заголовок» и ввести произвольную последовательность. | | | | | 2. Выбранная область ввода подсветилась, а введённая последовательность отображается корректно. | | | | 2. Выполнено. |
| 3. Нажать левой кнопкой мыши в меню «Редактирование» кнопку «создать статью» | | | | | 3. На странице будут подсвечены незаполненные поля. | | | | 3. Выполнено. |
| Продолжение таблицы 7 | | | | | | | | | |
| 4. Нажать левой кнопкой мыши на поле ввода «Дата составления» и выбрать произвольную дату. | | | | | 4. В выбранной области ввода отобразилось меню с возможностью ввода даты строго в формате даты. | | | | 4. Выполнено. |
| 5. Повторить действие 3. | | | | | 5. Результат действия 3. | | | | 5. Выполнено. |
| 6. Нажать левой кнопкой мыши на поле ввода «Текст статьи» и ввести произвольную последовательность. | | | | | 6. Результат действия 2. | | | | 6. Выполнено. |
| 7. Нажать левой кнопкой мыши на поле ввода «Теги» и ввести произвольную последовательность. | | | | | 7. Результат действия 2. | | | | 7. Выполнено. |
| 8. Нажать левой кнопкой мыши на поле ввода «Ссылка на статью» и ввести произвольную последовательность. | | | | | 8. Результат действия 2. | | | | 8. Выполнено. |
| Повторить действие 3. | | | | | 9.Открывается страница «newsru | Создание статьи» веб-приложения. | | | | 9. Выполнено. |
|  | | | | | Результат | | | | Выполнено |

Таблица 8 – Тестирование процесса выхода со страницы «Создание статьи» во время заполнения полей

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| № сценария | | Т-3 | | | | Приоритет | Высокий | | |
| Что проверяется | | Процесс выхода со страницы «Создание статьи» во время заполнения полей | | | | | | | |
| Предусловия | | | | | | | | | |
| Использование определённого браузера, Переход на адрес 127.0.0.1:8000/add | | | | | | | | | |
| История изменений | | | | | | | | | |
| Создан | 08.12.2021 | | кем | Родионов Вадим | | | | Тестирование интерфейса веб-приложения | |
| Изменен | 08.12.2021 | | кем | Родионов Вадим | | | | Повторное тестирование интерфейса веб-приложения | |
| Продолжение таблицы 8 | | | | | | | | | |
| Закрыт | 08.12.2021 | | кем | Родионов Вадим | | | |  | |
| Тестовый сценарий | | | | | | | | | |
| Действие | | | | | Ожидаемый результат | | | | Фактический результат (выполнено, не выполнено, выполнено с ошибкой). |
| 1. Нажать левой кнопкой мыши на поле ввода «Текст статьи» и ввести произвольную последовательность. | | | | | 1. Выбранная область ввода подсветилась, а введённая последовательность отображается корректно. | | | | 1. Выполнено. |
| 2. Нажать левой кнопкой мыши в меню «Редактирование» кнопку «Вернуться на главную». | | | | | 2. Открывается страница «newsru | Создание статьи» веб-приложения | | | | 2. Выполнено. |
| 3. Нажать комбинацию клавиш «alt + стрелка влево». | | | | | 3. Возврат на страницу «newsru | Создание статьи» веб-приложения с сохранёнными данными в поле «Текст статьи». | | | | 3. Выполнено. |
| 4. Нажать левой кнопкой мыши на «любую» категорию в левом меню. | | | | | 4. Открывается страница «newsru | Коллекция» с коллекцией выбранных соответствующих статей и их количеством. | | | | 4. Выполнено. |
| 5. Повторить действие 3. | | | | | 5. Результат действия 3. | | | | 5. Выполнено. |
|  | | | | | Результат | | | | Выполнено |

Таблица 9 – Тестирование классификатора текста на странице «newsru | Коллекция»

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| № сценария | | Т-4 | | | | Приоритет | Высокий | | |
| Что проверяется | | Классификатор текста на странице «newsru | Коллекция» | | | | | | | |
| Предусловия | | | | | | | | | |
| Использование определённого браузера, Переход на адрес 127.0.0.1:8000 | | | | | | | | | |
| Продолжение таблицы 9 | | | | | | | | | |
| История изменений | | | | | | | | | |
| Создан | 08.12.2021 | | кем | Родионов Вадим | | | | Тестирование интерфейса веб-приложения | |
| Изменен | 08.12.2021 | | кем | Родионов Вадим | | | | Повторное тестирование интерфейса веб-приложения | |
| Закрыт | 08.12.2021 | | кем | Родионов Вадим | | | |  | |
| Тестовый сценарий | | | | | | | | | |
| Действие | | | | | Ожидаемый результат | | | | Фактический результат (выполнено, не выполнено, выполнено с ошибкой). |
| 1. Нажать левой кнопкой мыши в меню «Действие» кнопку «Классифицировать текст?». | | | | | 1. Выделенная кнопка открыла окно для ввода данных с клавиатуры. | | | | 1. Выполнено. |
| 2. Нажать левой кнопкой мыши на поле ввода текста и ввести произвольное предложение. | | | | | 2. Выбранная область ввода подсветилась, а введённая последовательность отображается корректно. | | | | 2. Выполнено. |
| 3. Нажать левой кнопкой мыши на кнопку «классифицировать». | | | | | 3. После нажатия на кнопку через небольшой промежуток времени появится сообщение с результатами классификации. | | | | 3. Выполнено. |
| 4. Нажать левой кнопкой мыши на кнопку ОК. | | | | | 4. После нажатия кнопки ОК, отображается поле ввода | | | | 4. Выполнено. |
| 5. Нажать левой кнопкой мыши на крестик. | | | | | 5. Окно ввода текста закрывается, появляется страница «newsru | Коллекция» веб-приложения.. | | | | 5. Выполнено. |
|  | | | | | Результат | | | | Выполнено |

Таблица 10 – Тестирование классификатора статей на странице «newsru | Коллекция»

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| № сценария | | Т-5 | | | | Приоритет | Высокий | | |
| Что проверяется | | Классификатор статей на странице «newsru | Коллекция» | | | | | | | |
| Предусловия | | | | | | | | | |
| Использование определённого браузера, Переход на адрес 127.0.0.1:8000 | | | | | | | | | |
| История изменений | | | | | | | | | |
| Создан | 08.12.2021 | | кем | | Родионов Вадим | | | Тестирование интерфейса веб-приложения | |
| Изменен | 08.12.2021 | | кем | | Родионов Вадим | | | Повторное тестирование интерфейса веб-приложения | |
| Закрыт | 08.12.2021 | | кем | | Родионов Вадим | | |  | |
| Тестовый сценарий | | | | | | | | | |
| Действие | | | | Ожидаемый результат | | | | | Фактический результат (выполнено, не выполнено, выполнено с ошибкой). |
| 1. Нажать левой кнопкой мыши в меню «Действие» кнопку «Классифицировать?». | | | | 1. Окно с предупреждением о том, что «Режим классификации включён» и кнопки ОК. Кнопка классифицировать стала голубой и переместилась к «Да». | | | | | 1. Выполнено. |
| 2. Нажать на произвольную статью на странице в меню «статьи». | | | | 2. После нажатия на статью через небольшой промежуток времени появится сообщение с результатами классификации. | | | | | 2. Выполнено. |
| 3. Нажать левой кнопкой мыши на кнопку ОК. | | | | 3. После нажатия кнопки ОК, отображается веб-страница «newsru | Коллекция» в режиме «классифицровать?» | | | | | 3. Выполнено. |
| 4. Повторить действие 1 | | | | 4. Окно с предупреждением о том, что «Режим классификации включён» и кнопки ОК. Кнопка классифицировать стала серой и переместилась к «Нет». | | | | | 4. Выполнено. |
|  | | | | Результат | | | | | Выполнено |

Таблица 11 – Тестирование процесса поиска статей на странице «newsru | Коллекция»

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| № сценария | | Т-6 | | | | Приоритет | Высокий | | |
| Что проверяется | | Поиск статей на странице «newsru | Коллекция» | | | | | | | |
| Предусловия | | | | | | | | | |
| Использование определённого браузера, Переход на адрес 127.0.0.1:8000 | | | | | | | | | |
| История изменений | | | | | | | | | |
| Создан | 08.12.2021 | | кем | Родионов Вадим | | | | Тестирование интерфейса веб-приложения | |
| Изменен | 08.12.2021 | | кем | Родионов Вадим | | | | Повторное тестирование интерфейса веб-приложения | |
| Закрыт | 08.12.2021 | | кем | Родионов Вадим | | | |  | |
| Тестовый сценарий | | | | | | | | | |
| Действие | | | | | Ожидаемый результат | | | | Фактический результат (выполнено, не выполнено, выполнено с ошибкой). |
| 1. Нажать левой кнопкой мыши в меню «Действие» кнопку «Найти статью». | | | | | 1. Появляется выпадающее меню с выбором из трёх полей и кнопкой «поиск». | | | | 1. Выполнено. |
| 2. Нажать левой кнопкой мыши на выпадающее меню «Категория» и выбрать «В мире» | | | | | 2. Открылось выпадающее меню с возможностью выбора «категории статьи», а после нажатия на выбранную «категорию», меню закрылось, и категория получила отображение выбранной категории. | | | | 2. Выполнено. |
| 3. Нажать левой кнопкой мыши на кнопку «Поиск!». | | | | | 3. Открывается страница «newsru | Создание статьи» веб-приложения с найденными статьями | | | | 3. Выполнено. |
| 4. Повторить действие 2. | | | | | 4. Результат действия 2. | | | | 4. Выполнено. |
| Продолжение таблицы 11 | | | | | | | | | |
| 5. Нажать левой кнопкой мыши на поле ввода «Теги» и ввести «Культура, Музыка». | | | | | 5. Выбранная область ввода подсветилась, а введённая последовательность отображается корректно. | | | | 5. Выполнено. |
| 6. Повторить действие 3. | | | | | 6. Результат действия 3. | | | | 6. Выполнено. |
| 7. Повторить действие 2. | | | | | 7. Результат действия 2. | | | | 7. Выполнено. |
| 8. Повторить действие 5. | | | | | 8. Результат действия 5. | | | | 8. Выполнено. |
| 9. Нажать левой кнопкой мыши на поле ввода «Название заголовка» и ввести «Моргенштерн». | | | | | 9. Результат действия 5. | | | | 9. Выполнено. |
| 10. Повторить действие 3. | | | | | 10. Результат действия 3. | | | | 10. Выполнено. |
|  | | | | | Результат | | | | Выполнено |

Таблица 12 – Тестирование редактирования статьи

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| № сценария | | Т-7 | | | | Приоритет | Высокий | | |
| Что проверяется | | Редактирование статьи | | | | | | | |
| Предусловия | | | | | | | | | |
| Использование определённого браузера, Переход на адрес 127.0.0.1:8000/id=77 | | | | | | | | | |
| История изменений | | | | | | | | | |
| Создан | 08.12.2021 | | кем | Родионов Вадим | | | | Тестирование интерфейса веб-приложения | |
| Изменен | 08.12.2021 | | кем | Родионов Вадим | | | | Повторное тестирование интерфейса веб-приложения | |
| Закрыт | 08.12.2021 | | кем | Родионов Вадим | | | |  | |
| Тестовый сценарий | | | | | | | | | |
| Действие | | | | | Ожидаемый результат | | | | Фактический результат (выполнено, не выполнено, выполнено с ошибкой). |
| Продолжение таблицы 13 | | | | | | | | | |
| 1. Нажать левой кнопкой мыши на текст в произвольном месте и ввести произвольную последовательность. | | | | | 1. Никаких изменений в содержании статьи не происходит. | | | | 1. Выполнено. |
| 2. Нажать левой кнопкой мыши в меню «Редактирование» на пункт «Режим редактора». | | | | | 2. Окно с предупреждением о том, что «Режим редактора включён» и кнопки ОК. | | | | 2. Выполнено. |
| 3. Нажать левой кнопкой мыши на кнопку ОК. | | | | | 3. Окно с предупреждением закрывается. | | | | 3. Выполнено. |
| 4. Повторить действие 1. | | | | | 4. Произвольная комбинация в произвольном месте добавлена. | | | | 4. Выполнено. |
| 5. Нажать левой кнопкой мыши в меню «Редактирование» на пункт «Режим просмотра». | | | | | 5. Окно с предупреждением о том, что «Режим редактора выключен. Не забудьте сохранить статью!» и кнопки ОК. | | | | 5. Выполнено. |
| 6. Нажать левой кнопкой мыши в меню «Редактирование» на пункт «Сохранить». | | | | | 6. Страница обновлена, произвольная комбинация в текст сохранена. | | | | 6. Выполнено. |
| 7. Повторить действие 2. | | | | | 7. Результат действия 2. | | | | 7. Выполнено. |
| 8. Повторить действие 3. | | | | | 8. Результат действия 3. | | | | 8. Выполнено. |
| 9. Повторить действие 1. | | | | | 9. Результат действия 4. | | | | 9. Выполнено. |
| 10. Нажать комбинацию клавиш «alt + стрелка влево». | | | | | 10. Результаты не сохранены. | | | | 10. Выполнено. |
|  | | | | | Результат | | | | Выполнено |

Таблица 12 – Тестирование удаления статьи

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| № сценария | | Т-8 | | | | Приоритет | Высокий | | |
| Что проверяется | | Удаление статьи | | | | | | | |
| Предусловия | | | | | | | | | |
| Использование определённого браузера, Переход на адрес 127.0.0.1:8000/id=78 | | | | | | | | | |
| История изменений | | | | | | | | | |
| Продолжение таблицы 12 | | | | | | | | | |
| Создан | 08.12.2021 | | кем | Родионов Вадим | | | | Тестирование интерфейса веб-приложения | |
| Изменен | 08.12.2021 | | кем | Родионов Вадим | | | | Повторное тестирование интерфейса веб-приложения | |
| Закрыт | 08.12.2021 | | кем | Родионов Вадим | | | |  | |
| Тестовый сценарий | | | | | | | | | |
| Действие | | | | | Ожидаемый результат | | | | Фактический результат (выполнено, не выполнено, выполнено с ошибкой). |
| 1. Нажать левой кнопкой мыши в меню «Редактирование» на пункт «Удалить». | | | | | 1. Окно с предупреждением под пунктом «удалить» с информацией о том, что «Данная статья будет удалена. Вы уверены?» и кнопками «Да» и «Нет». | | | | 1. Выполнено. |
| 2. Нажать левой кнопкой мыши на кнопку «Нет». | | | | | 2. Окно с предупреждением закрывается, статья не удалена. | | | | 2. Выполнено. |
| 3. Повторить действие 1. | | | | | 3. Результат действия 1. | | | | 3. Выполнено. |
| 4. Нажать левой кнопкой мыши на кнопку «Да». | | | | | 4. Статья удалена, переход на страницу «newsru | Коллекция» веб-приложения. | | | | 4. Выполнено. |
|  | | | | | Результат | | | | Выполнено |

## 6.2 Результаты тестирования

В результате автоматического тестирования не удалось выявить неполадки на этапе загрузки сервера и отправки POST\GET запросов. Случай верного ответа от сервера при верных и неверных запросах получены во всех случаях, тем самым, загружались требуемые страницы\ответы.

В результате ручного тестирования были доработаны недостатки, выявленные при отображении данных на страницах. Таким образом, удалось нивелировать последствия от работы растеризации страниц разными браузерами и в итоге достигнуть качественного результата.

# 7 ИНТЕРФЕЙС И ВОЗМОЖНОСТИ СИСТЕМЫ

Веб-интерфейс приложения условно можно разделить на:

* Страница создания статей (newsru | Создание статьи) – страница, отвечающая за создание статей пользователем и с внесения определённых признаков (таких как заголовок, дата создания, текст статьи, теги, ссылка на источник). Пример интерфейса веб-страницы представлен на рисунке 2.

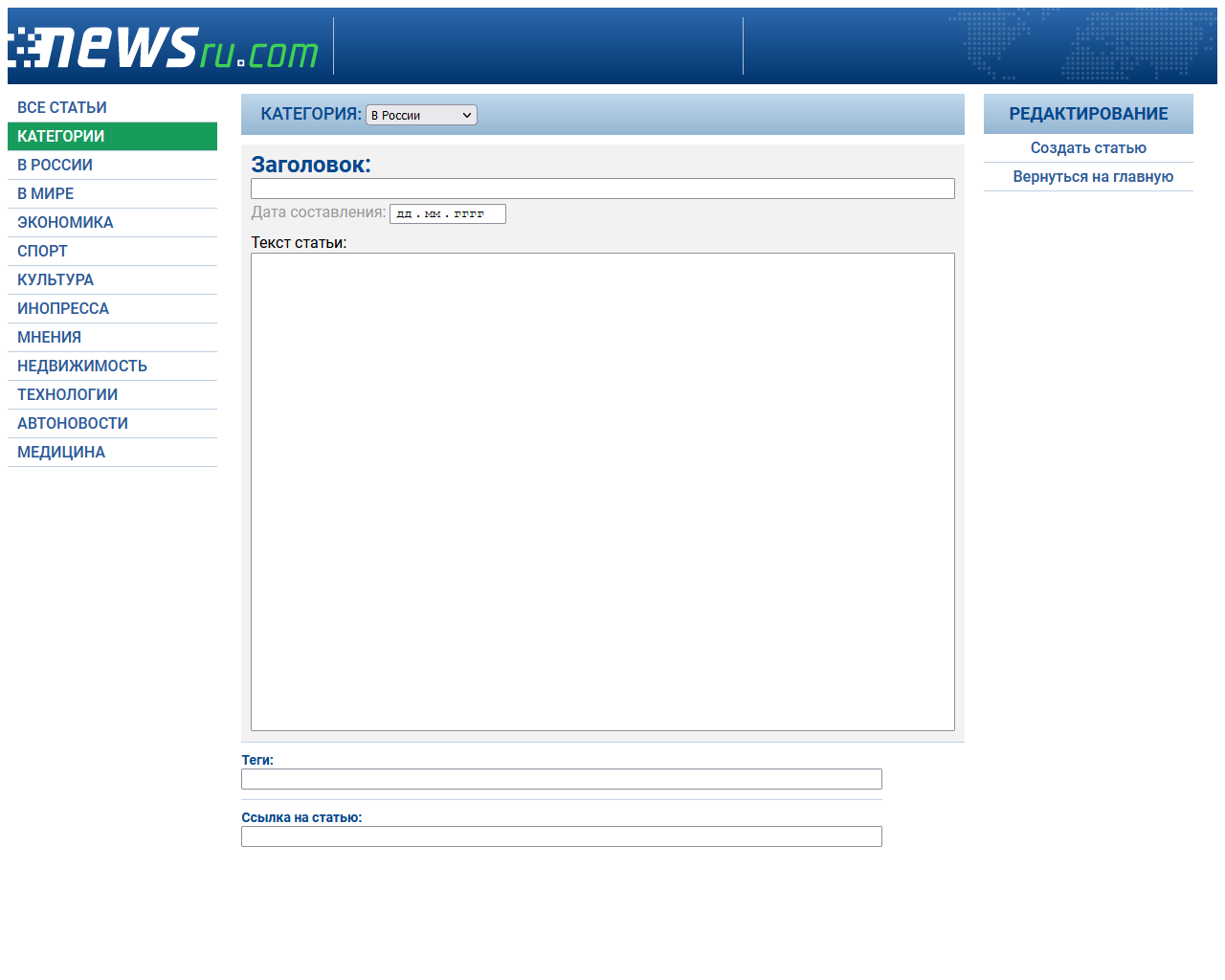


Рисунок 2 – Веб-интерфейс страницы «создания статей»

* Главная страница (newsru | Коллекция) – главная страница, на которой можно воспользоваться поиском нужных статей по заданным критериям, узнать результаты классификации текста, перейти на другую страницу со статьями или в другую категорию. Пример интерфейса веб-страницы представлен на рисунке 3.

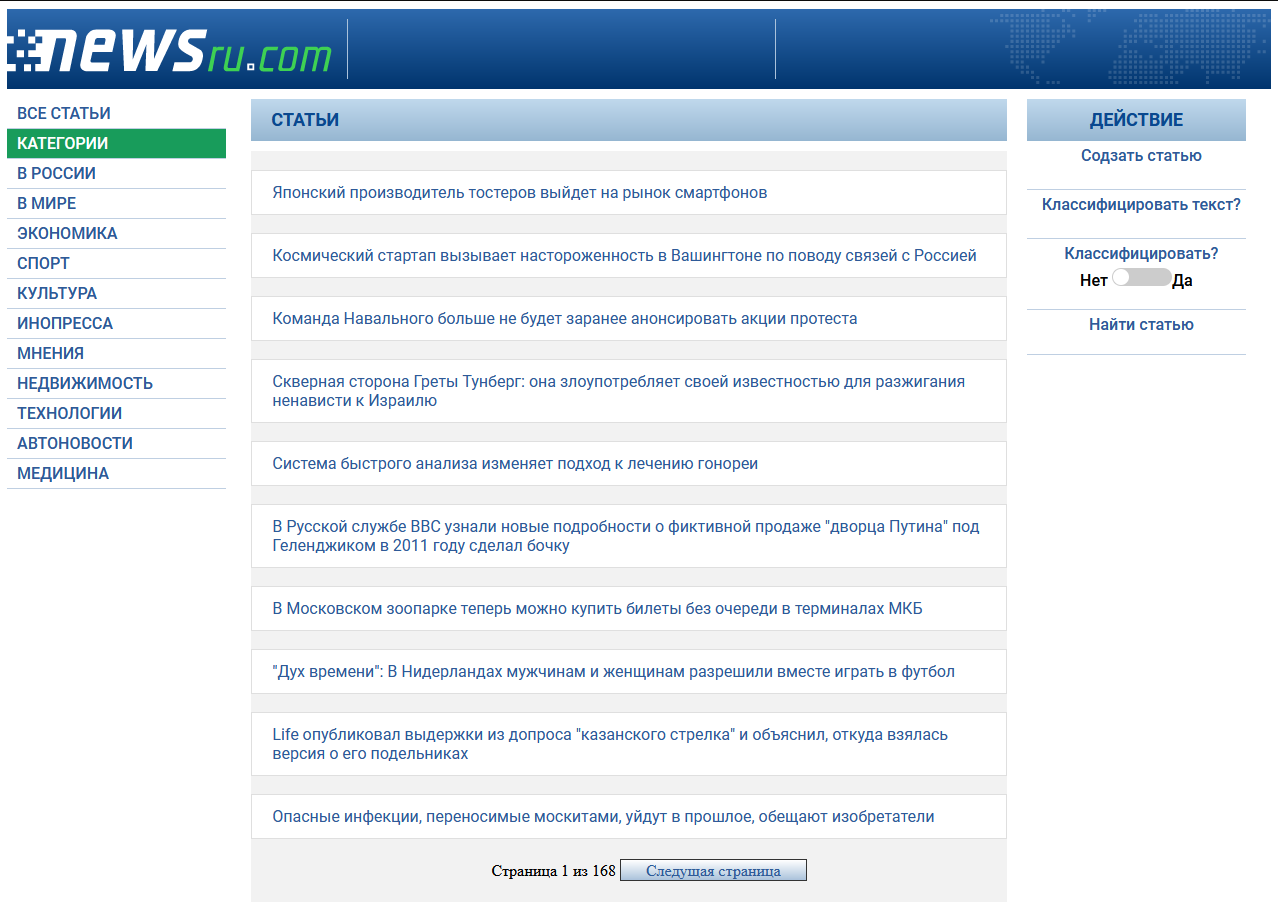


Рисунок 3 – Веб-интерфейс страницы «главная страница»

* Страница новости (newsru | 'название новости’) – страница, содержащая в себе саму новость. Обладает возможностью редактирования статьи и её удаления. С данной страницы можно так же перейти в другую категорию. Пример интерфейса веб-страницы представлен на рисунке 4.



Рисунок 4 – Веб-интерфейс страницы «новость»

* Страницы ошибок (ошибочка 'код ошибки') – служебные страницы. Отображаются в случае ошибок и\или некорректного запроса пользователя. Пример интерфейса веб-страницы представлен на рисунке 5.

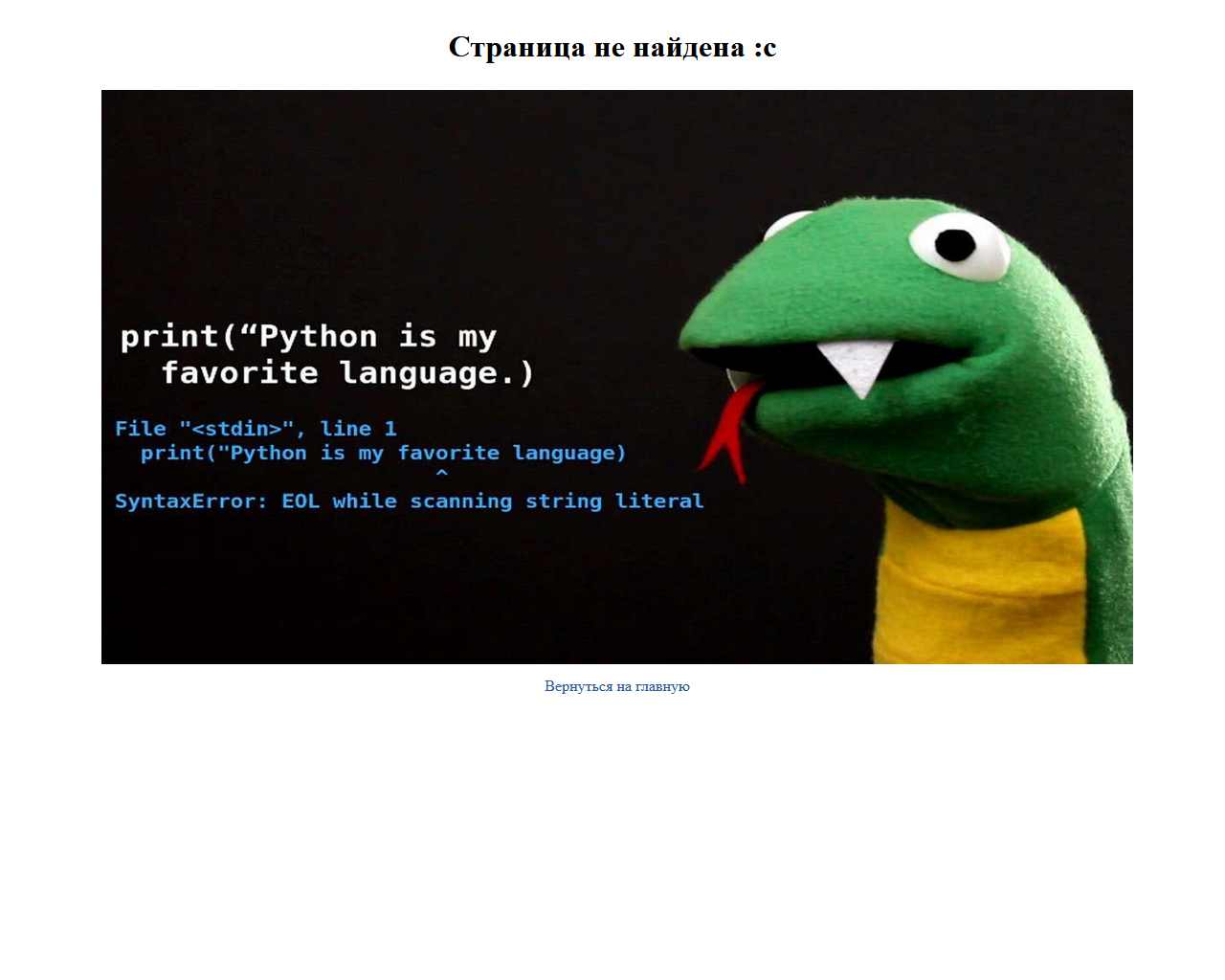


Рисунок 5 – Веб-интерфейс страницы «ошибка 404»

Сам веб-интерфейс наследован с сайта [2] newsru.com, с которого первоначально забирались статьи. Данный веб-ресурс отличается своей простотой в использовании и интуитивно понятным интерфейсом, а также поддерживает разно размерные устройства – поддержка в форматах txt, palm, pda, classic и без префикса третьего уровня. Интерфейс сайта не менялся с 2009 года. Но главным остаётся тот факт, что интерфейс не претерпевает существенных изменений. За основу был взят интерфейс без префикса третьего уровня.

# 8 АНАЛИЗ РЕЗУЛЬТАТОВ

В результате проделанной работы было разработано клиент-серверное веб-приложение, выполняющее хранение коллекций статей и классификацию как введённых статей, так и введённого текста. Программный продукт соответствует требованиям, описанными ранее в данной пояснительной записке.

С точки зрения сопровождения и продолжения поддержки продукта, данное приложение в настоящий момент не требует доработок, а введение нового функционала не скажется негативно на самом приложении, так как используемые методы отлажены, а входящие в них данные не смогут привести к краху приложения. Тем самым, появляется возможность свободного добавления нового функционала, который не затронет критически важные методы программы.

Возможность добавления новых моделей связанна с некоторым рядом трудностей, и потребует обновления файла моделей, что в свою очередь будет накладно в плане процессорного и физического времени, однако, данная операция может выполняться параллельно с другим функционалом приложения, тем самым нивелируя свой недостаток.

Клиентская часть требует некоторых доработок для работы с другими браузерами, и ввиду развития стандарта языки вёрстки HTML – и соответствующего обновления с выходом новых версий.

Отказоустойчивость приложения проверена множественными тест кейсами и нагрузочным тестированием (1000 запросов в секунду), при котором посылались запросы на отрисовку разных представлений страниц.

Точность классификатора находится на достаточном уровне, что позволяет его использовать для классификации разного рода статей и выводить процентное соотношение статьи к той или иной категории, тем самым, позволяя предположить, о чём в статье именно идёт речь.

# СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ

1. Проект DjangoCooledByCollection («Django Unchained для коллекций статей с сайта newsru.com. Просмотр, классификация, редактирование».) [Электронный ресурс]. – Режим доступа: github.com/Vadim000009/DjangoCooledByCollection Просмотрено: 20.12.2021

2. Новостной сайт newsru.com [Электронный ресурс]. – Режим доступа: www.newsru.com Просмотрено: 20.09.2021

3. Дронов, В. А. Django 3.0. Практика создания веб-сайтов на Python. / Дронов, В. А. - БХВ-Петербург, 2021 – 704 c.

4. Лутц, М. Изучаем Python / Лутц, М. – Символ-Плюс.:O``Reilly, 2010 – 1280 с.