

Комментарии и выводы к первому заданию  
Автор: Шаврина Татьяна Олеговна

### **Комментарии к коду**

Оба сайта имеют достаточно простую структуру исходного кода страницы, поэтому я осмысленно не буду использовать библиотеку lxml и разворачивать все дерево и делать поиск по нему, а буду “выкусывать” нужное мне содержимое страницы регулярными выражениями – это быстрее.

Я также заметила, что на сайте investfunds часто публикуются короткие новости о котировках, изменениях стоимости валют и пр, взятые из других источников без каких-либо изменений, при этом среди них могут быть и автоматически сгенерированные.

Еще немного к процессу скачивания: иногда некоторые тематики становятся доминирующими, в зависимости от даты (кризис, конец финансового года, публичный отчет крупной компании и пр.), но на самом деле их доля в общем объеме текстов меньше. Так как тематик к выделению предлагается 25, а статей - всего 2500, то есть мы заранее закладываем нежелательное смещение в двух выборках, и будем сравнивать не совсем реально существующие различия, так как все новости с обоих сайтов скачиваются с неодинакового временного периода – с 19 августа 2015 для prodengi и с 22 января 2016 для investfunds. При такой постановке задачи можно сохранять дату каждого текста и после скачивания делать анализ трендов.

### **Выводы**

Оба ресурса посвящены финансовым новостям, однако темы освещают немного разные и в разной пропорции. Посмотрим на таблицу ниже, где сведены все ключевые термины по 25 темам каждого ресурса: голубым цветом выделены уникальные для каждого ресурса темы, а разными оттенками теплых цветов – пары аналогичных тем.

topic number	investfunds	prodengi
0	экран, 11, изделие, сахар, новый, цена, товар, 10, услуга, 17"	страховой, страхование, рынок, развитие, тариф, рабочий, бюро, вид, группа, энергия"
1	акция, индекс, пункт, компания, британский, бумага, вырастить, торги, крупный, цена"	пенсионный, фонд, взнос, социальный, система, выплата, пенсия, обязательный, налог, уровень"
2	рубль, доллар, валюта, курс, нефть, рынок, руб, цена, российский, неделя"	страна, казахстан, бизнес, экономика, экономический, рынок, валюта, кризис, россия, становиться"
3	компания, доллар, прибыль, миллиард, составлять, акция, выручка, квартал, млрд, чистый"	монета, банкнота, казахстан, тенге, бензин, деньги, купюра, национальный, банкнот, станция"
4	компания, рост, 2016, вырастить, млрд, сектор, страна, составлять, данные, квартал"	цена, товар, продукт, услуга, изделие, подорожать, масло, продукция, инфляция, доллар"
5	пункт, акция, индекс, компания, сша, рынок, dow, 500, jones, бумага"	доллар, тенге, курс, тенге1, валюта, обменный, евро, пункт, банк, рубль"
6	компания, фас, mcdonald, россия, санкция, вбд, ресторан, москва, проект, контракт"	тенге, составлять, кредитный, рынок, кредит, 2015, объем, млрд, доля, портфель"
7	фонд, инвестор, акция, рынок, млрд, etf, актив, портфель, управление, инвестиция"	пенсионный, снпф, актив, фонд, накопление, облигация, средство, доходность, инвестиционный, вкладчик"
8	пенсионный, фонд, инфляция, актив, деньги, снпф, экономика, цена, казахстан, инвестиционный"	банк, кредит, банковский, казахстан, актив, сектор, тенге, млрд, кредитование, риск"
9	великобритания, ес, brexit, мэй, британский, выход, европейский, финансовый, заявлять, министр"	клиент, банк, система, новый, компания, банковский, онлайн, услуга, финансовый, сервис"
10	forex, club, сайт, компания, https, гк, www, торговый, турция, терминал"	ставка, банк, рынок, инфляция, рост, базовый, национальный, политика, страна, уровень"
11	биткоин, криптовалюта, amazon, биржа, становиться, новый, покупка, компания, стоимость, oculus"	банк, заем, заемщик, депозит, национальный, сумма, кредит, компенсация, информация, вкладчик"
12	автомобиль, компания, навappo, tesla, сша, китай, инвестиция, американский, ford, motor"	предприятие, инвестиция, отрасль, производство, миллиард, рост, приток, квартал, промышленность, экономика"
13	индекс, рынок, компания, акция, японский, фондовый, регион, китайский, цена, торги"	курс, нацбанк, тенге, национальный, рынок, банк, экономика, вопрос, политика, правительство"
14	тенге, доллар, торги, kase, составлять, тыс, сша, расчет, сделка, сессия"	страна, компания, казахстан, место, самый, автомобиль, рынок, мир, рейтинг, россия"
15	банк, 2016, млрд, кредит, уровень, капитал, составлять, актив, тенге, конец"	бумага, ценный, матч, инвестор, портфель, онлайн, ставка, биржа, казахстанский, китай"
16	рынок, страна, рост, трамп, китай, сша, экономика, торговый, президент, становиться"	процент, казахстан, интернет, платежный, услуга, 2015, составлять, платеж, товар, карточка"
17	акция, 01, 2017nasdaq, фб, йоркский, индекс, 2017ню, предыдущий, 03, 02"	тенге, валюта, рынок, валютный, депозит, банк, ставка, аналитик, нацбанк, иностранный"
18	ставка, фрс, повышение, процентный, экономика, рост, федеральный, сша, инфляция, рынок"	доллар, нефть, тенге, курс, цена, баррель, рубль, рынок, неделя, валюта"
19	франция, германия, выборы, евро, ес, европейский, ле, европа, пена, лебедь"	банк, казахстан, национальный, евразийский, развитие, требование, страна, капитал, проект, рк"
20	доллар, иена, курс, сша, евро, торги, валюта, трамп, страна, предыдущий"	банк, казахстан, национальный, председатель, глава, финансовый, компания, международный, центр, нацбанк"
21	компания, сделка, млрд, млн, сообщать, директор, заявлять, сша, банк, бизнес"	тенге, млрд, составлять, млн, 2016, тыс, 2015, объем, январь, сумма"
22	рост, ецб, 2017, инфляция, прогноз, цена, европейский, уровень, еврозона, экономика"	акция, акционер, банк, бта, сделка, самрук, ккб, коллекторский, простой, ао"
23	сеть, розничный, супермаркет, закрывать, тыс, банкнота, продажа, подделка, sears, рождество"	солнечный, проект, виз, магазин, интернет, мощность, энергетика, работа, электростанция, область"
24	boeing, самолет, компания, сша, новый, российский, производство, россия, тысяча, рабочий"	тенге, доллар, курс, составлять, валюта, торги, валютный, сентябрь, объем, сессия"

Так, например, у обоих ресурсов прослеживаются такие популярные темы, как курсы доллара, евро и тенге, а также котировки акций казахстанских компаний и деятельность казахстанских банков.

Однако, такие процессы, как влияние выборов в Америке на мировую экономику, макроэкономические процессы Азии, азиатские курсы валют, влияние брэкзита и выборов в европарламент на мировую экономику, отдельные процессы и случаи с крупнейшими мировыми компаниями (ford, amazon, mcdonalds) и их влияние на экономику обсуждается только в investfunds.kz. Также на этом сайте выделяются в отдельные темы котировки различных акций и ценных бумаг, игра на бирже форекс, а также акции доткомов и курсы криптовалют. Что касается prodengi.kz, то на этом ресурсе гораздо больше тематик посвящено экономическим процессам внутри самого Казахстана – освещается динамика цен на бензин, пенсионное страхование, инфляция, рост базовой ставки кредитования, подорожание продуктов и тому подобное, скорее более интересное населению, а не инвесторам и финансистам. Выделяются отдельные, более узкие, чем у investfunds, темы по Казахстанскому банкингу - базовые ставки, кредитование, госрегулирование, нацбанк. Единственная страна, кроме Казахстана, которая становится значимым термином в тематиках prodengi – это Россия (и еще совсем немного - Китай). Затрагиваются также темы локальной промышленности, казахстанских интернет-технологий и всевозможные рейтинги с участием Казахстана. Всплывают и отдельные

должностные лица – председатели и главы банков, компаний, и больше упоминаний “правительства” в ключевых терминах.

В целом 25 тем для этих ресурсов, пожалуй, многовато, и если полученные темы объединить логически в более крупные группы, то получается, что соотношение общих тем на ресурсах очень похоже: 12% - курсы валют, 8% - акции, от 12 до 16% - котировки, от 8 до 16% - промышленность Казахстана, 4-8% - пенсионный фонд и половина всех новостей специфицирована собственными темами ресурсов.

Макро-тема	Доля макро-темы в общем объеме		Лингв истиче ский комме нтари й: на сайте proden gi.kz
	investfunds	prodengi	
Курсы валют	3\25	3\25	
Промышленность Казахстана	4\25	2\25	
Акции и котировки, Казахстан	2\25	2\25	
Ценные бумаги, управление инвестициями	3\25	4\25	
Пенсионный фонд, Казахстан	1\25	2\25	
Остальное (разное на ресурсах)	12\25	12\25	

есть некоторая доля статей с кликбейтовыми заголовками, обычно проправительственными, с просмотрами сильно больше среднего числа – несколько тысяч вместо 300-400 средних.

source	title	views
<a href="https://prodengi.kz/lenta/kazahstanam-otkryli-dostup-k-deshevoj-ipoteke/">https://prodengi.kz/lenta/kazahstanam-otkryli-dostup-k-deshevoj-ipoteke/</a>	Казахстанцам открыли доступ к дешевой ипотеке	2369
<a href="https://prodengi.kz/lenta/v-kazahstane-s-1-iyulya-povysili-pensii-posobiya-i-zarplaty/">https://prodengi.kz/lenta/v-kazahstane-s-1-iyulya-povysili-pensii-posobiya-i-zarplaty/</a>	В Казахстане с 1 июля повысили пенсии, пособия и зарплаты	1344

Есть основания считать prodengi.kz гораздо менее политически нейтральным ресурсом, чем investfunds.kz, в котором заголовки очень формальные.

Хотя общая тематика сайтов пересекается, при подробном анализе становится очевидным, что тексты на них пишутся под разную аудиторию. investfunds.kz – информативный ресурс, для профессионального сообщества, с очень сдержанным, сухим и нейтральным стилем изложения, много коротких текстов, описывающих факты – изменения котировок, акций и пр. Возможно, некоторые такие тексты вообще автоматически сгенерированы. prodengi.kz – ресурс для широкой аудитории, где есть возможность расшарить новость в социальных сетях, есть счетчик просмотров, кликбейтовые заголовки и обилие оборотов, свойственных политическим текстам. В prodengi.kz также много прямой речи политиков, чего совсем нет в investfunds.kz.

Остается неясным, какая доля различий вызвана разными временными периодами собранных статей – 1,5 года для investfunds и 2 года для prodengi, но, по грубой смысловой оценке, к таким тематикам можно отнести что-то, случившееся за период августа 2015 по январь 2016, а это

резкое падение курса местных валют и тема “страна, казахстан, бизнес, экономика, экономический, рынок, валюта, кризис, россия, становиться” на сайте prodengi, а также, на этом же сайте, годовые отчеты с упоминанием 2015 года – темы “тенге, составлять, кредитный, рынок, кредит, 2015, объем, млрд, доля, портфель” и “процент, казахстан, интернет, платежный, услуга, 2015, составлять, платеж, товар, карточка”

.