

Текстовая аналитика в деле

Кейс в образовании

Алексей Пятов // Руководитель группы текстовой аналитики SAS Russia
Константин Дудников // Эксперт по текстовой аналитике SAS Russia

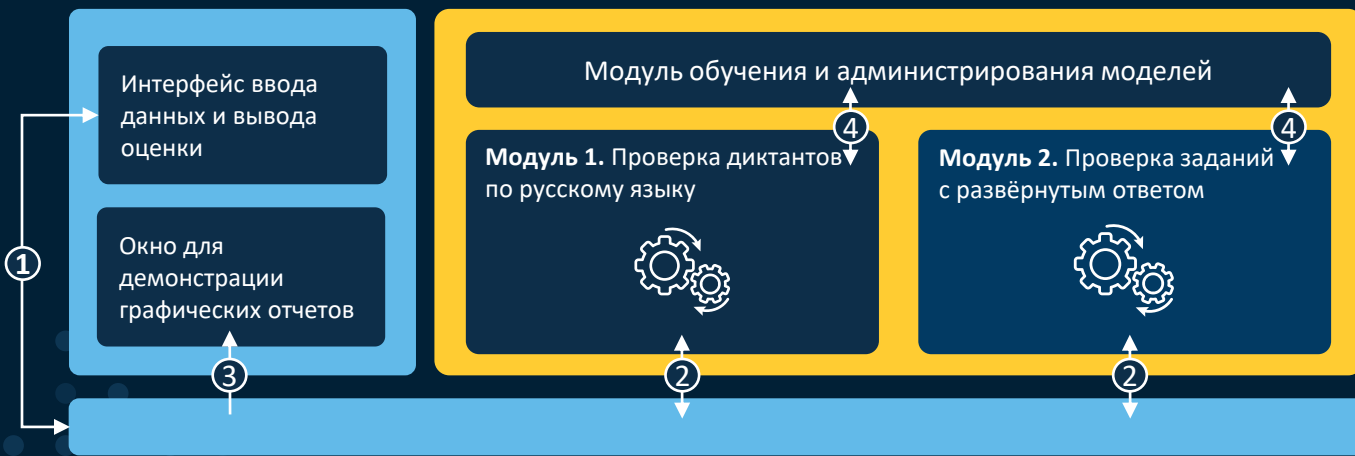
210510

Концепт-дизайн системы автоматической проверки

Цель системы – снизить финансовые и временные затраты на организацию качественной проверки диктантов и заданий с развёрнутым ответом по естественным и гуманитарным наукам.

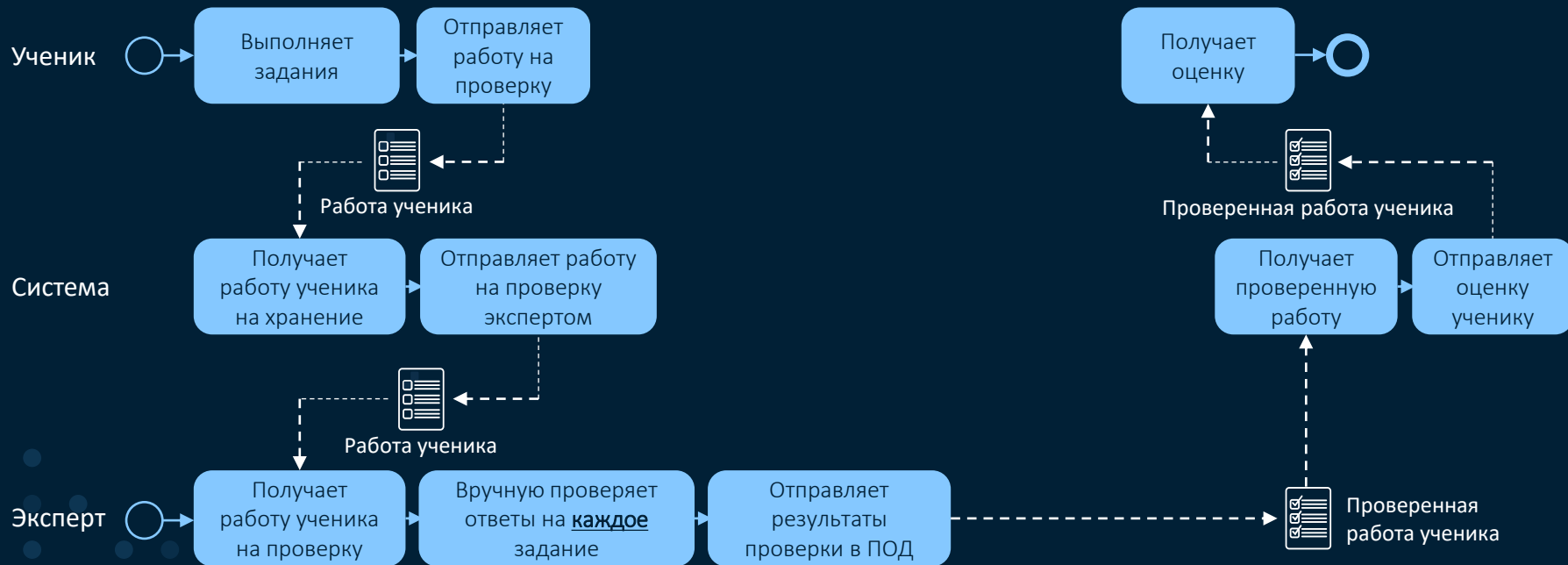
ИС ***
(1 этап – «zzz»)

Система автоматической проверки



- ① Отправка работ в хранилище. Вывод результатов проверки.
- ② Проверка полученных работ. Отправка результатов.
- ③ Подсчет статистики и демонстрация отчета
- ④ Обновление и мониторинг качества работы модулей

Процесс проверки заданий с развернутым ответом – AS IS



Процесс проверки заданий с развернутым ответом – **TO BE**

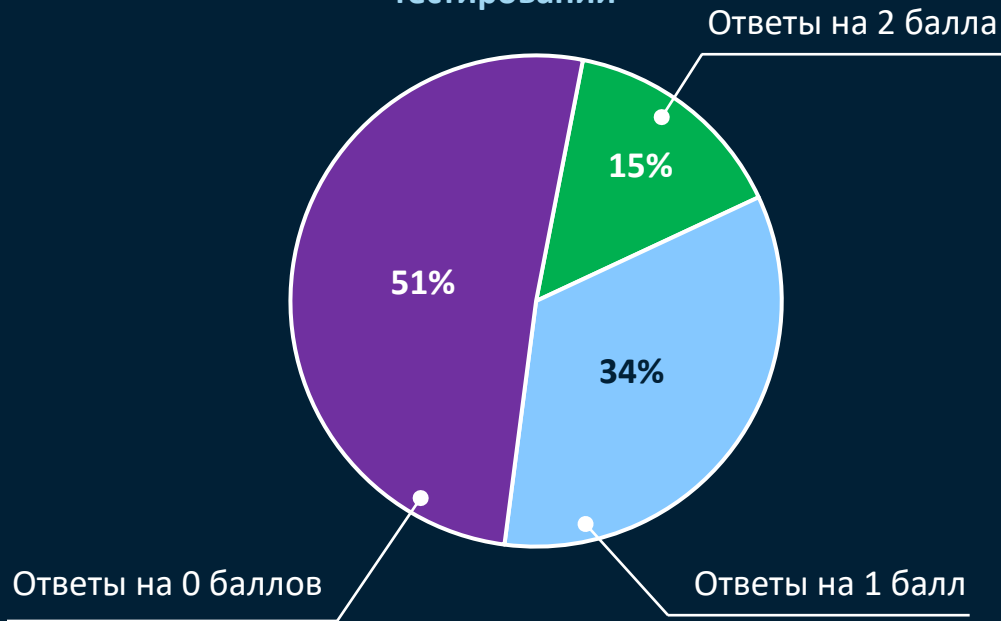


«0 баллов» – преобладающая оценка заданий

Более 75 тыс. развёрнутых ответов
было собрано на тестировании
по предмету *** в январе 2019 г.

Эксперты вручную проверяли
каждый ответ.

Распределение оценок на диагностическом
тестировании*



Результаты внутреннего тестирования

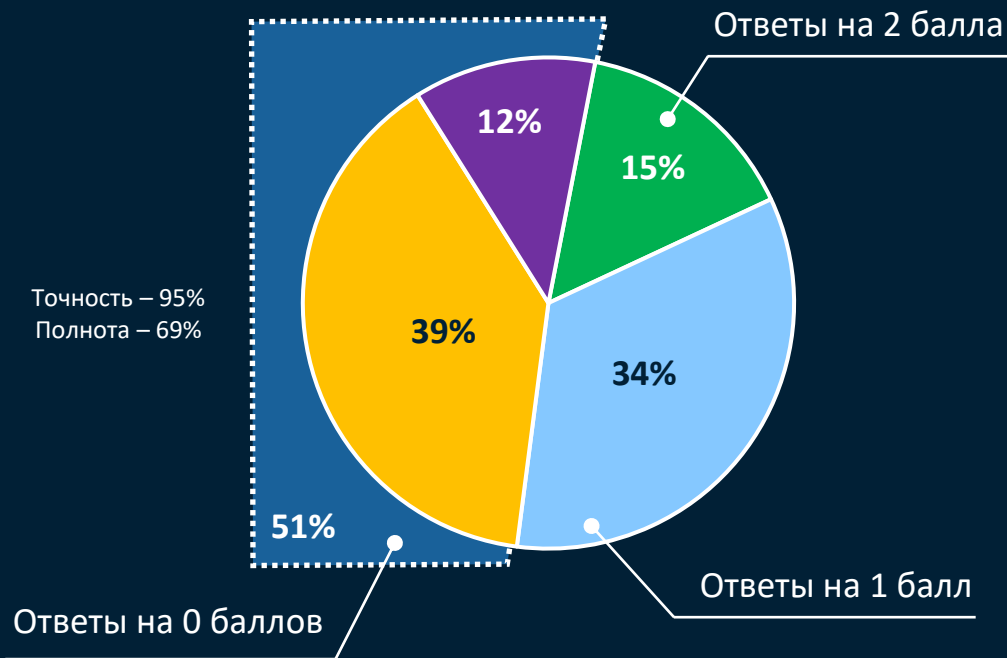
Более 75 тыс. развёрнутых ответов было собрано на тестировании по предмету *** в январе 2019 г.

Эксперты вручную проверяли каждый ответ.

39%

всех ответов модуль 2 способен проверять автоматически

Распределение оценок на диагностическом тестировании



Описание подхода

Описание подхода: проблема



Эксперт

1. Видит работу

2. Видит критерии

3. Понимает смысл
и выставляет оценку

Задача

Аэростат поднимается до определённой высоты. Запишите условие, при котором подъём аэростата прекратится.

Критерии

Элементы содержания верного ответа

(допускаются иные формулировки, не искажающие смысл ответа)

Подъём аэростата прекратится, когда сила Архимеда станет примерно равна или меньше силы тяжести /если Архимедова сила станет примерно равна или меньше суммы сил тяжести и сопротивления воздуха.

Указания к оцениванию	Баллы
Дан правильный ответ	1
Другие варианты ответа	0
Максимальный балл	1

Работа 1

1

Подъём аэростата прекратится тогда, когда архимедова сила, действующая на шар станет меньше силы тяжести аэростата.

Работа 2

1

Для поднятия шара в воздух, необходимо, чтобы архимедова сила, действующая на шар, была больше силы тяжести. Когда это условие перестает действовать то и подъём аэростата прекратится.

Работа 3

0

Подъём аэростата прекратится, если открыть клапан, и выпустить находящийся в шаре газ



Компьютер

1. Видит байты

2. Видит байты

3. Не понимает, что
происходит



Как донести до
компьютера смысл?



Вектора!

Векторное представление чего угодно

Как сравнить 2 яблока?



VS.



С помощью векторов



цвет	размер	вкус	страна	цена
0	1	0	0	1

VS.



цвет	размер	вкус	страна	цена
1	0	1	1	0

Признаки

(измерения):

цвет	размер	вкус	страна	цена
'зеленый': 1 'красный': 0	'большой': 1 'маленький': 0	'кислый': 1 'сладкий': 0	'Польша': 1 'Россия': 0	'дорогой': 1 'дешевый': 0

Векторное представление слов

one-hot encoding

Как сравнить 2 слова?



С помощью векторов

«Аэростат» VS. «Шар»

«Аэростат»

поднятие	воздух	шар	сила	аэростат	*
0	0	0	0	1	

VS.

«Шар»

поднятие	воздух	шар	сила	аэростат	*
0	0	1	0	0	

Признаки (измерения):

- слова из словаря. Количество признаков = размер словаря.

Настоящие вектора

поднятие	воздух	шар	сила	аэростат	...
'Аэростат': 0 'Шар': 0	'Аэростат': 0 'Шар': 0	'Аэростат': 0 'Шар': 1	'Аэростат': 0 'Шар': 0	'Аэростат': 1 'Шар': 0	'Аэростат': 0 'Шар': 0

→ Сотни тысяч признаков

* Сработает при условии, что в словаре 5 слов.

Векторное представление слов

one-hot encoding

Как сравнить 2 слова?



С помощью векторов

«Аэростат» VS. «Шар»

Признаки (измерения):

- слова из словаря. Количество признаков = размер словаря.

	поднятие	воздух	шар	сила	аэростат	*
«Аэростат»	0	0	0	0	1	

VS.

	поднятие	воздух	шар	сила	аэростат	*
«Шар»	0	0	1	0	0	

Проблемы подхода



Вектора большой размерности тяжелы для вычислений



One-hot вектора не отражают семантические отношения слов

* Сработает при условии, что в словаре 5 слов.

Сокращаем размерность вектора слова

Слова, встречающиеся в схожих контекстах, имеют схожее значение.

[Harris, Z. 1956]

Признаки (измерения):

- неинтерпретируемые характеристики.
Количество признаков произвольное,
обычно 200-700

«Аэростат»

0,25	1,04	-3,78	...	0,66	-0,13
------	------	-------	-----	------	-------

512 чисел (признаков)

«Шар»

0,34	0,99	-3,56	...	1,01	0,45
------	------	-------	-----	------	------

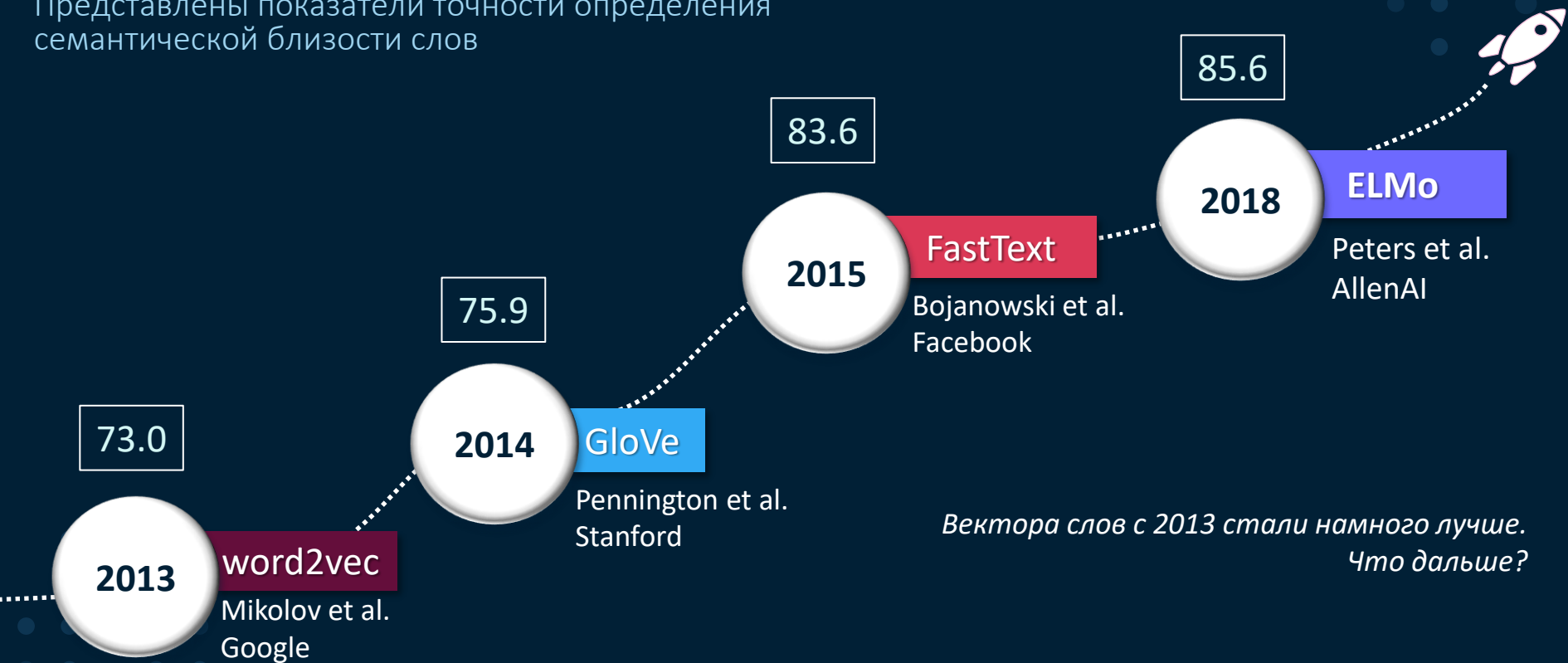
Слова «Шар» и «аэростат»
встречаются в схожих контекстах

Их вектора будут близкими



Хорошо ли это работает?

Представлены показатели точности определения семантической близости слов



*Вектора слов с 2013 стали намного лучше.
Что дальше?*

Как получить вектора предложений?

“Для поднятия аэростата в воздух, необходимо, чтобы архимедова сила, действующая на шар, была больше силы тяжести.”



для	в	е	к	т	о	р
поднятие	в	е	к	т	о	р
аэростат	в	е	к	т	о	р
в	в	е	к	т	о	р
	...					

Магия

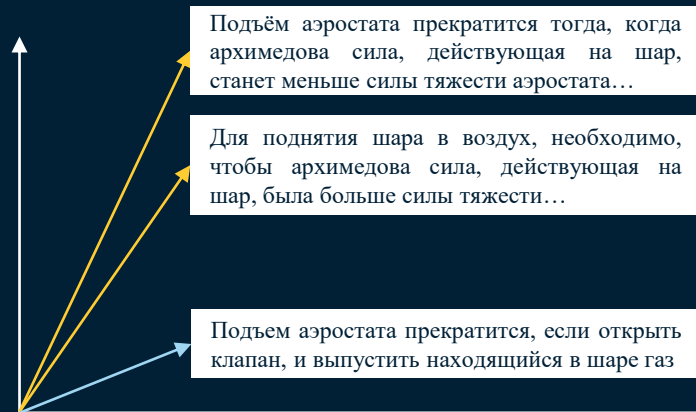


[0.12, 3.69, ..., 0.20, 3.17]

512 измерений

Виды магии

- Нормализация: усреднение, softmax
- Нейронные сети: CNN, RNN
- Механизм attention
- Позиционное кодирование
- и пр.



Описание подхода: а что дальше?



Эксперт

1. Видит работу

2. Видит критерии

3. Понимает смысл
и выставляет оценку

Задача

Аэростат поднимается до определённой высоты. Запишите условие, при котором подъём аэростата прекратится.

Критерии

Элементы содержания верного ответа

(допускаются иные формулировки, не искажающие смысл ответа)

Подъём аэростата прекратится, когда сила Архимеда станет примерно равна или меньше силы тяжести /если Архимедова сила станет примерно равна или меньше суммы сил тяжести и сопротивления воздуха.

Указания к оцениванию	Баллы
Дан правильный ответ	1
Другие варианты ответа	0
Максимальный балл	1

Работа 1

1

Подъём аэростата прекратится тогда, когда архимедова сила, действующая на шар станет меньше силы тяжести аэростата.

Работа 2

1

Для поднятия шара в воздух, необходимо, чтобы архимедова сила, действующая на шар, была больше силы тяжести. Когда это условие перестает действовать то и подъём аэростата прекратится.

Работа 3

0

Подъем аэростата прекратится, если открыть клапан, и выпустить находящийся в шаре газ



Компьютер

1. Видит вектор

2. Понимает смысл

3. Не может поставить
оценку



Как научить компьютер
выставлять оценку?



Машинное обучение!

Что такое машинное обучение?

Работа	Оценка
Для поднятия аэростата...	0
Подъем аэростата прек...	1
Если открыть клапан и ...	0
Когда сила архимеда, д...	1
...	
Сила тяжести должна ...	

Проверенные документы



Определение разницы
между двумя классами



Выставление оценки
новой работе

Подъем аэростата
прекратится, если открыть
клапан, и выпустить
находящийся в шаре газ

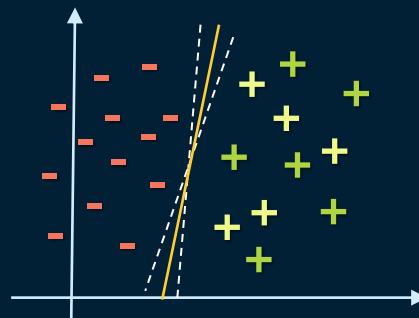
Оценка «0»

Концептуально

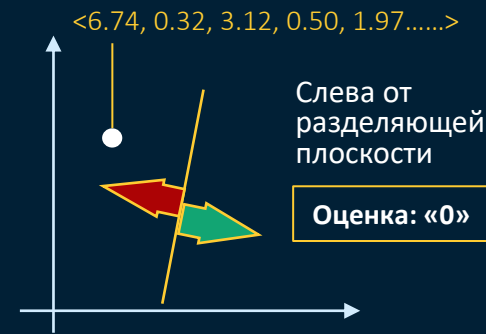
Под капотом

Вектор работы	Оценка
<0.15, 0.36, 1.12, 4.80...>	0
<9.72, 0.01, 3.41, 1.25...>	1
<0.50, 0.35, 0.97 5.12...>	0
<5.64, 0.11, 2.92, 1.32...>	1
...	
<6.45, 0.02, 4.09, 1.12...>	1

Вектора проверенных
документов



Поиск разделяющей поверхности
в многомерном пространстве



Определение класса
нового объекта

Слева от
разделяющей
плоскости

Оценка: «0»

Подготовка размеченных данных... 250MB
Построение векторной модели для коллекции докуме...
Сохранение векторной модели... done.

Векторизация текстов работ... done.

Обучение классификатора... done.
Инициализация случайной разделяющей гиперплоскости..
Эпоха обучения: 1...
Точность на тестовых данных: 70%.
Эпоха обучения: 2...
Точность на тестовых данных: 82%.
...
Эпоха обучения: 27...
Точность на тестовых данных: 95%.

Сохранение модели классификации ... done.

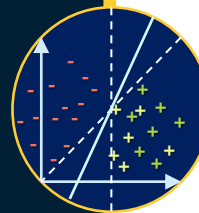
Обработка работы 1...
Векторизация документа...
Применение модели к вектору документа...
Класс документа: 0, score: 0.950.

Обработка работы 2...
Векторизация документа...
Применение модели к вектору документа...
Класс документа: 0, score: 0.962.



Анализ текстов работ

векторное представление
документов



Разделение на два класса

поиск разделяющей гиперплоскости
в 512-мерном пространстве



Обработка новых документов

векторное представление новых
работ и проставление им класса

Подготовка размеченных данных... 250MB
Построение векторной модели для коллекции докуме...
Сохранение векторной модели... done.

Векторизация текстов работ... done.

Обучение классификатора... done.
Инициализация случайной разделяющей гиперплоскости..
Эпоха обучения: 1...
Точность на тестовых данных: 70%.
Эпоха обучения: 2...
Точность на тестовых данных: 82%.
...
Эпоха обучения: 27...
Точность на тестовых данных: 95%.

Сохранение модели классификации ... done.

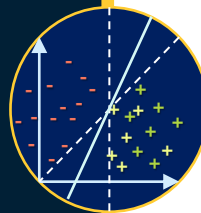
Обработка работы 1...
Векторизация документа...
Применение модели к вектору документа...
Класс документа: 0, score: 0.950.

Обработка работы 2...
Векторизация документа...
Применение модели к вектору документа...
Класс документа: 0, score: 0.962.



Анализ текстов работ

векторное представление документов



Разделение на два класса

поиск разделяющей гиперплоскости в 512-мерном пространстве



Обработка новых документов

векторное представление новых работ и проставление им класса

Подготовка размеченных данных... 250MB
Построение векторной модели для коллекции докуме...
Сохранение векторной модели... done.

Векторизация текстов работ... done.

Обучение классификатора... done.
Инициализация случайной разделяющей гиперплоскости..
Эпоха обучения: 1...
Точность на тестовых данных: 70%.
Эпоха обучения: 2...
Точность на тестовых данных: 82%.
...
Эпоха обучения: 27...
Точность на тестовых данных: 95%.

Сохранение модели классификации ... done.

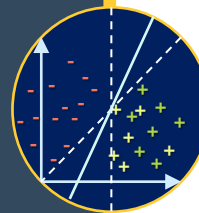
Обработка работы 1...
Векторизация документа...
Применение модели к вектору документа...
Класс документа: 0, score: 0.950.

Обработка работы 2...
Векторизация документа...
Применение модели к вектору документа...
Класс документа: 0, score: 0.962.



Анализ текстов работ

векторное представление документов



Разделение на два класса

поиск разделяющей гиперплоскости в 512-мерном пространстве



Обработка новых документов

векторное представление новых работ и проставление им класса

Подготовка размеченных данных... 250MB
Построение векторной модели для коллекции докуме...
Сохранение векторной модели... done.

Векторизация текстов работ... done.

Обучение классификатора... done.
Инициализация случайной разделяющей гиперплоскости..

Эпоха обучения: 1...

Точность на тестовых данных: 70%.

Эпоха обучения: 2...

Точность на тестовых данных: 82%.

...

Эпоха обучения: 27...

Точность на тестовых данных: 95%.

Сохранение модели классификации ... done.

Обработка работы 1...

Векторизация документа...

Применение модели к вектору документа...

Класс документа: 0, score: 0.950.

Обработка работы 2...

Векторизация документа...

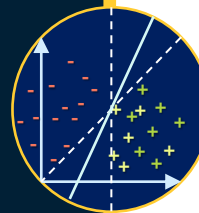
Применение модели к вектору документа...

Класс документа: 0, score: 0.962.



Анализ текстов работ

векторное представление
документов



Разделение на два класса

поиск разделяющей гиперплоскости
в 512-мерном пространстве



Обработка новых документов

векторное представление новых
работ и проставление им класса

Как оценить результаты?

Метрики качества

$$\text{Precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}$$

$$\text{Recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$

$$\text{Accuracy} = \frac{\text{true positives} + \text{true negatives}}{\text{true positives} + \text{false positives} + \text{true negatives} + \text{false negatives}}$$

		Данные экспертов	
		1	0
Предсказание системы	1	true positives	false positives
	0	false negatives	true negatives

Предсказываем класс «1»

Как оценить результаты?

Accuracy paradox



$$\text{Precision} = \frac{0}{0 + 125} = \mathbf{0}$$

$$\text{Recall} = \frac{0}{0 + 1} = \mathbf{0}$$

$$\text{Accuracy} = \frac{0 + 1125}{0 + 125 + 1125 + 1} = \mathbf{0,9}$$

*Предсказание
системы*

Данные экспертов

	1	0
1	0	125
0	1	1125

*Предсказываем
класс «1»*

Система не может предсказать класс «1», но точность всё равно высокая

Как оценить результаты?

В приоритете точность

95%

Насколько уверенным можно быть
в результатах работы системы?

*Какова вероятность,
что проверенным работам
выставлена правильная оценка?*

Precision

ТОЧНОСТЬ



69%

Сколько **нулевых** работ система
проверит за экспертов?

Recall

ПОЛНОТА



Результаты внутреннего тестирования

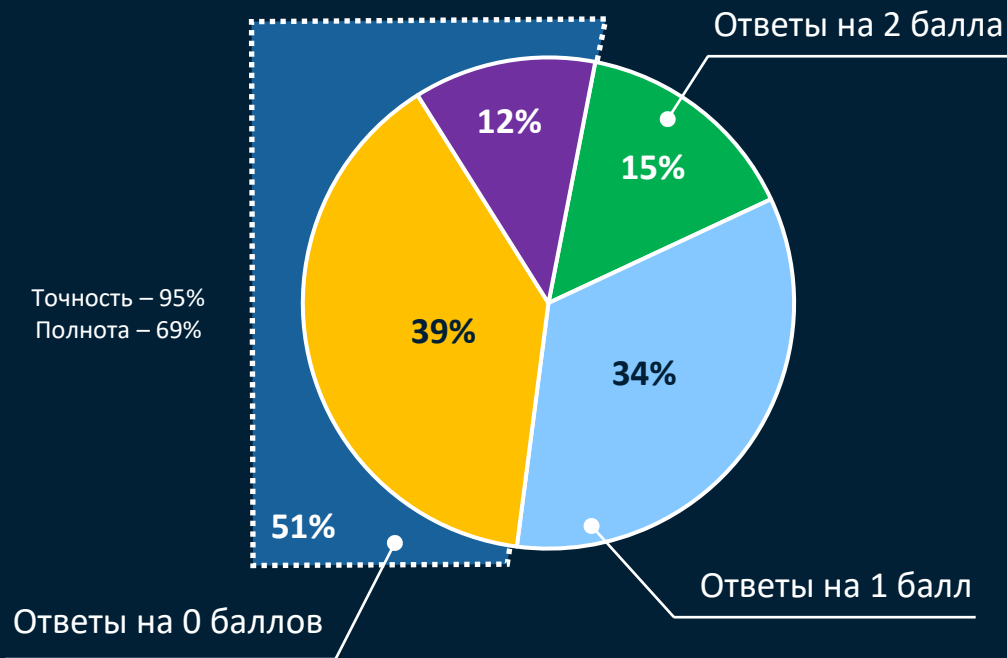
Более 75 тыс. развёрнутых ответов было собрано на тестировании по предмету *** в январе 2019 г.

Эксперты вручную проверяли каждый ответ.

39%

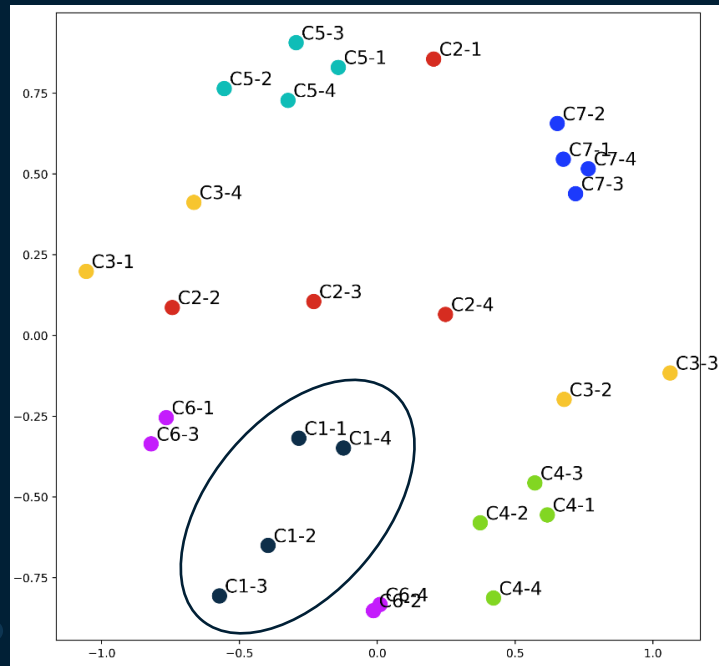
всех ответов модуль 2 способен проверять автоматически

Распределение оценок на диагностическом тестировании

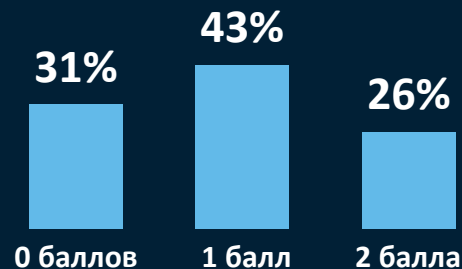


Семантический анализатор заданий

Точность – 99%
Полнота – 90%

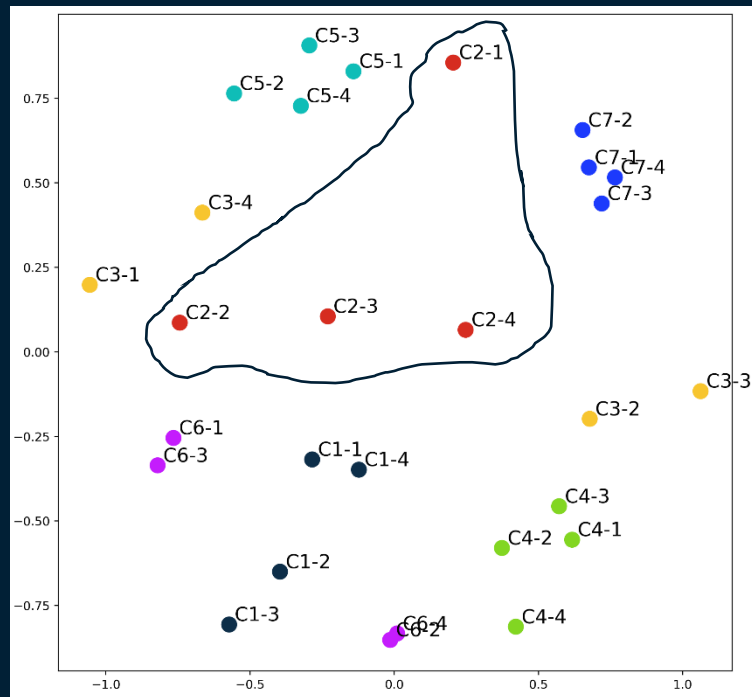


Карта семантической близости
вариантов заданий



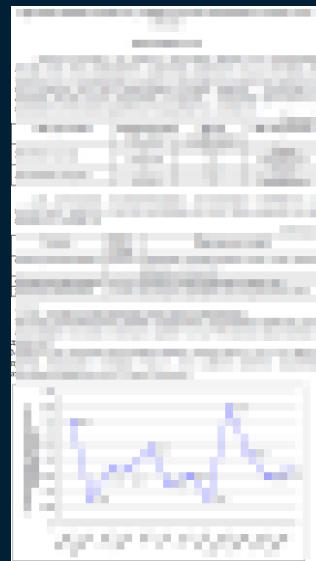
Семантический анализатор заданий

Точность – 92%
Полнота – 50%



Карта семантической близости
вариантов заданий

1
3



2



4



53%



0 баллов

36%



1 балл

11%

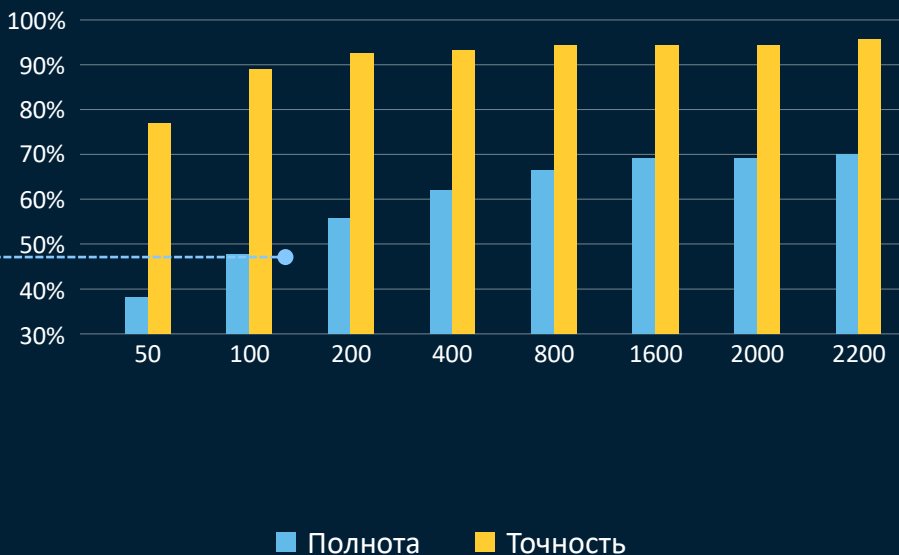


2 балла

Требования к запуску модуля 2

1. Обучить модели по всем заданиям с учетом имеющихся данных
2. Изменить существующие процессы проверки свободных ответов:
 - на первоначальном этапе – проводить ручную проверку заданий
 - после накопления **достаточного** числа работ обучать аналитическую модель
 - исключить из экспертной проверки задания, проверенные автоматически

Зависимость полноты и точности от размера обучающей выборки



Демонстрация работы системы

Вопросы можете задавать в Telegram:

@pyatov – Алексей Пятов

@krinistopen – Константин Дудников

sas.com

