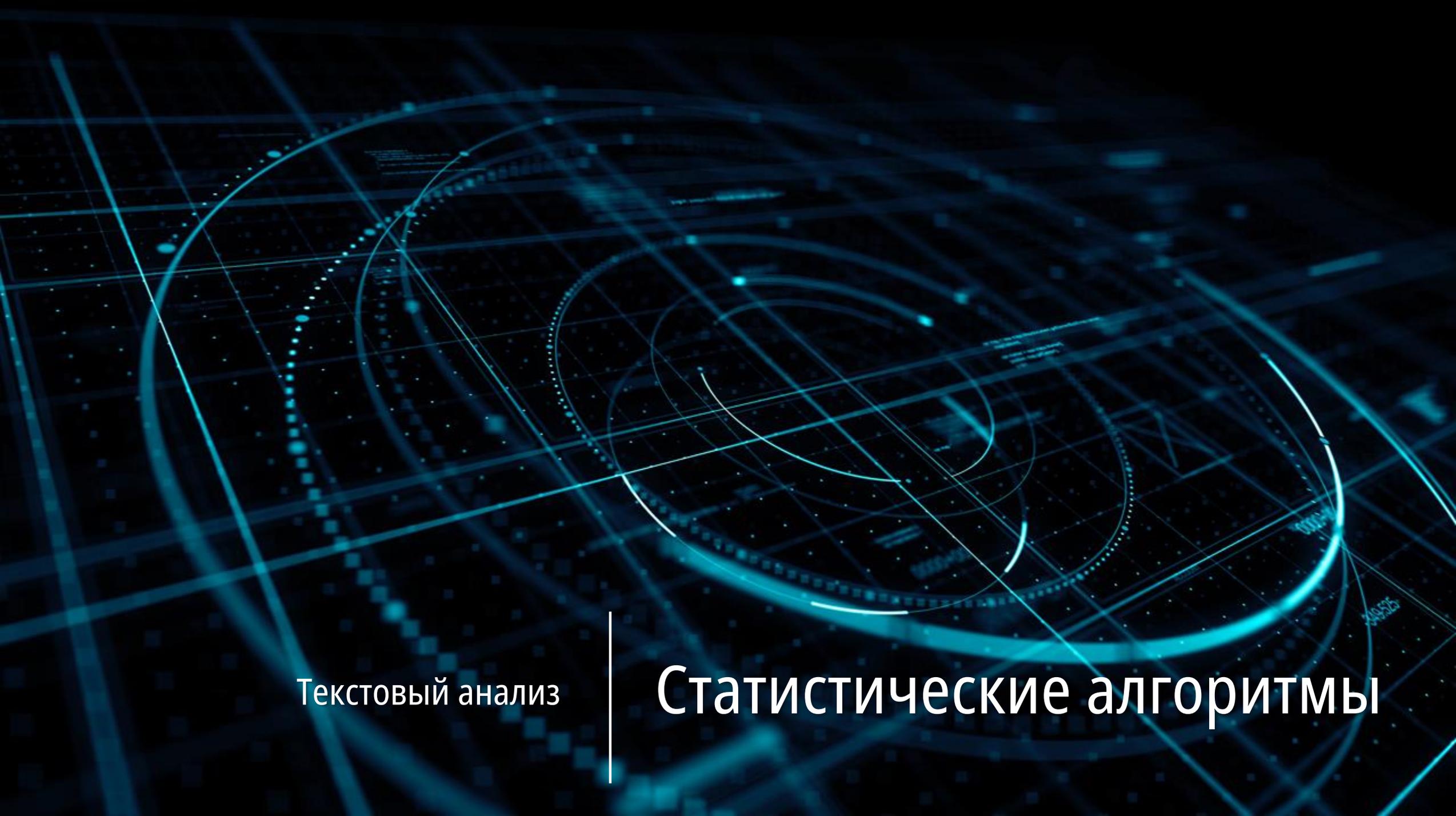




Как извлекать полезную  
информацию из текста -

Система PolyAnalyst™

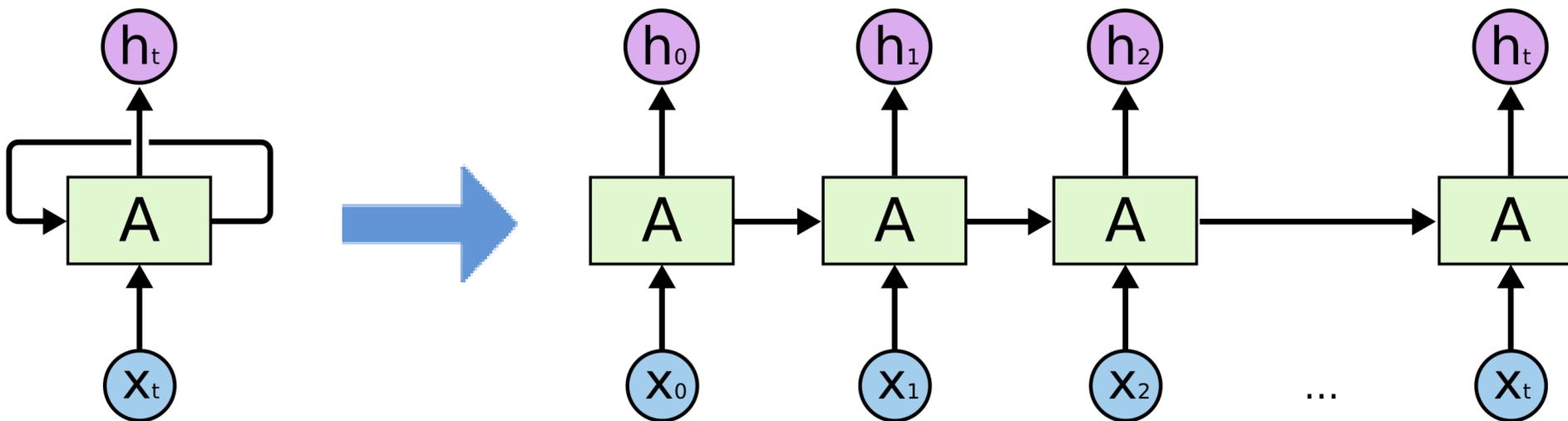




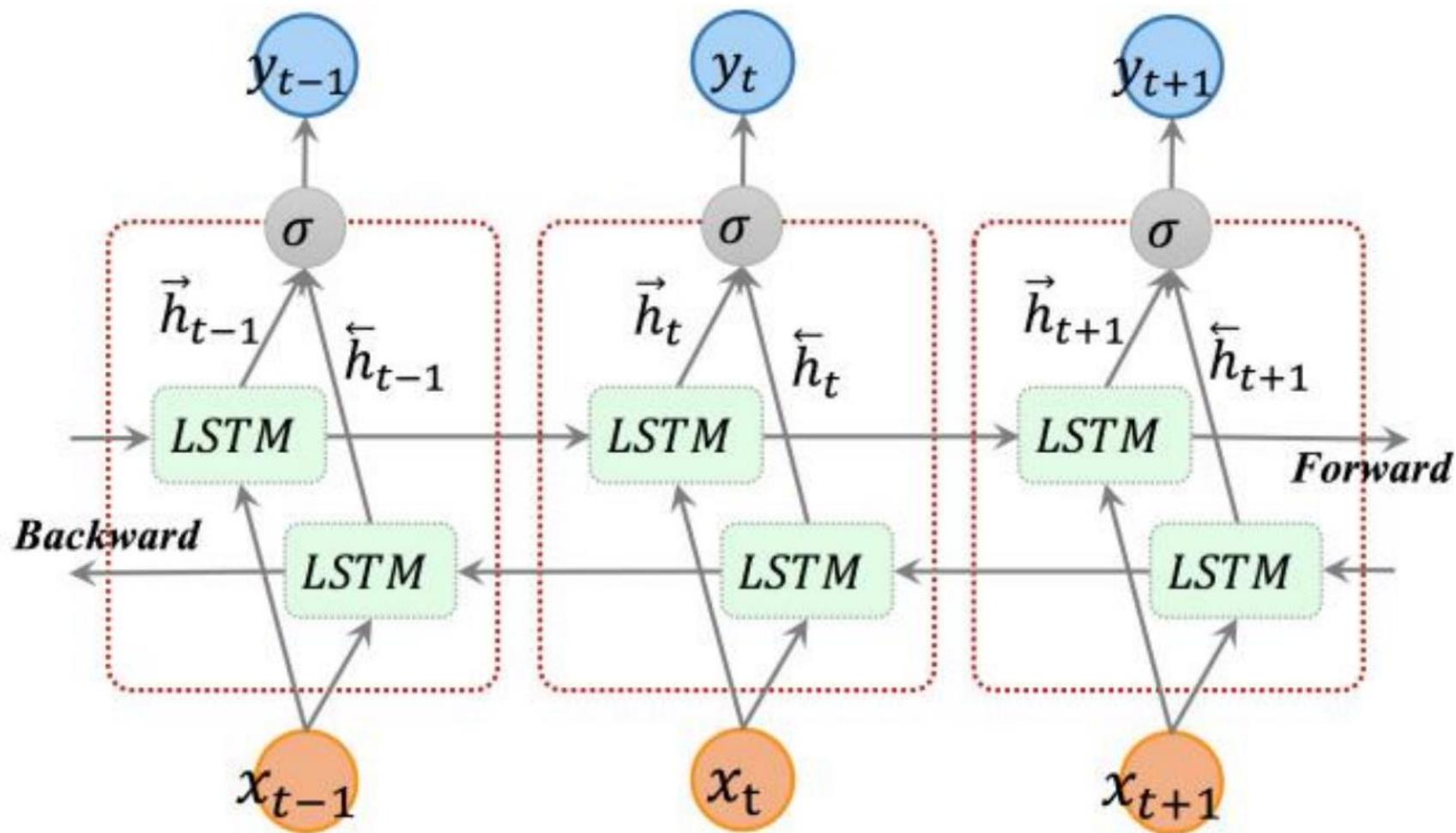
Текстовый анализ

Статистические алгоритмы

# РЕКУРРЕНТНЫЕ НЕЙРОННЫЕ СЕТИ – МАШИННОЕ ОБУЧЕНИЕ



# Bi-directional LSTM (BI-LSTM) – МАШИННОЕ ОБУЧЕНИЕ





## КЛАССИФИКАЦИЯ



группировка текстов на основе заранее  
определенного пользователем списка  
классов

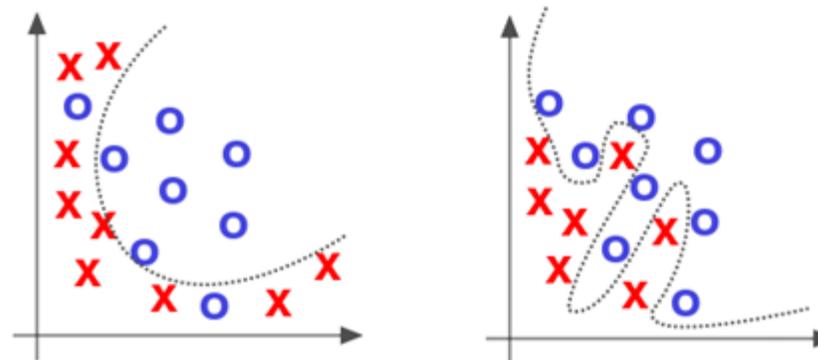


## КЛАСТЕРИЗАЦИЯ

группировка близких по содержанию текстов без  
предварительной информации

# ЭФФЕКТ ПЕРЕОБУЧЕНИЯ

Применяя машинное обучение на небольшом размеченном корпусе текстов, можно получить некорректные результаты.



Корректная модель

Переобученная модель

## ПРИМЕР ПЕРЕОБУЧЕНИЯ

Классификатор для события «банкротство», обученный на новостях за период банкротства компании UTair

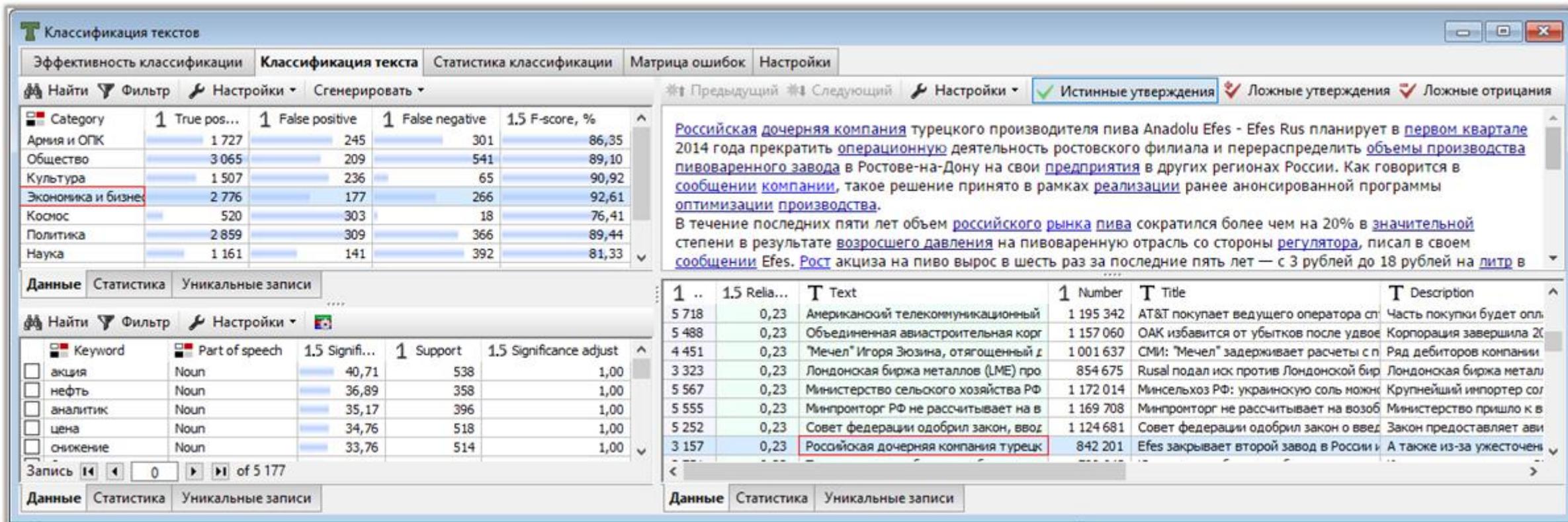
<input type="checkbox"/> <input checked="" type="checkbox"/> Токен	1.5 Значимость
<input type="checkbox"/> банкрот	100,00
<input checked="" type="checkbox"/> авиакомпания	80,89
<input type="checkbox"/> банкротство	70,83
<input type="checkbox"/> признание	67,70
<input type="checkbox"/> арбитражный	58,09
<input type="checkbox"/> суд	51,55

ЕКАТЕРИНБУРГ, 14 апреля. /ТАСС/. [Альфа-банк](#) подал в [Арбитражный суд Тюменской области](#) иск о банкротстве компании ["ЮТэйр-Лизинг"](#) (дочерняя структура [авиакомпаний UTair](#)), сообщили [ТАСС](#) в суде.

ЕКАТЕРИНБУРГ, 7 апреля. /ТАСС/. [Арбитражный суд](#) Ханты-Мансийского автономного округа-Югры назначил на 16 апреля рассмотрение первого иска [Альфа-банка](#) о признании банкротом [авиакомпаний UTair](#), сообщили [ТАСС](#) в суде.

# ПРИМЕР КЛАССИФИКАЦИИ ТЕКСТОВ

## СПИСОК КАТЕГОРИЙ



**Классификация текстов**

Эффективность классификации | **Классификация текста** | Статистика классификации | Матрица ошибок | Настройки

Найти | Фильтр | Настройки | Сгенерировать

Истинные утверждения | Ложные утверждения | Ложные отрицания

Category	True pos...	False positive	False negative	1.5 F-score, %
Армия и ОПК	1 727	245	301	86,35
Общество	3 065	209	541	89,10
Культура	1 507	236	65	90,92
Экономика и бизнес	2 776	177	266	92,61
Космос	520	303	18	76,41
Политика	2 859	309	366	89,44
Наука	1 161	141	392	81,33

Данные | Статистика | Уникальные записи

Найти | Фильтр | Настройки

Keyword	Part of speech	1.5 Signifi...	Support	1.5 Significance adjust
<input type="checkbox"/> акция	Noun	40,71	538	1,00
<input type="checkbox"/> нефть	Noun	36,89	358	1,00
<input type="checkbox"/> аналитик	Noun	35,17	396	1,00
<input type="checkbox"/> цена	Noun	34,76	518	1,00
<input type="checkbox"/> снижение	Noun	33,76	514	1,00

Запись 0 of 5 177

Данные | Статистика | Уникальные записи

1 .. 1.5 Reli... | T Text | 1 Number | T Title | T Description

5 718 0,23 Американский телекоммуникационный 1 195 342 AT&T покупает ведущего оператора сп Часть покупки будет опл.

5 488 0,23 Объединенная авиастроительная корг 1 157 060 ОАК избавится от убытков после удвое Корпорация завершила 2К

4 451 0,23 "Мечел" Игоря Эюзина, отягощенный д 1 001 637 СМИ: "Мечел" задерживает расчеты с п Ряд дебиторов компании

3 323 0,23 Лондонская биржа металлов (LME) про 854 675 Rusal подал иск против Лондонской бир Лондонская биржа метали

5 567 0,23 Министерство сельского хозяйства РФ 1 172 014 Минсельхоз РФ: украинскую соль можн Крупнейший инпортер сог

5 555 0,23 Минпромторг РФ не рассчитывает на в 1 169 708 Минпромторг не рассчитывает на возоб Минпротомрг пришло к в

5 252 0,23 Совет федерации одобрил закон, ввод 1 124 681 Совет федерации одобрил закон о введ Закон предоставляет ави

3 157 0,23 Российская дочерняя компания турецк 842 201 Efes закрывает второй завод в России и А также из-за ужесточени

Данные | Статистика | Уникальные записи

## СПИСОК ПРЕДИКТОРОВ

# ПРИМЕР КЛАСТЕРИЗАЦИИ ТЕКСТОВ



Классификация текстов

Классификация | Настройки

Найти | Фильтр | Настройки | Сгенерировать

Cluster #	Cluster name	Support
#1	президент, страна, государство, безопасность, украинс	3 5
#2	компания, турист, курорт, туристический, туроператор,	3 5
#3	космический, корабль, ракета, аппарат, спутник, запус	3 4
#4	театр, фильм, режиссёр, музей, картина, премия, куль	3 0
#5	выбор, образование, вуз, депутат, партия, государстве	2 8
#6	пожар, самолёт, человек, район, погибнуть, произойти	2 8
#7	суд, дело, уголовный, банк, следствие, лицо, преступле	2 4

Запись 1 of 10

Данные | Статистика | Уникальные записи

Найти | Фильтр | Настройки

Token	Part of speech	Significance	Support
<input type="checkbox"/> президент	Noun	100,00	1 756
<input type="checkbox"/> страна	Noun	85,52	2 327
<input type="checkbox"/> украинский	Adjective	81,53	878
<input type="checkbox"/> переговоры	Noun	78,48	1 001
<input type="checkbox"/> государство	Noun	77,63	1 502

Запись 0 of 593

Данные | Статистика | Уникальные записи

← Предыдущий | Следующий → | Настройки

МОСКВА, 9 августа. /ИТАР-ТАСС/. Глава МИД РФ Сергей Лавров и Госсекретарь США Джон Керри в ходе телефонного разговора высказали озабоченность обострившейся обстановкой в Ираке, выразив поддержку действиям иракских властей в борьбе с боевиками. Об этом сообщает МИД РФ.

США сообщили об успешном поражении ударами с воздуха позиций боевиков на севере Ирака

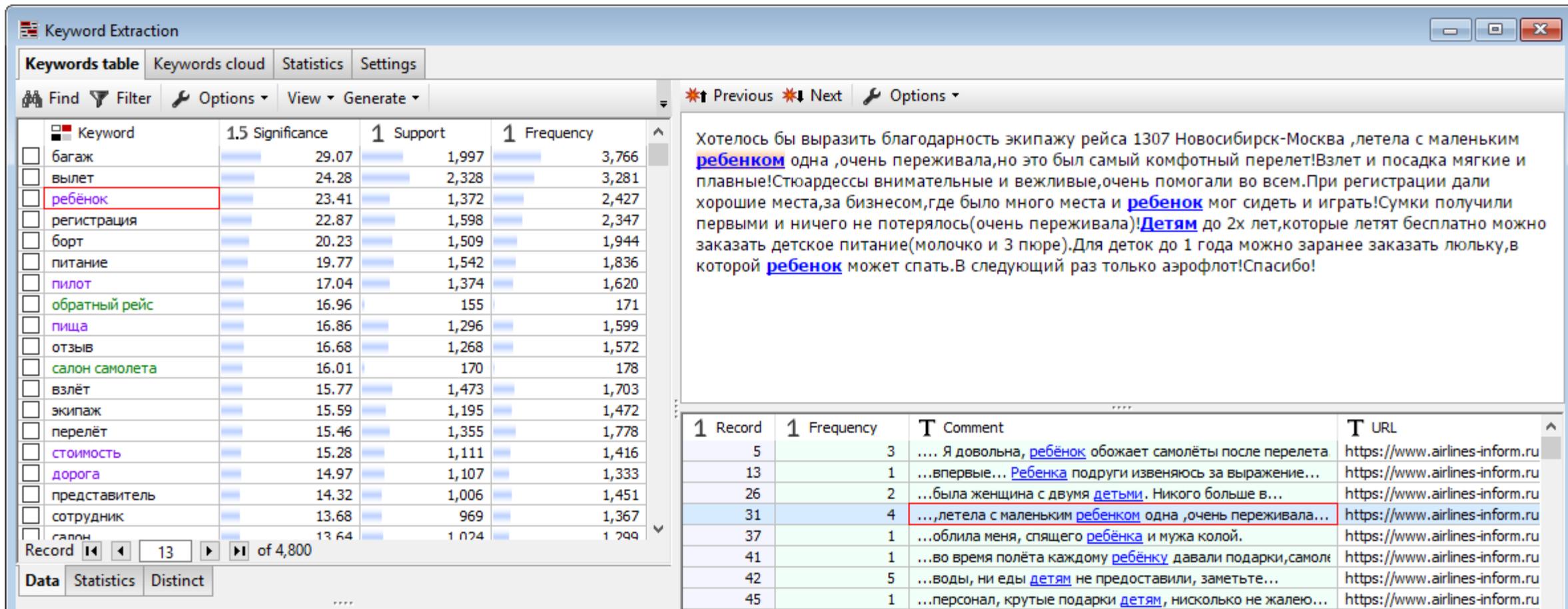
"Главы внешнеполитических ведомств высказали озабоченность резко обострившейся обстановкой в Ираке, выразив поддержку

1 ..	1.5 Reli...	T Text	1 Nu...	Subcategory	T Title	T Description
0 517	0,27	МОСКВА, 9 августа. /ИТАР-ТАСС/	1 370 906	Внешняя полити	Лавров и Керри высказали озабоч	Дипломаты выразили
7 328	0,25	МОСКВА, 19 декабря. /ТАСС/. Лк	1 659 199	Внешняя полити	МИД РФ: международное сообще	Любые планы по пост:
2 198	0,25	МОСКВА, 26 февраля. /ТАСС/. Вь	1 794 586	Внешняя полити	Лавров и Могерини обсудили вып	Телефонный разговор
1 139	0,24	ГАВАНА, 13 февраля. /ТАСС/. Мь	1 765 076	Вооружённые си	Шойгу прибыл с официальным ви	Российский министр of
2 281	0,24	МОСКВА, 27 февраля. /ТАСС/. Ф	1 796 857	Безопасность	ФСБ опубликовала список из 22 о	В списке террористиче
5 222	0,24	БЕЙРУТ, 3 января. /Корр. ИТАР-Т	870 483	Вооружённые ко	В Ираке застрелен главарь экстре	Его ликвидировали бо
8 264	0,24	МОСКВА, 8 января. /ТАСС/. Глав	1 687 608	Внешняя полити	Лавров и Фабиус: ЧП в Париже по	Глава Пентагона ране
954	0,24	ТЕГЕРАН, 5 февраля. /Корр. ИТА	558 121	Безопасность	Иран нанесет ответный удар стра	Глава Пентагона ране
0 327	0,24	Мюнхенская конференция по без	1 743 668		События недели 2-8 февраля в ф	Самые важные момен
3 637	0,24	МОСКВА, 15 октября. /ТАСС/. Пр	1 509 494	Внешняя полити	Ушаков: встреча на саммите АСЕМ	Договоренности о дву
2 348	0,24	МОСКВА, 22 сентября. /ИТАР-ТА	1 459 006	Внешняя полити	Путин обсудил с членами Совбеза	Также члены Совбеза
6 134	0,24	МОСКВА, 8 апреля. /ТАСС/. През	1 889 003	Внешняя полити	Путин: Россия заинтересована в г	"Российская сторона п

Запись 1 of 3 599

Данные | Статистика | Уникальные записи

# ПОИСК КЛЮЧЕВЫХ СЛОВ



The screenshot shows the 'Keyword Extraction' software interface. The main window is divided into several sections:

- Keywords table:** A table listing extracted keywords with their significance, support, and frequency. The word 'ребёнок' is highlighted in red.
- Text preview:** A snippet of text from a source document, containing the word 'ребёнок' in blue, indicating it is a keyword.
- Record list:** A table showing the source records for the keyword 'ребёнок', including record number, frequency, comment, and URL.

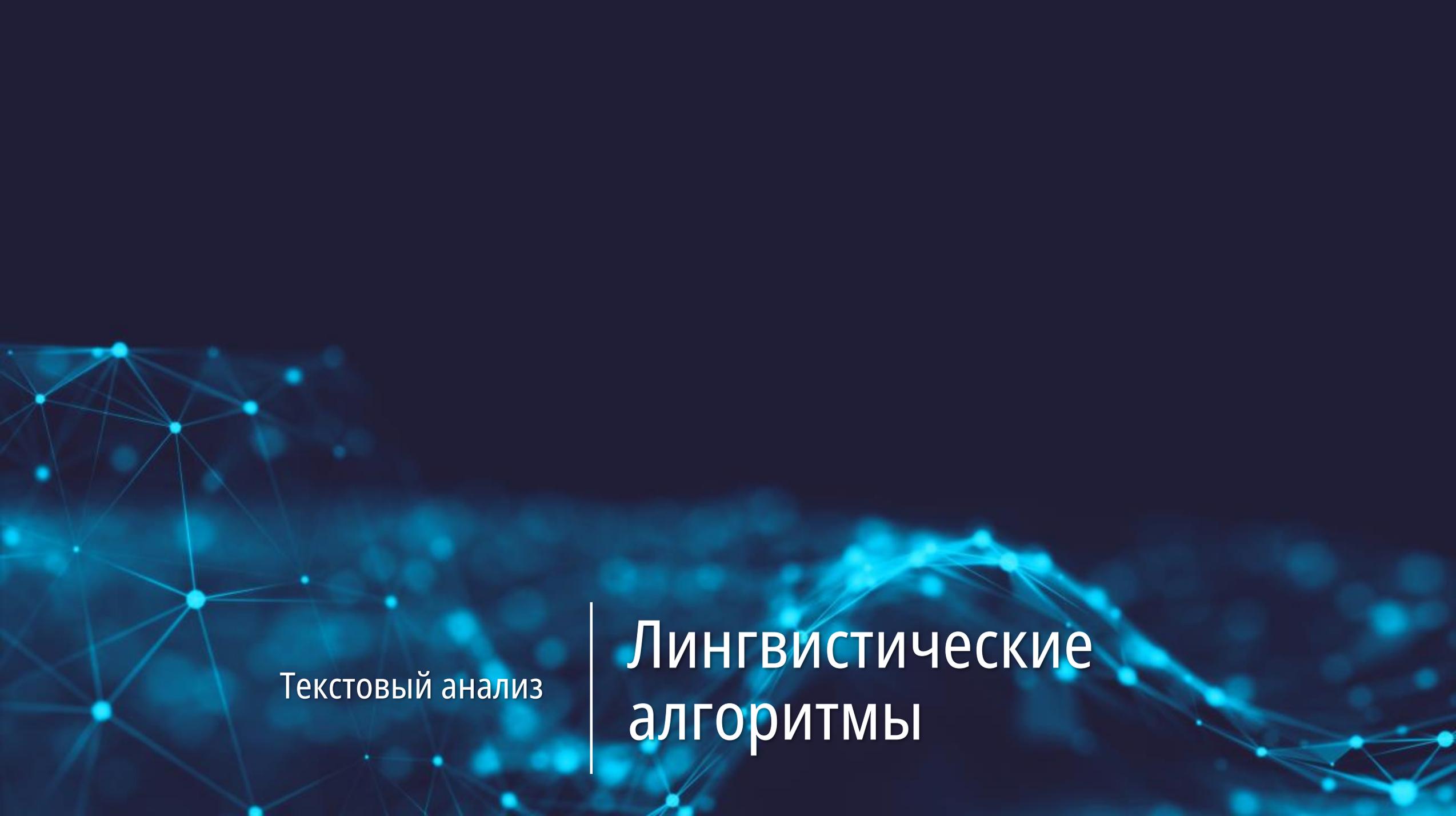
Keyword	1.5 Significance	1 Support	1 Frequency
багаж	29.07	1,997	3,766
вылет	24.28	2,328	3,281
ребёнок	23.41	1,372	2,427
регистрация	22.87	1,598	2,347
борт	20.23	1,509	1,944
питание	19.77	1,542	1,836
пилот	17.04	1,374	1,620
обратный рейс	16.96	155	171
пицца	16.86	1,296	1,599
отзыв	16.68	1,268	1,572
салон самолета	16.01	170	178
взлёт	15.77	1,473	1,703
экипаж	15.59	1,195	1,472
перелёт	15.46	1,355	1,778
стоимость	15.28	1,111	1,416
дорога	14.97	1,107	1,333
представитель	14.32	1,006	1,451
сотрудник	13.68	969	1,367
салон	13.64	1 074	1 299

Record 13 of 4,800

1 Record	1 Frequency	T Comment	T URL
5	3	.... Я довольна, <u>ребёнок</u> обожает самолёты после перелета	https://www.airlines-inform.ru
13	1	...впервые... <u>Ребенка</u> подруги извиняюсь за выражение...	https://www.airlines-inform.ru
26	2	...была женщина с двумя <u>детьми</u> . Никого больше в...	https://www.airlines-inform.ru
31	4	...,летела с маленьким <u>ребенком</u> одна ,очень переживала...	https://www.airlines-inform.ru
37	1	...облила меня, спящего <u>ребёнка</u> и мужа колой.	https://www.airlines-inform.ru
41	1	...во время полёта каждому <u>ребёнку</u> давали подарки,самолк	https://www.airlines-inform.ru
42	5	...воды, ни еды <u>детям</u> не предоставили, заметьте...	https://www.airlines-inform.ru
45	1	...персонал, крутые подарки <u>детям</u> , нисколько не жалею...	https://www.airlines-inform.ru

# ИЗВЛЕЧЕНИЕ ИНФОРМАЦИИ: ПОДХОДЫ

Статистический (ML)	Лингвистический (Rule-based)	Гибридный
<b>Размеченный корпус</b>		
Да	Нет	Нет
<b>Коррекция ошибок</b>		
Сложная	Легкая	Легкая
<b>Изменение постановки задачи</b>		
Сложное	Легкое	Легкое
<b>Качество решения</b>		
~0.8-0.85%	~0.9-0.95	~0.9-0.95
<b>Трудоемкость</b>		
Высокая	Высокая	Высокая
<b>Квалификация</b>		
Низкая	Высокая	Низкая



Текстовый анализ

# Лингвистические алгоритмы

# ОПРЕДЕЛЕНИЕ ЯЗЫКА

POR

## португальский

A história russa inicia-se com os eslavos do leste, que surgiram como um grupo étnico...

ES

## испанский

La historia rusa se inicia con los eslavos del este, que surgieron como un grupo étnico...

BL

## белорусский

Масква — горад у цэнтры еўрапейскай часткі Расіі, сталіца краіны.

UKR

## украинский

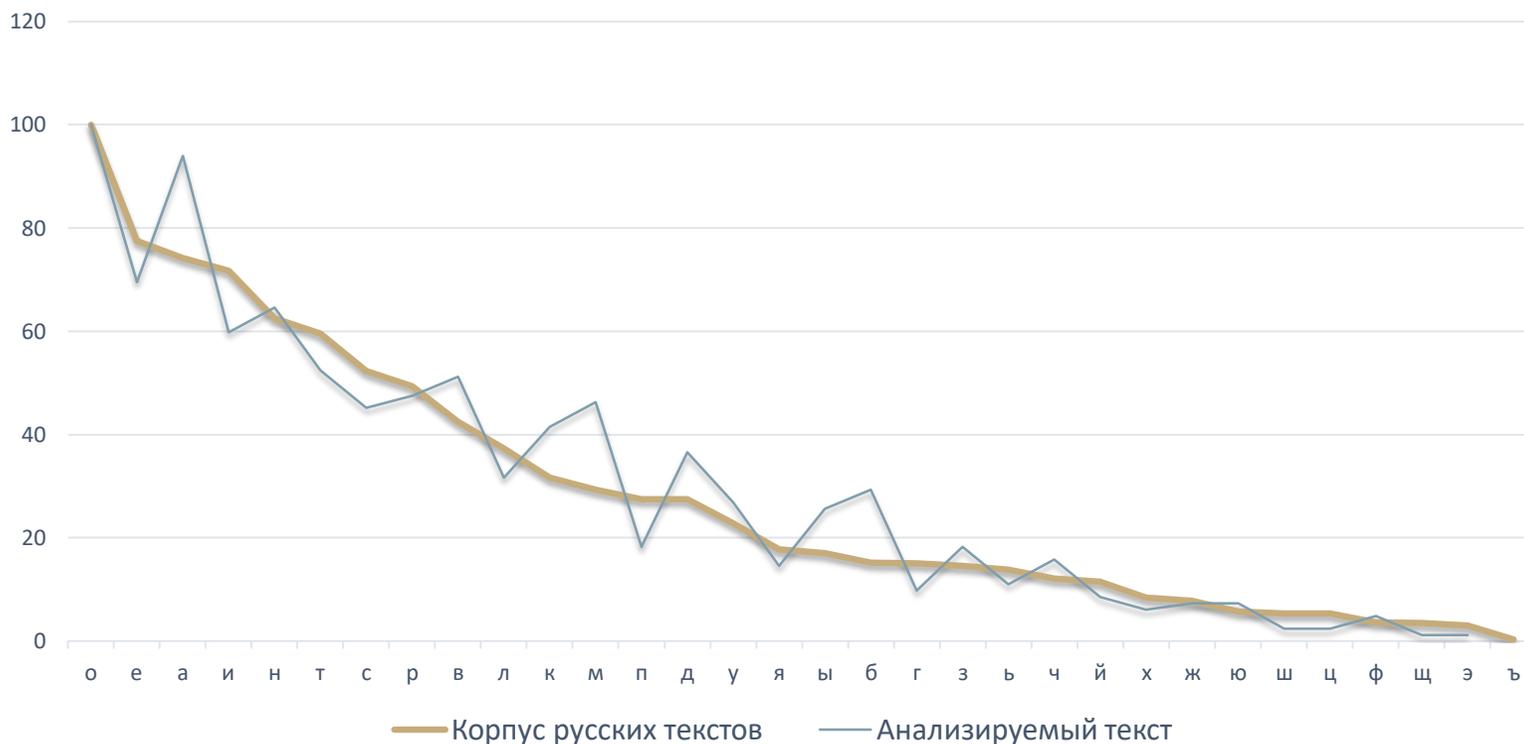
Москва - місто в центрі європейської частини Росії, столиця країни.

# ОПРЕДЕЛЕНИЕ ЯЗЫКА

МОСКВА, 23 марта. /ИТАР-ТАСС/. Двумя матчами в Химках и Грозном в воскресенье продолжится 22-й тур чемпионата России по футболу. Фаворитами являются хозяева - московские динамовцы продолжают борьбу за чемпионство, тогда как "Терек" в случае победы над "Кубанью" сможет выбраться из зоны вылета.

В 13:30 мск на стадионе "Арена-Химки" начнется матч между "Динамо" и "Рубином". Еще в прошлом сезоне обе команды боролись за самые высокие места, однако в 2014 году, после смены тренера Курбана Бердыева на Рината Билялетдинова, волжане фактически лишены турнирных задач: борьба за выживание им вряд ли грозит, тогда как до еврокубковой зоны - 10 очков. Билялетдинов, по сути, наигрывает новую команду, в которой есть место как звездным новичкам вроде форварда сборной Украины Марко Девича, так и совсем юным дебютантам, как 19-летний Эльмир Набиуллин. Уверенная победа над "Волгой" в последнем туре со счетом 3:1 показала, что казанцы на верном пути, и они вполне способны дать бой "бело-голубым".

## Частотное распределение букв (юниграммы)



# ОПРЕДЕЛЕНИЕ ЯЗЫКА

Language Detection Properties	
Select columns	Settings
Tags	General
Language	Append
Arabic	<input checked="" type="checkbox"/>
Chinese	<input checked="" type="checkbox"/>
Chinese-Traditional	<input checked="" type="checkbox"/>
Czech	<input checked="" type="checkbox"/>
Danish	<input checked="" type="checkbox"/>
Dutch	<input checked="" type="checkbox"/>
English	<input checked="" type="checkbox"/>
Finnish	<input checked="" type="checkbox"/>
French	<input checked="" type="checkbox"/>
German	<input checked="" type="checkbox"/>
Greek	<input checked="" type="checkbox"/>
Italian	<input checked="" type="checkbox"/>
Japanese	<input checked="" type="checkbox"/>
Korean	<input checked="" type="checkbox"/>
Norwegian	<input checked="" type="checkbox"/>
Polish	<input checked="" type="checkbox"/>
Portuguese	<input checked="" type="checkbox"/>
Romanian	<input checked="" type="checkbox"/>
Russian	<input checked="" type="checkbox"/>
Slovak	<input checked="" type="checkbox"/>
Slovenian	<input checked="" type="checkbox"/>
Spanish	<input checked="" type="checkbox"/>
Swedish	<input checked="" type="checkbox"/>
Thai	<input checked="" type="checkbox"/>

Banca diz que casas estão mais caras desde dezembro

O valor médio da avaliação bancária na habitação subiu 9%, para 1100 euros por metro quadrado, segundo o Institut

T Text	Text_Language
A organização do Global Investment Challenge (GIC) entregou no início deste mês os prémios aos vencedores da quinta	Portuguese
Banca diz que casas estão mais caras desde dezembro. O valor médio da avaliação bancária na habitação subiu 9%, pa	Portuguese

(东莞韶关讯) 广东一名自闭症少年去年在深圳走失, 被送往韶关新丰练溪托养中心一个多月后死亡, 媒体调查发现, 称已严厉查处该中心, 并对涉嫌违纪违规的公职人员开展调查。

官方通报称, 已对该托养中心负责经营管理的四名主要责任人采取强制措施, 正在对托养对象的死亡人数和死亡原因组

T Text	Text_Language
(联合早报网讯)最新调查显示, 美韩在韩国境内部署萨德反导系统的决定所引发的韩中外交争端, 使得中国取	Chinese
(东莞韶关讯)广东一名自闭症少年去年在深圳走失, 被送往韶关新丰练溪托养中心一个多月后死亡, 媒体调查	Chinese

Jeudi durant la nuit, la nébulosité augmentera à partir de l'ouest, mais le temps restera sec, sous un vent modéré de si degrés dans les autres régions.

Vendredi, il fera très nuageux avec de la pluie ou des averses dans le courant de la journée. Les maxima oscilleront au

T Text	Text_Language
[jeudi 3 novembre 2016 20h14] Les rhétos sont nombreux dans les auditorios universitaires pour les cours ouverts Per	French
[jeudi 3 novembre 2016 08h40] Des périodes ensoleillées alterneront avec des champs nuageux ce jeudi après la dissip	French

## ЦИФРО-БУКВЕННЫЕ ПОСЛЕДОВАТЕЛЬНОСТИ

нет границы слова

свой собственный **3D** кинотеатр  
автоинспекторы применяют «Визир **2М**»

есть граница слова

масса по сертификату составляет **2495кг**  
состав грунта на глубине до **2м**

## ДЕФИС

нет границы слова

Цены на бензин **АИ-92**  
Большегруз перекрыл трассу **М-7**

есть граница слова

ограничение скорости в 20 км/ч вместо **60-80**.  
на глубине **2-3м**

## ЗНАКИ ПУНКТУАЦИИ

есть граница  
предложения

Результаты исследования удивили специалистов.  
Как вы оцениваете перспективы проекта?

нет границы  
предложения

... выступит **к.т.н. Иванов Д.Н.**  
«**М.Видео**» покупает «Эльдорадо» за 45,5 млрд рублей  
тариф «Включайся!» стал еще доступнее



По результатам этих исследований, оценка Японии упала с 3.56 до 3.33 баллов.

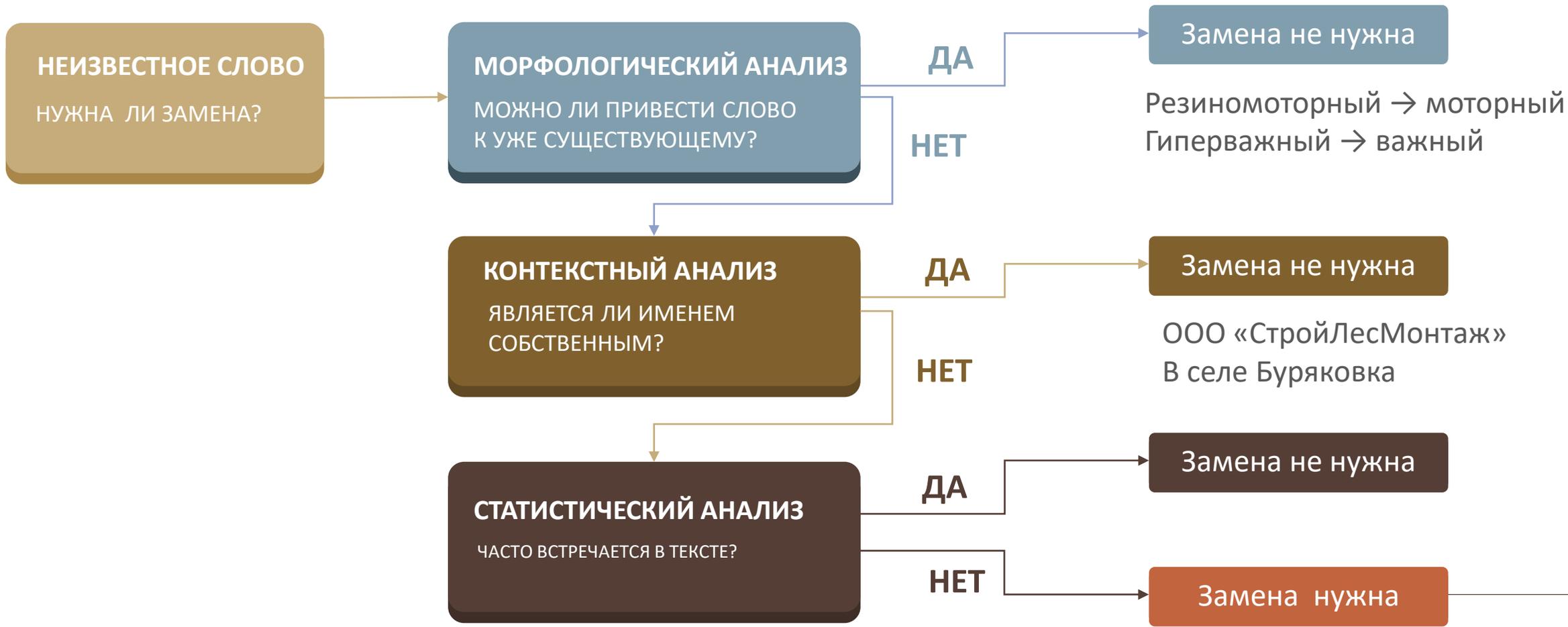
Исходный текст

日本在此次调查中的分数也从3.56下降至3.33。

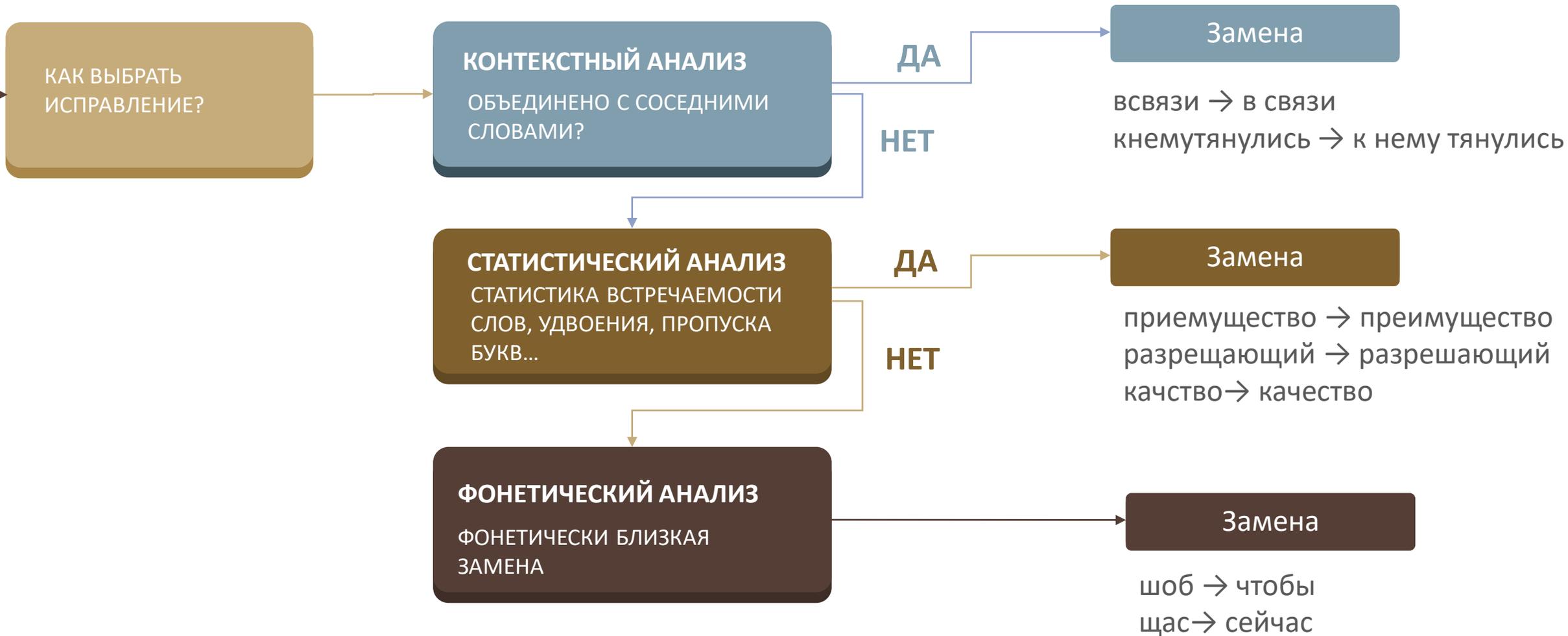
Токенизированный текст

日本在此次调查中的分数也从3.56下降至3.33。

# КОРРЕКЦИЯ ОРФОГРАФИЧЕСКИХ ОШИБОК



# КОРРЕКЦИЯ ОРФОГРАФИЧЕСКИХ ОШИБОК



«За песчаной косой лопаухий косой пал под острой косой косой бабы с косой.»

What is that nonsense?..



# ОПРЕДЕЛЕНИЕ ЧАСТЕЙ РЕЧИ



«За песчаной **косой** лопухий **косой** пал под острой **косой** **косой** бабы с **косой**»



# ОПРЕДЕЛЕНИЕ ЧАСТЕЙ РЕЧИ

Закрыв измученные веки,

Миг отошедший берегу.

О если б так стоять вовеки

На этом тихом берегу.

беречь

берег



\*Исследование неоднозначности употребления слов в европейских языках /  
Э.С.Клышинский // Препринты ИПМ им. М.В.Келдыша. 2015.

# ПАРАДИГМА СЛОВОИЗМЕНЕНИЯ

Словоформа	Унцией		Штормов	
Лемма	[Унци]я		[Шторм]	
Парадигма	-я	-и	«-»	-ы
	-и	-ий	-а	-ов
	-и	-ям	-у	-ам
	-ю	-ии	«-»	-ы
	-ей	-ями	-ом	-ами
	-и	-ях	-е	-ах

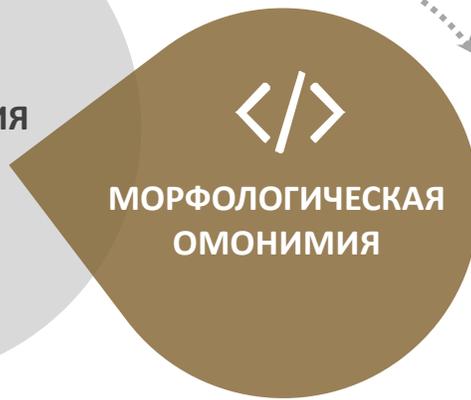
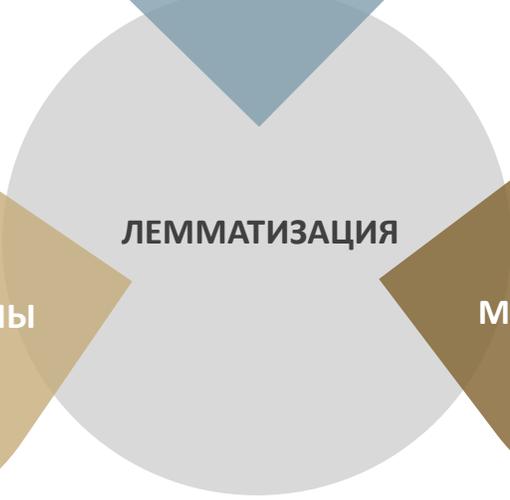
Словоформа	Выпадал	
Лемма	[Выпад]ать	
Парадигма	-аю	-ем
	-ай	-айте
	-аешь	-аете
	-ает	-ают
	-ает	-али
	-ает	

# ЛЕММАТИЗАЦИЯ

ЧЕЛОВЕК	
ЧЕЛОВЕК	<b>людям</b>
<b>люди</b>	ЧЕЛОВЕКА
ЧЕЛОВЕКА	<b>людей</b>
<b>людей</b>	ЧЕЛОВЕКУ
ЧЕЛОВЕКУ	...



КАРТА	
КАРТА	КАРТАМ
КАРТЫ	КАРТУ
КАРТЫ	КАРТЫ
КАРТ	КАРТОЙ
КАРТЕ	...



ЖЕЧЬ (ГЛАГОЛ)  
зачем **жгут** траву весной?

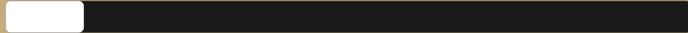
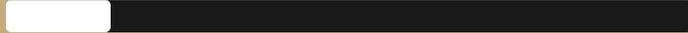
ЖГУТ (СУЩЕСТВИТЕЛЬНОЕ)  
СЛЕДУЕТ НАЛОЖИТЬ  
МЕДИЦИНСКИЙ **ЖГУТ** ИЛИ ПОВЯЗКУ

Что делать,  
если слово  
неизвестно?

Он и сам не **выланда**л на ногах и покорновался по размежному **плою** горы.

выланда

плою

Выланда (глагол)	= страдать	90%
		
Выланда (существительное)	= сандал, вандал	10%
		
Плоить (глагол)	= строю, ною	15%
		
Плой (существительное)	= строй, рой	85%
		

# АНАЛИЗ НЕИЗВЕСТНЫХ СЛОВ

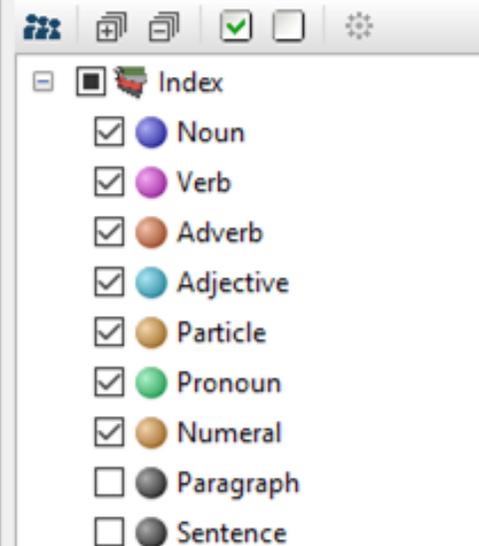
Заглотитель Ланс Оливер **чуть не погиб** в результате напличения турма. Он **ехал ласкунно** на лошади **покровнательно** от Мэнсфилда (Австралия) и **увидел вахню турмов**, в которой было **кастожно 15** животных. Столенно, **ничего бы и не случилось**, если бы собака Оливера **не начала порочить** на вахню.

Один из турмов – **старый, крупный лователь**, **выбатушенный** корочением собаки, **бросился за ней**. Та **отпешила** **сумановаться** за лошадью, на которой сидел Оливер. Тогда турм **бросился уже на Оливера**. Он **схватил** **подвешенца** **отмалснными** твинами за плечи и **вытокнул его** на землю.

Оливер **вскочил на ноги** и **схватил турма**, **пытаясь подочевать ему** **обкап-лужиться** на хвост и **таким образом** **принять доланское** положение. Турм **подстоголками** **укивовых лап** в **ропли** **раскелепил** на заглотителе **лахту** и **толстый тарат**. **Наконец**, Оливеру **удалось** **зайти сзади турма** и **поклочить его**. Однако он и сам **не выландал на ногах** и **вместе с турмом** **покорновался** по **размежному** **плюю горы** **кастожно 35** метров и **упал в сорот**. Шатировка **продолжалась** и в сороте, пока Оливер **не слеменился** **схватить камень** и **обвинить** **несколько сильных ударов** по голове турма. Турм **был помотрен**.

**Рустуя кулировку** с Оливером, **мельбурнский падакователь** **Кэвенег** **заявил**, что турмы, **особенно старые лователи**, **напличивают** на человека, **когда они**, как **говорится**, **доспонарены** к стене и **им некуда сополиться**.

**Я не помню**, **сказал Кэвенег**, **чтобы такие** **напличения** **скамничались** **торцией** человека, **но имеется много случаев**, **когда турм** **обвичивает** человеку **серподины**.

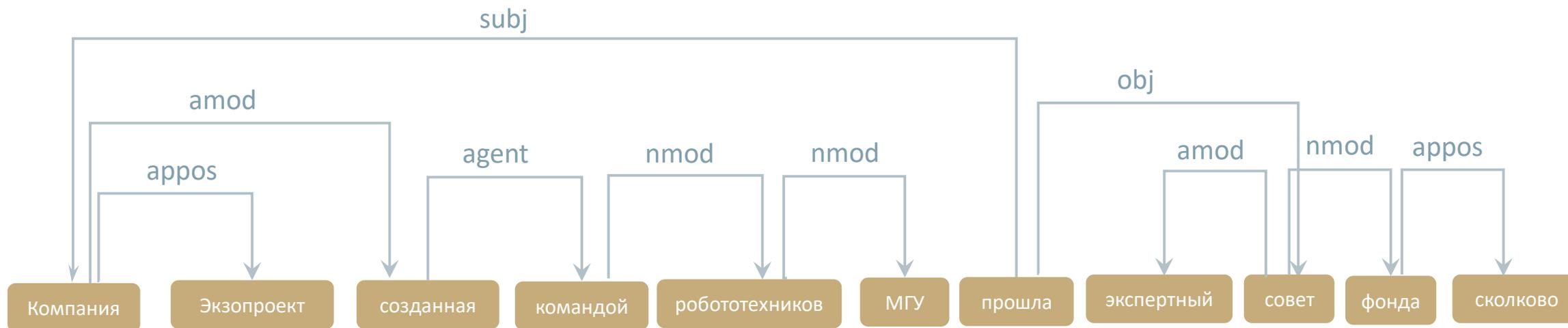


Index

- Noun
- Verb
- Adverb
- Adjective
- Particle
- Pronoun
- Numeral
- Paragraph
- Sentence

Пример заимствован из  
[Фрумкина Р.М. Психоллингвистика]

# СИНТАКСИЧЕСКИЙ АНАЛИЗ



Полный синтаксический разбор

Компания «ЭкзоПроект», созданная командой робототехников МГУ, прошла экспертный совет Фонда «Сколково»

Неполный синтаксический разбор

Компания «Экзопроект»,

созданная

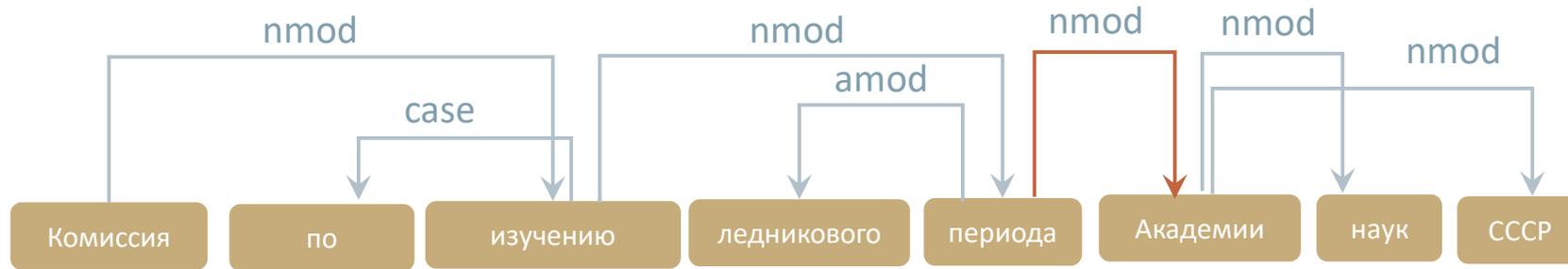
командой робототехников МГУ,

прошла

экспертный совет

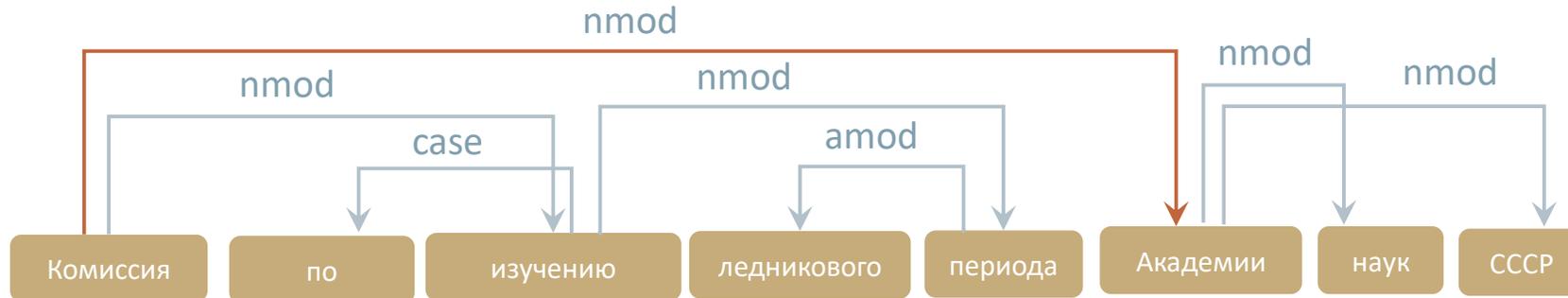
фонда «Сколково».

# СИНТАКСИЧЕСКАЯ МНОГОЗНАЧНОСТЬ

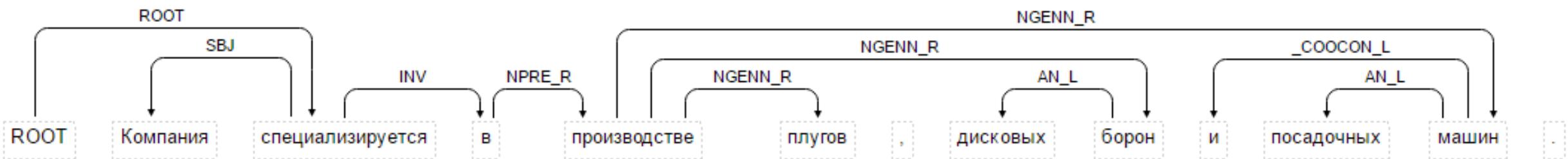
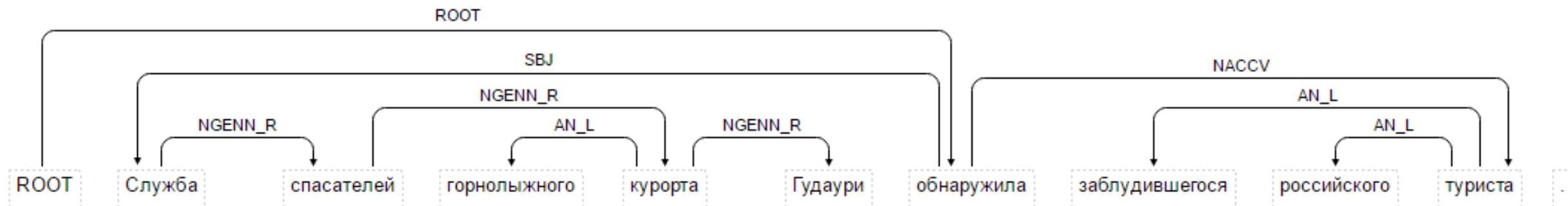


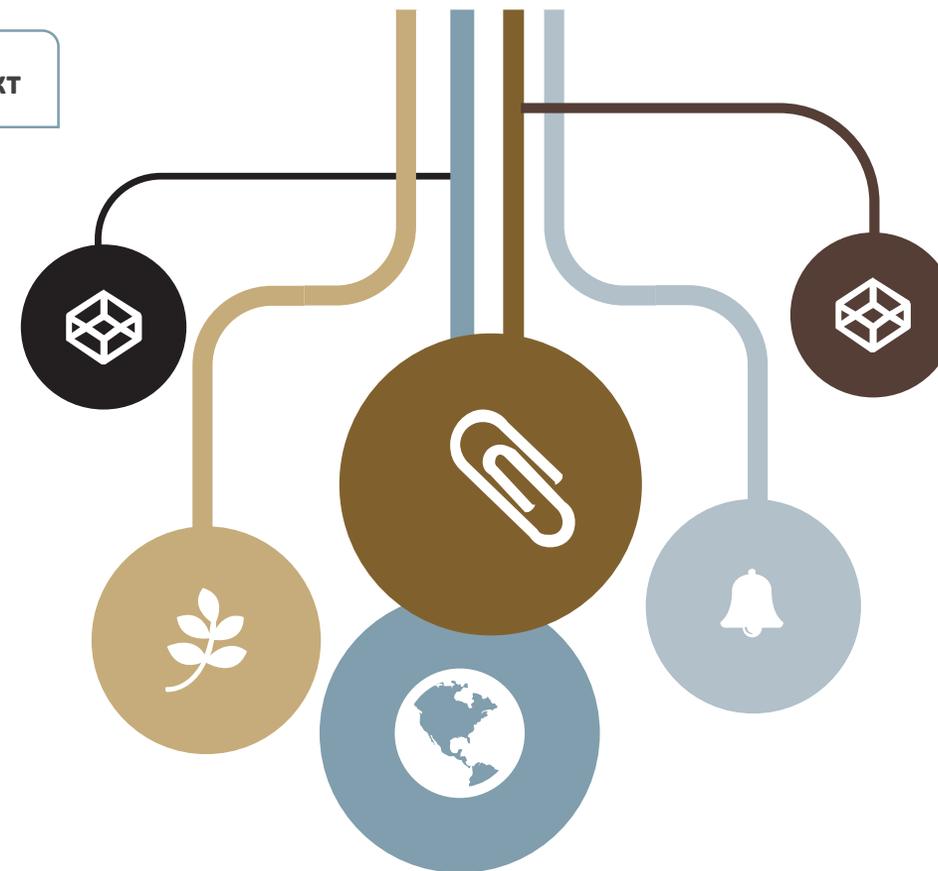
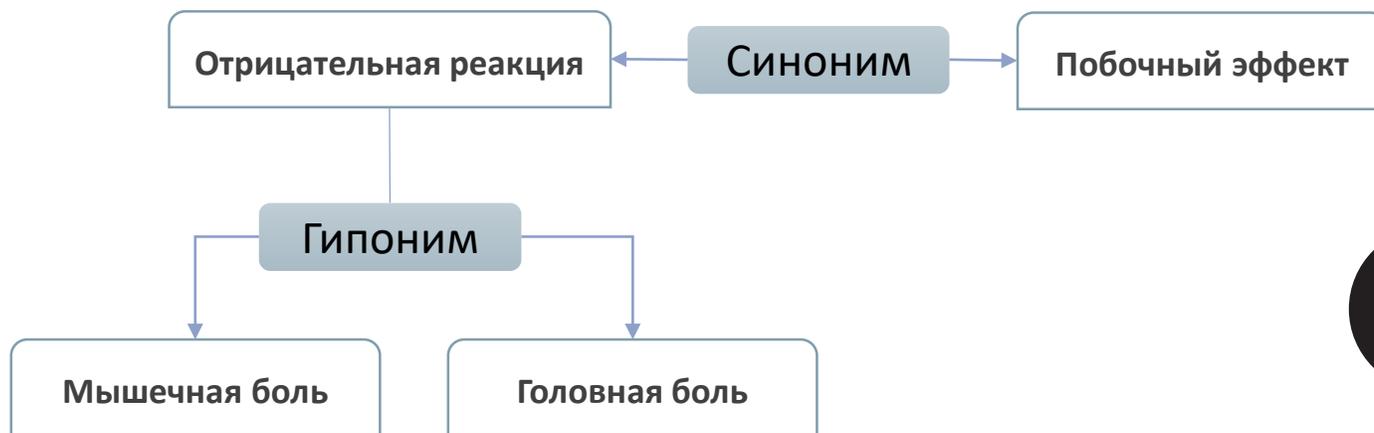
Вариант синтаксического разбора

Комиссия по изучению Ледникового периода Академии наук СССР



# СИНТАКСИЧЕСКИЙ АНАЛИЗ





Побочный эффект

Головная боль после приема ингибитора фосфодиэстеразы.

Побочный эффект

Пациент жалуется на мышечную боль и усиление боли в шее.

Relations

Synonym	Query
▼ [Industry Entity]	
▶ ↓ Company	
▶ ↓ Component	
▶ ↓ Device	
▼ ↓ Healthcare	
▼ ↓ Drug	
▶ ↓ Branded Drug	
▶ ↓ Effect	
▼ ↓ Substance	
↓ aspergillus	
↓ aspidin	
▼ ↓ Aspirin	
▶ ↓ Aggrenox	
↓ Alka Seltzer	
↓ Alka Seltzer Original	
↓ Alka-seltzer Extra Strength	
↓ Alka-seltzer Plus	
↓ Anacin	
↓ Anacin Af	
↓ Anacin Es	
▶ ↓ Ascomp With Codeine	
↓ Ascriptin Maximum Strength ...	
↓ Ascriptin Regular Strength Bu...	

Show Relations

<input checked="" type="checkbox"/> ↓ develops	<input type="checkbox"/> ↑ developed by
<input checked="" type="checkbox"/> ↓ has brand	<input type="checkbox"/> ↑ is brand
<input checked="" type="checkbox"/> ↓ has category	<input type="checkbox"/> ↑ is category
<input checked="" type="checkbox"/> ↓ has effect	<input checked="" type="checkbox"/> ↑ is effect
<input checked="" type="checkbox"/> ↓ has instance	<input type="checkbox"/> ↑ is instance
<input checked="" type="checkbox"/> ↓ has name	<input type="checkbox"/> ↑ is name
<input checked="" type="checkbox"/> ↓ has os	<input type="checkbox"/> ↑ is os
<input checked="" type="checkbox"/> ↓ has synonym	<input type="checkbox"/> ↑ is synonym
<input checked="" type="checkbox"/> ↓ Hyponym	<input type="checkbox"/> ↑ Hypernym
<input checked="" type="checkbox"/> ↓ is aspect	<input type="checkbox"/> ↑ has aspect

Part of speech filter

Noun Verb Adjective Adverb

# ОТРИЦАНИЯ И МОДАЛЬНОСТЬ

Мне **понравилось**  
обслуживание

утверждение

Мне **не понравилось**  
обслуживание.

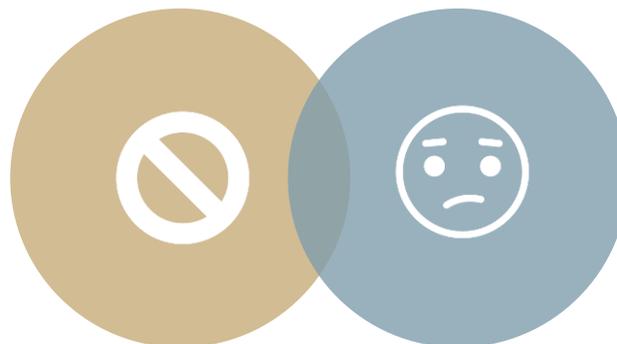
отрицание

В аптеке **было** нужное  
лекарство.

утверждение

В аптеке **не хватает**  
нужных лекарств.

отрицание



факт

В банке **работают**  
компетентные сотрудники.

гипотеза

**Думал, что** в банке  
**работают** компетентные  
сотрудники.

факт

Раньше в банке  
**работали** компетентные  
сотрудники.

гипотеза

**Если бы** в банке  
**работали** компетентные  
сотрудники...

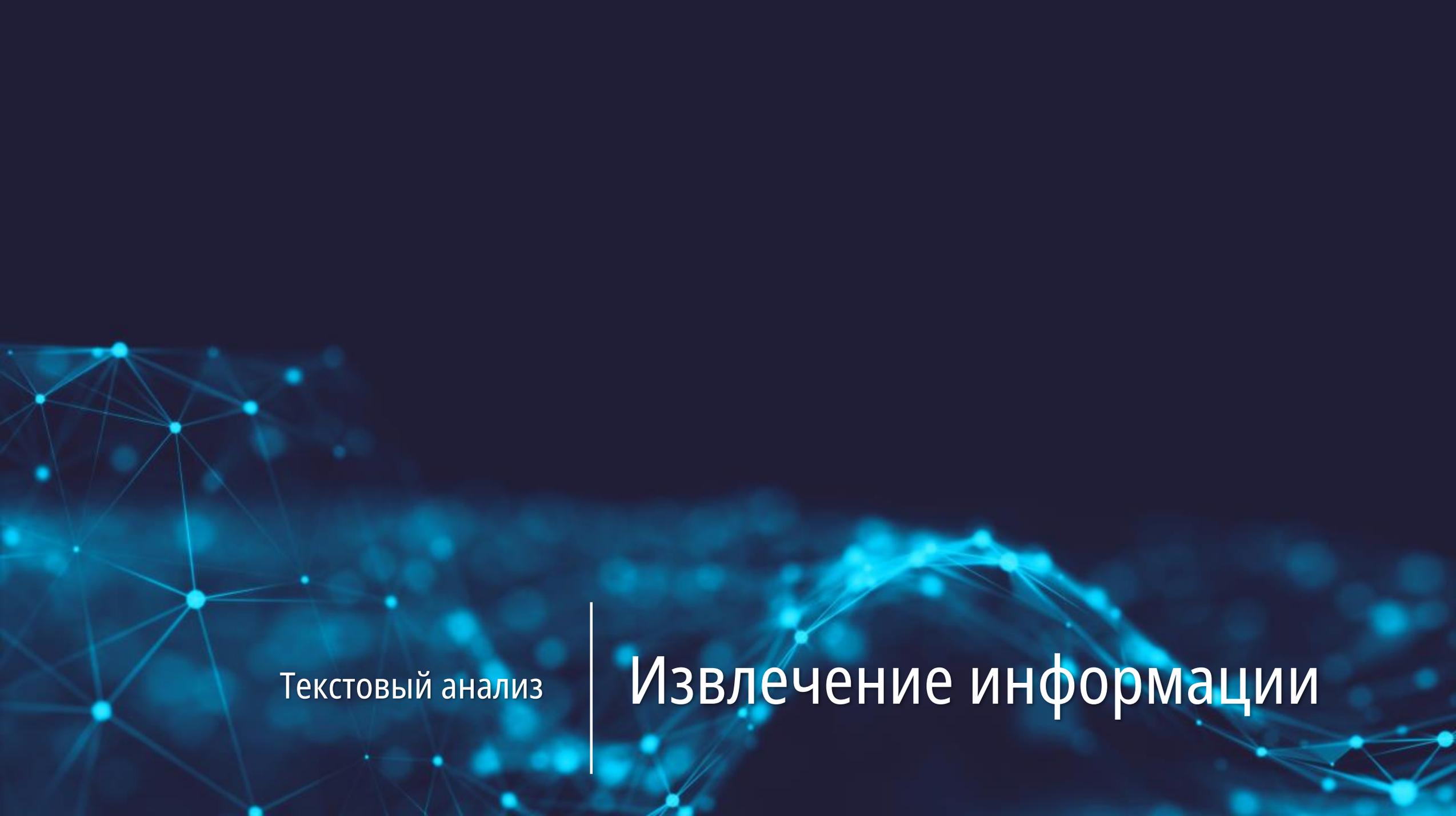
# АНАЛИЗ ТОНАЛЬНОСТИ

Со всех сторон хороший отель! Номера небольшие, но хорошие. Если относится к номеру как к месту для сна в поездке/путешествии, все в порядке. Без излишеств и пафоса. Если хотите, чтобы было фешенебельно, ищите другой отель! Очень удачное местоположение, все близко или в 5-15 минутах езды на такси. бесплатный Wi-Fi, внизу можно взять кофе с собой.

- ▼  **Standard (5)**
- ▼  **Positive (4)**
  -  бесплатный Wi-Fi (1)
  -  Номера небольшие, но хорошие (1)
  -  Очень удачное местоположение (1)
  -  хороший отель (1)
- ▼  **Negative (1)**
  -  Номера небольшие (1)

Отличное соотношение цена/качество. Отель неплохой. Персонал очень приветливый, номера чистые. Не понравилась парковка (маленькая и трудно доступная как для проезда, так и далеко от входа в гостиницу - с вещами очень не удобно идти).

- ▼  **Standard (5)**
- ▼  **Positive (4)**
  -  номера чистые (1)
  -  Отель неплохой (1)
  -  Отличное соотношение цена качество
  -  Персонал приветливый (1)
- ▼  **Negative (1)**
  -  Не понравилась парковка (1)



Текстовый анализ

Извлечение информации

## Морфологический анализ

В Майкопе 13 апреля произошёл пожар на заводе «Картон» .

предлог

существ

числит

существ

глагол

существ

предлог

существ

пункт

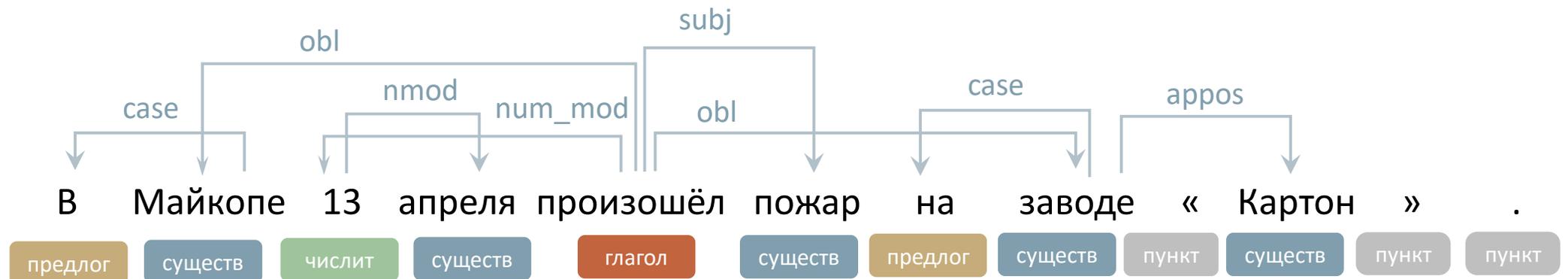
существ

пункт

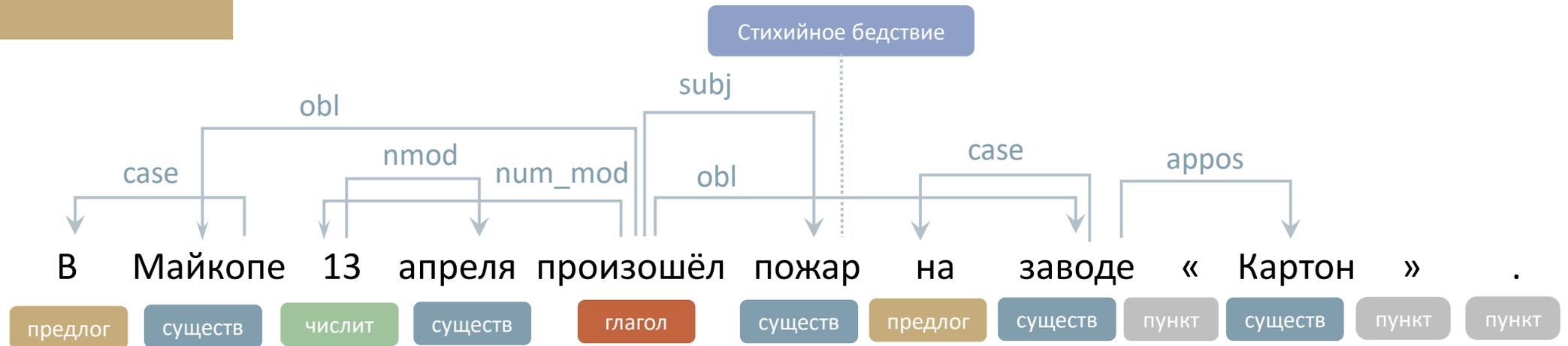
пункт

Синтаксический анализ

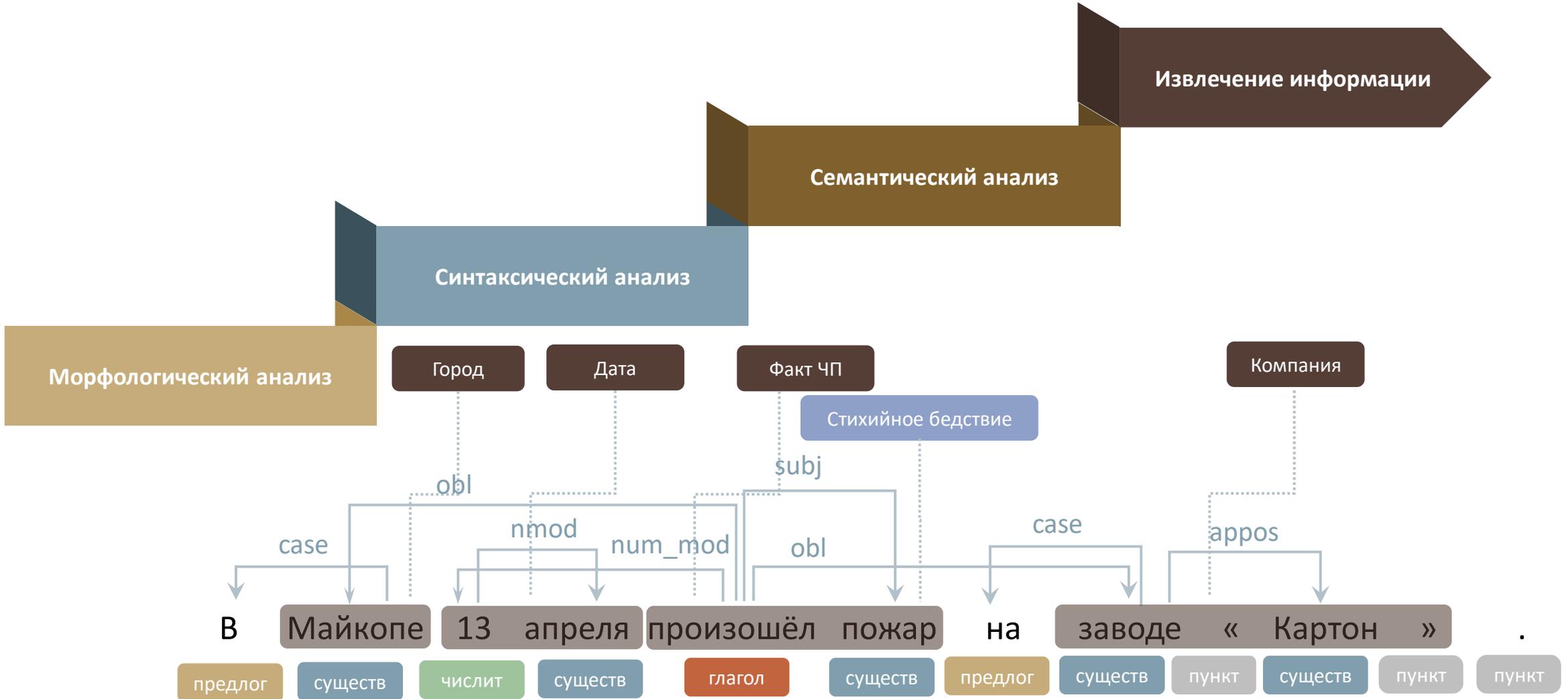
Морфологический анализ



# ИЗВЛЕЧЕНИЕ ИНФОРМАЦИИ



# ИЗВЛЕЧЕНИЕ ИНФОРМАЦИИ



## 01 ИМЕНОВАННЫЕ СУЩНОСТИ

Имена людей

Названия компаний, организаций

Географические объекты

Даты

...

Первым космическим туристом, побывавшим на МКС, стал **американский бизнесмен итальянского происхождения Деннис Тито**, который отправился на орбиту **28 апреля 2001 года**.

## 02 ФАКТЫ И ОТНОШЕНИЯ

Смена менеджмента

Банкротства

Слияния и поглощения

Партнерские соглашения

Совместные предприятия

...

Компании **Singulato Motors** и **Beijing Automotive Group (BAIC)** заключили партнерское соглашение по разработке и запуску в серийное производство нового электрокара.

## 03 ТОНАЛЬНОСТИ

Оценочные и эмоциональные высказывания

Отзывы клиентов

Был в гостинице с 30-3 июня. Из хорошего: **чисто, хорошая кровать, удобно расположен телевизор, хорошие завтраки**. Из плохого: **нет питьевой воды** вообще.

## Подходы к решению задачи

- ✓ Правила
- ✓ Машинное обучение
- ✓ Гибридные подходы

Дата

Персона

9 декабря министр финансов США Джейкоб Лью

Организация

сообщил, что правительство США продало частным

Компания

инвесторам свой пакет акций General Motors,

которой была оказана государственная поддержка

Сумма

на сумму \$49,5 млрд.

# ИЗВЛЕЧЕНИЕ СУЩНОСТЕЙ

## Стандартные

Более 30 типов стандартных сущностей (имена людей, названия компаний, географических объектов, адреса, номера телефонов и т.д.)

## Пользовательские

Возможность добавления произвольных сущностей (страховые инциденты, лекарственные препараты и т.д.)

Договор подряда на выполнение работ по строительству № 03Т11-78  
 г. **Москва**  
 06.04.2018

**Общество с ограниченной ответственностью "СТРОИТЕЛЬНАЯ ХИМИЯ"** в лице генерального директора **Максима Викторовича Черевковского**, действующего на основании Устава, в дальнейшем именуемый "Подрядчик", с одной стороны, и **ООО "ИНТЕРТЕХ"** в лице **директора Степана Осиповича Макарова**, действующего на основании Устава, в дальнейшем именуемый "Заказчик", с другой стороны, заключили настоящий договор о нижеследующем:

### 1. ПРЕДМЕТ ДОГОВОРА

1.1. В соответствии с настоящим Договором, Подрядчик обязуется выполнить строительно-монтажные работы по реконструкции объекта недвижимости, расположенного по адресу: **г. Москва, ул. Строителей, д. 15** (далее - "Работы") в полном объеме по Техническому заданию Заказчика (Приложение №1 к настоящему Договору) в установленный настоящим Договором срок собственными силами. Подрядными организациями по утвержденным Заказчиком Графикам выполнения работ.  
 1.2. Технические, экономические и другие требования к Работам, являющимся предметом настоящего Договора, должны соответствовать требованиям СНиПам, ГОСТам и другим нормативным актам **Российской Федерации**.

### 2. СРОК ДЕЙСТВИЯ ДОГОВОРА, ВЫПОЛНЕНИЯ РАБОТ

2.1. Сроки начала и окончания, промежуточные сроки в определяются Графиком выполнения работ по строите. объекта недвижимости (Приложение № 2 к настоящему

### 3. СТОИМОСТЬ РАБОТ И ПОРЯДОК РАСЧЕТОВ

3.1. Стоимость Работ по настоящему Договору составляет **девять миллионов четыреста восемьдесят две тысячи двести** копеек, в том числе НДС 18% - **12 124 377** (Двенадцать и **четыре тысячи триста семьдесят семь) рублей 22 копейки.**

3.2. Стоимость работ, указанная в п. 3.1. настоящего Договора, включает в себя стоимость: строительных, монтажных, ремонтных, транспортно-заготовительских расходов, специальных, пуско-наладочных работ, временных зданий и сооружений, зимнего удорожания, резерва средств на непредвиденные работы и затраты, все

## Обогащение

данных из локальных и внешних источников, а также нормализация извлеченных сущностей

Standard/Locations/GeoAdministrative
Name: <i>Россия</i>
Category: <i>Country</i>
Country: <i>Россия</i>
WorldPart: <i>Европа</i>
Area: <i>17 102 345,00</i>
Population: <i>145 452 581</i>

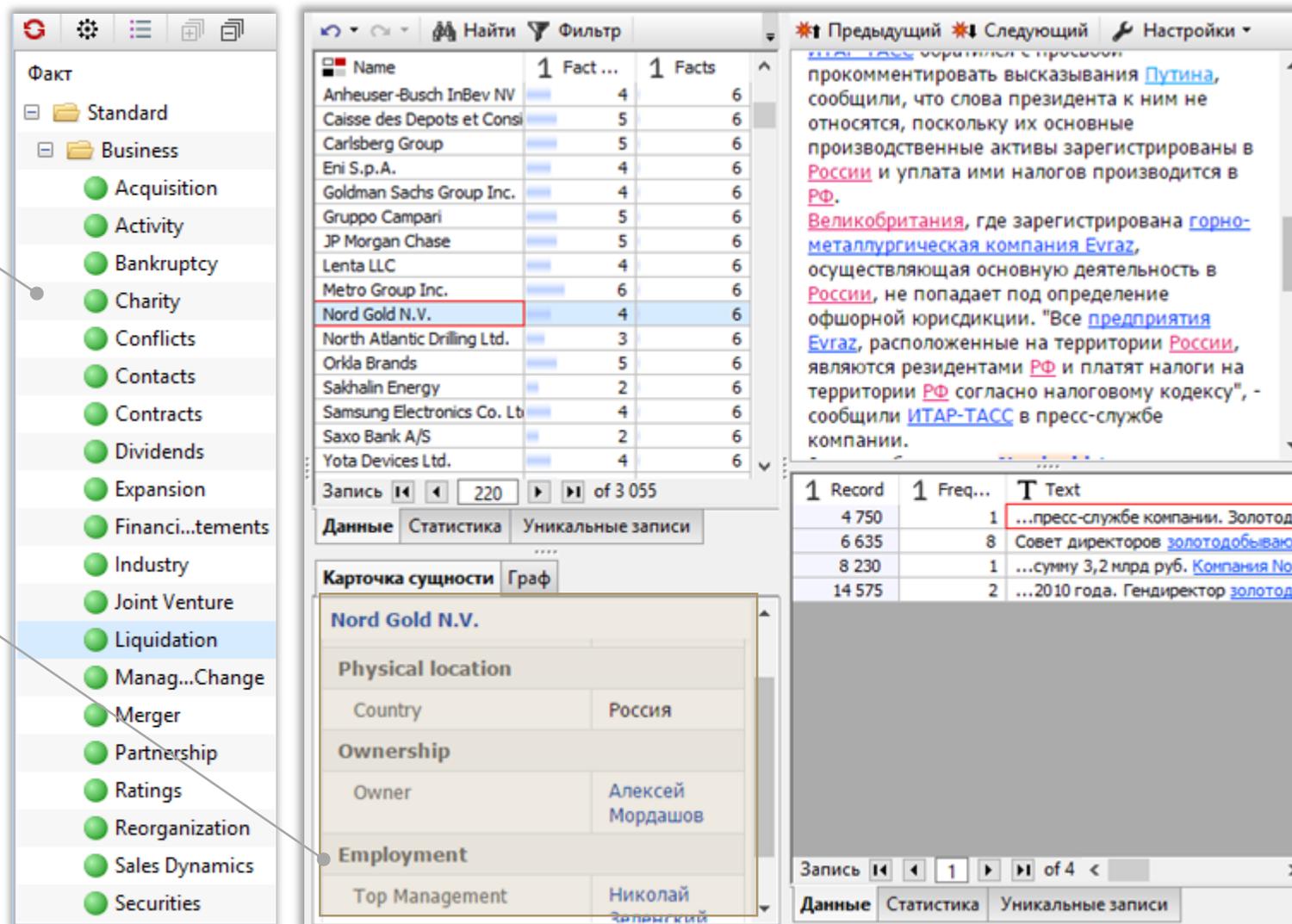
# ИЗВЛЕЧЕНИЕ ФАКТОВ

## Стандартные

Более 40 типов стандартных фактов (назначения, сделки, уголовные дела, банкротства и т.д.)

## Карточка сущности

Информация о человеке, компании или организации, найденная в документе. Все сведения сопровождаются фрагментами документов.



The screenshot displays the Megaputer software interface. On the left, a sidebar lists various fact types under 'Факт' (Fact), including 'Standard' and 'Business' categories. The 'Business' category is expanded, showing a list of fact types such as Acquisition, Activity, Bankruptcy, Charity, Conflicts, Contacts, Contracts, Dividends, Expansion, Financi...tements, Industry, Joint Venture, Liquidation, Manag...Change, Merger, Partnership, Ratings, Reorganization, Sales Dynamics, and Securities. The 'Liquidation' type is currently selected.

The main window shows a table of extracted facts. The table has columns for 'Name', 'Fact ...', and 'Facts'. The 'Nord Gold N.V.' entry is highlighted in red. Below the table, there are navigation controls and tabs for 'Данные', 'Статистика', and 'Уникальные записи'. A 'Карточка сущности' (Entity Card) is displayed for 'Nord Gold N.V.', showing details such as 'Physical location' (Country: Россия), 'Ownership' (Owner: Алексей Мордашов), and 'Employment' (Top Management: Николай).

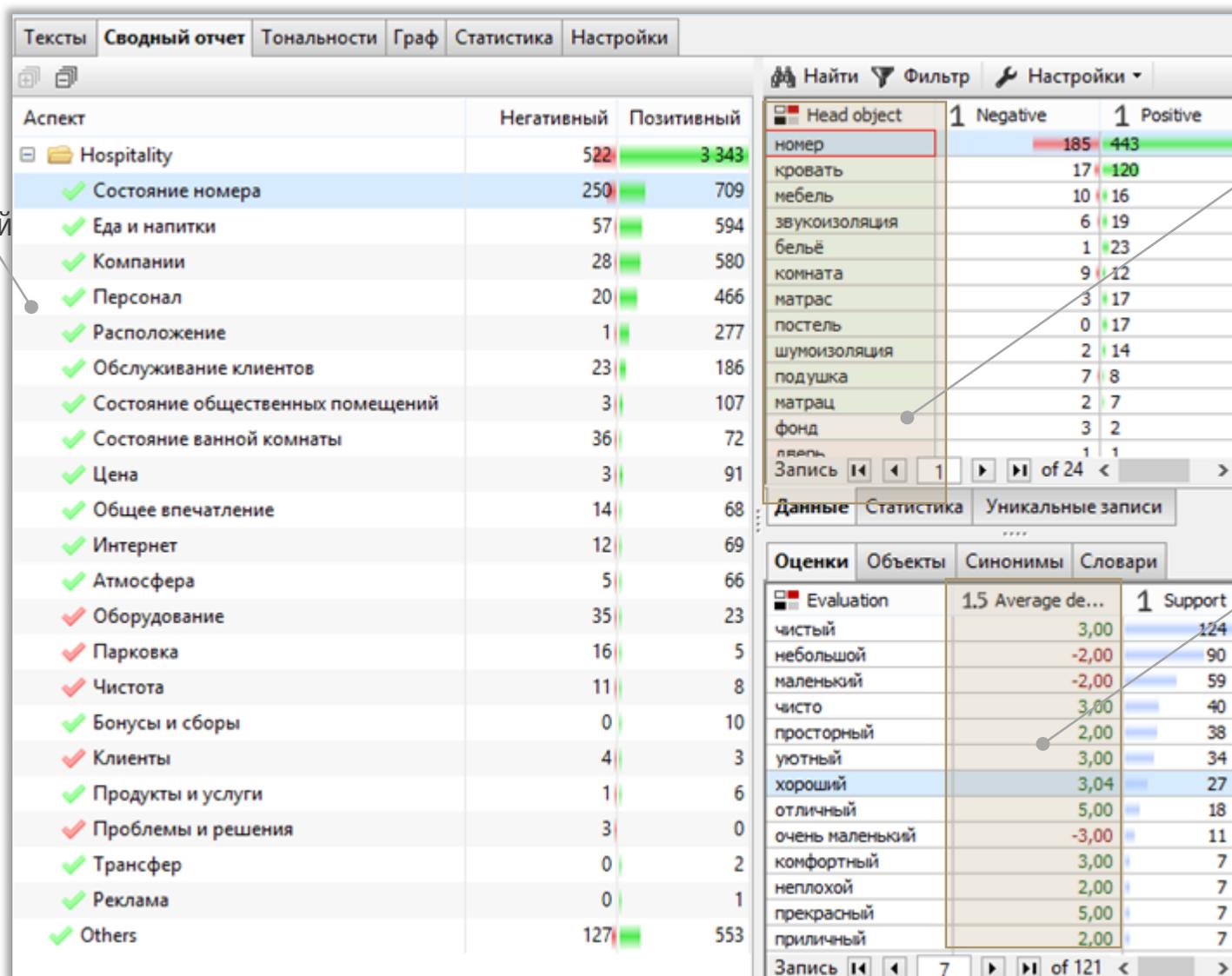
On the right, a text snippet is shown, likely a document fragment containing the extracted fact. It mentions 'ИТАР-ТАСС' and 'пресс-службе компании'. Below the text snippet, there is a table with columns for 'Record', 'Freq...', and 'Text'. The table shows four records, with the first record highlighted in red.

Record	Freq...	Text
4 750	1	...пресс-службе компании. Золотодо
6 635	8	Совет директоров <a href="#">золотодобывающ</a>
8 230	1	...сумму 3,2 млрд руб. <a href="#">Компания Норс</a>
14 575	2	...2010 года. Гендиректор <a href="#">золотодо</a>

# АНАЛИЗ ТОНАЛЬНОСТЕЙ

## Аспекты

отражают специфику конкретной предметной области (технический, финансовый, гостиничный домен и т.д.)



Аспект	Негативный	Позитивный
Hospitality	522	3 343
Состояние номера	250	709
Еда и напитки	57	594
Компании	28	580
Персонал	20	466
Расположение	1	277
Обслуживание клиентов	23	186
Состояние общественных помещений	3	107
Состояние ванной комнаты	36	72
Цена	3	91
Общее впечатление	14	68
Интернет	12	69
Атмосфера	5	66
Оборудование	35	23
Парковка	16	5
Чистота	11	8
Бонусы и сборы	0	10
Клиенты	4	3
Продукты и услуги	1	6
Проблемы и решения	3	0
Трансфер	0	2
Реклама	0	1
Others	127	553

Head object	1 Negative	1 Positive
номер	185	443
кровать	17	120
мебель	10	16
звукоизоляция	6	19
бельё	1	23
комната	9	12
матрас	3	17
постель	0	17
шумоизоляция	2	14
подушка	7	8
матрац	2	7
фонд	3	2
льед	1	1

Оценки	Объекты	Синонимы	Словари
Evaluation	1.5 Average de...	1 Support	
чистый	3,00	124	
небольшой	-2,00	90	
маленький	-2,00	59	
чисто	3,00	40	
просторный	2,00	38	
уютный	3,00	34	
хороший	3,04	27	
отличный	5,00	18	
очень маленький	-3,00	11	
комфортный	3,00	7	
неплохой	2,00	7	
прекрасный	5,00	7	
приличный	2,00	7	

## Объект оценки

Выделение объекта, к которому относится оценка

## Ранжирование

Высказываний по степени эмоциональной окраски



# РАСЧЕТ ЗНАЧИМОСТИ КЛЮЧЕВЫХ ФРАЗ

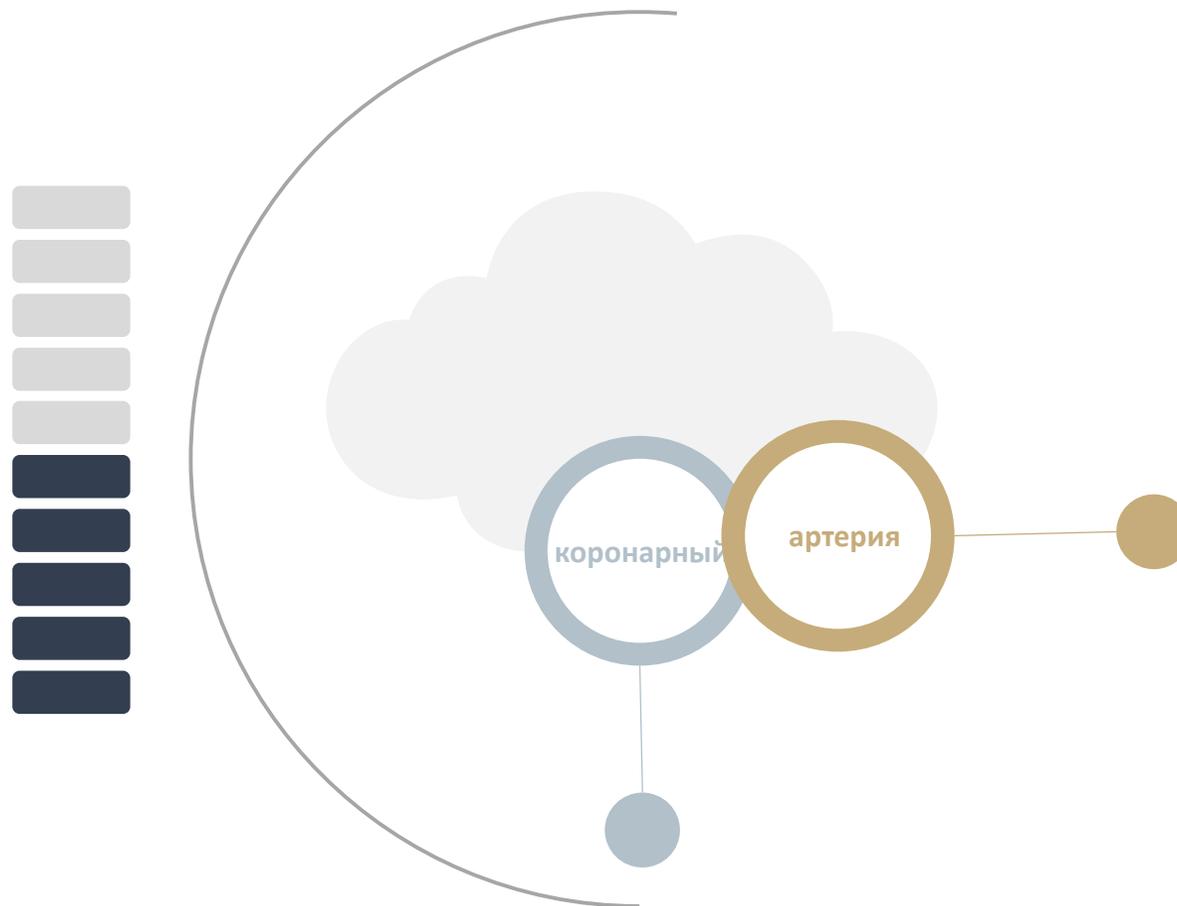
---



# РАСЧЕТ ЗНАЧИМОСТИ КЛЮЧЕВЫХ ФРАЗ

Факторы, влияющие на расчет значимости

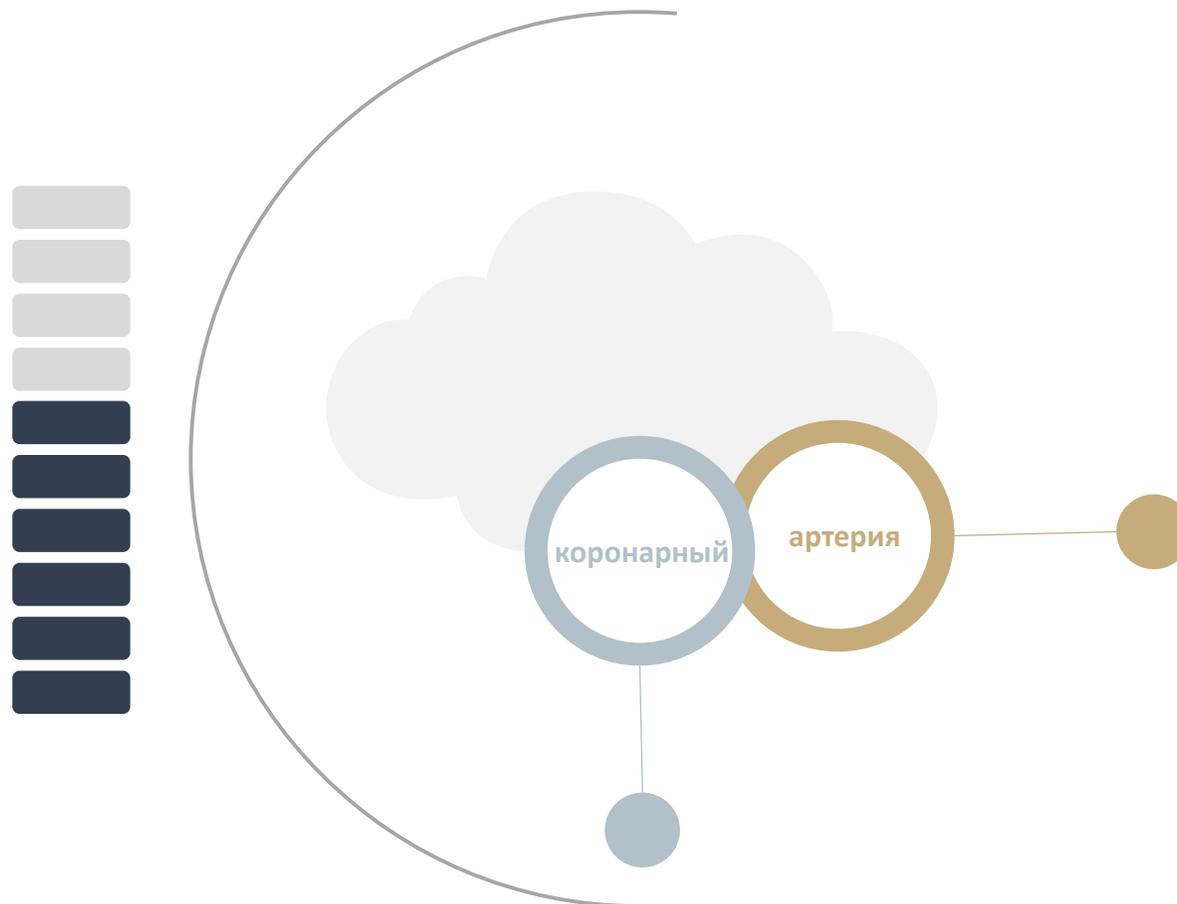
- ✓ Встречаемость компонентов фразы в корпусе



# РАСЧЕТ ЗНАЧИМОСТИ КЛЮЧЕВЫХ ФРАЗ

Факторы, влияющие на расчет значимости

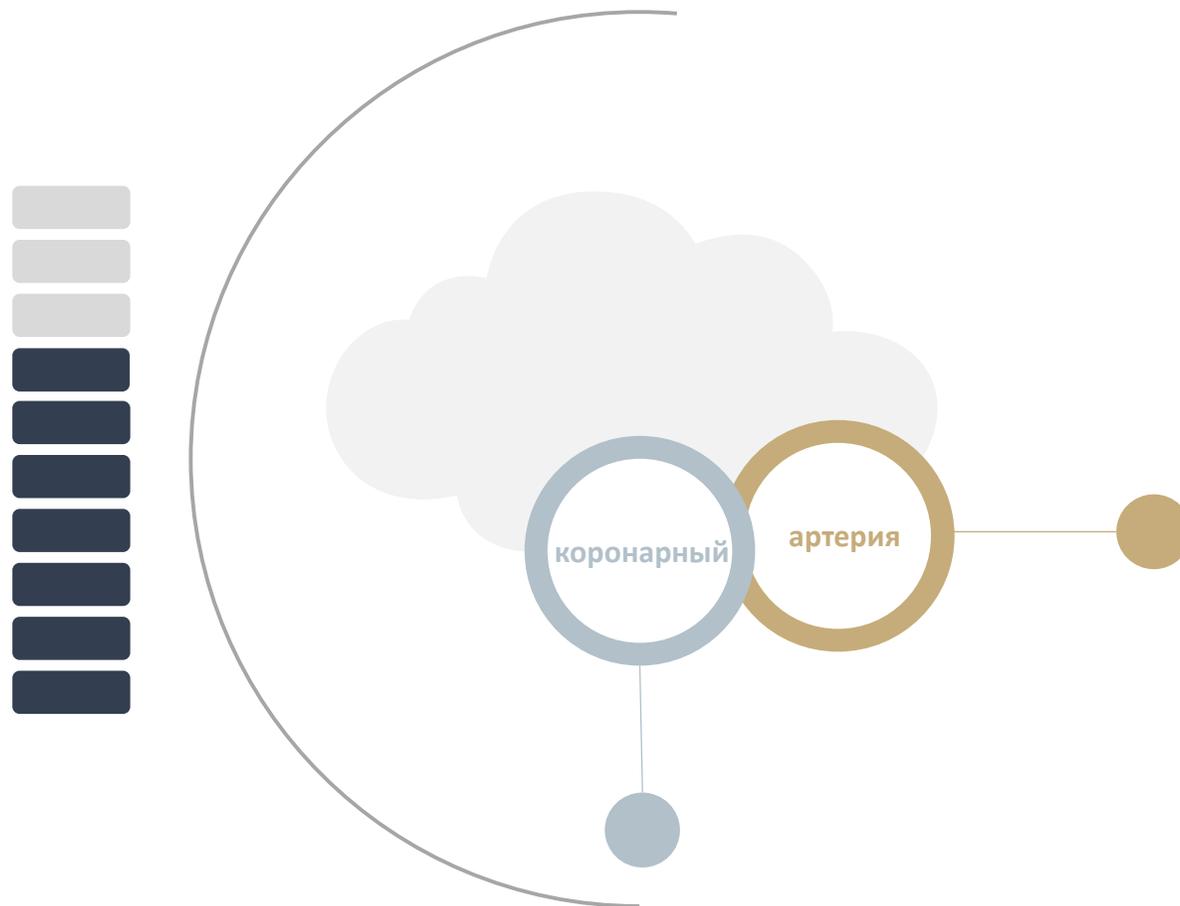
- ✓ Встречаемость компонентов фразы в корпусе
- ✓ Значимость компонентов фразы в документе



# РАСЧЕТ ЗНАЧИМОСТИ КЛЮЧЕВЫХ ФРАЗ

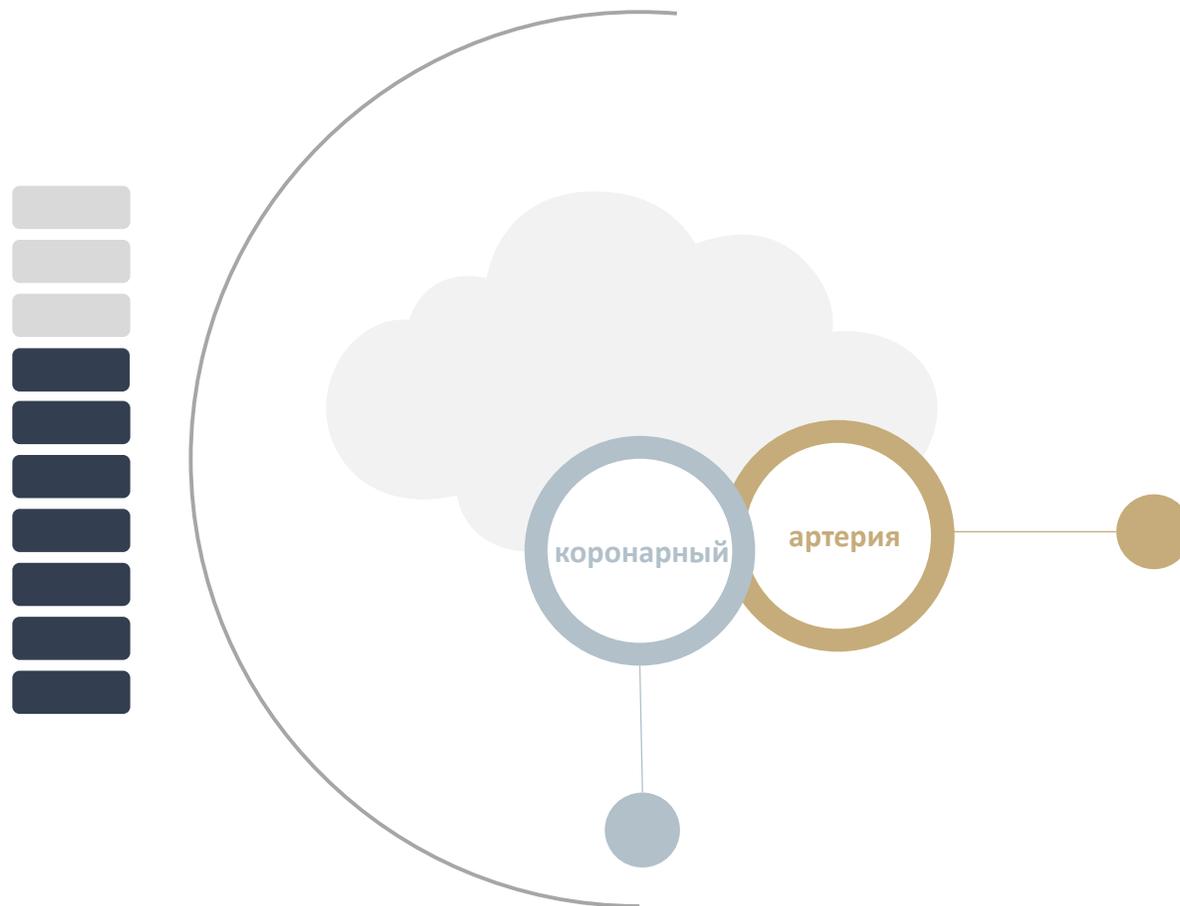
Факторы, влияющие на расчет значимости

- ✓ Встречаемость компонентов фразы в корпусе
- ✓ Значимость компонентов фразы в документе
- ✓ Совместная встречаемость компонентов в документе



## Факторы, влияющие на расчет значимости

- ✓ Встречаемость компонентов фразы в корпусе
- ✓ Значимость компонентов фразы в документе
- ✓ Совместная встречаемость компонентов в документе





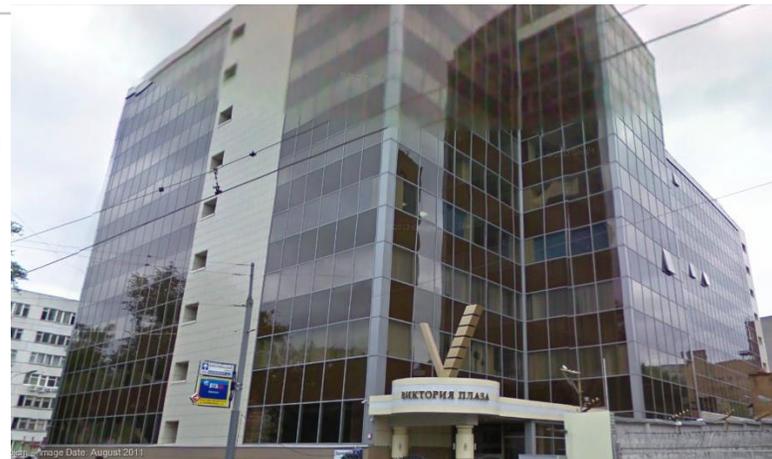
PolyAnalyst 6.5

О компании  
Megaruter Intelligence

# Основные факты

- **Мегапьютер основан:** 1997
- **Финансирование:** органический рост
- **Офисы расположены:**
  - Москва; Чебоксары; Блумингтон
- **Размер:** 120+ сотрудников
- **Клиенты:** 500+ , включая
  - 25 компаний из Fortune100
  - 200 университетов
- **Представители:** в 15-ти странах

Москва, Россия



Блумингтон, США



Чебоксары, Россия



# О КОМПАНИИ MEGAPUTER INTELLIGENCE

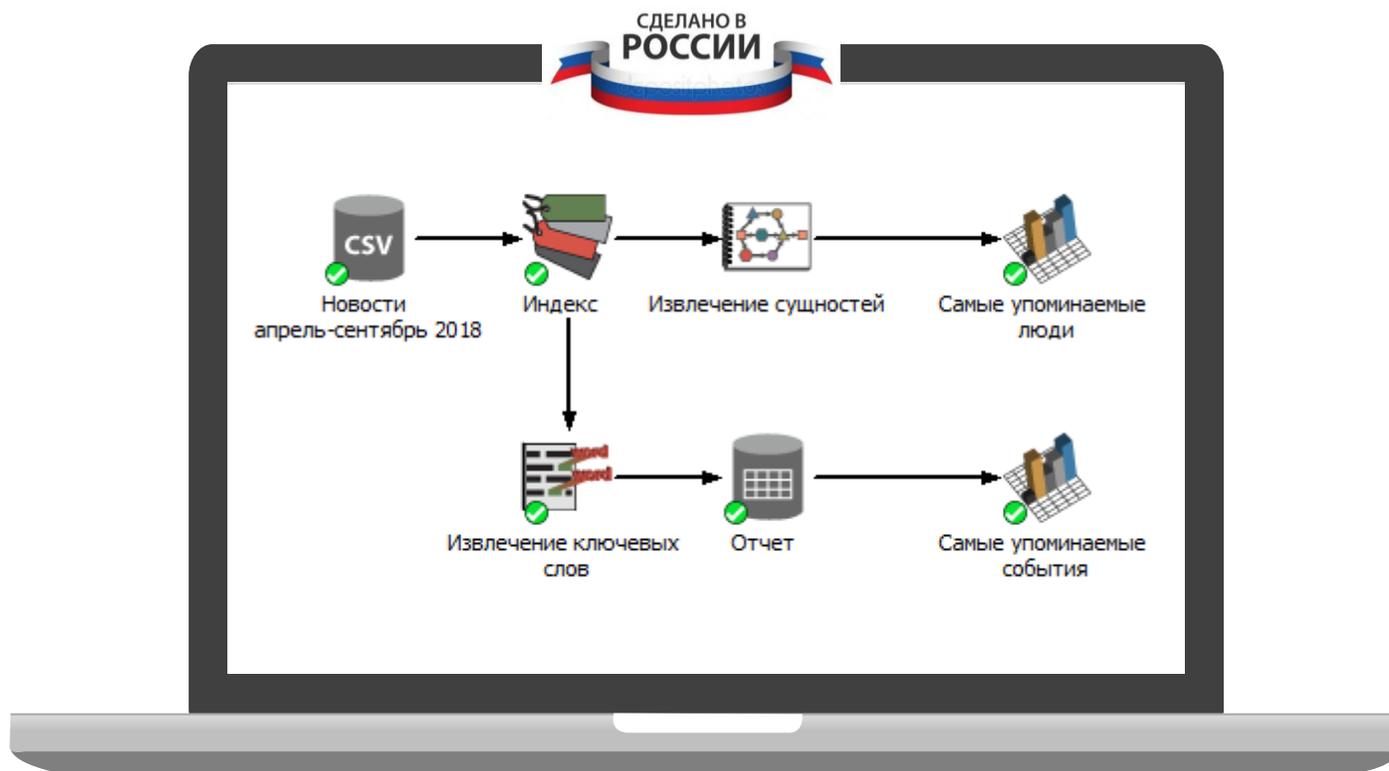


Среди наших клиентов – четверть компаний из списка Fortune 100

На рынке с 1997 года

PolyAnalyst™ включен в Единый реестр российских программ для ЭВМ и баз данных (реестровый номер 4414)





## АНАЛИТИЧЕСКАЯ ПЛАТФОРМА

PolyAnalyst – это клиент-серверная система анализа структурированных и текстовых данных.

## ГРАФИЧЕСКИЙ ИНТЕРФЕЙС

Построение решения на платформе PolyAnalyst осуществляется с помощью соединения готовых блоков.

## ПОСТРОЕНИЕ ОТЧЁТОВ

В платформу встроены средства для визуализации результатов в виде интерактивных веб-отчётов.

# Этапы работы с данными

## ПОДГОТОВКА ДАННЫХ

Загрузка данных из файлов, баз данных, документов, электронной почты, социальных сетей



## ЗАГРУЗКА ДАННЫХ

Очистка, преобразование данных, агрегация, проверка орфографии и грамматики



Классификация, кластеризация, извлечение сущностей и фактов, выявление тенденций



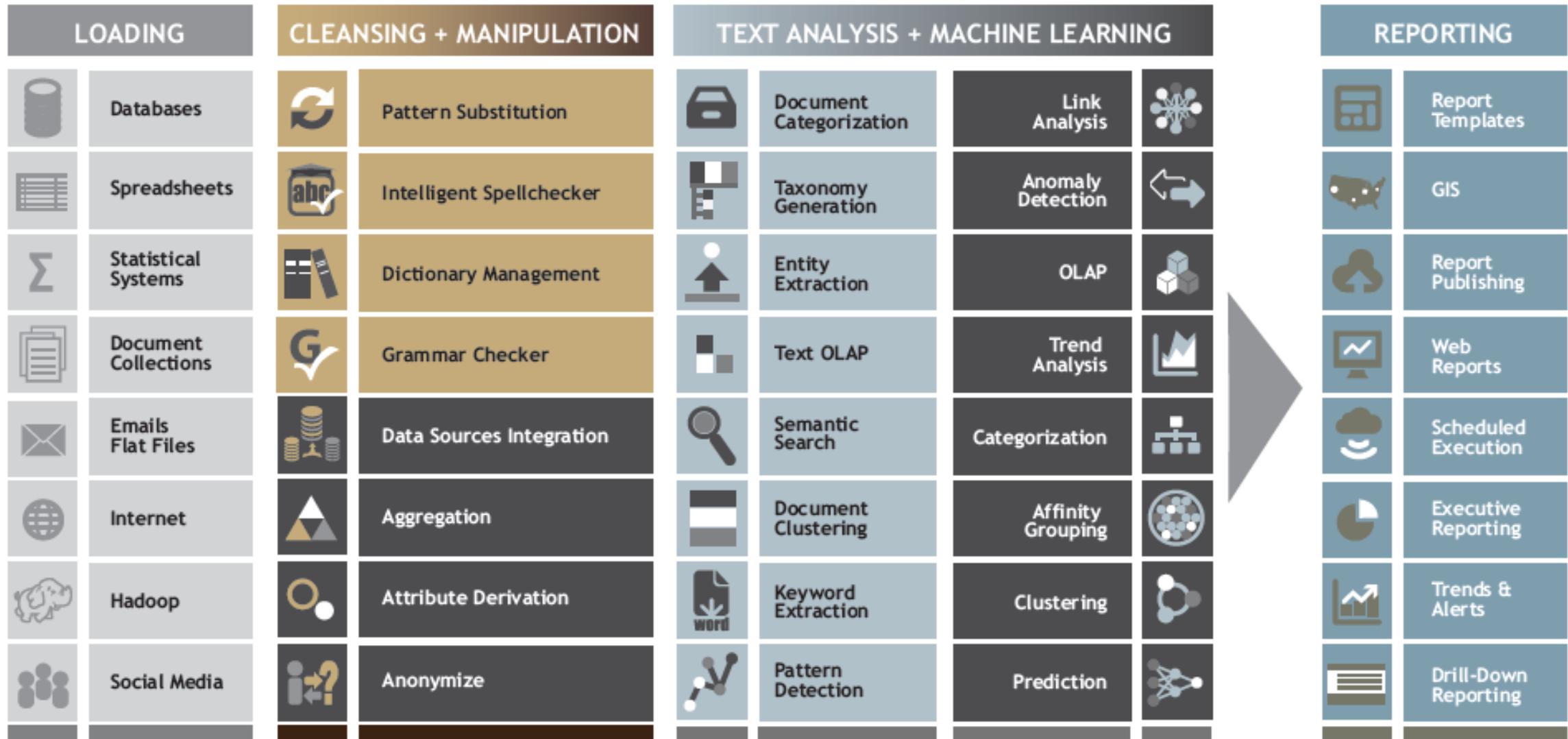
## МАШИННОЕ ОБУЧЕНИЕ И ТЕКСТОВЫЙ АНАЛИЗ

## ВИЗУАЛИЗАЦИЯ И ЭКСПОРТ

Представление найденной информации в удобном для восприятия виде



# Система PolyAnalyst™



# Анализ текста на 16-ти языках



**11**  
европейских

**5**  
азиатских

# ПОДДЕРЖИВАЕМЫЕ ЯЗЫКИ

PolyAnalyst поддерживает обработку текстов на 16 языках:

## Европа

Английский (English)  
Испанский (Español)  
Французский (Français)  
Португальский (Português)  
Русский (Русский)  
Немецкий (Deutsch)  
Итальянский (Italiano)  
Голландский (Nederlands)  
Польский (Polski)  
Греческий (Ελληνικά)  
Турецкий (Türk)



## Азия

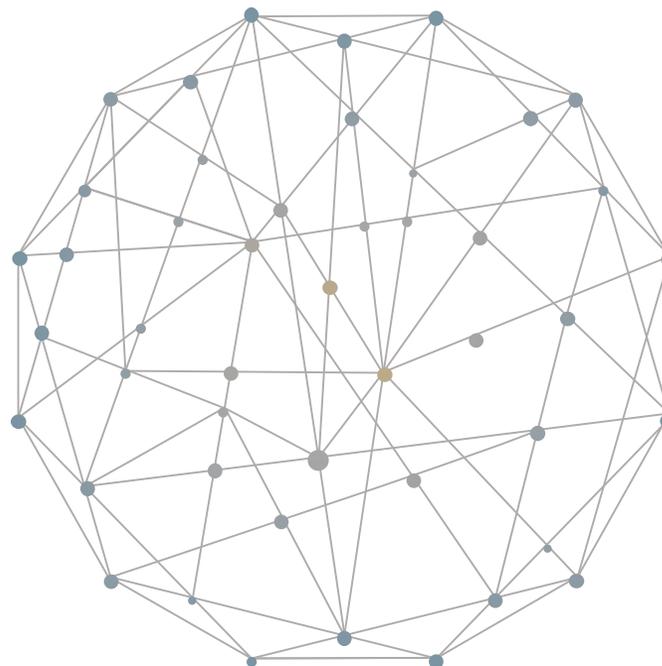
Арабский (العربية)  
Китайский (简体中文 | 繁体中文)  
Японский (日本語)  
Корейский (한국어)  
Тайский (ไทย)

## ИЗВЛЕЧЕНИЕ ИНФОРМАЦИИ

Извлечение ключевых слов,  
сущностей, фактов, отношений  
Анализ тональностей

## КЛАССИФИКАЦИЯ

Построение таксономий,  
Текстовая классификация  
и кластеризация



и многое другое...

## ПОДГОТОВКА ДАННЫХ

Определение языка,  
Индексирование,  
Проверка орфографии,  
Удаление дубликатов,  
Поиск общих фрагментов

## ПОИСК ВЗАИМОСВЯЗЕЙ

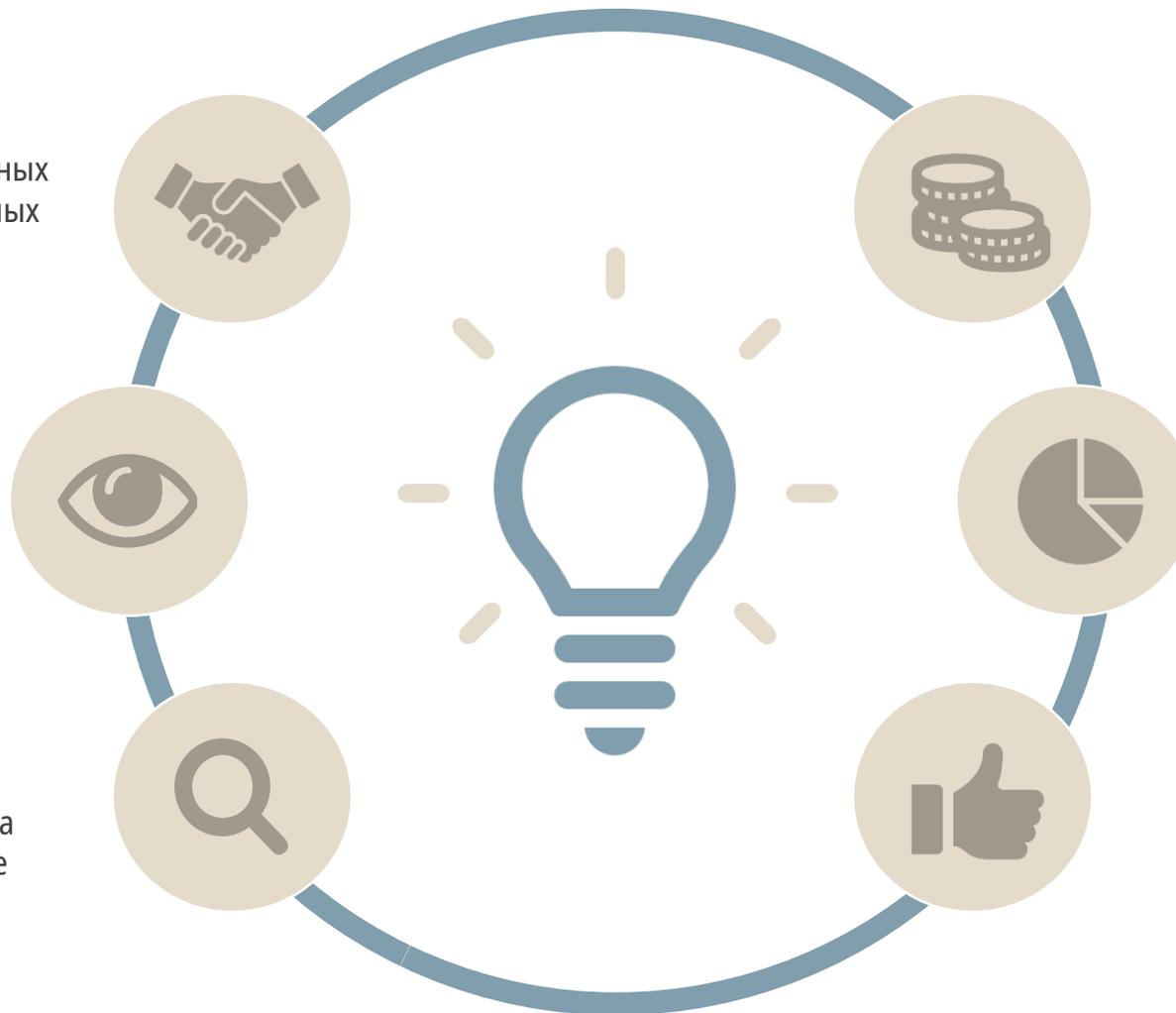
Связи терминов,  
Мониторинг упоминаемости

# Примеры применений PolyAnalyst™

**Аналитика call-центра**  
Систематизация записей из контактных центров и анализ полученных данных

**Выявление рисков**  
Анализ гарантийных претензий, записей службы поддержки и выявление закономерностей

**Анализ исков**  
Раннее выявление и прогноз риска судебных споров, резюмирование ключевых фактов



**Выявление мошенничества**  
Определение нормальных моделей поведения и анализ системных отклонений

**Анализ данных опросов**  
Анализ потребностей и ожиданий клиентов и сотрудников, совершенствование стратегии

**Аналитика социальных сетей**  
Изучение мнения потребителей, конкурентная разведка, выявление потребностей клиентов

# Аналитический клиент PolyAnalyst™

## Список инструментов

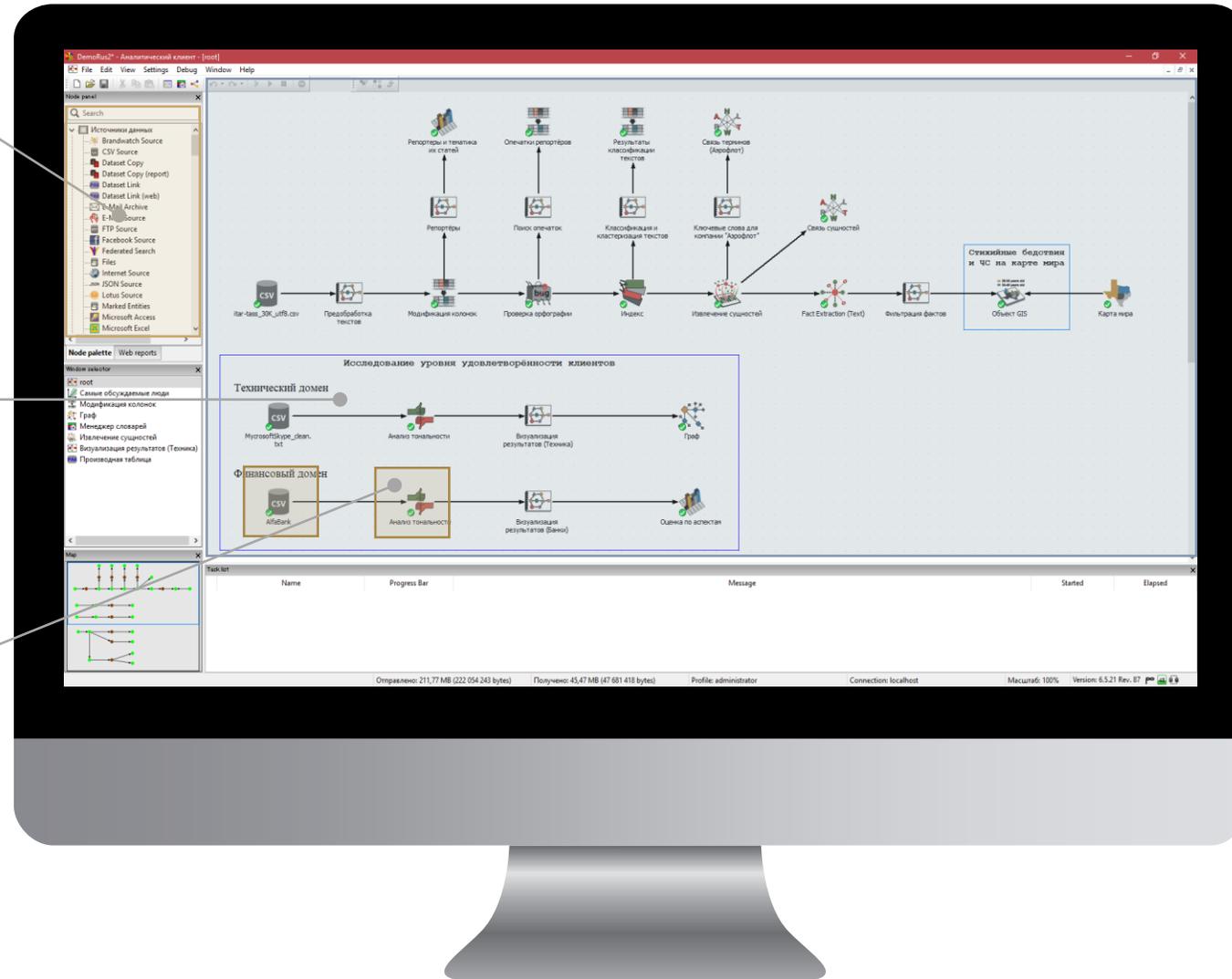
Полный доступ к инструментарию PolyAnalyst

## Схема проекта

Наглядное изображение структуры проекта

## Узлы проекта

Графический интерфейс для настройки и отображения результатов

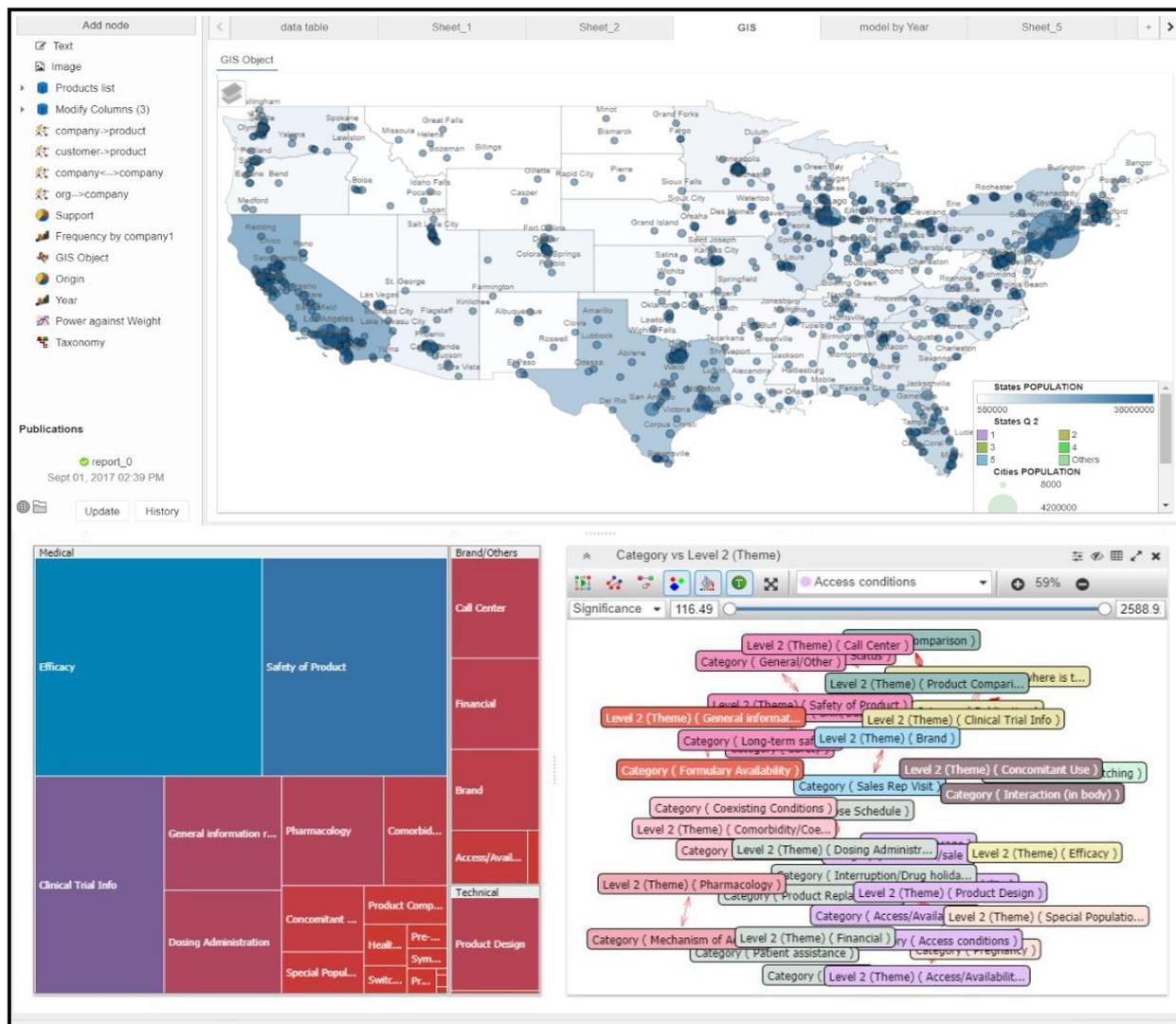


Все возможности программы в одном окне

Одновременная работа над несколькими проектами

Параллельная работа над одним проектом

# PA Web Reports™ - построение графических отчетов



- Помощь в принятии решений бизнес-пользователями
- Интерактивность
- Тонкий клиент (доступ через браузер)
- Возможность работы с текстовыми данными
- Масштабируемость

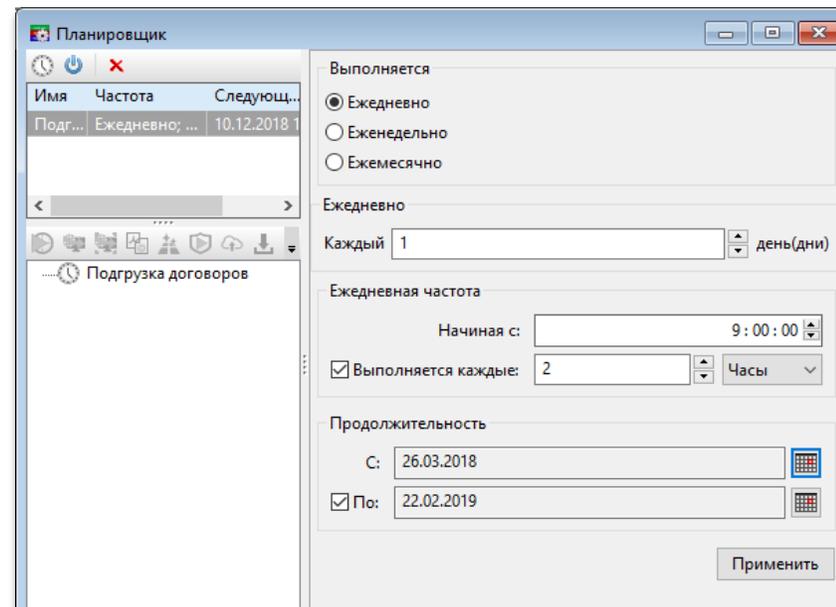
# Способы интеграции PolyAnalyst™



## Планировщик заданий

Выполнение задач при запуске сервера или по расписанию

Уведомления о результатах выполнения



## Командная строка

Выполнение проекта, узла или задачи из планировщика

Уведомления о результатах выполнения

```
my_Batch — Блокнот
Файл  Правка  Формат  Вид  Справка
<?xml version="1.0" encoding="UTF-8" standalone="no" ?>
<Node>
  <Server Server_name="localhost" Port="5063" User_name="administrator" Password=""></Server>
  <Object Project_ID="73e6fcf5-ae05-47bb-aae9-e4d93e906915">
    <Node_name></Node_name>
  </Object>
</Node>

C:\Users\ [redacted] >"C:\Megaputer Intelligence\PolyAnalyst 6.5 Server 64-bit\Bin64\
PABatExec.exe" C:\PABatches\my_Batch.xml
Starting execution, project ID: 73e6fcf5-ae05-47bb-aae9-e4d93e906915
Execution successfully completed
```



## Интерфейс Прикладного Программирования (API)

PolyAnalyst 6.5

# Загрузка и подготовка данных

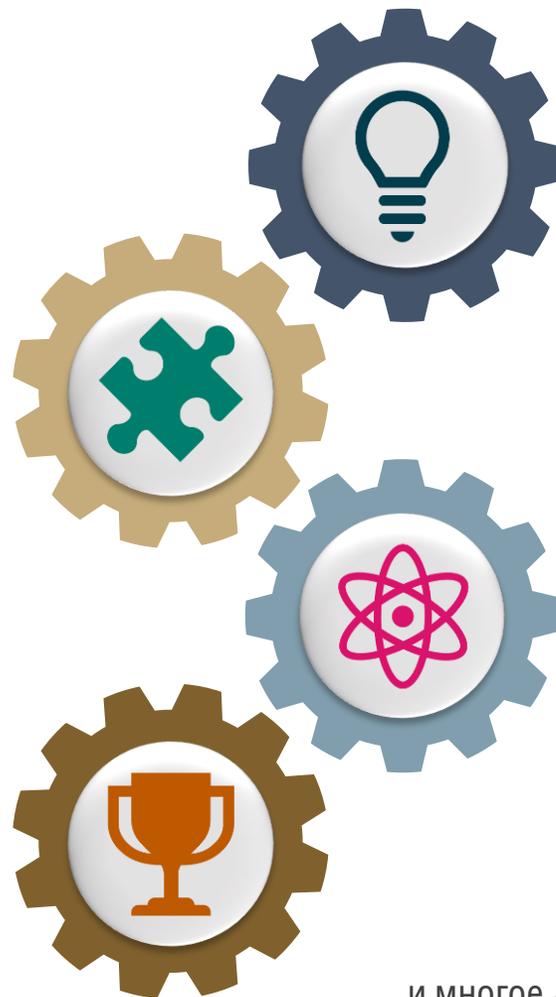


## ОПЕРАЦИИ С КОЛОНКАМИ

Дискретизация, Извлечение и замена терминов, Модификация, Разбиение, Создание производных колонок, Фильтрация

## ЭКСПОРТ ДАННЫХ

В файлы различных форматов, ODBC, словарь, отдельный каталог



## ОПЕРАЦИИ С ТАБЛИЦАМИ

Агрегирование, Конкатенация, Объединение, Сравнение, Транспонирование

## ОПЕРАЦИИ СО СТРОКАМИ

Выборка, Дополнение, Обогащение, Сортировка, Уникальные записи, Фильтрация

и многое другое...

# Пример подготовки данных

## Импорт

данных из файла

T Text	Place	Source
МОСКВА, 11 марта. /ИТАР-ТАСС/		
ТОКИО, 26 апреля. /Корр. ИТАР-Т		
МОСКВА, 31 января. /Корр. ИТАР-		
НЬЮ-ЙОРК, 17 февраля. (АРМС-Т		
МОСКВА, 19 апреля. /ИТАР-ТАСС/		
ЖЕНЕВА, 9 августа. /Корр.ИТАР-Т		
НАЛЬЧИК, 18 августа. /Корр.ИТАР		
МАЙКОП, 16 августа. /Корр. ИТАР-		
КАИР, 27 августа. /Корр. ИТАР-ТА		
БРЮССЕЛЬ, 29 октября. /Корр. ИТ	БРЮССЕЛЬ	Дубровин Денис
ВЕНА, 29 октября. /Корр.ИТАР-ТА		
Ранее был объявлен состав сборн	МОСКВА	Лазорин Игорь
МОСКВА, 14 ноября. /ИТАР-ТАСС/		
НЬЮ-ЙОРК, 16 ноября. /ИТАР-ТАС	НЬЮ-ЙОРК	
ЛОНДОН, 17 ноября. /Корр. ИТАР-	ЛОНДОН	
МОСКВА, 17 ноября. (АРМС-ТАСС/		
МОСКВА, 21 ноября. /ИТАР-ТАСС/	МОСКВА	
ЛОНДОН, 22 ноября. /Корр. ИТАР-	ЛОНДОН	Виталий Макарач
МОСКВА, 26 ноября. /ИТАР-ТАСС/	МОСКВА	
МОСКВА, 27 ноября. /ИТАР-ТАСС/	МОСКВА	
ОТТАВА, 30 ноября. /ИТАР-ТАСС/	ОТТАВА	
НИЖНИЙ НОВГОРОД, 2 декабря. /	НИЖНИЙ НОВГО	Андрей Сорваче
МОСКВА, 4 декабря. /ИТАР-ТАСС/	МОСКВА	

Новости с указанием автора.csv

## Фильтрация

статей с указанным автором

T Text	Place	Source
БРЮССЕЛЬ, 29 октября. /Корр. ИТ	БРЮССЕЛЬ	Дубровин Денис
Ранее был объявлен состав сборн	МОСКВА	Лазорин Игорь
ЛОНДОН, 22 ноября. /Корр. ИТАР-	ЛОНДОН	Виталий Макарач
НИЖНИЙ НОВГОРОД, 2 декабря. /	НИЖНИЙ НОВГО	Андрей Сорваче
ТВЕРЬ, 11 декабря. /Корр. ИТАР-Т	ТВЕРЬ	Александр Харч
НОВОГОРСК /Московская область	НОВОГОРСК	Михаил Зислис
БРАТИСЛАВА /Словакия/, 11 янва	БРАТИСЛАВА	Константин Нуж
ПХЕНЬЯН, 13 января. /Корр. ИТАР	ПХЕНЬЯН	Юрий Сидоров
БРЮССЕЛЬ, 21 января. /Корр. ИТ	БРЮССЕЛЬ	Александр Шиш
МОСКВА, 21 января. /Корр. ИТАР-	МОСКВА	Руслан Алиев
СОЧИ, 14 февраля. /Спец. корр. И	СОЧИ	Вероника Роман
ВЕЛИКИЙ НОВГОРОД, 26 февраля	ВЕЛИКИЙ НОВГО	Юлия Генерозов
КАИР, 12 марта. /Корр. ИТАР-ТАС	КАИР	Денис Соловых
ХОЛМЕНКОЛЛЕН /Норвегия/, 22 м	ХОЛМЕНКОЛЛЕН	Елена Соболев
ГОНКОНГ /СЯНГАН/, 28 марта. /К	ГОНКОНГ	Александр Эозин
НИЖНИЙ НОВГОРОД, 1 апреля. /К	НИЖНИЙ НОВГО	Андрей Сорваче
НОВОГОРСК /Московская область	НОВОГОРСК	Михаил Зислис
МАГНИТОГОРСК, 27 апреля. /Спец	МАГНИТОГОРСК	Игорь Какурин
ГААГА, 3 мая. /Корр. ИТАР-ТАСС/	ГААГА	Виталий Чугин
МОСКВА, 5 апреля. /Корр. ИТАР-Т	МОСКВА	Артем Калинин
МОСКВА, 6 мая. /Корр. ИТАР-ТАС	МОСКВА	Игорь Лазорин
НИЖНИЙ НОВГОРОД, 13 мая. /Кор	НИЖНИЙ НОВГО	Наталья Чистяк
АРХАНГЕЛЬСК, 16 мая. /Корр. ИТ	АРХАНГЕЛЬСК	Софья Ануфрие

Filter Rows Фильтруем строки, где указан автор

## Агрегирование

статей по автору

Source	T Text_ConcatUnique	Count
Олег Зеленин	ООН, 4 декабря. /Корр. ТАСС Олег	1
Наталья Баринава	МОСКВА, 24 июня. /Корр. ИТАР-ТА	2
Константин Нужденев	БРАТИСЛАВА /Словакия/, 11 январ	1
Александр Эозин	ГОНКОНГ /СЯНГАН/, 28 марта. /Кор	1
Юрий Сидоров	ПХЕНЬЯН, 13 января. /Корр. ИТАР-	1
Вероника Романенкова	СОЧИ, 14 февраля. /Спец. корр. ИТ	1
Ольга Свистунова	МОСКВА, 25 декабря. /Корр. ТАСС	2
Александр Харченко	ТВЕРЬ, 11 декабря. /Корр. ИТАР-Т	1
Виктор Хребтов	КРАСНОЯРСК, 5 ноября. /Корр. ТАС	1
Юлия Генерозова	ВЕЛИКИЙ НОВГОРОД, 26 февраля.	1
Виталий Чугин	ГААГА, 3 мая. /Корр. ИТАР-ТАСС В	1
Вероника Советова	МОСКВА, 15 апреля. /Корр. ТАСС В	3
Олег Сердобольский	САНКТ-ПЕТЕРБУРГ, 13 марта. /Кор	1
Руслан Алиев	МОСКВА, 21 января. /Корр. ИТАР-Т	1
Даниил Студнев	ОТТАВА, 12 декабря. /Корр. ТАСС	1
Алан Бадов	ЛОНДОН, 11 июля. /Корр. ИТАР-ТА	1
Андрей Серов	ЯРОСЛАВЛЬ, 8 апреля. /Корр. ТАС	1
Денис Соловых	КАИР, 12 марта. /Корр. ИТАР-ТАСС	1
Софья Ануфриева	АРХАНГЕЛЬСК, 16 мая. /Корр. ИТА	1
Александр Шишло	БРЮССЕЛЬ, 21 января. /Корр. ИТА	1
Владимир Нуякшев	КАЛИНИНГРАД, 18 марта. /Корр.ТА	1
Вениамин Демидецкий	КИШИНЕВ, 9 декабря. /Корр. ТАСС	1
Елена Соболев	ХОЛМЕНКОЛЛЕН /Норвегия/, 22 ма	1
Наталья Чистякова	НИЖНИЙ НОВГОРОД, 13 мая. /Корр	1
Игорь Какурин	МАГНИТОГОРСК. 27 апреля. /Спец	2

Aggregate Агрегирование статей по автору

## Фильтрация

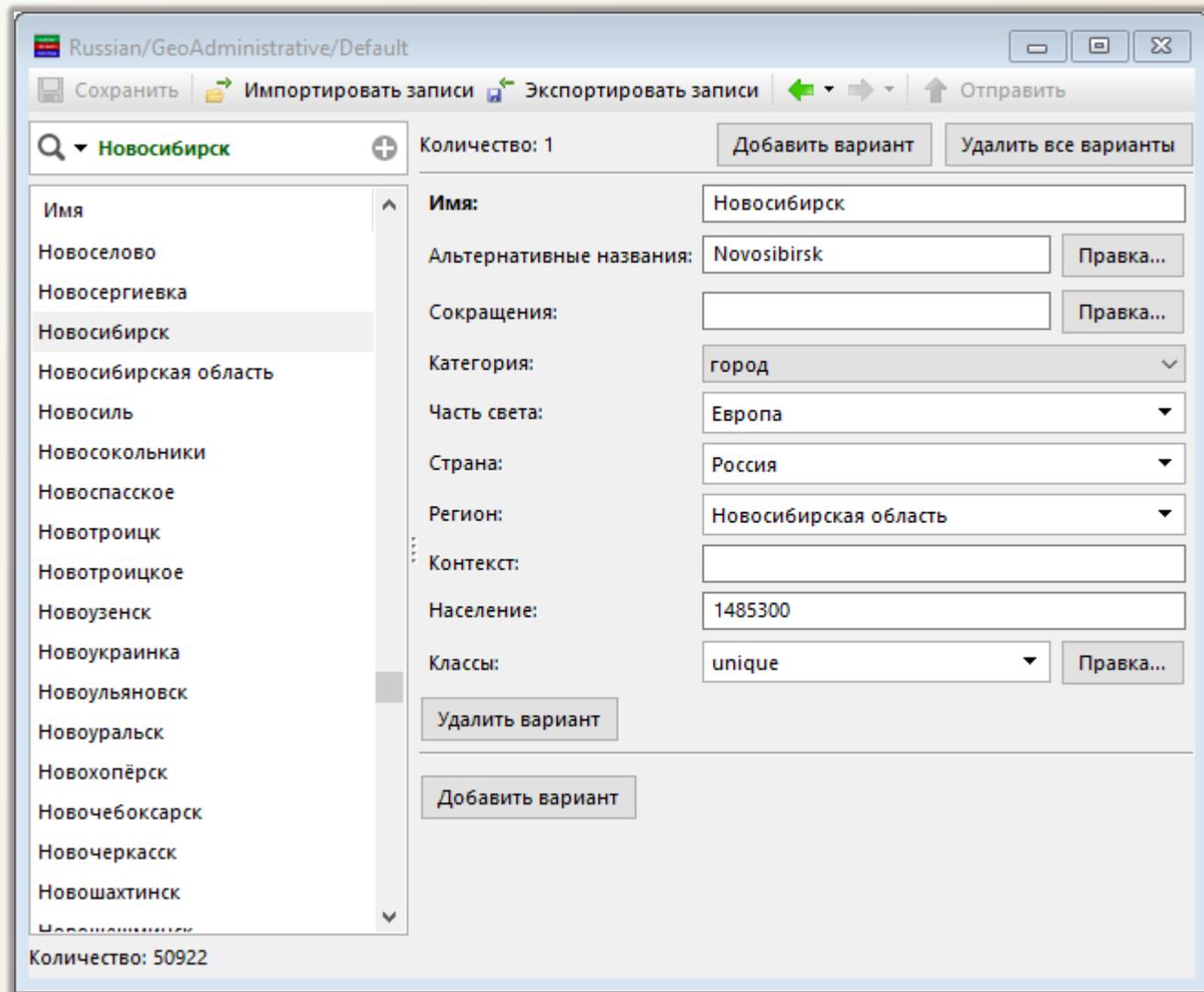
авторов, опубликовавших более 1 статьи

Source	T Text_ConcatUnique	Count
Наталья Баринава	МОСКВА, 24 июня. /Корр. ИТАР-ТАСС Наталья Баринав	2
Ольга Свистунова	МОСКВА, 25 декабря. /Корр. ТАСС Ольга Свистунова/.	2
Вероника Советова	МОСКВА, 15 апреля. /Корр. ТАСС Вероника Советова/.	3
Игорь Какурин	МАГНИТОГОРСК, 27 апреля. /Спец. корр. ИТАР-ТАСС И	2
Илья Андреев	МОСКВА, 14 октября. /Корр. ТАСС Илья Андреев/. "Уфе	3
Игорь Лазорин	МОСКВА, 27 октября. /Корр. ТАСС Игорь Лазорин/. Рос	2
Олег Кошелев	ЕКАТЕРИНБУРГ, 19 октября. /Спец. корр. ТАСС Олег Ко	2
Михаил Зислис	НОВОГОРСК /Московская область/, 18 декабря. /Корр.	2
Андрей Сорвачев	НИЖНИЙ НОВГОРОД, 1 апреля. /Корр. ИТАР-ТАСС Анд	2
Светлана Янкина	МОСКВА, 1 сентября. /Корр. ИТАР-ТАСС Светлана Янк	2

Filter Rows Авторы, опубликовавшие 2 и более статьи

# РАБОТА СО СЛОВАРЯМИ

- ✓ поддержка различных типов словарей (морфологии, фраз, сущностей, синонимов и другое)
- ✓ графический редактор (поиск, просмотр и редактирование словарных статей)
- ✓ система контроля версий и разграничения прав пользователей
- ✓ экспорт и импорт словарей
- ✓ операции с несколькими словарями: объединение, пересечение, вычитание
- ✓ создание пользовательских словарей с произвольными полями



# ВАРИАНТЫ АНОНИМИЗАЦИИ

г. Саранск 16 мая 2018 года  
ООО "ЗАЩИТАИНФОТРАНС",  
именуемое в дальнейшем  
"Исполнитель", в лице генерального

## ЗАМЕНА НА КОД

г. Город1932173537v1  
Дата3007097787v1  
ЮрЛицо1739268359v1, именуемое в  
дальнейшем "Исполнитель", в лице  
ФизЛицо1758330182v1,

Замена информации на  
уникальные коды

## ЗАМЕНА НА ТЕКСТ

г. [Город] дд.мм.гггг  
[Юр.Лицо], именуемое в дальнейшем  
"Исполнитель", в лице [Физ.Лицо],  
действующего на основании Устава,

Замена сущности на заданный  
пользователем текст

## ПЕРЕИМЕНОВАНИЕ

г. Тула 02.08.2017  
ООО "Арсенал+", именуемое в  
дальнейшем "Исполнитель", в лице

Замена сущности на случайно  
выбранную сущность того же типа

# МОНИТОРИНГ УПОМИНАЕМОСТИ

## Изменение интереса

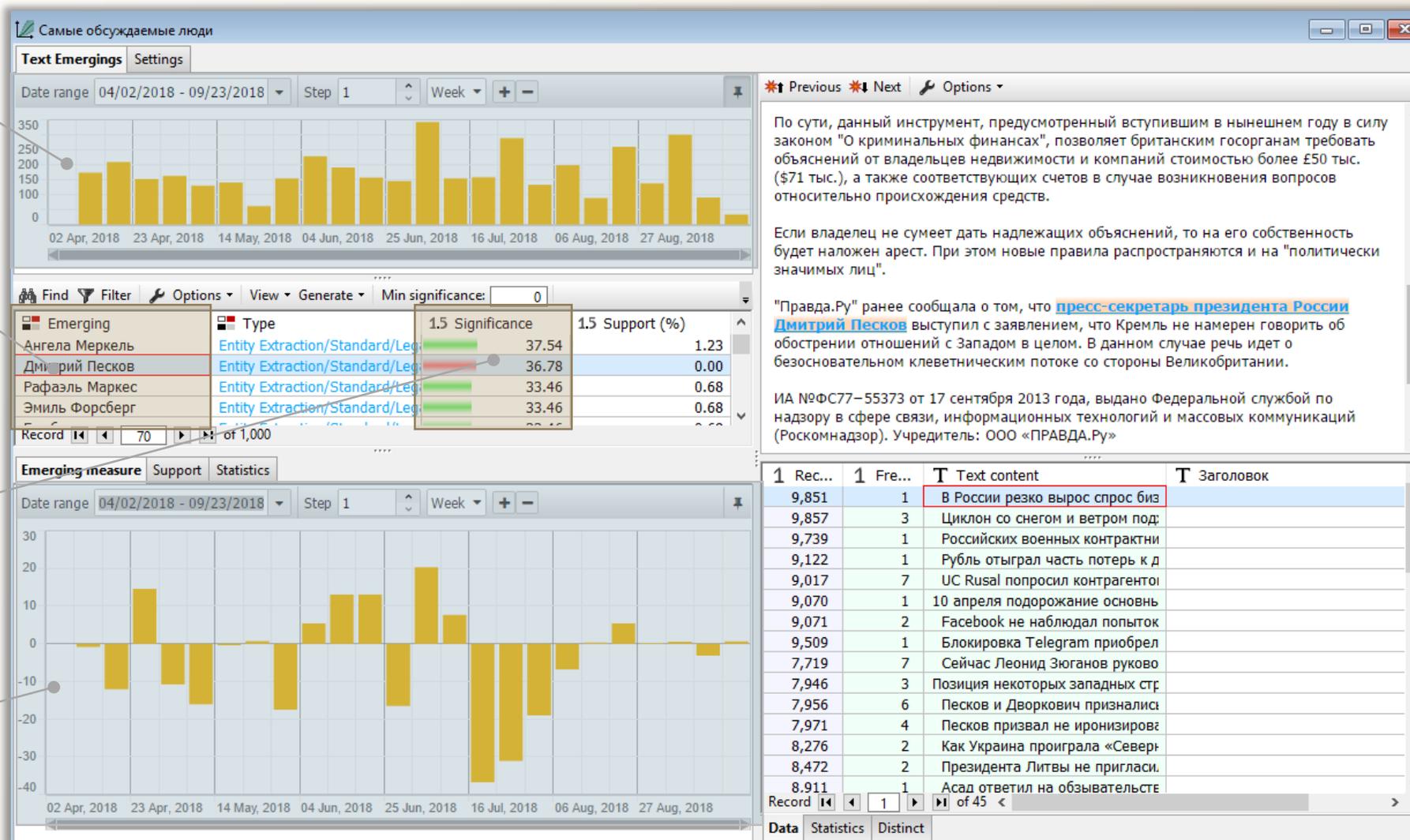
к тематике со временем

## Обсуждаемые объекты

## Показатель интереса

## Изменение интереса

к объекту со временем



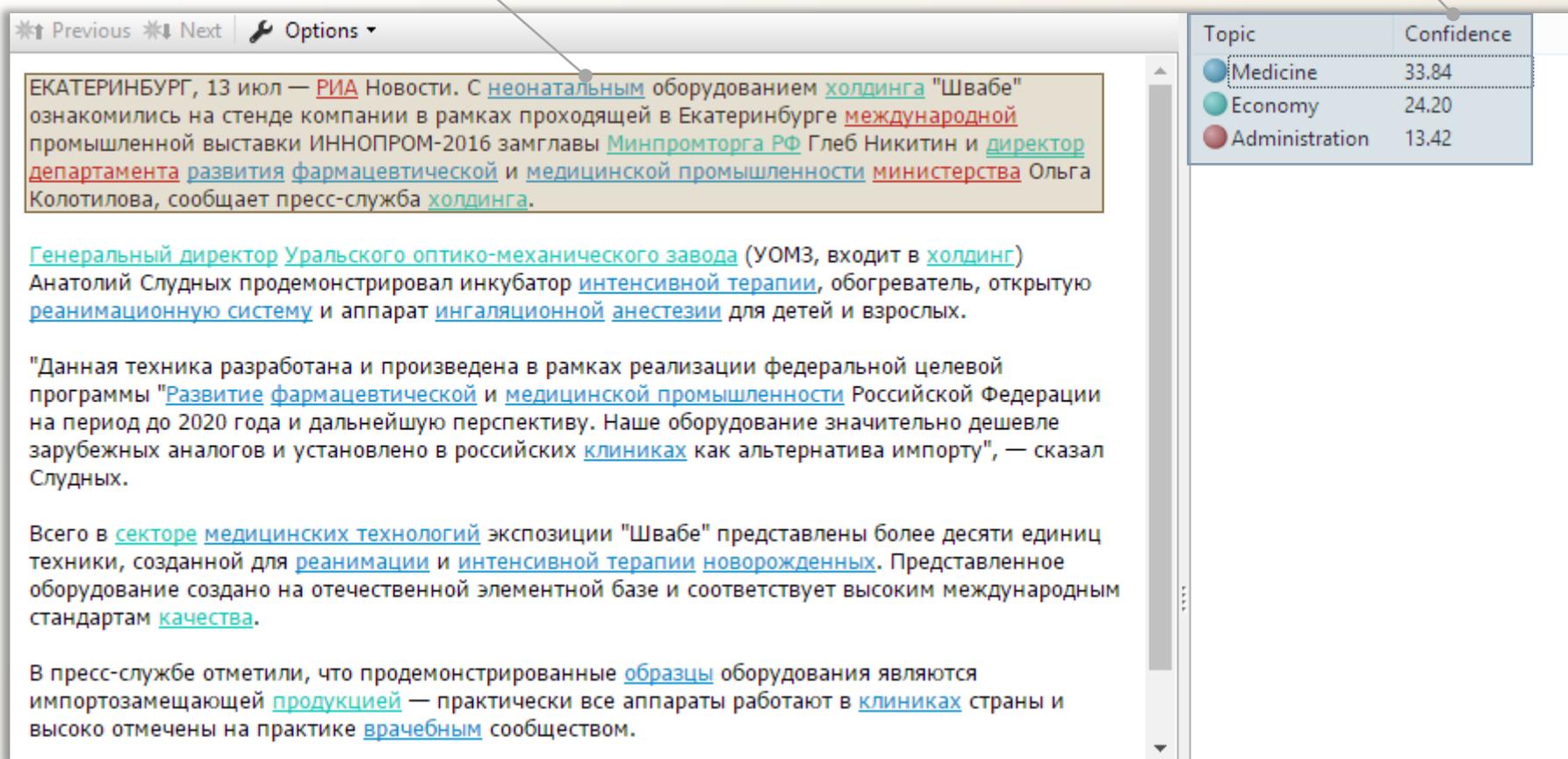
# АНАЛИЗ ТЕМАТИКИ ТЕКСТА

## Маркеры тематики

Выделяются в тексте

## Список тематик

Ранжирование по  
достоверности



Previous Next Options

ЕКАТЕРИНБУРГ, 13 июл — [РИА](#) Новости. С [неонатальным](#) оборудованием [холдинга](#) "Швабе" ознакомились на стенде компании в рамках проходящей в Екатеринбурге [международной](#) промышленной выставки ИННОПРОМ-2016 замглавы [Минпромторга РФ](#) Глеб Никитин и [директор](#) [департамента](#) [развития](#) [фармацевтической](#) и [медицинской промышленности](#) [министерства](#) Ольга Колотилова, сообщает пресс-служба [холдинга](#).

[Генеральный директор Уральского оптико-механического завода](#) (УОМЗ, входит в [холдинг](#)) Анатолий Слудных продемонстрировал инкубатор [интенсивной терапии](#), обогреватель, открытую [реанимационную систему](#) и аппарат [ингаляционной анестезии](#) для детей и взрослых.

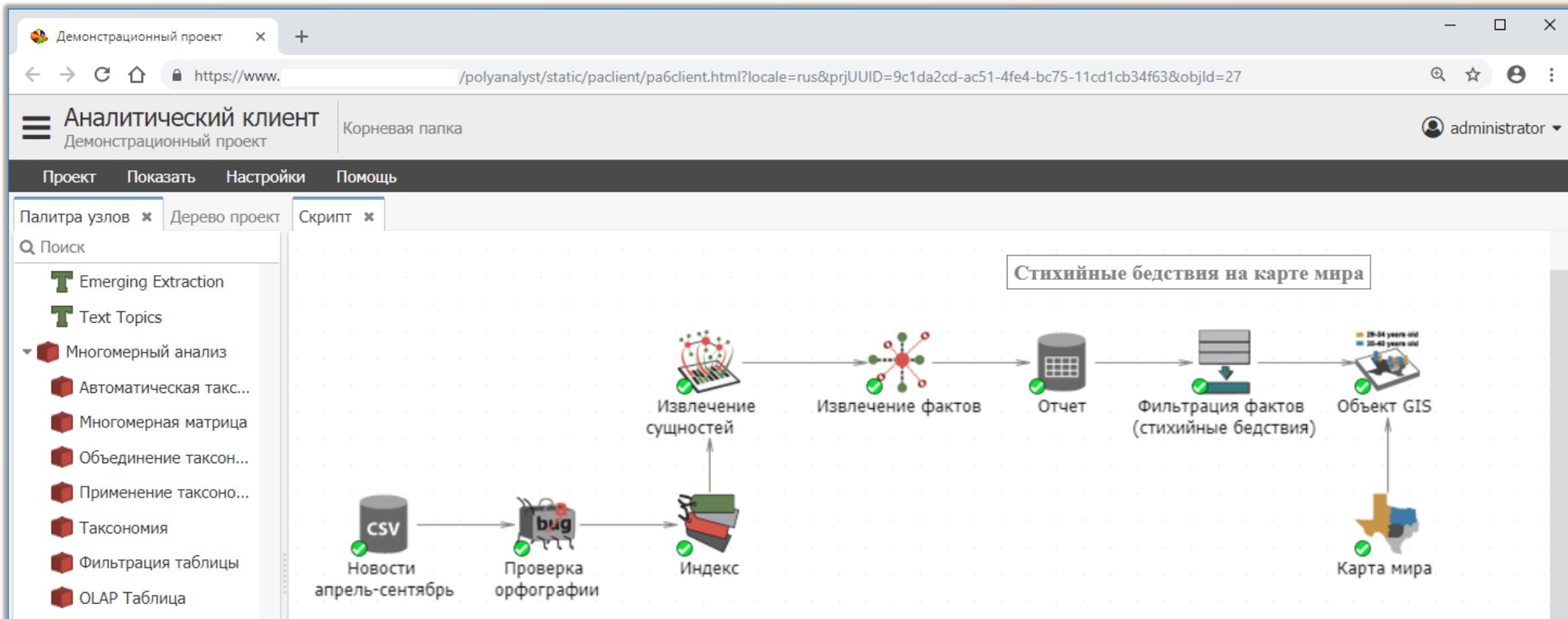
"Данная техника разработана и произведена в рамках реализации федеральной целевой программы "[Развитие фармацевтической и медицинской промышленности](#) Российской Федерации на период до 2020 года и дальнейшую перспективу. Наше оборудование значительно дешевле зарубежных аналогов и установлено в российских [клиниках](#) как альтернатива импорту", — сказал Слудных.

Всего в [секторе медицинских технологий](#) экспозиции "Швабе" представлены более десяти единиц техники, созданной для [реанимации](#) и [интенсивной терапии новорожденных](#). Представленное оборудование создано на отечественной элементной базе и соответствует высоким международным стандартам [качества](#).

В пресс-службе отметили, что продемонстрированные [образцы](#) оборудования являются импортозамещающей [продукцией](#) — практически все аппараты работают в [клиниках](#) страны и высоко отмечены на практике [врачебным сообществом](#).

Topic	Confidence
● Medicine	33.84
● Economy	24.20
● Administration	13.42

# ТЕКСТОВЫЙ АНАЛИЗ В ПРОЕКТЕ



Демонстрационный проект

https://www. /polyanalyst/static/paclient/pa6client.html?locale=rus&prjUUID=9c1da2cd-ac51-4fe4-bc75-11cd1cb34f63&objId=27

Аналитический клиент  
Демонстрационный проект

Корневая папка administrator

Проект Показать Настройки Помощь

Палитра узлов x Дерево проект Скрипт x

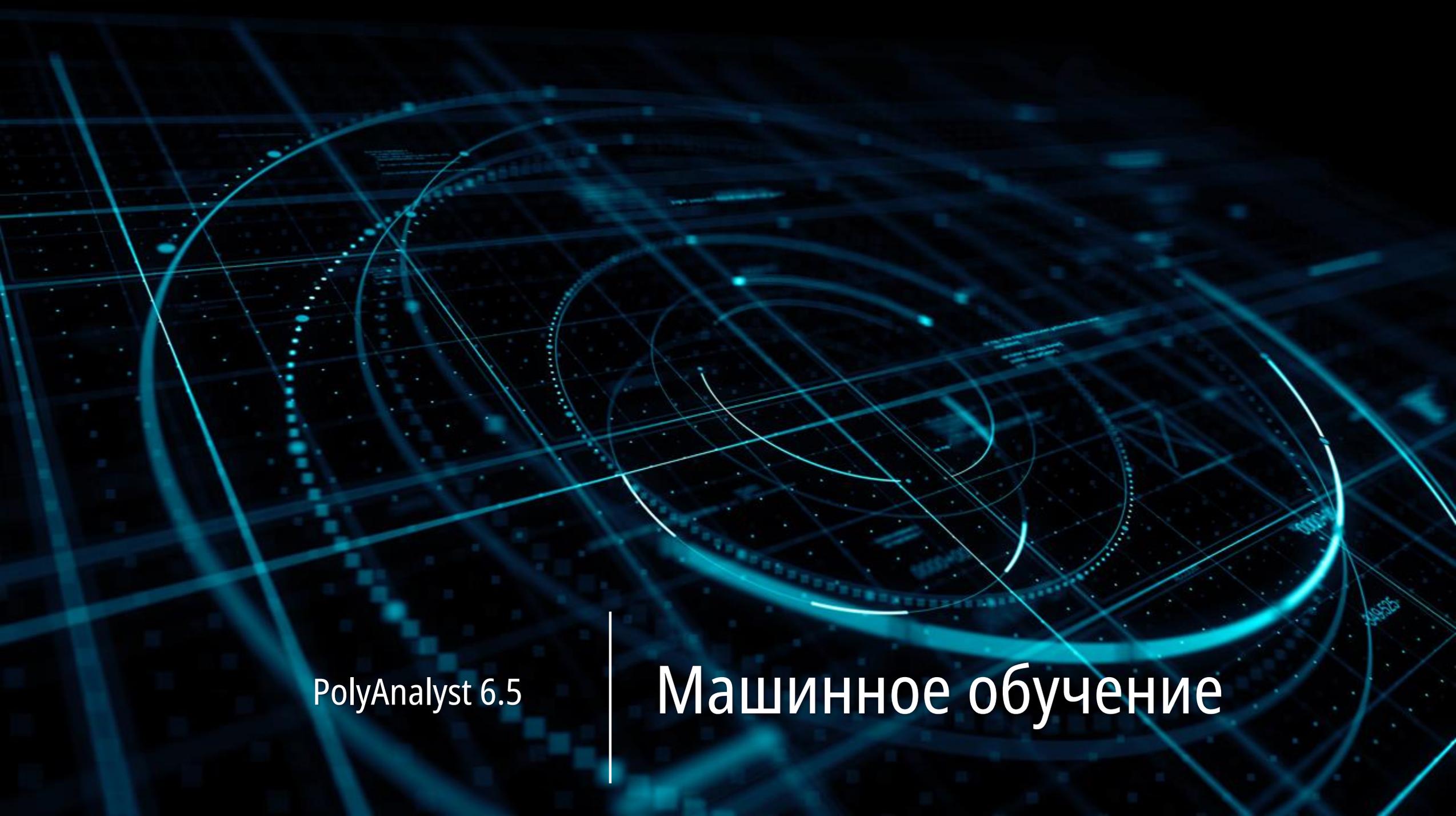
Поиск

- Emerging Extraction
- Text Topics
- Многомерный анализ
  - Автоматическая такс...
  - Многомерная матрица
  - Объединение таксон...
  - Применение таксоно...
  - Таксономия
  - Фильтрация таблицы
  - OLAP Таблица

### Стихийные бедствия на карте мира

Новости апрель-сентябрь → Проверка орфографии → Индекс → Извлечение сущностей → Извлечение фактов → Отчет → Фильтрация фактов (стихийные бедствия) → Объект GIS → Карта мира

20-34 years old  
35-40 years old



PolyAnalyst 6.5

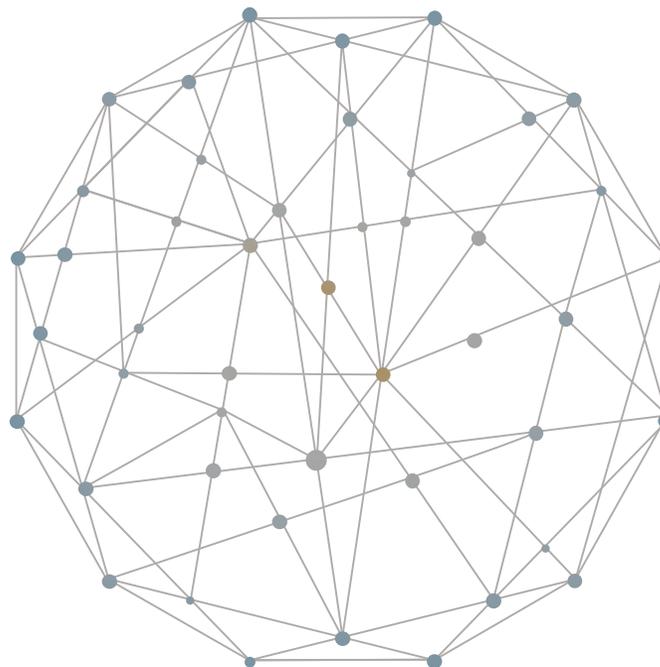
Машинное обучение

## ДЕРЕВЬЯ РЕШЕНИЙ

Деревья решений,  
Таблица решений,  
Случайный лес, CHAID

## АНАЛИЗ СВЯЗЕЙ

Анализ социальной сети,  
Корреляционный анализ,  
Факторный анализ



и многое другое...

## КЛАССИФИКАЦИЯ

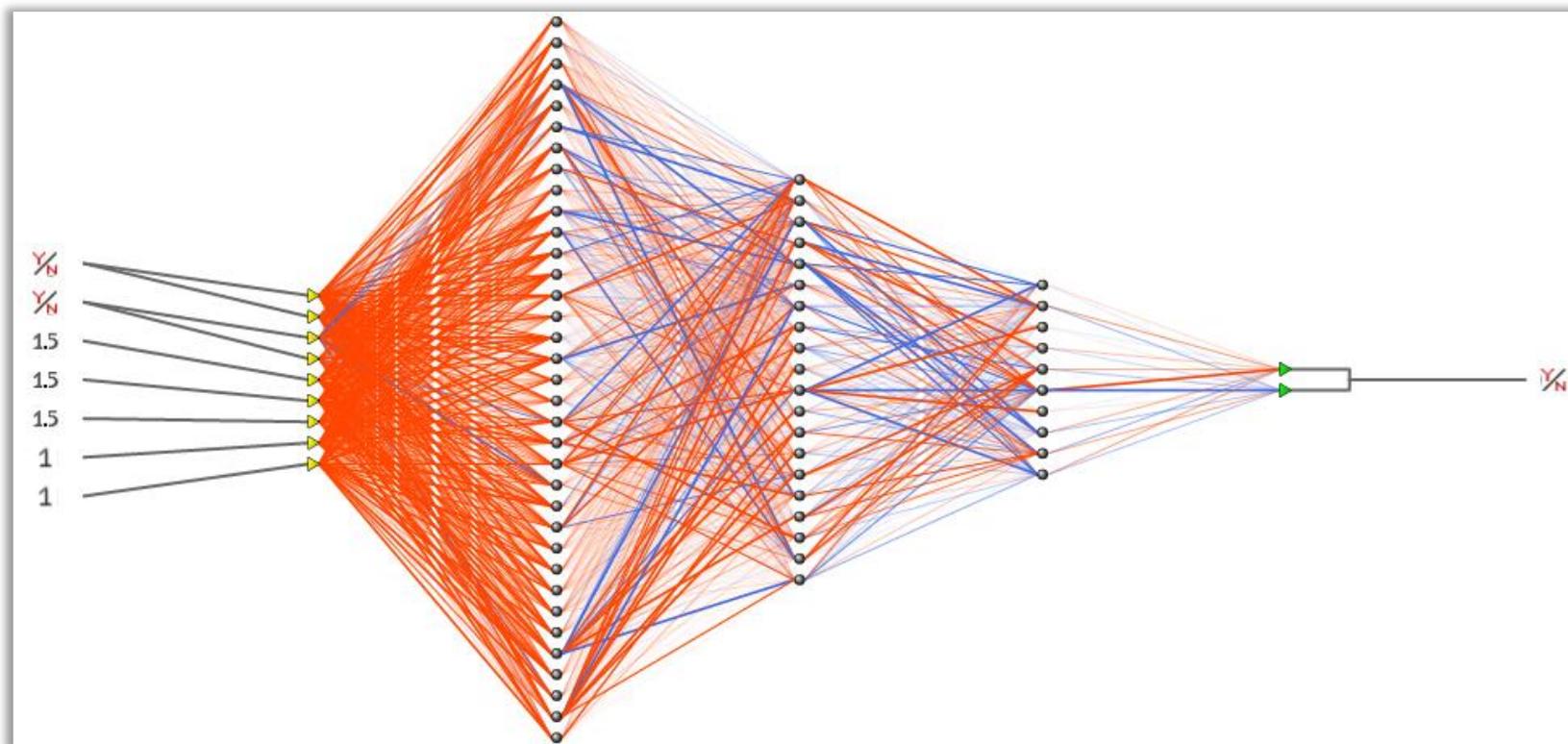
SVM, Логистическая регрессия,  
Байесовская классификация,  
Метод ближайших соседей,  
Дискриминантный анализ корзин

## НЕЙРОННЫЕ СЕТИ

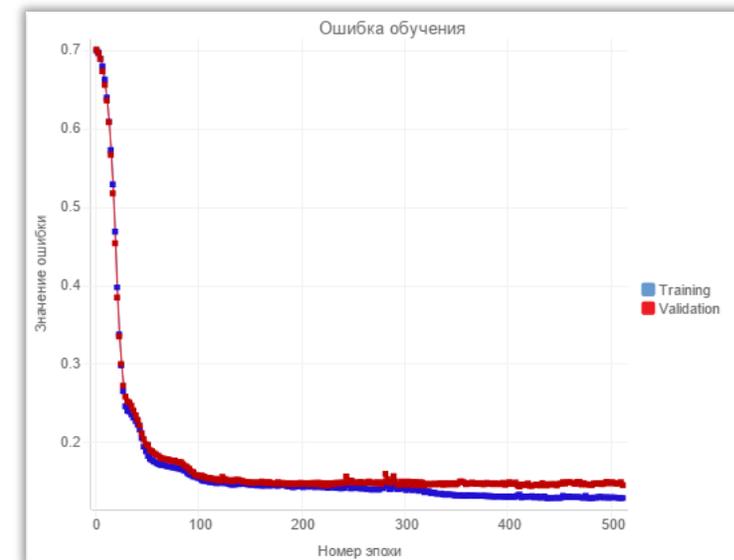
Нейронная сеть,  
Сеть Кохонена

# Пример нейронной сети

## Визуализация



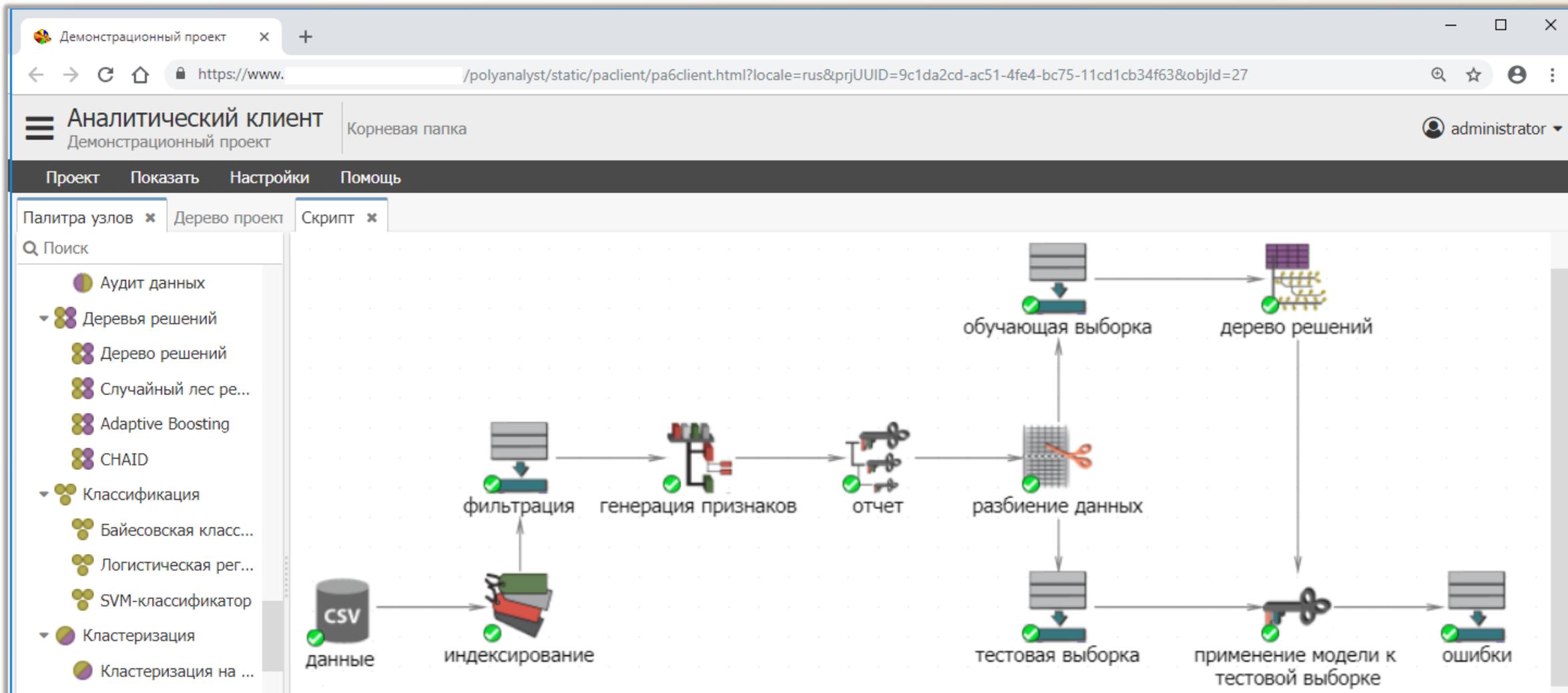
## Ошибка обучения



## Матрица ошибок

	Прогнозируемое		
Действительное	No	Yes	False negative
No	84,55	0,96	0,96
Yes	6,63	7,86	6,63
False positive	7,27	10,88	7,59

# МАШИННОЕ ОБУЧЕНИЕ В ПРОЕКТЕ



# КОНТАКТЫ



## **Сергей Ананян**

Генеральный директор

+7 (977) 104-60-80 (Москва)

+1 (812) 330-0110 (США)

sananyan@megaputer.com

1600 W Bloomfield Rd.

Bloomington, IN 47403 USA

[www.megaputer.com](http://www.megaputer.com)



## **Алексей Русских**

Генеральный директор (Россия)

+7 (915) 424-23-45

russkikh@megaputer.ru



## **Давид Сазонов**

Руководитель направления текстового анализа

+7 (903) 789-72-37

sazonov@megaputer.ru

## **Мегапьютер Интеллидженс**

Москва, ул. Бауманская 6, офис 723

+7 (499) 753-01-29

[www.megaputer.ru](http://www.megaputer.ru)