

Техническое задание

импорт контента из соцмедиа по API.

Ramin Aliev

АННОТАЦИЯ

Документ описывает техническое задание по парсингу контента из соцмедиа и распределение его в базе данных.

1. ТЕРМИНАЛОГИЯ.

- 1.1. Сообщество – открытая группа в соцмедиа, которое содержит контент в виде публикаций и комментариев к публикациям. Пример сообществ: (а) [Высокие технологии](#) Facebook; (б) [Шедевры рекламы](#) Facebook; (в) [Вульгарные](#) vk; (г) [Трейлеры HD](#) vk.
- 1.2. Контент – набор текста/картинок/видео/ссылок/хештегов, используемых в публикациях и комментариях.
- 1.3. Лента – настроенные критерии импорта контента из определенного сообщества с определенными настройками.
- 1.4. ID сообщества – определенный идентификатор сообщества в соцмедиа.

2. ВВОДНОЕ.

- 2.1. **Задача:** Импорт контента из соцмедиа vk/facebook и распределение его в базе данных MySQL по категориям и пользователям.
- 2.2. Скрипт должен быть разработан в виде плагина на Wordpress и быть совместимым с базой данных MySQL.
- 2.3. Тип импортируемого контента: (а) текст; (б) фото; (в) ссылки(картинка, текст, описание); (г) видео.
- 2.4. Форма импортируемого контента: (а) Публикации из сообществ; (в) Комментарии к публикациям.
- 2.5. Система должна позволять импортировать контента сразу из нескольких сообществ и при этом соответствовать всем ограничениям соцмедиа. Т.е. если социальная сеть ограничивает количество запросов на контент, значит мы должны его соблюдать.
- 2.6. Разработанный плагин должен соответствовать стандартам WordPress Codex.

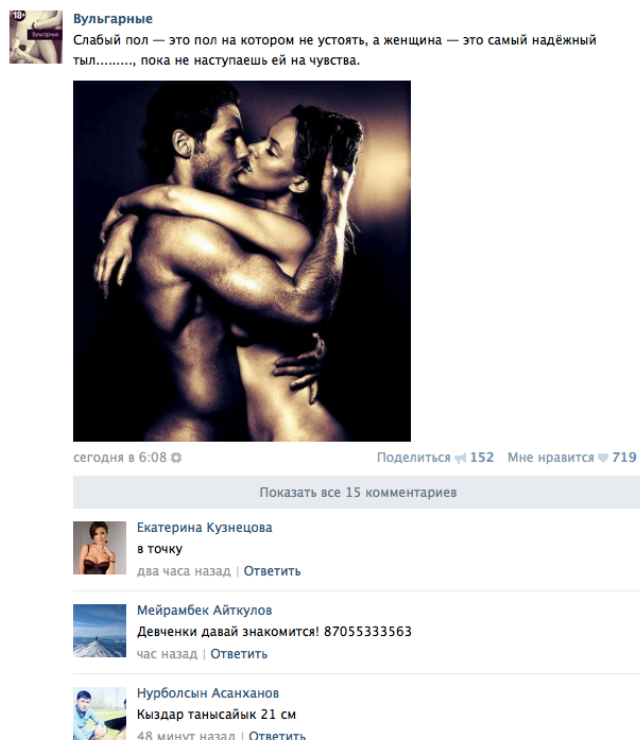
3. ФУНКЦИОНАЛЬНОСТЬ СИСТЕМЫ.

- 3.1. Импортируемый контент должен закрепляться за определенными пользователями в системе, в определенных рубриках и с определенным статусом публикации(стандарт WordPress). Скрипт должен позволять выбирать существующих пользователей или создавать новых.
- 3.2. Модератор должен иметь возможность задавать настройки плагина и импорта через админку Wordpress: API ключ, секретный ключ, id сообщества.
- 3.3. Система должна позволять настраивать критерии импорта контента: (а) минимальное кол-во лайков/комментариев/репостов публикации; (б) выбор даты начала публикаций; (в) количество импортируемого контента, (г) тип и форму импортируемого контента.
- 3.4. Система должна позволять фильтровать импортируемый контент по: (а) блок словам; (б) ссылкам; (в) хештегам; (г) уникальность контента.
- 3.5. Импортируя контент система задает заголовок и миниатюру публикации.
- 3.5.1. Логика формирования заголовка: (а) кол-во первых символов и/или (б) \n(переход строки). Стараемся выбирать заголовок законченный. К примеру если заголовок имеет 150 символов, но при этом он не закончен, мы сокращаем его до первой точки, к примеру до 132 символов.
- 3.5.2. Импортируемая картинка/видео автоматически назначается в качестве миниатюры публикации.

- 3.6. Система должна позволять настраивать таймер импорта контента по каждой ленте, а также позволять включать и выключать ленту.
- 3.7. Система должна позволять задавать дату публикации: (а) дата импорта или (б) дата публикации поста в соцмедиа.
- 3.8. Импорт видео должен осуществляться по двум опциям: (а) Импорт видео на сервер; (б) Сохранение ссылки на видео в базе данных с возможностью воспроизведения его на сайте.
- 3.9. Система должна позволять выбирать место размещения медиа контента: (а) в начале публикации; (б) в конце публикации.
- 3.10. Система должна позволять добавлять метки ко всем импортируемым публикациям. Метки задаются модератором через административную панель.

4. СЦЕНАРИИ ИСПОЛЬЗОВАНИЯ ПЛАГИНА.

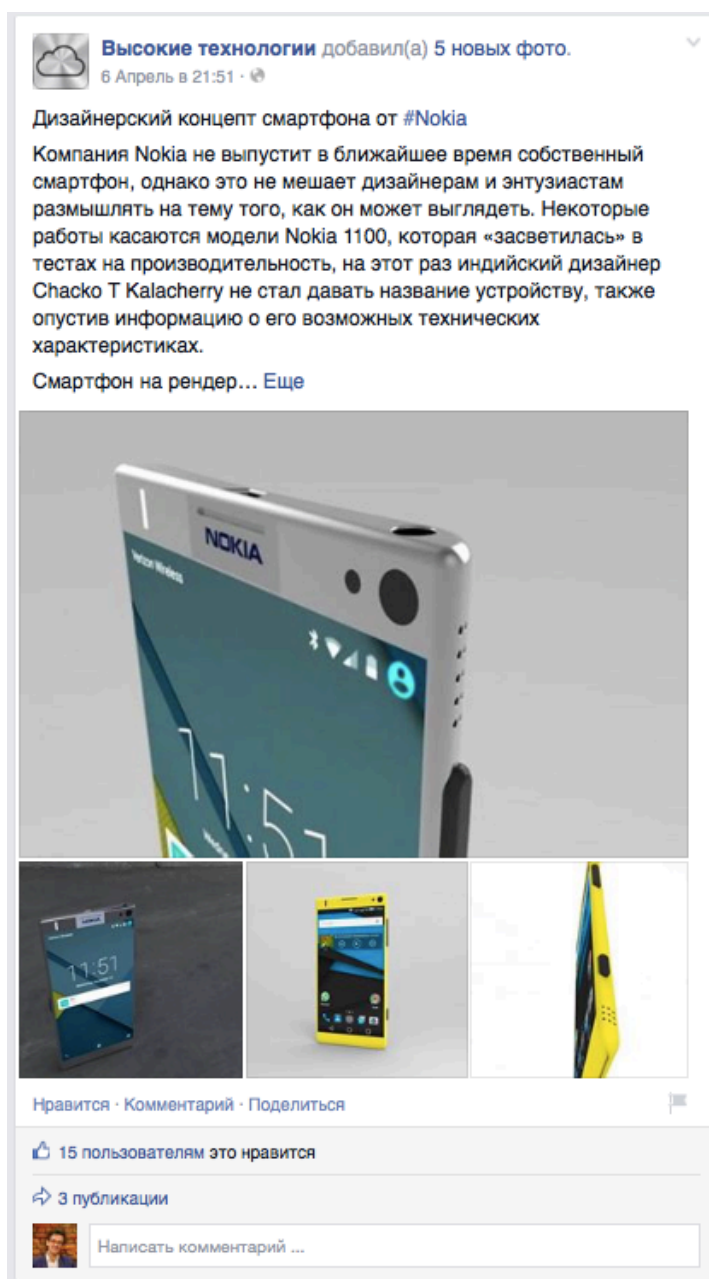
- 4.1. Импортируем контента из группы «[Вульгарные](#)». Критерии импорта: (а) начинаем импорт публикаций с 1 января 2015г; (б) посты, которые набрали более 50 лайков; (в) тип контента: текст+картинка; (г) скачиваем комментарии к посту и создаем под каждый комментарий нового пользователя; (д) блокируем импорт комментариев если есть ссылка; (е) максимальное кол-во символов в заголовке не должно превышать 150. Переход на новую строку не учитываем; (ж) импортируемый контент закрепляем за пользователем ХХХ, публикуем в рубрику Вульгарные со статусом опубликовано; (з) в качестве даты публикации берем исходную дату создания поста в соцмедиа;



В публикации есть 122 символа, 1 картинка. Пост набрал 719 лайков и был опубликован 18 апреля 2015. Таким образом мы импортируем публикацию,

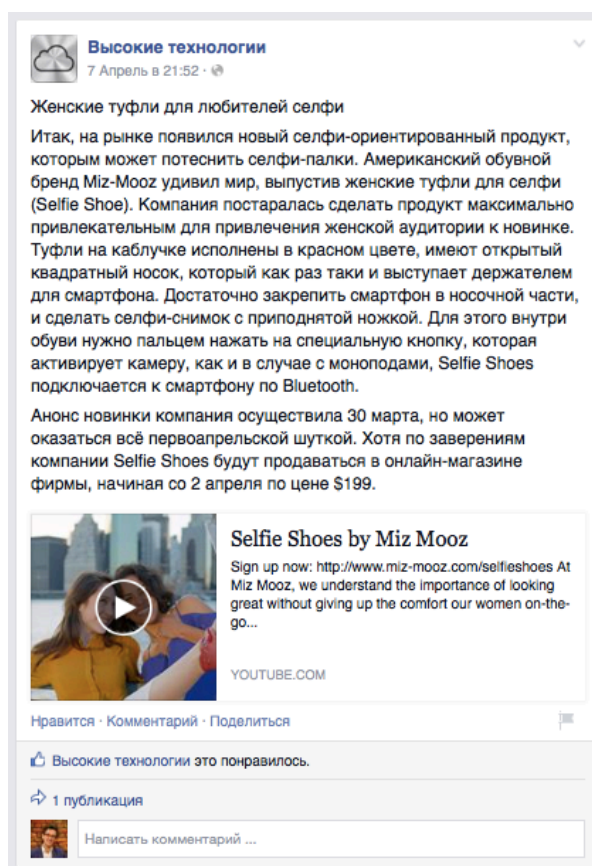
задаем в качестве миниатюры картинку, в качестве заголовка задаем все 122 символа. Контент опубликован в рубрике Вульгарные за пользователем ХХХ со статусом опубликовано. В комментариях отсутствуют внешние ссылки. Отдельно создаем 18 новых пользователей с ФИ и аватаркой и за каждым закрепляем комментарий со статусом опубликовано.

- 4.2. Импортируем контент из сообщества «[Высокие технологии](#)». Критерии импорта: (а) начинаем импорт публикаций с 1 февраля 2014г; (б) посты, которые набрали более 10 лайков; (в) тип контента: текст+картинка, текст+ссылка; (г) скачиваем комментарии к посту и создаем под каждый комментарий нового пользователя; (д) блокируем импорт комментариев если есть ссылка; (е) максимальное кол-во символов в заголовке не должно превышать 150. Учитываем переход на новую строку; (ж) импортируемый контент закрепляем за пользователем ХХХ, публикуем в рубрику Технологии со статусом опубликовано; (з) в качестве даты публикации берем дату импорта; (к) Удаляем хештеги # из поста.



В публикации 884 символа, 15 лайков, текст и 5 картинок, один хештег, опубликован 6 апреля 2015. Публикация нам подходит. В качестве заголовка берем 43 символа «Дизайнерский концепт смартфона от #Nokia», при этом удаляем хештег и остается 42 символа. В качестве миниатюры выбираем первую картинку. Оставшиеся картинки размещаются в галерею поста. Так как комментариев к посту нет, мы пропускаем этот шаг.

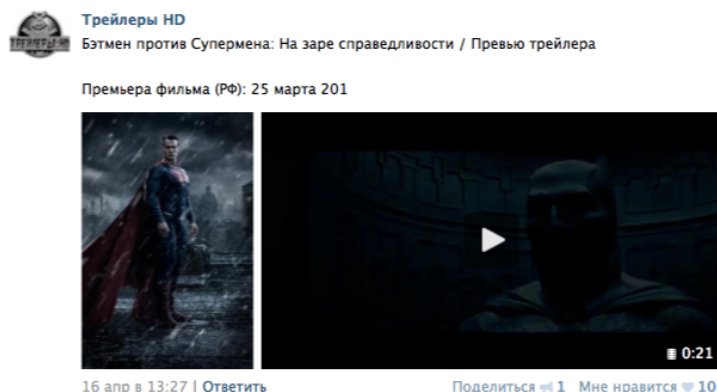
- 4.3. Импортируем контент из сообщества «**Высокие технологии**». Критерии импорта: (а) начинаем импорт публикаций с 1 марта 2014г; (б) посты, которые набрали более 1 лайка; (в) тип контента: текст+картинка, текст+ссылка; (г) скачиваем комментарии к посту и создаем под каждый комментарий нового пользователя; (д) блокируем импорт комментариев если есть ссылка; (е) максимальное кол-во символов в заголовке не должно превышать 150. Учитываем переход на новую строку; (ж) импортируемый контент закрепляем за пользователем ХХХ, публикуем в рубрику Технологии со статусом опубликовано; (з) в качестве даты публикации берем дату импорта; (к) Удаляем хештеги # из поста.



В публикации 952 символа, текст + ссылка на видео в Youtube. Публикация собрала один лайк и не содержит хештегов. Таким образом она подходит для импорта. В качестве заголовка используем «Женские туфли для любителей селфи». В качестве миниатюры используем картинку ссылки. На странице публикации мы размещаем видео Youtube.

- 4.4. Импортируем контент из сообщества «**Трейлеры HD**». Критерии импорта: (а) начинаем импорт публикаций с 1 октября 2014г; (б) посты, которые набрали более 10 лайков; (в) тип контента: текст+видео+картинка; (г) скачиваем комментарии к посту и создаем под каждый комментарий нового

пользователя; (д) блокируем импорт комментариев если есть ссылка; (е) максимальное кол-во символов в заголовке не должно превышать 150. Учитываем переход на новую строку; (ж) импортируемый контент закрепляем за пользователем ХХХ, публикуем в рубрику Трейлеры со статусом опубликовано; (з) в качестве даты публикации берем дату импорта; (к) Удаляем хештеги # из поста; (л) Сохраняем видео на сервер; (м) исключить картинку из поста.



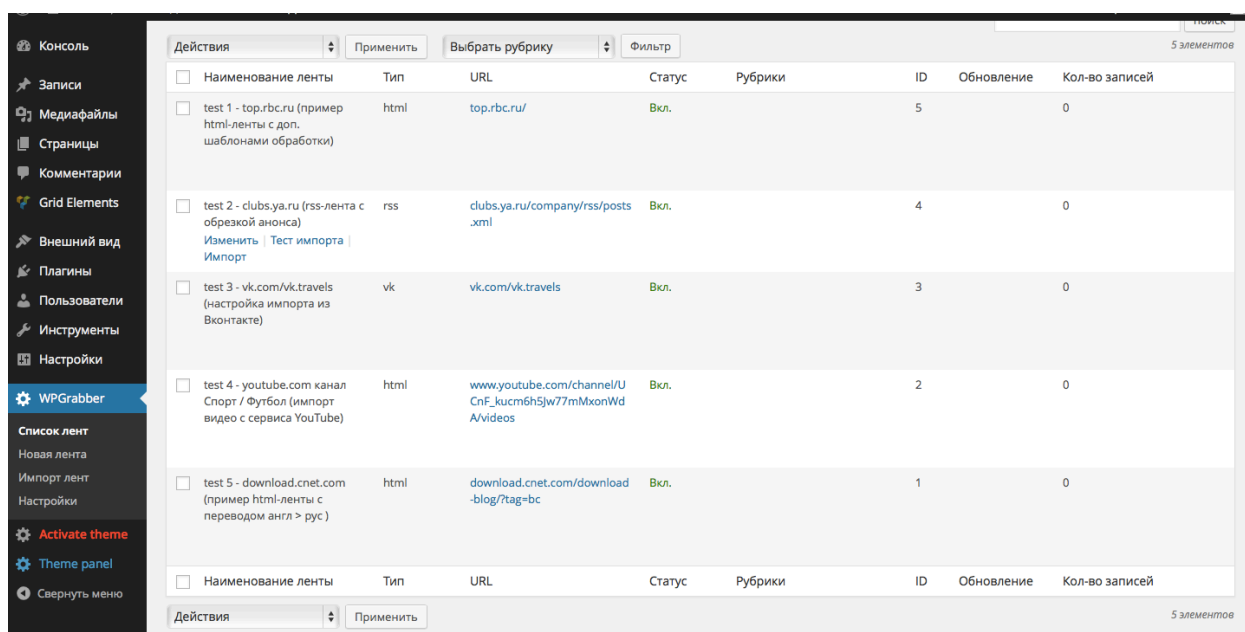
В публикации 101 символ, картинка + текст + видео. Пост собрал 10 лайков. Публикация подходит под критерии импорта. В качестве заголовка используем «Бэтмен против Супермена: На заре справедливости / Превью трейлера». В качестве миниатюры используем видео. Если у нас есть сложности в использовании видео в миниатюре, используем картинку.

5. ПОЛЕЗНЫЕ МАТЕРИАЛЫ.

5.1. Плагин, который имеет часть реализованного бекенд.

<http://www.wpallimport.com/>

5.2. Рекомендую посмотреть как реализован плагин WPGrabber. Скрин ниже.



Основные	Контент	Картинки	Перевод	Обработка	Вид	Дополнительно
Наименование ленты	test 3 - vk.com/vk.travels (настройка импорта из Вконтакте)					
Тип ленты	vk					
URL VK-стены	http://vk.com/vk.travels					
Шаблон заголовка	Путешествия vk.travels					
Кол-во слов в заголовке	5 (если не указан Шаблон заголовка или заголовок не определен!)					
Просмотр ссылок сверху вниз	<input type="radio"/> Да <input checked="" type="radio"/> Нет					
Начать с ссылки	0 0 - с первой ссылки					
Пропускать ранее не загруженные (ошибочные) ссылки	<input type="radio"/> Да <input checked="" type="radio"/> Нет					
Включить ленту	<input checked="" type="radio"/> Да <input type="radio"/> Нет					
<div>Сохранить изменения</div> <div>Применить</div> <div>Тест импорта</div> <div>Импорт</div> <div>Отмена</div>						

Основные	Контент	Картинки	Перевод	Обработка	Вид	Дополнительно
За один запуск сохранять не более (записей)	0 (0 - неограничено)					
Сохранять записи только уникальными (не повторяющимися) заголовками	<input type="radio"/> Да <input checked="" type="radio"/> Нет					
Сохранять записи в Рубрике	<div><input type="checkbox"/> Бизнес</div> <div><input type="checkbox"/> Гаджеты</div> <div><input type="checkbox"/> Новости</div> <div><input type="checkbox"/> Стиль</div>					
Тип	Запись					
Автор записей	THELIFENEWS (thelifenew)					
Статус создаваемых записей	Опубликовано					
Настройка генерации анонса						
Для выделения анонса вставлять тег Далее	<input type="radio"/> Да <input checked="" type="radio"/> Нет					
Размер анонсовой части текст (кол-во символов)	500					
Конечный символ для отделения анонса	<div></div> - пустое значение в этом поле заменяется на пробел (для обрезки по предложению вставьте точку.)					