



XXIII научно-практическая конференция «Комплексная защита информации»



МИРЭА – РОССИЙСКИЙ ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ
Институт комплексной безопасности и специального приборостроения
Кафедра КБ-3 «Управление и моделирование систем»

ИНЖИНИРИНГ НОВЫХ ЗНАНИЙ ПРИ АНАЛИЗЕ ТЕКСТОВЫХ СООБЩЕНИЙ В СОЦИАЛЬНЫХ СЕТЯХ

Самолетова К.С.
Руководитель:
д.в.н., проф. Лось В.П.



МНОГООБРАЗИЕ И ОБЪЕМ ЦИФРОВОЙ ИНФОРМАЦИИ

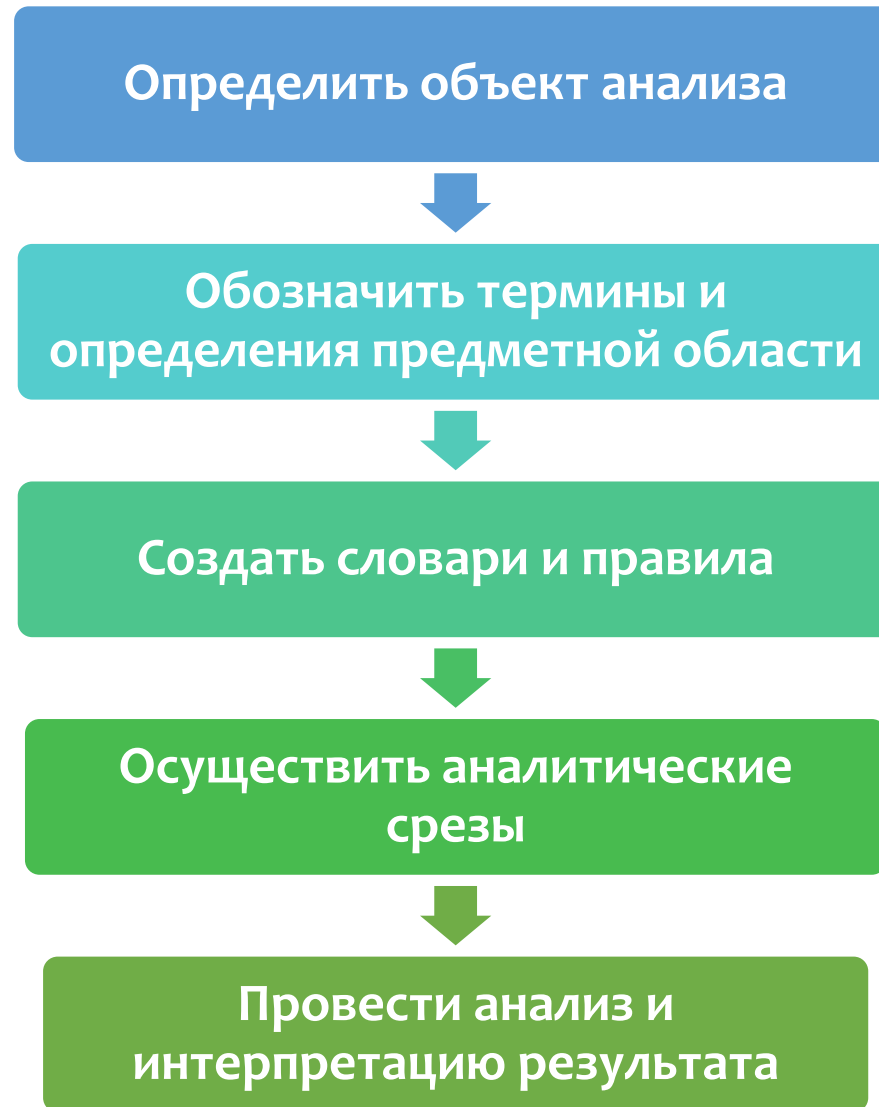


более **80%** информации
не структурировано

Способы работы с неструктурированными данными



Этапы обработки неструктурированной информации



«МИРЭА – Российский технологический университет»



Было проанализировано **37 376 записей** из
24 сообществ ВУЗа социальной сети Вконтакте:

https://vk.com/priem_mirea

<https://vk.com/club1384>

<https://vk.com/club4117>

<https://vk.com/public147961686>

https://vk.com/spravochnaya_rsuitre

<https://vk.com/podslushanomirea>

https://vk.com/overhear_mtu

https://vk.com/spravochnaya_rsuitre

<https://vk.com/public147961686>

<https://vk.com/public84761699>

https://vk.com/podslushano_mitht

<https://vk.com/public99845649>

<https://vk.com/public80799731>

<https://vk.com/public79544978>

https://vk.com/ask_mirea

<https://vk.com/board42869722>

<https://vk.com/board5411326>

<https://vk.com/mgupipoit6>

https://vk.com/mgup_it6

https://vk.com/kb_3_mgupi

https://vk.com/bestkb_3

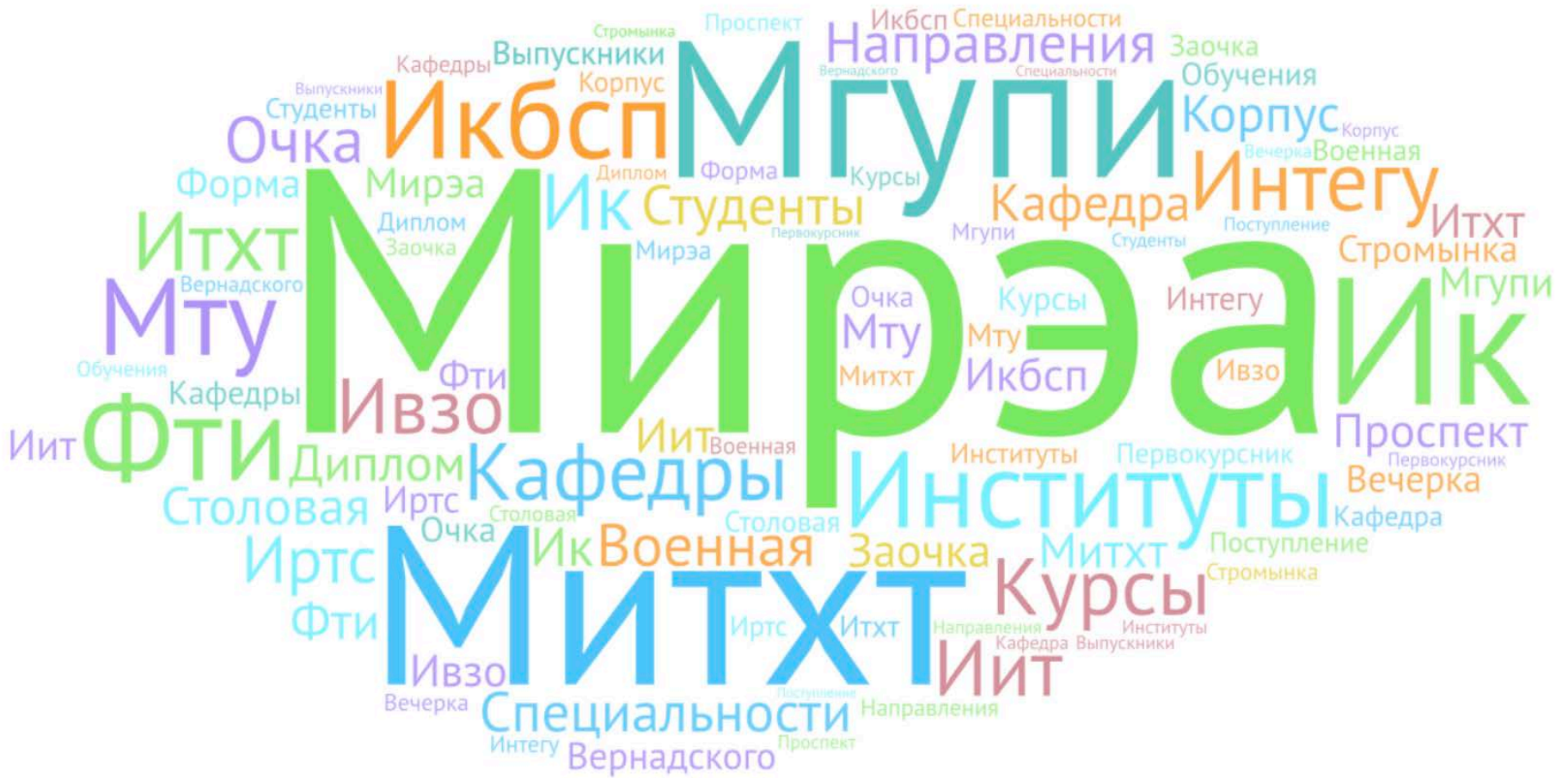
https://vk.com/mgupi_bkbo_01_14

https://vk.com/mgupi_it6_2013

https://vk.com/mag_kb3

ТЕРМИНЫ И ОПРЕДЕЛЕНИЯ

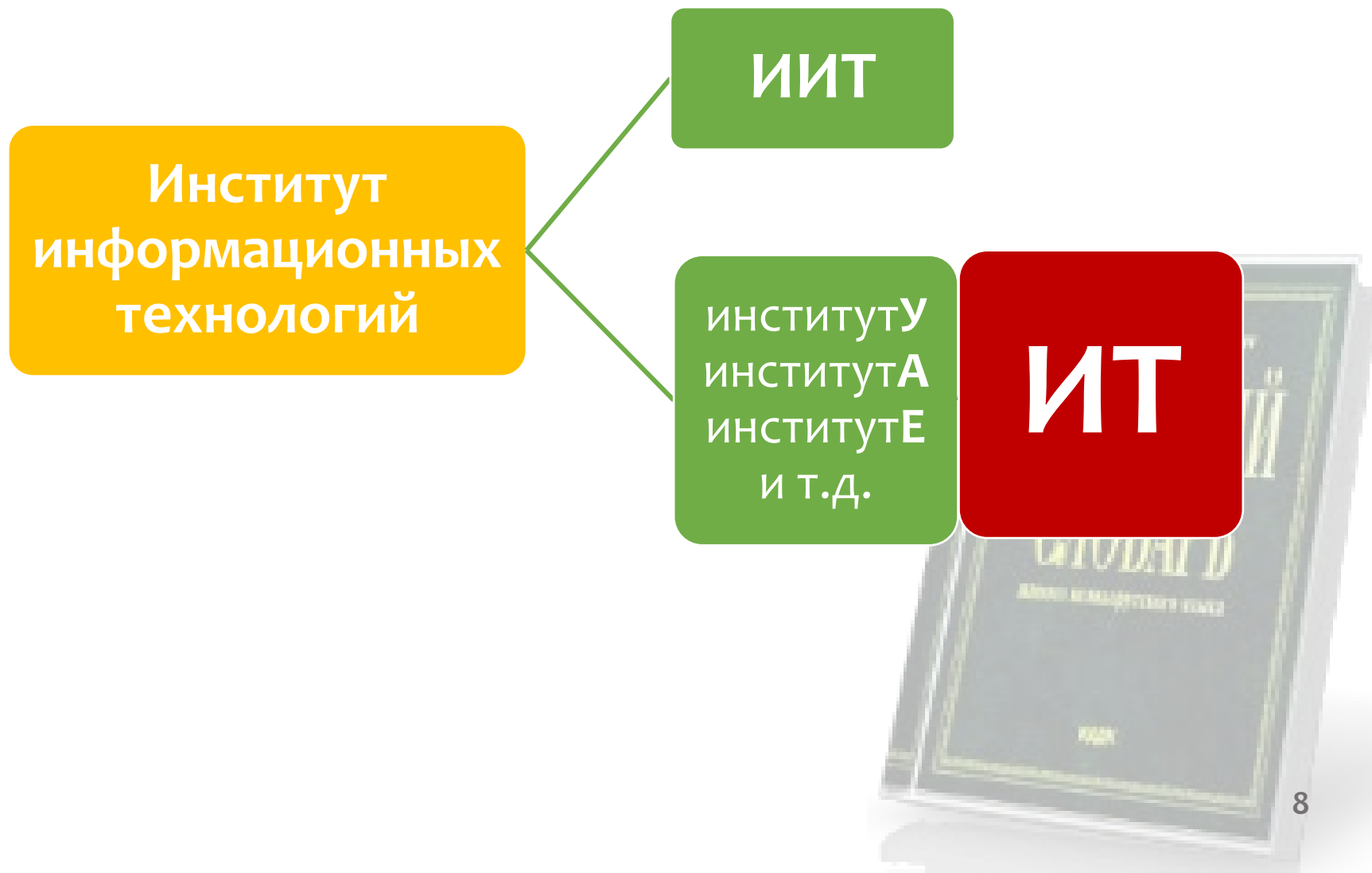
ПРЕДМЕТНОЙ ОБЛАСТИ



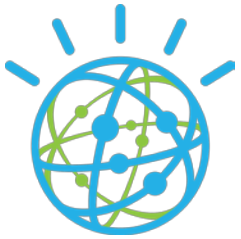
СЛОВАРИ И ПРАВИЛА



СЛОВАРИ И ПРАВИЛА



АНАЛИТИЧЕСКИЕ СРЕЗЫ



IBM WATSON CONTENT ANALYTICS

Представляет необработанные данные
в удобном для восприятия виде

Осуществляет синтаксический
и морфологический разбор

Позволяет экспортировать результаты
в различные форматы

Обладает открытым инструментарием
для создания пользовательских аналитических средств



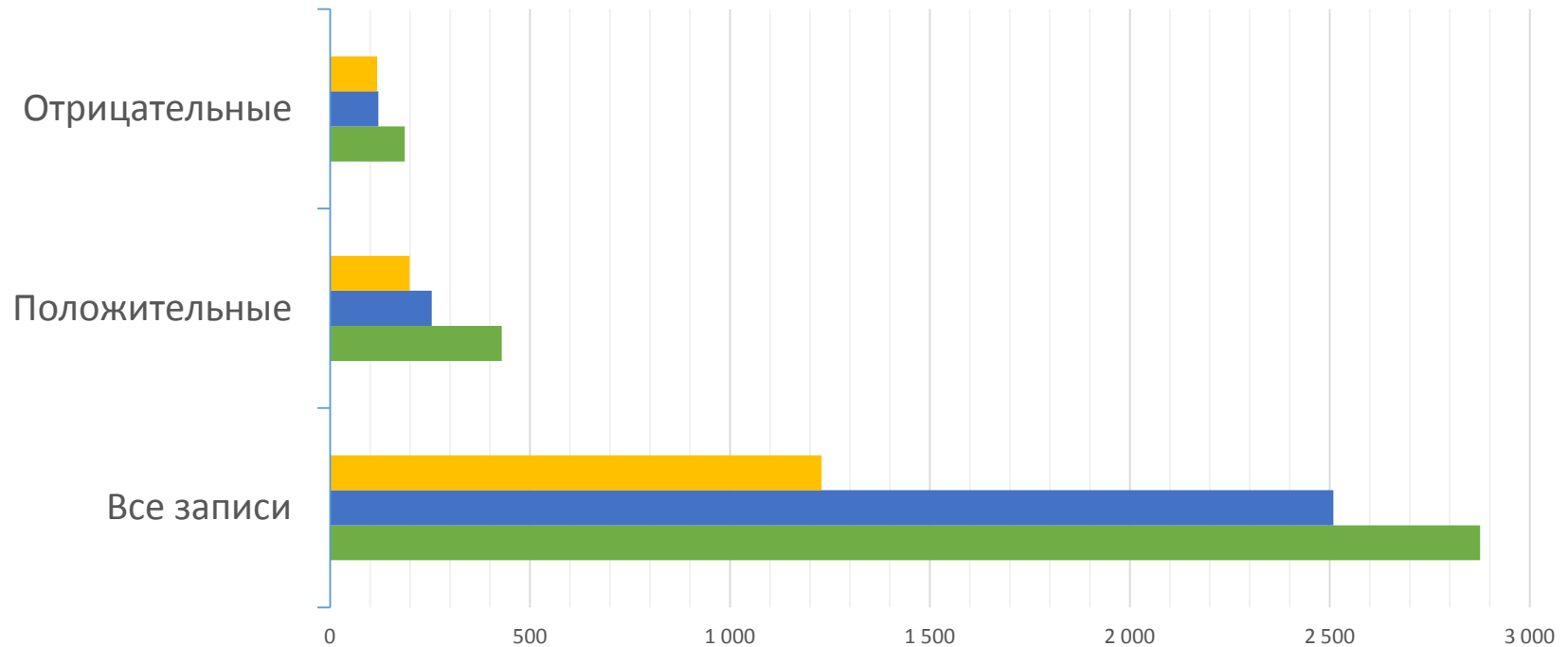
**Аналитические
срезы**

**Анализ и
интерпретация
результата**

- ✓ частота употреблений
- ✓ уровень корреляции
- ✓ сочетания частей речи

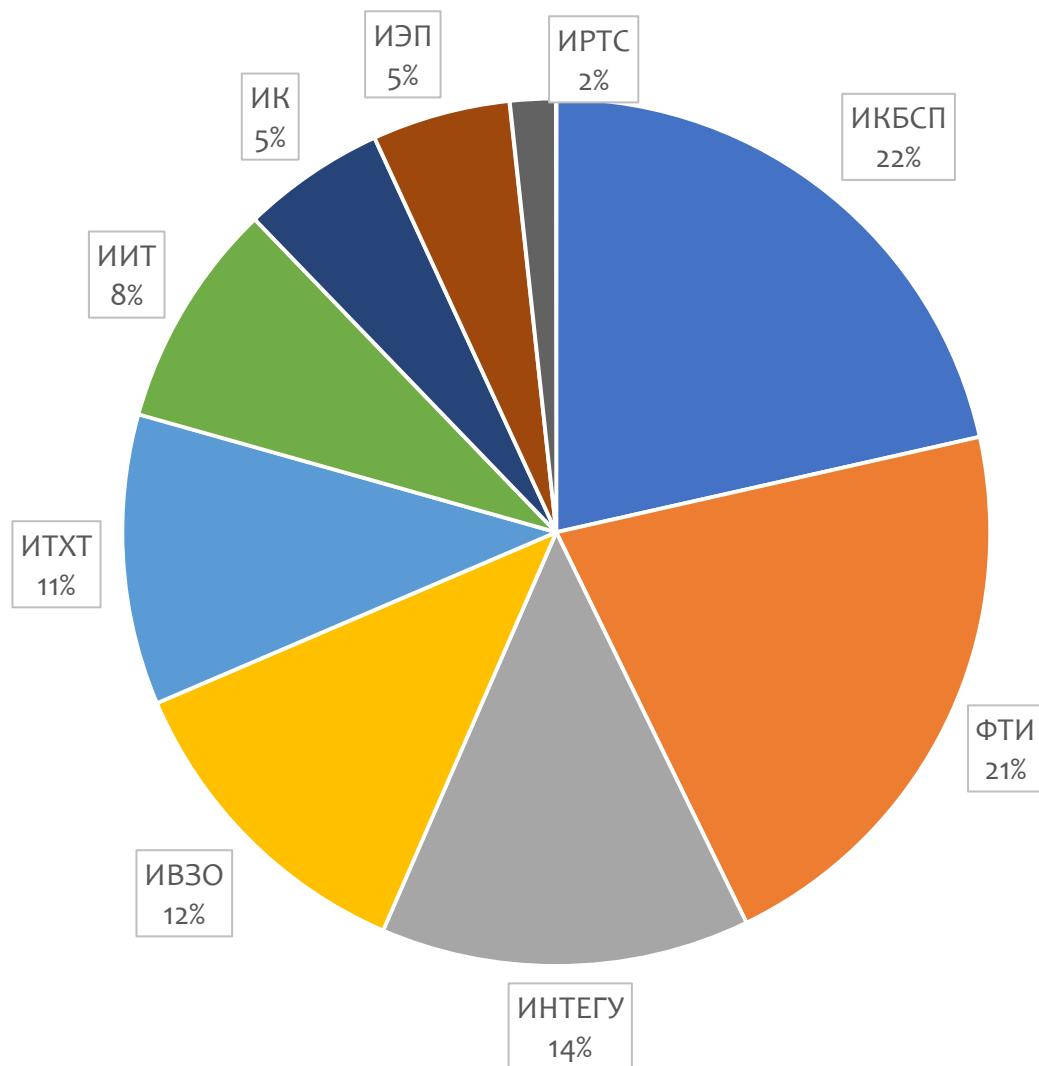
5 759 записей

о кампусах университета: МИРЭА, МГУПИ, МИТХТ



	Все записи	Положительные	Отрицательные
■ МИТХТ	1229	199	118
■ МГУПИ	2 509	254	121
■ МИРЭА	2 876	429	187

Самый популярный институт?



■ ИКБСП ■ ФТИ ■ ИНТЕГУ ■ ИБЗО ■ ИТХТ ■ ИИТ ■ ИК ■ ИЭП ■ ИРТС

Кто задает больше вопросов?



О чем думает первокурсник?

Значения	Частота	Корреляция
перевестись	90	4,7
отчислить	72	4,3
учиться	219	3,6
получать	63	2,4
поступать	91	2,4
закрывать	57	2,3
ходить	100	2,3
сдать	155	2,2
остаться	72	2,2
помочь	114	2,1

Навигация по фасетам

Отфильтровать

- Часть речи
 - Существительное
 - Нарицательное существительное
 - Другое
 - Глагол
 - Прилагательное
 - Наречие
 - Союз
 - Междометие
 - Числительное
- Составляющая фразы
- Выражение отношений
 - Магистратура
 - Преподаватели
 - Бакалавриат
 - Прочее

Документы

Фасеты

Временной ряд

Отклонения

Тенденции

Пары фасет

Просмотреть: Значения Оператор: OR

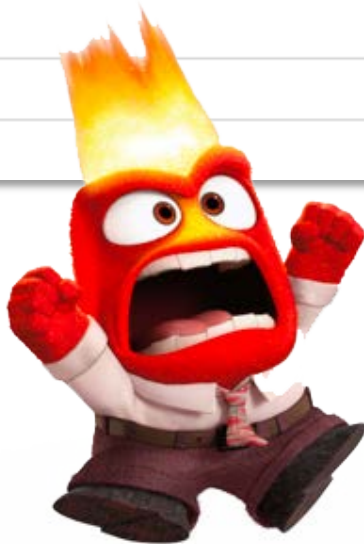
Отфильтровать:

Значения	Частота	Корреляция
☐ линалу	22	5,3
☐ матану	43	4,2
☐ перваков	37	3,6
☐ заочка	20	3,4
☐ перваки	26	3,2
☐ лабы	33	3,2
☐ Скай	31	3,0
☐ преподы	28	2,6
☐ первак	18	2,6
☐ универ	58	2,4
☐ физру	25	2,3



Второй курс ненавидит первый?

☐ Значения	Частота	Корреляция ▼
☐ хт	12	7,6
☐ матану	20	3,7
☐ преподы	19	3,6
☐ хим	11	3,4
☐ первак	12	3,4
☐ перваков	17	3,0
☐ перваки	13	2,9
☐ изи	14	2,8
☐ заочка	9	2,6
☐ лабы	13	2,2





а **ВЫ** пойдете
на военную кафедру?

☐ Значения	Частота	Корреляция
☐ третий	78	21,9
☐ военный	24	7,0
☐ очный	19	6,1
☐ второе	32	5,6
☐ целое	10	3,9
☐ старший	10	3,8

☐ Значения	Частота	Корреляция
☐ третий ... курс	72	97,9
☐ на ... курс	35	7,8
☐ на ... кафедра	26	4,7
☐ военный ... кафедра	21	7,8
☐ в ... вуз	20	3,0



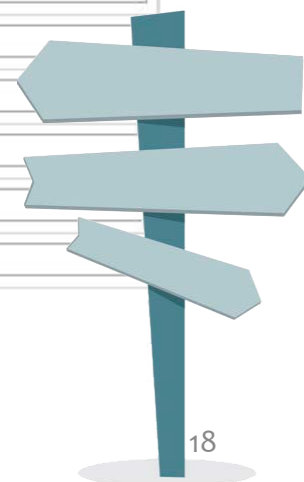
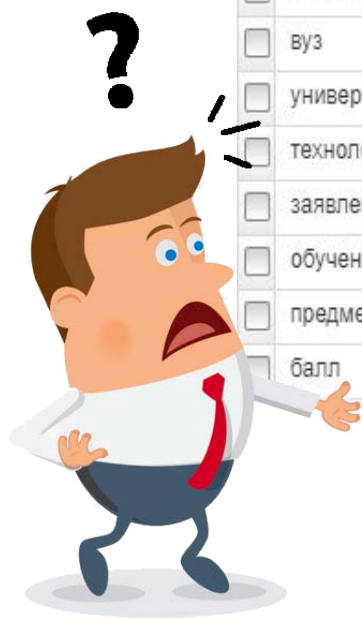
Студенты хотят учиться?

☐ Значения	Частота	Корреляция ▼
☐ курс	173	10,7
☐ долг	24	6,0
☐ отчисление	10	4,4
☐ магистратура	18	4,3
☐ диплом	20	4,0



Что волнует абитуриента?

☐ Значения	Частота	Корреляция
☐ зачисление	252	10,9
☐ согласие	150	9,6
☐ поступление	159	7,9
☐ оригинал	74	4,3
☐ бюджет	88	3,8
☐ специальность	118	3,7
☐ факультет	96	3,6
☐ учеба	69	3,5
☐ направление	149	3,2
☐ вуз	178	3,2
☐ университет	99	2,9
☐ технология	84	2,9
☐ заявление	108	2,8
☐ обучение	139	2,7
☐ предмет	81	2,7
☐ балл	97	2,7



Заключение

Полученные результаты могут:

- ✓ выявить острые темы и предпосылки к разного рода негативным явлениям
- ✓ усовершенствовать F.A.Q.
- ✓ сформировать облако тегов



СПАСИБО ЗА ВНИМАНИЕ!

