

Национальный исследовательский университет
«Высшая школа экономики»
Институт статистических исследований и экономики знаний

Экологический чат-бот

Техническое задание на создание чат-бота в Telegram

Содержание

1	Введение	3
2	Требования к программе	3
2.1	Требования к функциональности	3
2.1.1	База знаний	3
2.1.2	Сбор новостей	4
2.1.3	Новостная рассылка	5
2.1.4	Сервисные уведомления	5
2.1.5	Управление подпиской на рассылки	6
2.1.6	Управление правами пользователей	6
2.2	Требования к интерфейсу	6
2.3	Требования к безопасности	6
2.4	Требования к надёжности	6
2.5	Условия эксплуатации	6
2.6	Требования к составу и параметрам технических средств	7
2.7	Требования к информационной и программной совместимости	7
2.8	Требования к маркировке и упаковке	7
2.9	Требования к транспортированию и хранению	7
2.10	Специальные требования	7
3	Требования к программной документации	7
4	Стадии и этапы разработки	7
5	Порядок контроля и приёмки	7

1. Введение

Программа «Экологический чат-бот» предназначена для информирования клиентов (сотрудников и студентов НИУ ВШЭ, а также других лиц) об экологической повестке внутри НИУ ВШЭ. Данный чат-бот предполагает пользовательское взаимодействие через мессенджер Telegram¹. Основанием для разработки данного чат-бота является (**TODO**: документ).

2. Требования к программе

2.1. Требования к функциональности

Чат-бот должен поддерживать следующие функции:

1. Доступ пользователей к внутренней базе знаний
2. Редактирование базы знаний пользователями, наделёнными на это правом
3. Рассылка новостей, связанных с экологической повесткой, подписавшимся на рассылку пользователям
4. Автоматический сбор новостей с интернет-ресурсов для их дальнейшей рассылки

2.1.1. База знаний

Внутренняя база знаний представляет из себя иерархическую систему из разделов и заметок. Вне зависимости от наполнения базы знаний информацией, в ней существует единственный корневой раздел, не имеющий имени. Каждый раздел, не являющийся корневым, а также каждая заметка принадлежат тому или иному разделу (в том числе корневому). Циклическая принадлежность разделов (например, раздел *A* принадлежит разделу *B*, который, в свою очередь, принадлежит разделу *A*) запрещена; также никакой раздел не может принадлежать самому себе. Корневой раздел не принадлежит никакому разделу. Таким образом, разделы, заметки и отношение принадлежности образуют иерархическую структуру — дерево разделов. Эта структура напоминает дерево файлов и папок в файловой системе. Схематичное изображение примера дерева разделов показано на рис. 1.

Каждый не-корневой раздел имеет имя. Каждая заметка имеет имя и содержимое. Именем является строка символов Юникода, поддерживаемых мессенджером, длиной не более 300 символов. У разделов и заметок, непосредственно принадлежащих одному и тому же разделу, имена должны различаться. Таким образом, запрещено наличие в

¹<https://telegram.org>

```

Корневой раздел
+---- Раздел «Экология в Вышке»
|      +---- Раздел «Переработка отходов»
|      |      +---- Заметка «Раздельный сбор мусора»
|      |      +---- Заметка «Места сбора батареек»
|      +---- Заметка «Планы развития»
+---- Раздел «О боте»
|      +---- Заметка «Справка»
+---- Заметка «Основная информация»

```

Рис. 1: Пример дерева разделов.

одном разделе (а) двух разделов, (б) двух заметок и (в) раздела и заметки с одинаковыми именами.

Содержимое каждой заметки является текстом длиной не более 3500 символов, а также может содержать не более 10 вложений. Каждое вложение — это либо изображение разрешением не больше 1920×1080 пикселей, либо видео размером не более 50 МБ², либо файл размером не более 50 МБ. Текст и вложения должны быть доступны пользователю через стандартные средства мессенджера. Допускается отправка текста и вложений отдельными соседними сообщениями.

Навигация и просмотр базы знаний доступны всем пользователям. Пользователи, имеющие на это право, могут также редактировать базу знаний. Способ навигации по базе знаний, просмотра заметок и редактирования базы знаний описан в разделе 2.2. Правила определения пользователей, имеющих право на редактирование, описаны в разделе 2.3.

2.1.2. Сбор новостей

Не реже, чем один раз в день, бот совершает запросы на Интернет-ресурсы с целью получения с них новостей, связанных с экологической повесткой. Перечень собираемой информации:

1. Заголовок новости
2. Текст новости
3. Прямая ссылка на источник новости (ресурс, с которого новость получена)
4. Автор новости (если указан)
5. Дата публикации (если указана)

Данная операция происходит автоматически, без участия человека. При невозможности сбора данных с одного или нескольких ресурсов по той или иной причине бот посылает сервисное уведомление с описанием ошибки, а также записывает информацию об ошибке в системный журнал.

²Здесь и далее подразумеваются двоичные единицы измерения: 1 МБ = 1024 КБ = 1048576 Б.

При риске возникновения препятствий к автоматическому сбору новостей (например, CAPTCHA-проверок) или если правила пользования ресурса, с которого собирается информация, запрещает автоматизированный доступ к новостям, допускается брать новости из RSS-ленты соответствующего ресурса при её наличии. В таком случае, допускается не получать полный текст новости, сохраняя лишь заголовок, ссылку на источник и иную информацию при её наличии.

Полученные новости сохраняются во внутреннюю базу данных бота и потом используются для новостных рассылок, описанных в разделе [2.1.3](#).

TODO: Перечень ресурсов, с которых необходимо собирать новости.

2.1.3. Новостная рассылка

Один раз в сутки (**TODO:** возможно, стоит дать пользователю возможность настройки?) бот должен рассылать всем пользователям, подписанным на новостную рассылку, новости с внешних ресурсов, которые были собраны за этот день. Пользователю должна выводиться вся информация, сохранённая в базе данных, перечисленная в разделе [2.1.2](#). Если текст новости длиннее 3500 символов или не помещается в стандартное сообщение мессенджера, допускается приводить только его начало, указывая, что по ссылке на источник откроется полный текст новости.

Каждая новость присылается одним сообщением, не содержит вложений и должна быть маркирована хэштегом #новость.

Пользователям, не подписанным на новостную рассылку на момент её проведения, не должны рассылаться новостные сообщения.

Бот не должен присылать сообщения с новостями пользователям, которые в данный момент находятся в интерактивном состоянии (см. раздел ??), однако он может уведомить их о том, что новостная рассылка для них готова. После выхода пользователя из интерактивного состояния бот должен повторить попытку рассылки не ранее, чем через 30 секунд, но не позднее, чем через 30 минут. Если, руководствуясь данными правилами, бот не может отправить пользователю рассылку более 24 часов, то допускается не отправлять данную рассылку этому пользователю в этот день.

2.1.4. Сервисные уведомления

При возникновении определённых ситуаций, описанных в техническом задании, бот рассылает сервисные уведомления тем пользователям, которые на них подписаны и имеют право их получать. Сервисное уведомление представляет из себя одно сообщение с текстом не длиннее 3500 символов и не более, чем одним вложением.

Каждое сервисное уведомление должно быть промаркировано хэштегом #сервисное. Также применимы правила отправки рассылок пользователям в интерактивном состоянии, описанные в разделе [2.1.3](#).

2.1.5. Управление подпиской на рассылки

Любой пользователь имеет возможность подписаться, отписаться и проверить статус подписки на любые рассылки, доступные этому пользователю. Перечень рассылок, поддерживаемых ботом:

1. Новостная рассылка, описанная в разделе [2.1.3](#). Доступна всем пользователям, по умолчанию подписка неактивна.
2. Рассылка сервисных уведомлений, описанная в разделе [2.1.4](#). Доступна только пользователям, наделённым правом получения сервисных уведомлений, подписка автоматически становится активна, когда пользователя в первый раз наделяют необходимыми правами.

2.1.6. Управление правами пользователей

Пользователи, наделённые правами администратора, могут выдавать другим пользователям права и отзывать их. Также они могут редактировать свои права. Поддерживаемые права пользователей описаны в разделе [2.3](#). После любого редактирования прав пользователей должен остаться хотя бы один пользователь с правами администратора. Попытка изменения прав, не соблюдающая это ограничение, заканчивается ошибкой и отсутствием изменений в правах.

2.2. Требования к интерфейсу

Интерфейс взаимодействия с ботом (отправка текста, изображений, видео и файлов) должен обеспечиваться стандартными средствами мессенджера.

2.3. Требования к безопасности

TODO: Описание прав пользователей и управления ими.

2.4. Требования к надёжности

TODO: Требования к обеспечению надёжного функционирования, контроль входных и выходных данных, время восстановления после отказа, etc.

2.5. Условия эксплуатации

TODO: Физические условия, вид обслуживания, количество и квалификация персонала для обслуживания

2.6. Требования к составу и параметрам технических средств

TODO: Требования к серверу (аппаратные и программные) и к клиентским приложениям для мессенджера

2.7. Требования к информационной и программной совместимости

TODO: Формат данных, методы решения, языки программирования, библиотеки, etc.

2.8. Требования к маркировке и упаковке

TODO: Поставка программы заказчику, взаимодействие пользователей с программой

2.9. Требования к транспортированию и хранению

TODO: Условия хранения и передачи данных

2.10. Специальные требования

TODO: Другие требования при наличии

3. Требования к программной документации

TODO: Предварительный состав документации, специальные требования к ней

4. Стадии и этапы разработки

TODO: Этапы разработки и их содержание

5. Порядок контроля и приёмки

TODO: Виды испытаний, другие требования к приёмке