

Mindblowing Internet
Revolution

Идея. Выделение трендов

**1. Выгружаем наборы новостей за разные периоды
(например, за каждый месяц на протяжении года).**

2 Для каждого набора новостей:

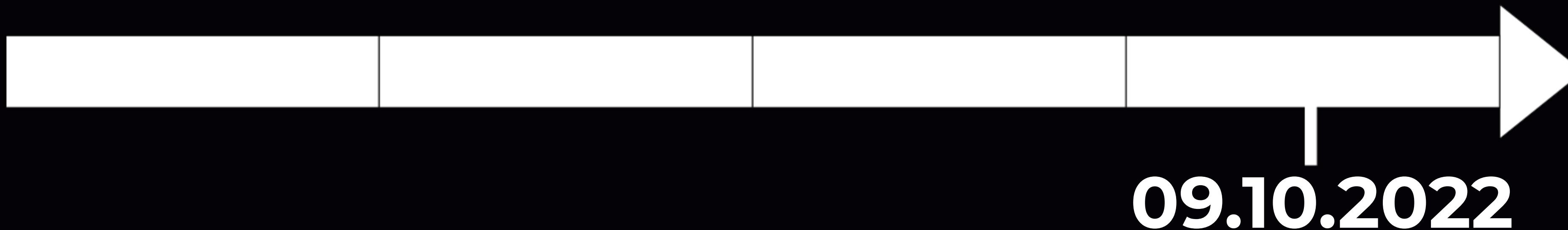
- a) Переводим каждую новость в пространство векторов;**
- b) Проводим кластеризацию полученных векторов по метрике близости;**
- c) Определяем наибольшие кластеры и их центры;**

3. Полученные центры кластеров сравнить по метрике близости.

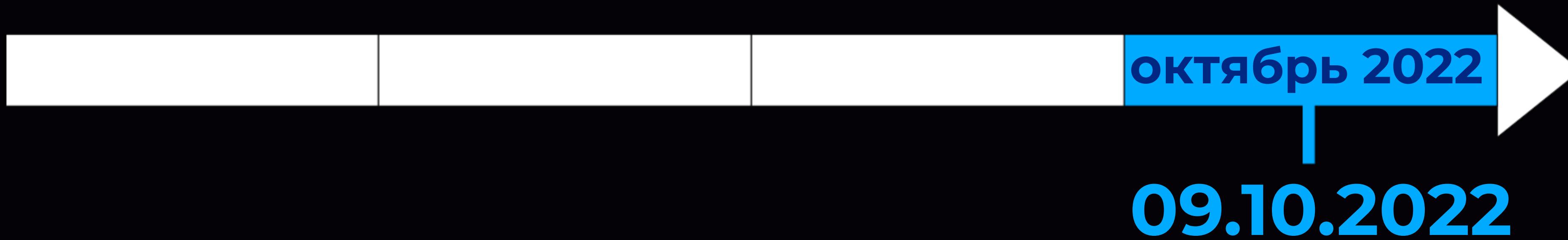
**Если вокруг кластера “свежих” новостей нет кластера “старых” новостей,
то это нарастающий тренд. Посмотрим статистику.**

**4. Для кластеров, показывающих тренды, извлечь новость,
соответствующую центру кластера.**

Выбираем начальную дату



Кластеризуем новости за 1 месяц, определяем кластеры.
Выделяем N наибольших кластеров по количеству эмбейдингов.
Это наши кластеры-гипотезы.



**Выбираем более ранний месяц, подсчитываем эмбендинги для него.
Определяем к каким кластерам-гипотезам они относятся.
Подсчитываем размеры кластеров-гипотез для этого месяца.**



Повторяем N раз для более ранних периодов



График размеров кластера №1 за выбранные периоды

Видим явно возрастание – новость является трендом

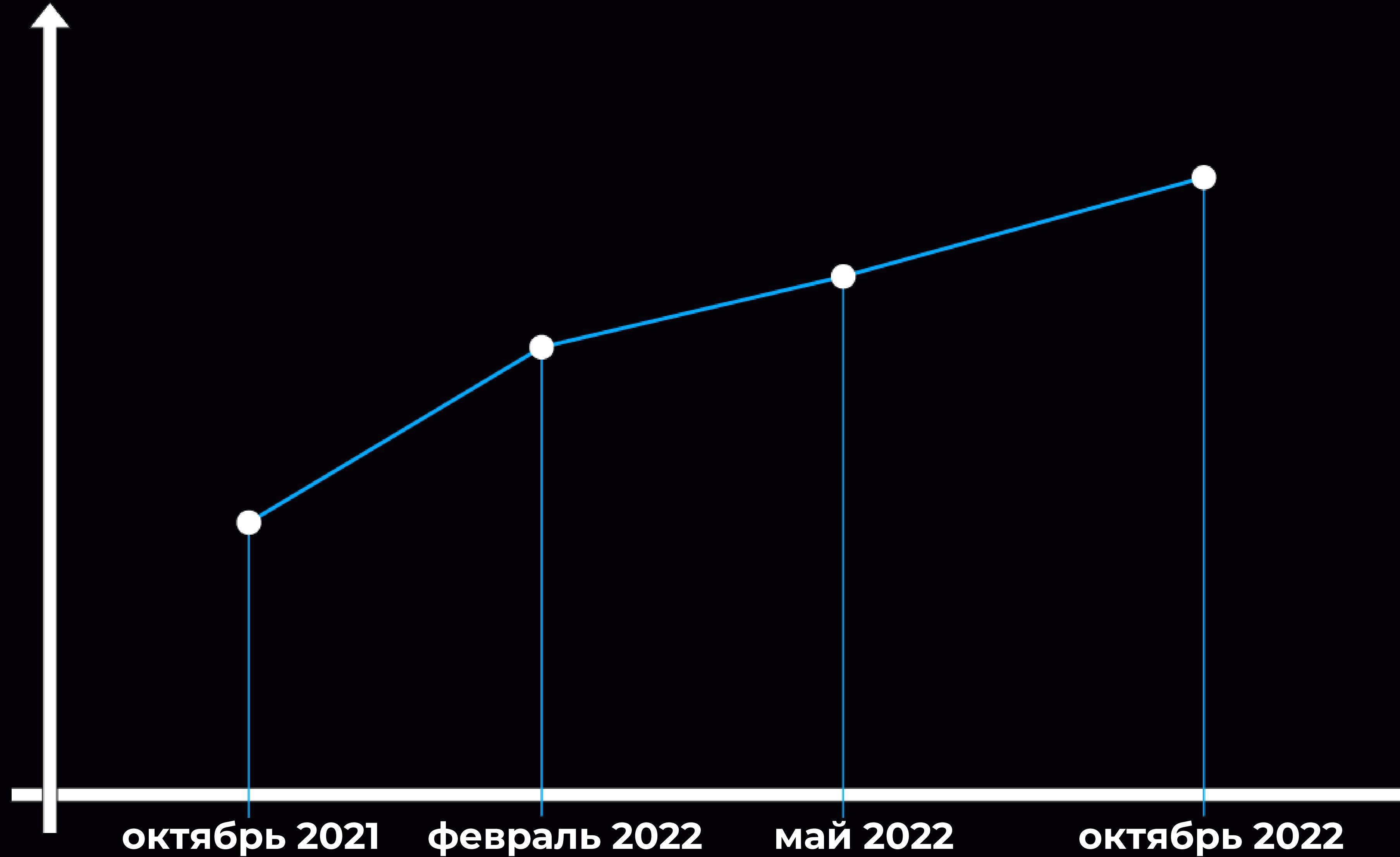
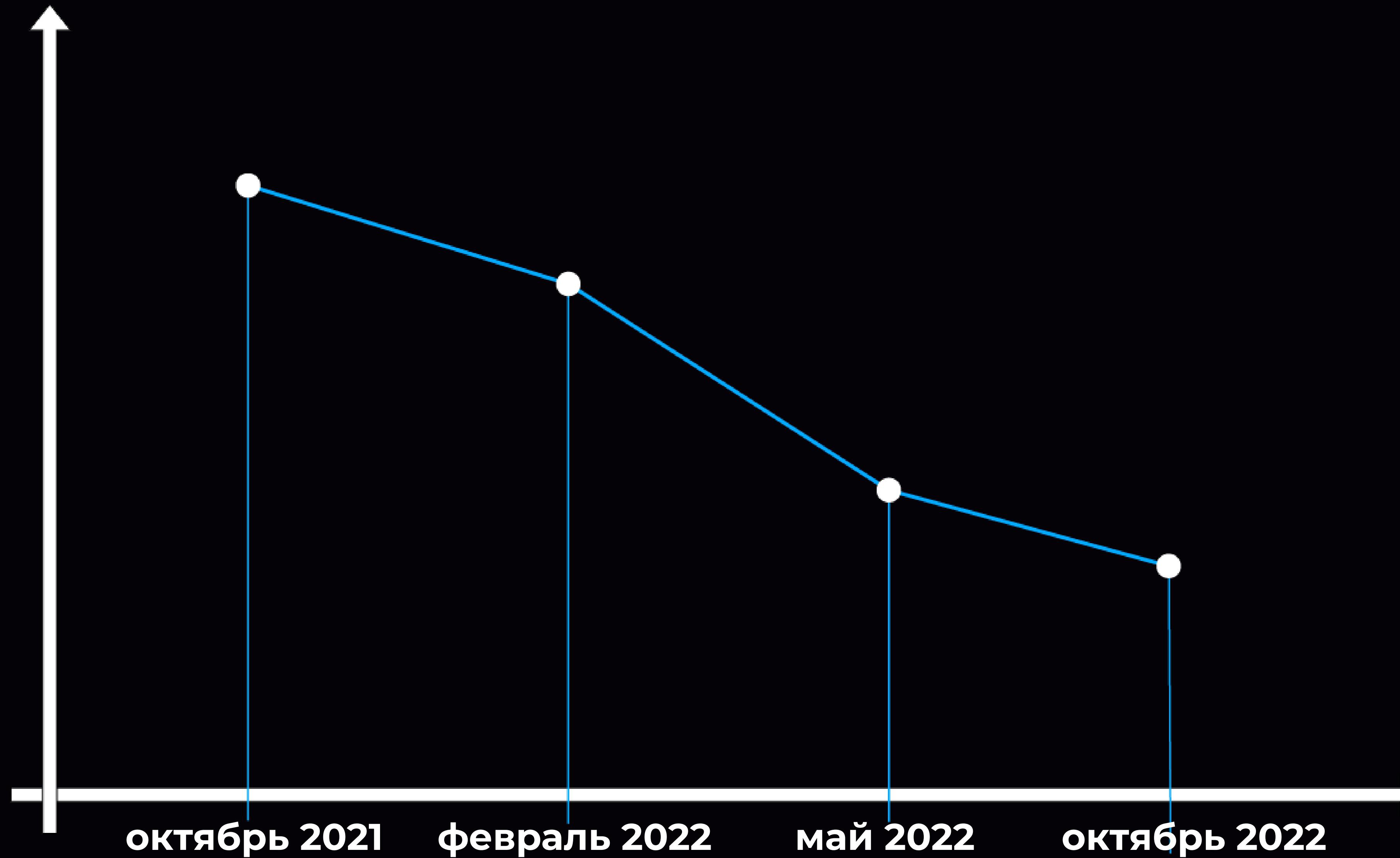


График размеров кластера №1 за выбранные периоды

Видим явное убывание – новость не является трендом



Идея. Дайджесты

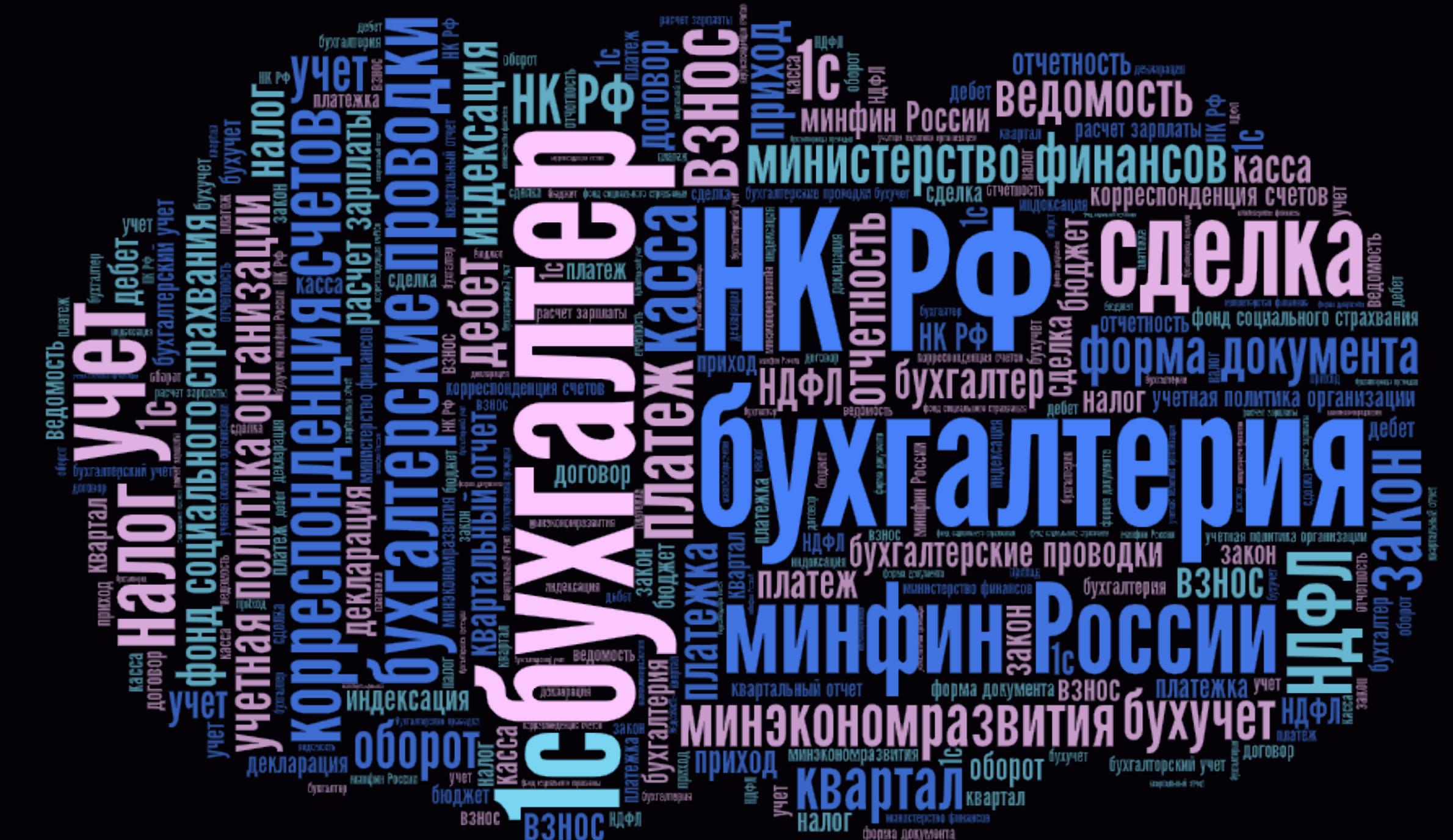
1. Парсим новости из непрофильных источников
2. Собираем массив слов и словосочетаний для каждой из ролей
3. Выделяем из новостей ключевые слова и переводим их в пространство векторов
4. Выбираем новости, ключевые слова которых ближе всего к массиву слов роли.



Ключевые слова для ролей

бизнесмен

бухгалтер



Ключевые слова по новостям

index	title	keywords
0	Минфин США: запрет на морские перевозки российской нефти начнет действовать с 5 декабря	['импорт российской нефти', 'цены российской нефти', 'нефти декабря нефтепродуктов', 'закупленной нефти ранее', 'цену закупленной нефти']
1	Bloomberg: производство в Германии упадет на 65% без газа из «Северного потока»	['отсутствия газа трубопровода', 'ускорится до расходы', 'энергоресурсов приведут сокращению', 'за отсутствия газа', 'приведет сокращению производств']
2	Дегтярев: провинциям КНР предложили перевозить их внутренние грузы через Хабаровский край	['россии дмитрий биричевский', 'сотрудничеству хабаровским краем', 'дмитрий биричевский заявлял', 'интерес сотрудничеству хабаровским', 'хабаровским краем проявляют']
3	Министр Кордейру: Португалия не согласна с ЕС по обязательной экономии электроэнергии	['электроэнергии евросоюза добившись', 'еврокомиссия намерена предложить', 'заявил что евросоюз', 'что еврокомиссия обязала', 'что еврокомиссия намерена']
4	Аналитик Citigroup Морс допустил ухудшение энергокризиса в ЕС из-за потолка цен на нефть из РФ	['усугубить энергокризис евросоюзе', 'мировой рынок скептически', 'усугубить энергетический кризис', 'цены на российскую', 'энергокризис евросоюзе ес']
5	Украина запросила \$1,5 млрд у Экспортно-импортного банка США на закупку газа	['газа нафтогаз украины', 'нафтогаз украины анонсировал', 'природного газа нафтогаз', 'транспортировке переработке нефти', 'на закупку газа']
6	В Италии растет недовольство ценами на электроэнергию	['сторонники молдавской оппозиционной', 'президента кишиневе объявили', 'оппозиционной партии шор', 'администрации президента кишиневе', 'молдавской оппозиционной партии']
7	Пищепром Украины готовится к повышению цен на продукты из-за подорожания энергоресурсов	['рублей производители сахара', 'пищевая промышленность готовится', 'портов пшеницу остальные', 'даст сахарозаводчикам газовые', 'производители сахара также']
8	Evgensel: Европе грозит холодная зима из-за газового противостояния с Россией	['ущерб понесут европейцы', 'введут потолок цен', 'антироссийские санкции бумерангом', 'экономике европы при', 'зимы газовое противостояние']
9	В Госдепе США прокомментировали заявление России относительно получателей зерна	['продовольствием украины российский', 'его словам развивающиеся', 'форуме заявил что', 'экономическом форуме заявил', 'чем это глобальные']

Дайджест новостей. Бухгалтер.

- “FT: ЕС окажет помощь пострадавшим от антироссийских санкций компаниям”
- “Аналитик Мехтиев назвал обесценивание одним из основных рисков при снятии...”
- “Гутериш встретится с Байденом для обсуждения цен на энергоносители”
- “Россияне стали покупать системы видеонаблюдения почти вдвое чаще за...”
- “Транзит контейнеров по железным дорогам России упадет на 25%”
- “Переориентация на внутренний рынок: новые реалии мебельной отрасли России”
- “Bloomberg: европейцы сорвали план ЕС по энергосбережению из-за отказа экон...”
- “Банки зафиксировали увеличение спроса россиян на вложения в драгметаллы”

Дайджест новостей. Бизнесмен.

- “Аналитик Мехтиев назвал обесценивание одним из основных рисков при снятии...”
- “Путин призвал производить конкурентоспособные двигатели для гражданской...”
- “Силуанов назвал предстоящий бюджет России самым сложным в его карьере”
- “Вице-премьер Молдавии сообщил о сокращении «Газпромом» поставок газа...”
- “Вице-премьер Молдавии Спину заявил о сокращении потребления газа в стране...”
- “Процедуру банкротства российской «дочки» Google приостановили из-за неуплаты...”
- “Президент Путин заявил, что экономический «блицкриг» ряда стран против России...”
- “Министр Польши Москва заявила, что предприятиям порекомендуют уменьшить...”
- “Международные резервы России сократились за неделю на \$4,5 млрд или на 0,8%”

Дайджест новостей. Пояснения.

Это результат за последний месяц из неспециализированных источников.
Но когда мы включили специализированный источник в данные,
получили 6 из 8 релевантных новостей из специализированного источника.

Источники для парсинга

- РБК.Бизнес – 766 новостей
- Газета.ру – 22452 новости
- Forbes.ru – 999 новостей
- Коммерсантъ – 1764 новостей

Технологии:

- **парсинг: Scrapy**
- **векторное представление слов: ‘rubert-tiny’**
- **кластеризация: AgglomerativeClustering**
- **выделение ключевых слов: ‘rubert-tiny’,
‘DeepPavlov/rubert-base-cased’**
- **API: FastAPI**

Работа API. Бизнесмен.



Filter

History Collections Clear all

No Environment

GET http://127.0.0.1:8000/get_digest?role=businessman Params Send Save

Pretty Raw Preview JSON

1 * {}
2 * "digest": [
3 * {
4 * "title": "Аналитик Мехтиев назвал обесценивание одним из основных рисков при снятии наличных со счета ",
5 * "content": "Основными рисками при снятии средств со счета являются их обесценивание, вопрос хранения и искушение их потратить. Об этом радио Sputnik заявил заместитель председателя Экспертного совета Банка России по защите прав потребителей финансовых услуг Эльман Мехтиев. «Если вы считаете, что нужно снимать деньги, подумайте о том, где и как вы будете хранить сумму, которую сняли в банке. Основной риск – не у каждого дома есть сейф. Кроме того, есть риск обесценивания из-за инфляции», – отметил Мехтиев. Он добавил, что еще одним основным риском является искушение потратить снятые средства. То есть возникает вопрос, который заключается в умении контролировать свои расходы. Однако лучше всего хранить свои сбережения в банке, так как они надежно застрахованы. По его словам, не стоит забывать, что от государства есть гарантии по вкладам размером до 1,4 млн руб. Причем с 2004 года эти гарантии работают без ошибок и остановок. Кроме того, около 99% вкладов населения как раз входят в эту сумму. 26 сентября основатель компании Anderida Financial Group Алексей Тараповский заявил, что россиянам следует задуматься о возможности открытия валютных счетов в «дружественных» странах на фоне усиления геополитической напряженности в мире. Допустимыми вариантами станут банки Белоруссии, Армении и Казахстана.",
6 * "url": "https://www.gazeta.ru/business/news/2022/09/28/18666733.shtml",
7 * "news_idx": 1515
8 * },
9 * {
10 * "title": "Путин призвал производить конкурентоспособные двигатели для гражданской авиации ",
11 * "content": "Президент России Владимир Путин на встрече с руководителями передовых инженерных школ в Великом Новгороде заявил, что Россия с советских времен не производила конкурентоспособные двигатели для гражданской авиации, сообщает ТАСС. «Никто ничего не делал все время. Конечно, лучше покупать, дешевле даже и комфортнее, «Боинги» и «Эйрбасы». Но если сами не будем делать – никогда не появятся [свои авиадвигатели для гражданской авиации – «Газета.Ru»]», – заявил Путин. Он также подчеркнул, что санкции Запада против России подстегнули производство авиадвигателей. «Здесь вот не было бы счастья, да несчастье помогло», – отметил глава государства. Путин также отметил, что из-за антироссийских санкций Минтранс оставил в России все переданные в лизинг самолеты. При этом президент отметил, что ресурс этих самолетов не бесконечен, и через несколько лет они начнут выходить из строя. «К этому времени нам нужны свои машины. Они у нас есть, но их нужно тиражировать и улучшать по качеству, в том числе и прежде всего по двигателенным установкам. Поэтому это очень важно, это востребовано прямо сейчас, сейчас нужны эти разработки», – заявил Путин. Он добавил, что несмотря на то, что Россия производит качественные двигатели для боевой авиации, на гражданском авиарынке совсем другие требования. Ранее стало известно, что США внесут самолеты авиакомпаний Ирана в список нарушителей санкций против России.",
12 * "url": "https://www.gazeta.ru/business/news/2022/09/21/18617047.shtml",
13 * "news_idx": 968
14 * },
15 * {
16 * "title": "Силуанов назвал предстоящий бюджет России самым сложным в его карьере ",
17 * "content": "Глава министерства финансов России Антон Силуанов считает, что российский бюджет на предстоящие три года станет самым сложным в его карьере. Об этом министр заявил в эфире телеканала «Россия-24» в фильме, посвященном 220-летию ведомства, которое он возглавляет. «Я думаю, что бюджет на предстоящую трехлетку будет самый сложный в моей профессиональной карьере», – сказал Силуанов. В декабре 2021 года президент России Владимир Путин подписал закон о федеральном бюджете на 2022 год и плановый период 2023–2024 года. В сентябре 2021 года глава кабмина Михаил Мишустин говорил, что следующие два года бюджет будет самым сложным в истории страны."}]

Today

GET http://127.0.0.1:8000/get_digest?role=businessman

POST https://api.aicloud.sbercloud.ru/public/v2/summarizer/predict

POST https://api.aicloud.sbercloud.ru/public/v2/summarizer/predict

POST https://api.aicloud.sbercloud.ru/public/v2/summarizer/predict

POST https://api.aicloud.sbercloud.ru/public/v2/summarizer/predict

Yesterday

GET https://api.vedomosti.ru/v2/lists/rubrics-business-top?offset=33&limit=20

GET https://api.vedomosti.ru/v2/lists/rubrics-business-top?offset=33&limit=5

GET https://api.forbes.ru/api/pub/lists/biznes?list%5Blimit%5D=2&list%5Boffset%5D=0

GET https://api.forbes.ru/api/pub/lists/biznes?list%5Blimit%5D=1&list%5Boffset%5D=11

GET https://api.forbes.ru/api/pub/lists/biznes?list%5Blimit%5D=20&list%5Boffset%5D=14

Работа API. Бухгалтер.



Filter

History Collections Clear all

No Environment

GET http://127.0.0.1:8000/get_digest?role=bookkeeper Params Send Save

Pretty Raw Preview JSON

1 {
2 "digest": [
3 {
4 "title": "FT: ЕС окажет помощь пострадавшим от антироссийских санкций компаниям ",
5 "content": "Европейский союз занимается разработкой серии мер, направленных на прямую поддержку компаний, пострадавших от санкций против России, которые были введены из-за ситуации на Украине. Об этом пишет британская газета Financial Times (FT), ссылаясь на черновой вариант готовящихся документов. По данным издания, Евросоюз намерен субсидировать европейские предприятия, страдающие от роста цен на энергоносители. Регулирующие органы в ЕС при этом отмечают, что помочь должна быть «точечной и пропорциональной» с соблюдением действующей нормативной базы, чтобы избежать «гонки субсидий» и нарушения механизмов честной конкуренции на общем рынке. Ранее руководитель направления «Финансы и экономика» Института современного развития Никита Масленников заявил, что Европа самостоятельно впадает в глубокий экономический кризис, который создала своими санкциями против РФ. До этого эксперт, глава нефтесервисной компании Canary Дэн Эберхарт выразил мнение, что Западу грозит «завершающий смертельный удар» от введения ограничений на стоимость энергоресурсов из России. Он назвал ценовой потолок на российскую нефть смертельным ударом для Запада. По его оценке, санкции против Москвы «не сулят ничего хорошего архитекторам этой политики». В настоящее время на Украине продолжается специальная операция Вооруженных сил РФ. О ее начале 24 февраля объявил президент России Владимир Путин . По его словам, цель военных действий заключается в демилитаризации соседней страны и денацификации ее властей. Решение о проведении операции стало причиной новых санкций против России со стороны США и их союзников.",
6 "url": "https://www.gazeta.ru/business/news/2022/10/08/18746803.shtml",
7 "news_idx": 2439
8 },
9 {
10 "title": "Аналитик Мехтиев назвал обесценивание одним из основных рисков при снятии наличных со счета ",
11 "content": "Основными рисками при снятии средств со счета являются их обесценивание, вопрос хранения и искушение их потратить. Об этом радио Sputnik заявил заместитель председателя Экспертного совета Банка России по защите прав потребителей финансовых услуг Эльман Мехтиев. «Если вы считаете, что нужно снимать деньги, подумайте о том, где и как вы будете хранить сумму, которую сняли в банке. Основной риск – не у каждого дома есть сейф. Кроме того, есть риск обесценивания из-за инфляции», – отметил Мехтиев. Он добавил, что еще одним основным риском является искушение потратить снятые средства. То есть возникает вопрос, который заключается в умении контролировать свои расходы. Однако лучше всего хранить свои сбережения в банке, так как они надежно застрахованы. По его словам, не стоит забывать, что от государства есть гарантии по вкладам размером до 1,4 млн руб. Причем с 2004 года эти гарантии работают без ошибок и остановок. Кроме того, около 99% вкладов населения как раз входят в эту сумму. 26 сентября основатель компании Anderida Financial Group Алексей Тараповский заявил, что россиянам следует задуматься о возможности открытия валютных счетов в «дружественных» странах на фоне усиления геополитической напряженности в мире. Допустимыми вариантами станут банки Белоруссии, Армении и Казахстана.",
12 "url": "https://www.gazeta.ru/business/news/2022/09/28/18666733.shtml",
13 "news_idx": 1515
14 },
15 {
16 "title": "Гуттерриш встретится с Байденом для обсуждения цен на энергоносители ",
17 "content": "Генсек ООН Антониу Гуттерриш заявил в интервью РИА «Новости», что планирует встретиться с президентом США Джо Байденом 21 сентября в Нью-Йорке на полях недели высокого уровня сессии Генассамблеи ООН и обсудить рост цен на энергоносители и продукты. По словам Гуттерриша, в беседе с американским президентом будут рассмотрены главные темы международной повестки, в том числе стороны коснутся «проблем, связанных с устойчивым развитием, ситуации в

Today

GET http://127.0.0.1:8000/get_digest?role=bookkeeper

GET http://127.0.0.1:8000/get_digest?role=businessman

POST https://api.aicloud.sbercloud.ru/public/v2/summarizer/predict

POST https://api.aicloud.sbercloud.ru/public/v2/summarizer/predict

POST https://api.aicloud.sbercloud.ru/public/v2/summarizer/predict

POST https://api.aicloud.sbercloud.ru/public/v2/summarizer/predict

Yesterday

GET https://api.vedomosti.ru/v2/lists/rubrics-business-top?offset=33&limit=20

GET https://api.vedomosti.ru/v2/lists/rubrics-business-top?offset=33&limit=5

GET https://api.forbes.ru/api/pub/lists/biznes?list%5Blimit%5D=2&list%5Boffset%5D=0

GET https://api.forbes.ru/api/pub/lists/biznes?list%5Blimit%5D=1&list%5Boffset%5D=11

Работа API. Тренды.



```
1  {
2    "trends": [
3      "экоактивистке",
4      "twitter",
5      "tesla",
6      "евросоюзе",
7      "газопровода",
8      "обеспечению",
9      "изъятые",
10     "отмывании",
11     "похищенных",
12     "минпромторг",
13     "свекловичного",
14     "превышающей"
15   ]
16 }
```

Наша команда

Mindblowing Internet Revolution



Булат
Вахитов
4 курс СПбГУ



Сарана
Абидуева
4 курс СПбГУ



Максим
Кораблев
4 курс СПбГУ



Дмитрий
Личко
4 курс МИСИС



Анастасия
Шконда
4 курс СПбГУ

Спасибо за внимание!

