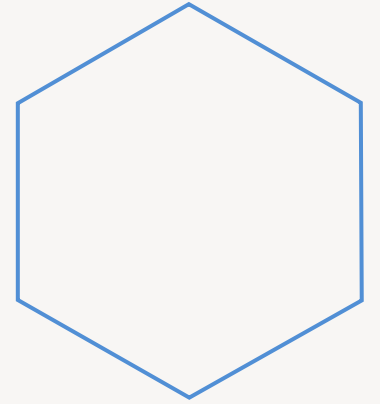


Общественная безопасность Таиланда

Харитонов
Дмитрий



Поиск источников

Характеристика	<u>The Thaiger</u>	Bangkok Post	<u>Khaosod English</u>	<u>Nation Thailand</u>
Политическая позиция	Сбалансированная, иногда с уклоном на туризм	Умеренная, традиционно консервативная	Независимая, освещает разные точки зрения	Умеренная, бизнес-ориентированная, может быть <u>про-правительственной</u>
Охват	Таиланд (с акцентом на популярные места)	Таиланд (с акцентом на Бангкок и бизнес)	Таиланд (с акцентом на социальные вопросы)	Таиланд (с акцентом на бизнес и политику)
Целевая аудитория	Экспаты, туристы, интересующиеся Таиландом	Бизнесмены, экспаты, интересующиеся политикой	Местные жители, интересующиеся новостями Таиланда	Бизнесмены, политики, экспаты



Владелец	<u>Thaiger Co., Ltd</u>	Post Publishing PCL, которая является частью Central Group.	<u>Matichon Group</u>
Проверка на человека		-	+
Постраничная навигация	+	+	-
Возможность выбрать регион <u>Таиланд</u>	+	+	+/- (Бангкок)

Парсинг данных

21 000 статей из двух источников:

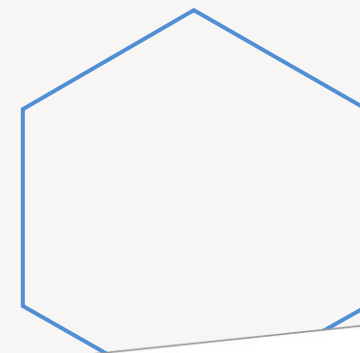
- Текст
- Ссылка
- Автор
- Темы
- Заголовок

#	Column	Non-Null Count	Dtype
0	title	8910 non-null	object
1	link	8910 non-null	object
2	text	8910 non-null	object

#	Column	Non-Null Count	Dtype
0	title	11891 non-null	object
1	link	11891 non-null	object
2	text	11891 non-null	object
3	date	11891 non-null	object
4	author	11891 non-null	object
5	main_categories	11891 non-null	object

Подготовка текста

- Удаление лишних символов
- Лемматизация



lemma_text
tragedy strike quiet village Suphan Buri year old boy kill brutal hit
rict Don Chedi district boy identify Aye lose life hit ten wheeler tru
ur yesterday April around Aye ride motorcycle reach notoriously sharp
port truck driver make attempt slow assist instead speed disappear sc
tal collision Nearby CCTV footage capture truck involve Nissan UD mod
e additional information could help track driver Adding tragedy Aye f
r a local Tik step help the family sell charcoal make end meet they s
er province even contact yet the story go viral online spark outpour
olice act swiftly report KhaoSod Police continue investigation vow b
Name: 0, dtype: object

Векторизация текста

- Эмбендинги
- Частность TF-IDF

	length	word_count
count	20801.000000	20801.000000
mean	1831.915437	256.941541
std	930.879580	127.138497
min	127.000000	19.000000
25%	1392.000000	200.000000
50%	1604.000000	225.000000
75%	1985.000000	277.000000
max	13825.000000	1879.000000

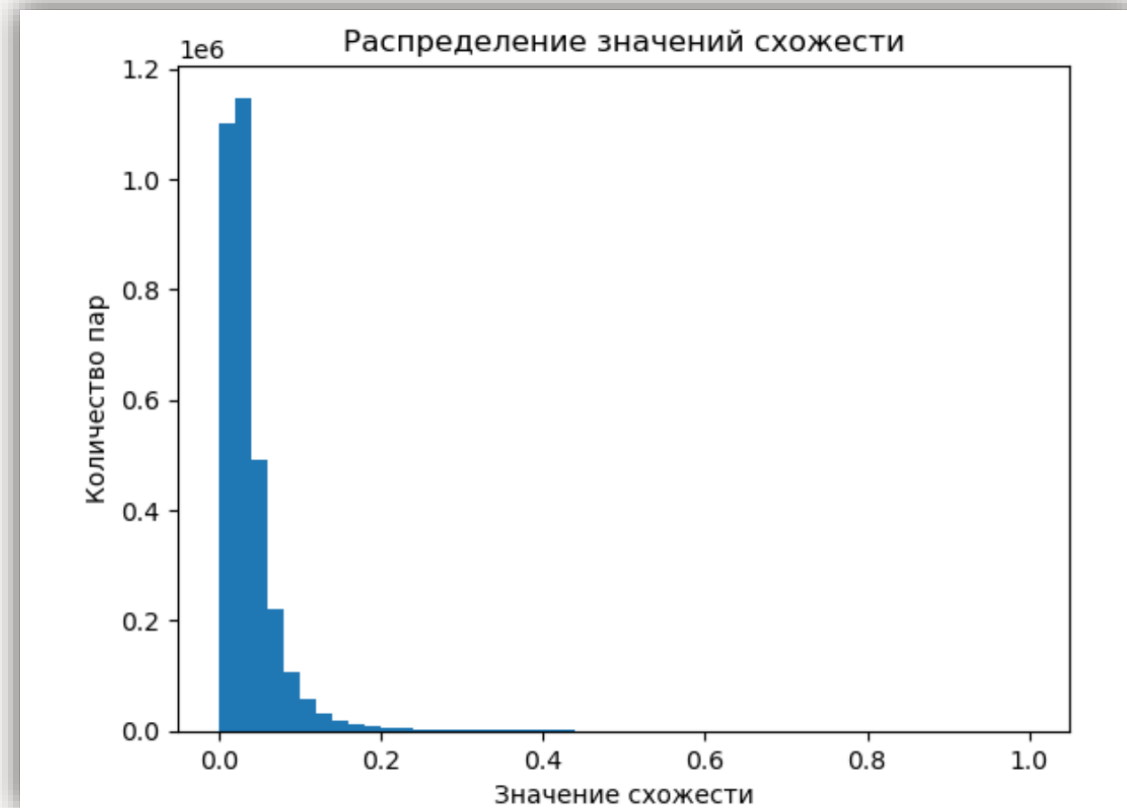
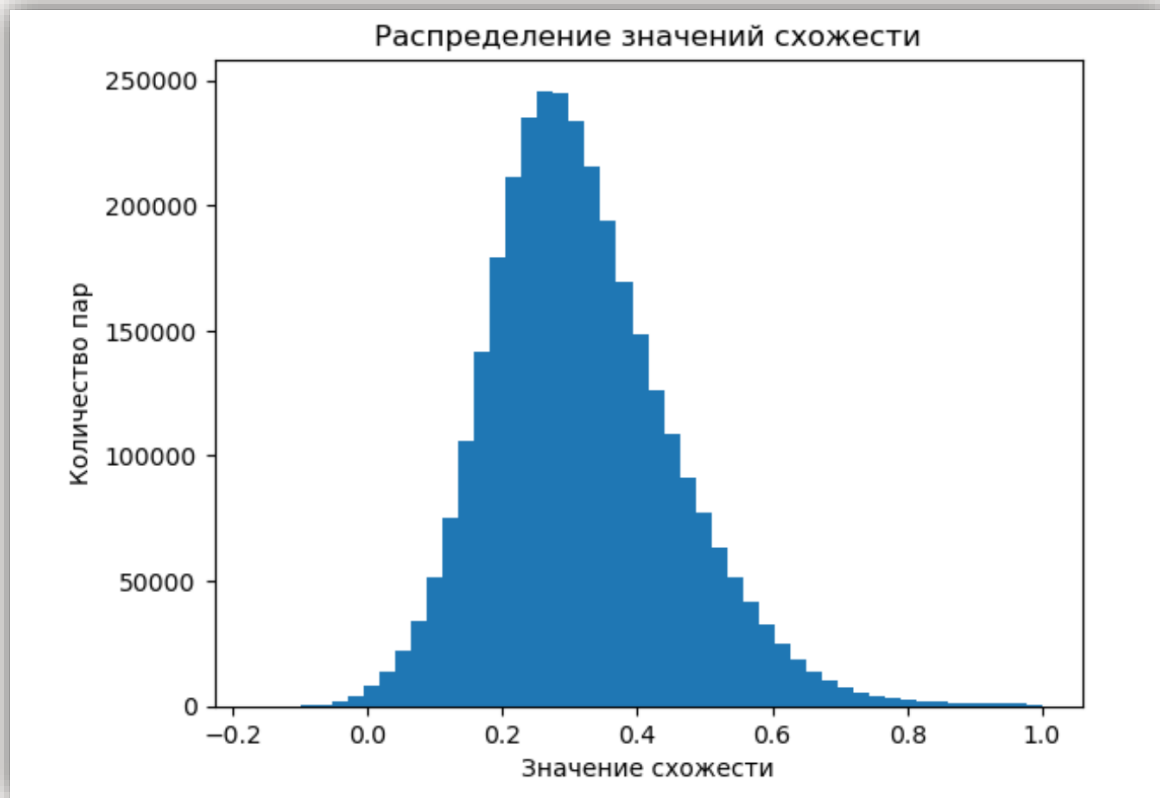
Размерность эмбедингов bangkok: (8910, 768)
Размерность эмбедингов thaiger: (11891, 768)

Размерность эмбедингов bangkok: (8910, 384)
Размерность эмбедингов thaiger: (11891, 384)

TF-IDF матрица успешно загружена.
Словарь успешно загружен.
Размерность TF-IDF матрицы: (20801, 14331)

TF-IDF матрица успешно загружена.
Словарь успешно загружен.
Размерность TF-IDF матрицы: (20801, 100000)

Поиск схожих статей



Количество пар, найденных по Эмбендингам: 7140
Количество пар, найденных по TF_IDF: 7944
Количество пар, найденных обоими методами: 5231

Разметка статей


- Классификация
- Ключевые слова
- Локация
- Категория

```
38 death_mentioned
39 connotation
40 Disasters
41 tourism
42 animal_attacks
43 fires
```

```
8 pattaya_news_label
9 phuket_news_label
10 eastern_thailand_news_label
11 northern_thailand_news_label
12 central_thailand_news_label
13 south_thailand_news_label
14 bangkok_news_label
15 crime_news_label
16 road_deaths_label
17 transport_news_label
18 environment_news_label
19 economy_news_label
20 politics_news_label
21 health_label
22 Tourism
23 Natural Disasters
24 Animal Attacks
25 Drugs
26 region
27 location
```

Коннотация

- Vader
- Bert
- Zero-Shot



vader_neg	vader_neu	vader_pos	vader_compound	positive_sentiment	negative_sentiment
20801.000000	20801.000000	20801.000000	20801.000000	20801.000000	20801.000000
0.076029	0.839188	0.084783	0.002663	0.313750	0.686250
0.058480	0.057170	0.055792	0.906549	0.432149	0.432149
0.000000	0.523000	0.000000	-1.000000	0.000327	0.000167
0.025000	0.802000	0.044000	-0.970200	0.004405	0.062841
0.067000	0.843000	0.073000	-0.012900	0.015958	0.984042
0.117000	0.880000	0.113000	0.977300	0.937159	0.995595
0.333000	1.000000	0.436000	0.999900	0.999833	0.999673

```
for label, score in zip(result['labels'], result['scores']):  
    if label == "was a serious threat to public safety":  
        threat_to_public_safety_prob = score  
    elif label == "the crime is under police control and no longer poses a threat":  
        police_control_prob = score
```


Расчет индексов

- Расчет индивидуальной негативной значимости

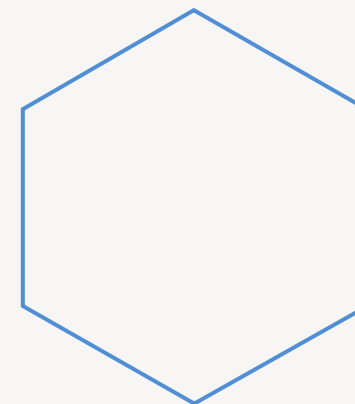
$$ICS = (\text{Влияние на общественную безопасность} \\ * (1 - \text{Успех полиции}) + \text{смертность}) \\ * \text{Коннотация})$$

- Суммирование

$$TCS = \text{сумма}(ICS)$$

- Нормализация

$$TCS_{norm} = \frac{TCS}{\text{Количество целевых} * \text{Макс. негатив. оценка}} * \frac{\text{Количество целевых}}{\text{Всего новостей}}$$



Crime Index: 0.16

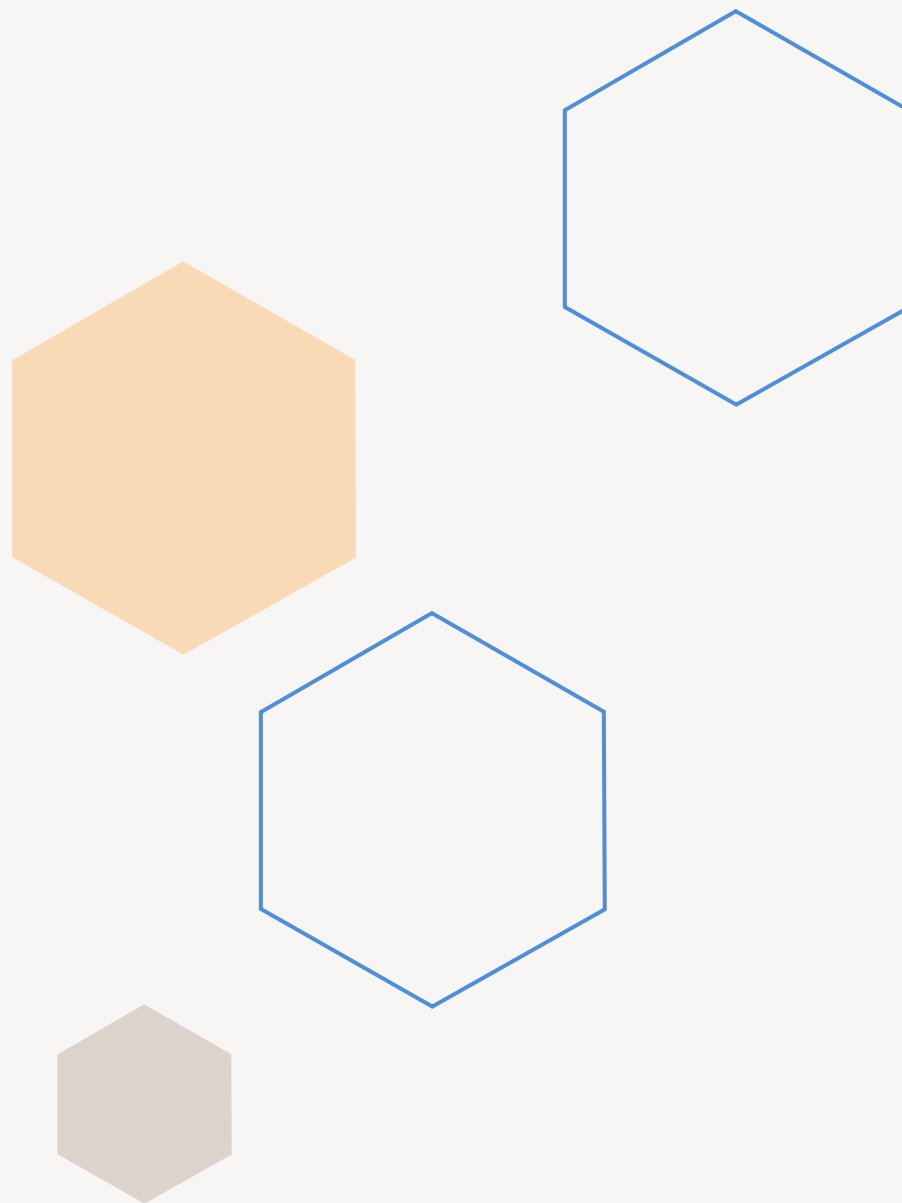
Crime Index: 0.46



Кластеризация

- K-means
- По эмбендингам
- Ключевые слова по TF-IDF

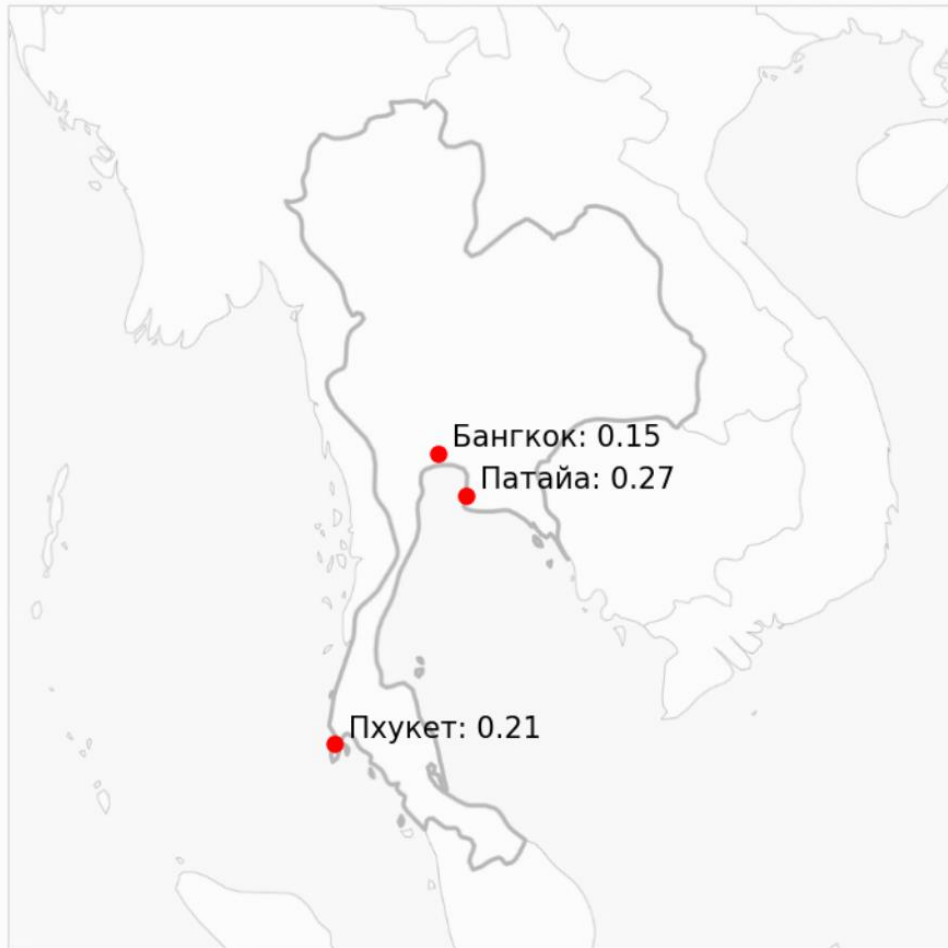
Топ-15 слов: say, court, mr, case, gen, pol, corrupti
Топ-15 слов: drug, pill, methamphetamine, arrest, ope
Топ-15 слов: man, assault, woman, victim, incident, c
Топ-15 слов: body, old, year old, man, incident, scen
Топ-15 слов: baht, arrest, say, scam, chinese, suspec



Представление результатов

Криминал

Глобальный индекс проблемы: 0.17.
Новостной фон: весь.



Основные темы:

Тема 1. Ключевые слова: say, court, mr, gen, case, pol, nacc, corrup
tion, pol gen, minister, party, former, baht, office, general
Количество статей: 846

Тема 2. Ключевые слова: baht, arrest, million, money, scam, million baht,
gambling, victim, suspect, account, say, gold, steal, online, bank
Количество статей: 1302

Тема 3. Ключевые слова: driver, road, accident, motorcycle, pattaya, inc
ident, vehicle, car, fire, man, truck, tourist, taxi, crash, scene
Количество статей: 864

Тема 4. Ключевые слова: body, old, year old, man, find, scene, suspec
t, death, shoot, incident, murder, hospital, officer, he, stab
Количество статей: 1216

Тема 5. Ключевые слова: assault, woman, man, victim, old, girl,
year old, she, incident, video, child, thai, he, take, say
Количество статей: 1020

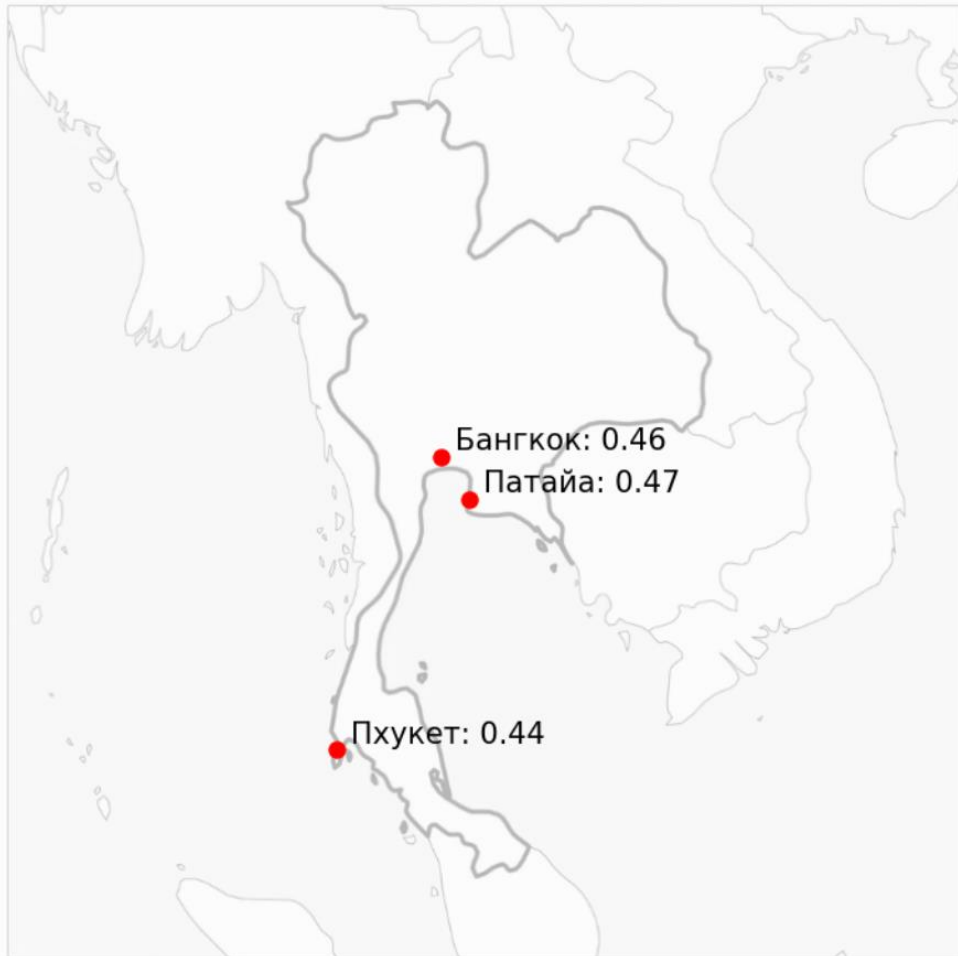
Тема 6. Ключевые слова: drug, pill, methamphetamine, arrest, operation, meth, s
eize, suspect, cigarette, say, million, cannabis, crystal, baht, narcotic
Количество статей: 782

Тема 7. Ключевые слова: myanmar, chinese, say, thailand, thai, immigration,
national, arrest, border, tourist, illegal, foreign, work, man, phuket
Количество статей: 920

Представление результатов

Криминал

Глобальный индекс проблемы: 0.46.
Новостной фон: только проблема.



Основные темы:

Тема 1. Ключевые слова: say, court, mr, gen, case, pol, nacc, corruption, pol gen, minister, party, former, baht, office, general
Количество статей: 846

Тема 2. Ключевые слова: baht, arrest, million, money, scam, million baht, gambling, victim, suspect, account, say, gold, steal, online, bank
Количество статей: 1302

Тема 3. Ключевые слова: driver, road, accident, motorcycle, pattaya, incident, vehicle, car, fire, man, truck, tourist, taxi, crash, scene
Количество статей: 864

Тема 4. Ключевые слова: body, old, year old, man, find, scene, suspect, death, shoot, incident, murder, hospital, officer, he, stab
Количество статей: 1216

Тема 5. Ключевые слова: assault, woman, man, victim, old, girl, year old, she, incident, video, child, thai, he, take, say
Количество статей: 1020

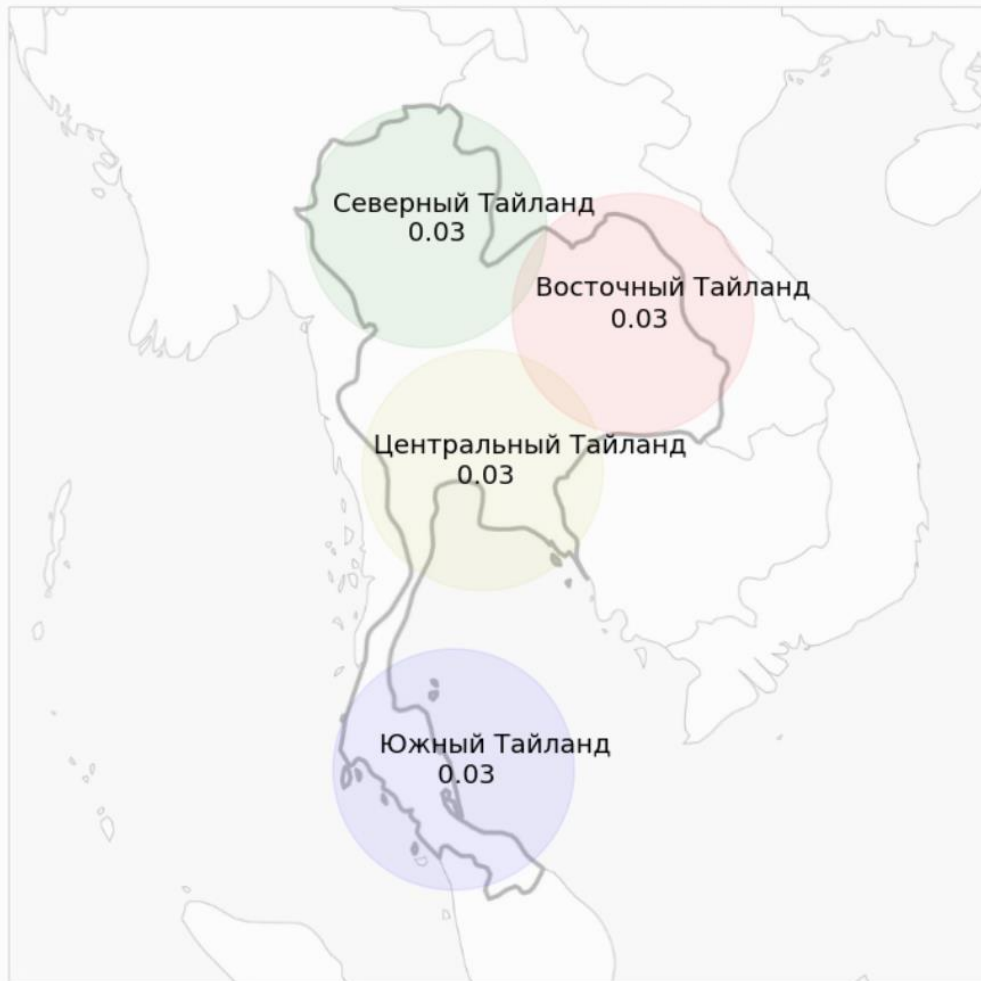
Тема 6. Ключевые слова: drug, pill, methamphetamine, arrest, operation, meth, seize, suspect, cigarette, say, million, cannabis, crystal, baht, narcotic
Количество статей: 782

Тема 7. Ключевые слова: myanmar, chinese, say, thailand, thai, immigration, national, arrest, border, tourist, illegal, foreign, work, man, phuket
Количество статей: 920

Представление результатов

Стихийные бедствия

Глобальный индекс проблемы: 0.03.
Новостной фон: весь



Основные темы:

Тема 1. Ключевые слова: fire, building, earthquake, say, collapse, rescue, bangkok, damage, report, hospital, cause, safety, construction, worker, district

Количество статей: 406

Тема 2. Ключевые слова: police, man, thailand, arrest, incident, year, woman, officer, video, old, year old, thailand, drug, pattaya, suspect

Количество статей: 294

Тема 3. Ключевые слова: water, flood, river, district, chiang, say, flooding, area, province, level, rain, water level, resident, mai

Количество статей: 439

Тема 4. Ключевые слова: phuket, pattaya, tourist, beach, water, island, rescue, local, boat, marine, year, say, fish, news, safety

Количество статей: 216

Тема 5. Ключевые слова: thunderstorm, wind, temperature, region, area, metre, wave, degree celsius, celsius, weather, degree, sea, heavy, low, high

Количество статей: 241

Тема 6. Ключевые слова: party, minister, say, pheu, pheu thai, government, thaksin, mr, prime, prime minister, election, paetongtarn, political, thai, bill

Количество статей: 145

Тема 7. Ключевые слова: thailand, year, say, price, thai, baht, government, market, us, economic, global, rice, product, tourism, million

Количество статей: 303