Анализ тональности аспектов товаров на основе отзывов покупателей

Киреева Алена Николаевна

Аспекты — это характеристики продукта/услуги, к которым пользователь выражает своё отношение (позитивное/негативное/нейтральное).

Amazon:

Read reviews that mention

sound quality	battery life	transparency mode	airpod pros
much better	original airpods	highly recommend	regular airpods
worth the money	noise cancelation	stopped working	airpod

Аспекты — это характеристики продукта/услуги, к которым пользователь выражает своё отношение (позитивное/негативное/нейтральное).

Яндекс Маркет:

Вот о чём пишут чаще всего

Этот отзыв написал наш умный алгоритм — он всё прочитал и выделил главное

Достоинства

«Шумоподавление, хороший звук.» «Неплохой звук, качественная сборка, хорошо сидят в ушах.»

«Качество звука, удобство.» «Шумоподавление.»

Недостатки

«Не заметил.» «Бывает выпадают из уха.» «Быстро царапается корпус, даже в чехле.» «Звук.»

Полезная информация? Да Нет

Цель: автоматическое определения тональности (положительной или отрицательной) выделенных аспектов товаров.

BERT fine-tuning:

- предобработка текста;
- дообучение на размеченных данных.

Практическое применение: генерация отзыва-резюме

- предобработка текста;
- выделение аспектов товаров (неразмеченные данные);
- предсказание тональности на дообученном BERT;
- выбор наиболее частотных аспектов.

Ограничения

- Выделение аспектов товаров не является основной целью проекта, но является необходимой задачей.
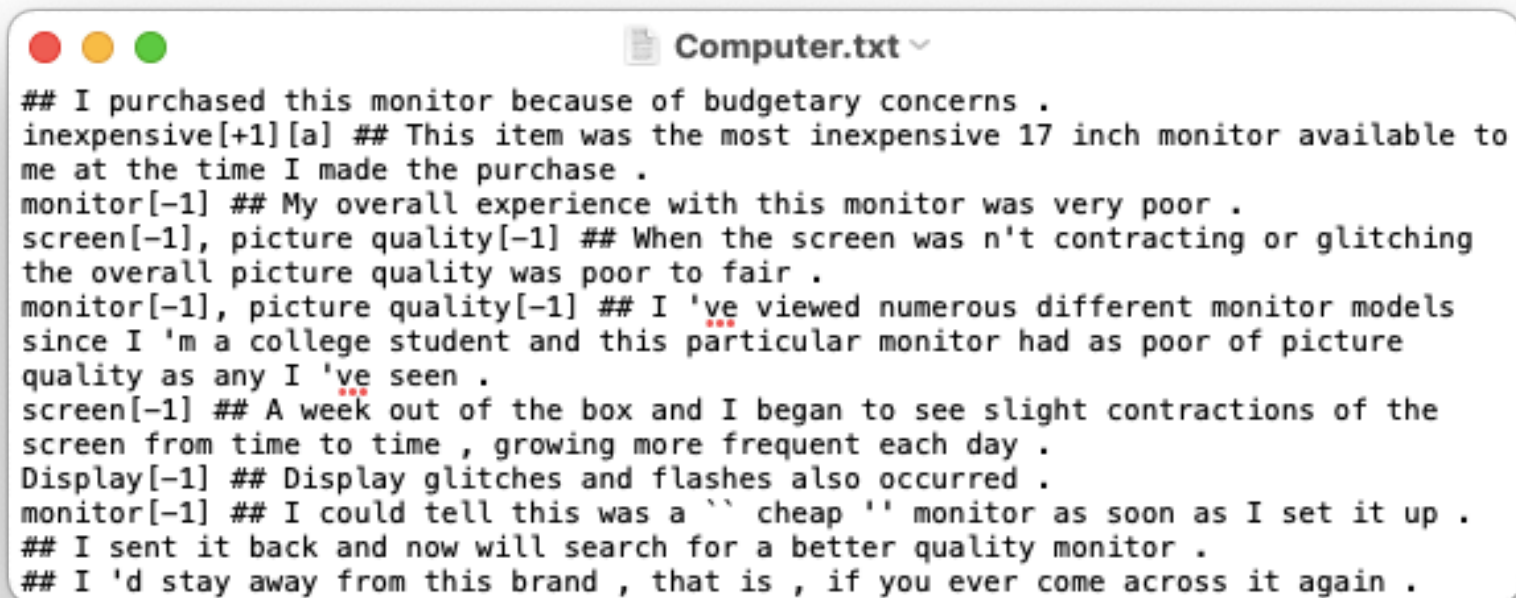
  —> простой, но менее эффективный подход

- Аспекты могут быть выражены имплицитно (неявно).

  —> исключены из обучающих данных

Наборы данных

Отзывы пользователей Amazon

- Всего 14 товаров
  из 3 датасетов:
  2004, 2008, 1015 гг.

- Аспекты - цитаты из
  предложений.

I purchased this monitor because of budgetary concerns .
inexpensive[+1][a] ## This item was the most inexpensive 17 inch monitor available to
me at the time I made the purchase .
monitor[-1] ## My overall experience with this monitor was very poor .
screen[-1], picture quality[-1] ## When the screen was n't contracting or glitching
the overall picture quality was poor to fair .
monitor[-1], picture quality[-1] ## I 've viewed numerous different monitor models
since I 'm a college student and this particular monitor had as poor of picture
quality as any I 've seen .
screen[-1] ## A week out of the box and I began to see slight contractions of the
screen from time to time , growing more frequent each day .
Display[-1] ## Display glitches and flashes also occurred .
monitor[-1] ## I could tell this was a `` cheap '' monitor as soon as I set it up .
I sent it back and now will search for a better quality monitor .
I 'd stay away from this brand , that is , if you ever come across it again .

Customer Review Datasets by Bing Liu et al.

Отзывы пользователей Amazon

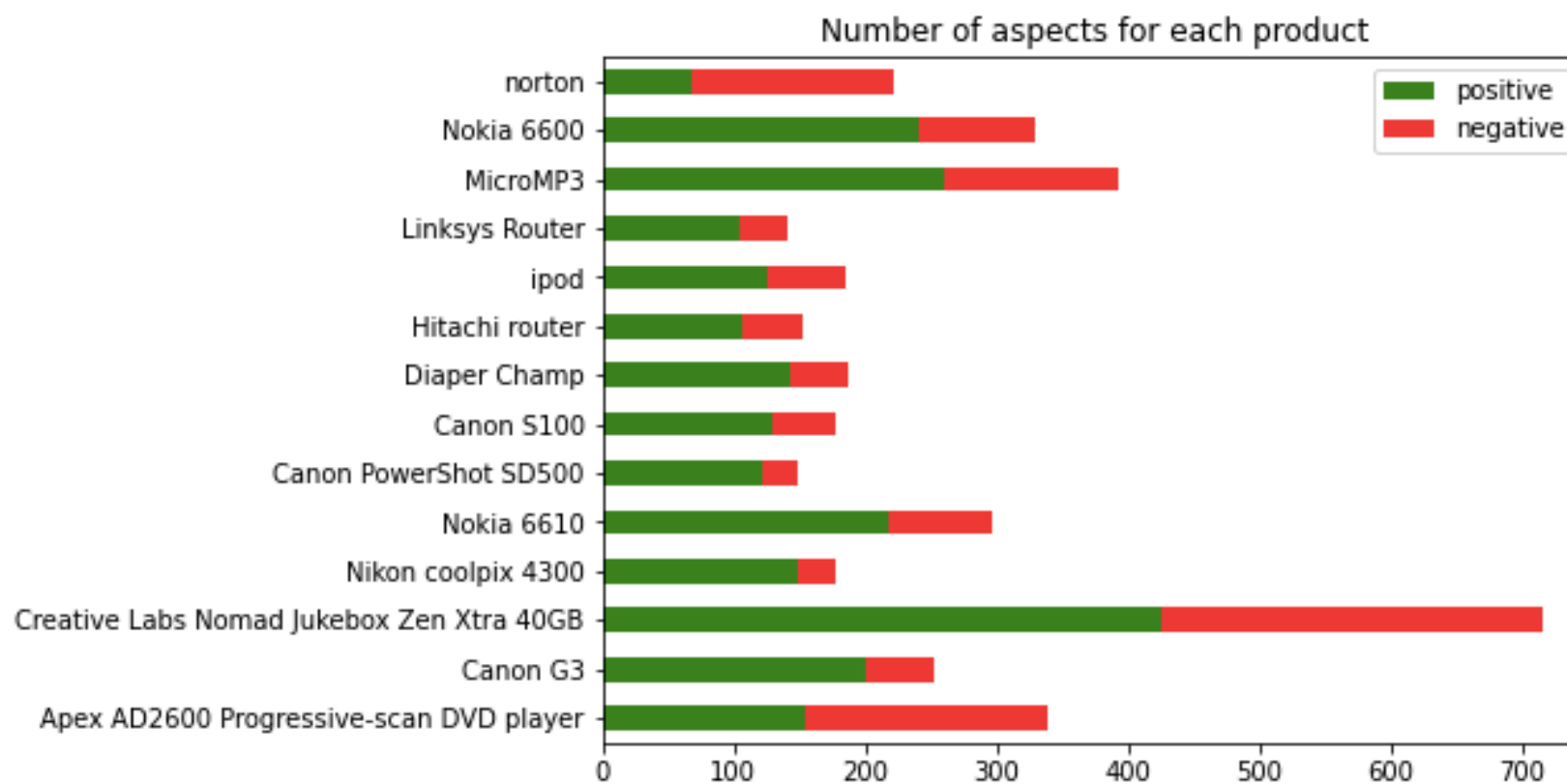|	review	aspect	sentiment	product
0	im a more happier person after discovering the...	i/p button	1	Apex AD2600 Progressive-scan DVD player
1	but , if you 're looking for my opinion of the...	dvd player	1	Apex AD2600 Progressive-scan DVD player
2	i 've had the player for about 2 years now and...	player	1	Apex AD2600 Progressive-scan DVD player
3	i 've had the player for about 2 years now and...	sound	0	Apex AD2600 Progressive-scan DVD player
4	for the price it is a well spent investment !	price	1	Apex AD2600 Progressive-scan DVD player
...	...	...	...	...
3711	I had tried to uninstall Norton Internet Secur...	uninstall	0	norton
3712	Of course, if you try to reach Symantec custom...	Symantec customer service	0	norton
3713	This time, it started the uninstallation proce...	uninstallation	0	norton
3714	I simply hate Symantec. I swear, if I could ha...	Symantec	0	norton
3715	Their products have been great and have saved ...	products	1	norton

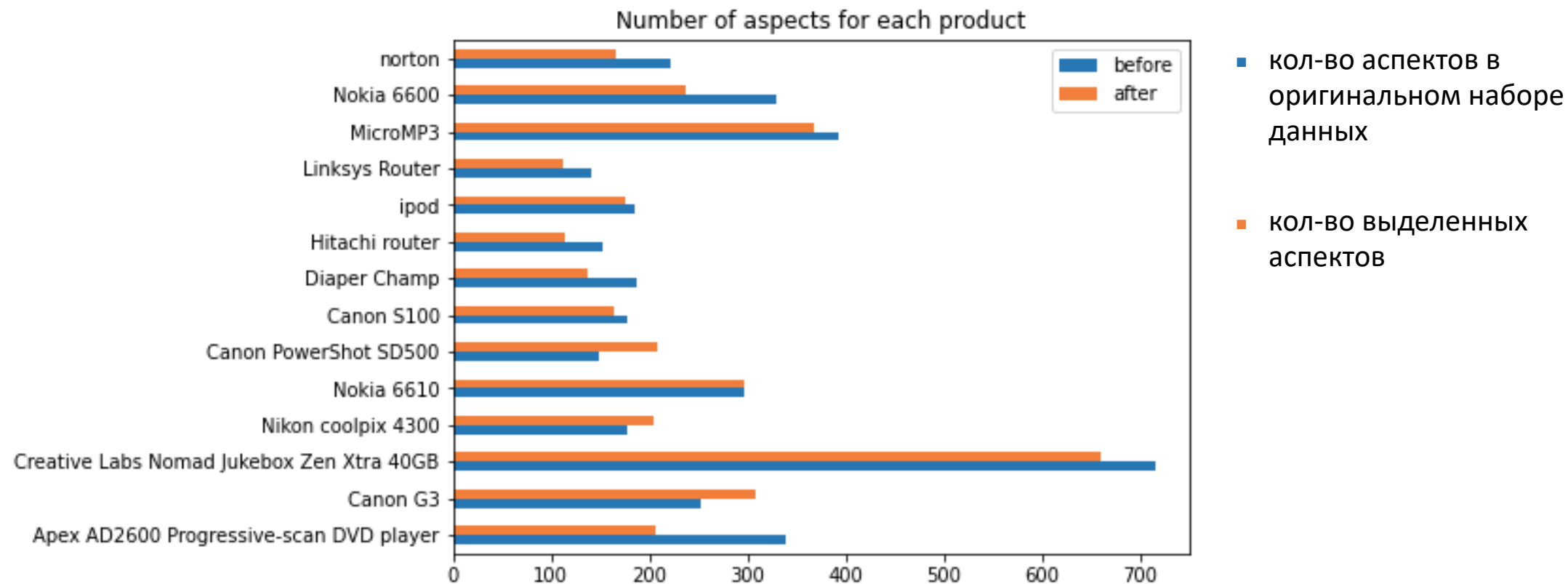3716 rows × 4 columns

Отзывы пользователей Amazon


```
Total number of aspects: 3716
Positive: 2448
Negative: 1268
```

Number of aspects for each product

Отзывы пользователей Amazon

Number of aspects for each product
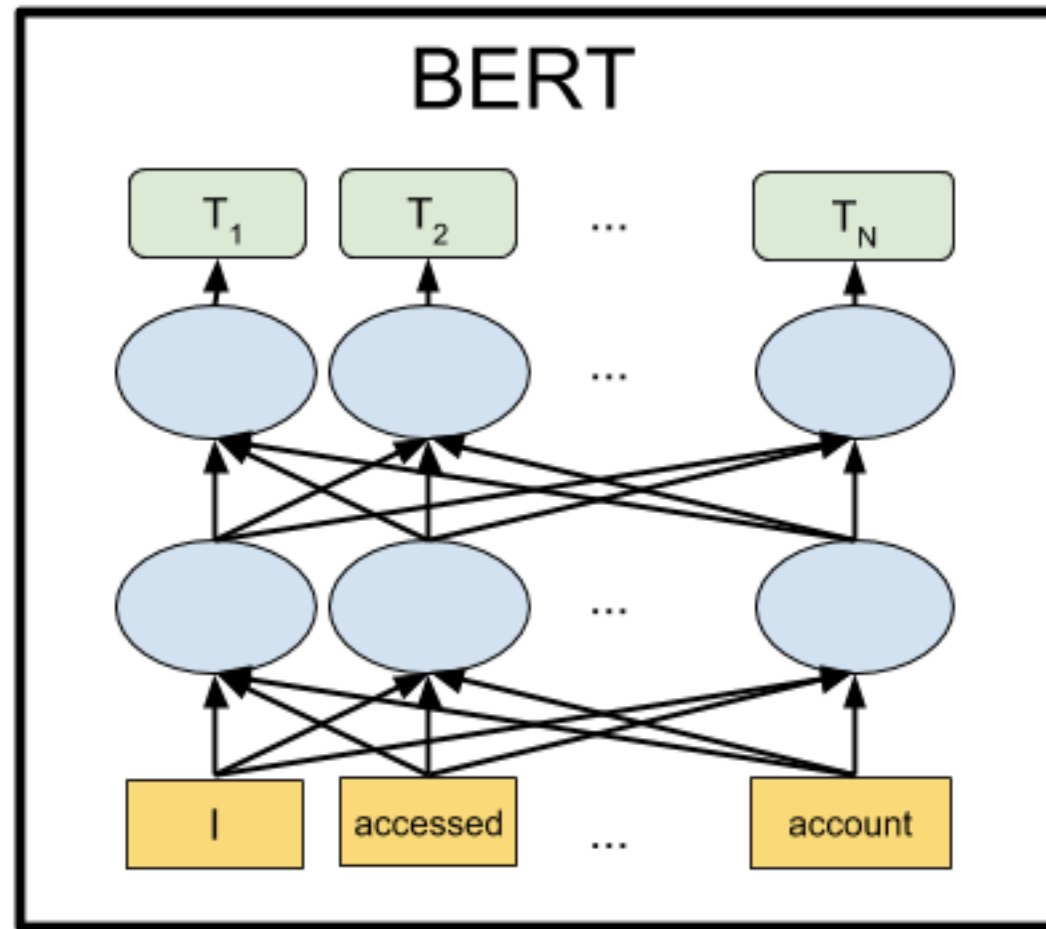
- кол-во аспектов в оригинальном наборе данных

- кол-во выделенных аспектов

Бинарная классификация тональности аспектов

BERT (Bidirectional Encoder Representations from Transformers)

BERT (Bidirectional Encoder Representations from Transformers)

- PyTorch API от HuggingFace

- Модель **bert-base-uncased**

- Архитектура **BertForSequenceClassification**

BERT fine-tuning

- Train 60%, validation 20%, test 20%

- Токенизация (отзыв + аспект)

- Класс torch dataset

- Функция расчёта метрик

- Аргументы и параметры функции Trainer()

- Обучение

- Предсказание

```python
epochs = 4
batch_size = 32 # or 24
num_steps = len(train_dataset) * epochs // batch_size
warmup_steps = num_steps // 10
num_classes = 2


training_args = TrainingArguments(
    output_dir = 'model',
    num_train_epochs = epochs,
    per_device_train_batch_size = batch_size,
    per_device_eval_batch_size = batch_size,
    warmup_steps = warmup_steps,
    weight_decay = 0.01,
    logging_dir = 'logs',
    logging_steps = 10,
    evaluation_strategy = 'epoch',
    learning_rate = 2e-5,
)

config = BertConfig.from_pretrained(
    'bert-base-uncased',
    architectures = ['BertForSequenceClassification'],
    hidden_size = 768,
    num_hidden_layers = 12,
    num_attention_heads = 12,
    hidden_dropout_prob = 0.1,
    num_labels = num_classes
)
```

BERT fine-tuning

- Высокие показатели метрик качества

- Аспекты в обучающей выборке отличаются (!)

Epoch	Training Loss	Validation Loss	Accuracy	Precision	Recall	F1
1	0.476900	0.352714	0.869630	0.915433	0.900208	0.907757
2	0.234100	0.249085	0.906667	0.960352	0.906445	0.932620
3	0.099600	0.254763	0.911111	0.950749	0.923077	0.936709
4	0.050900	0.261605	0.914074	0.949045	0.929314	0.939076

```
***** Running Evaluation *****
  Num examples = 676
  Batch size = 32
                                                    [22/22 00:02]
eval_loss 0.24553631246089935
eval_accuracy 0.9215976331360947
eval_precision 0.9372197309417041
eval_recall 0.9435665914221218
eval_f1 0.9403824521934758
eval_runtime 2.1936
eval_samples_per_second 308.168
eval_steps_per_second 10.029
epoch 4.0
```

Выделение аспектов товаров

Поиск словосочетаний в текстах отзывов

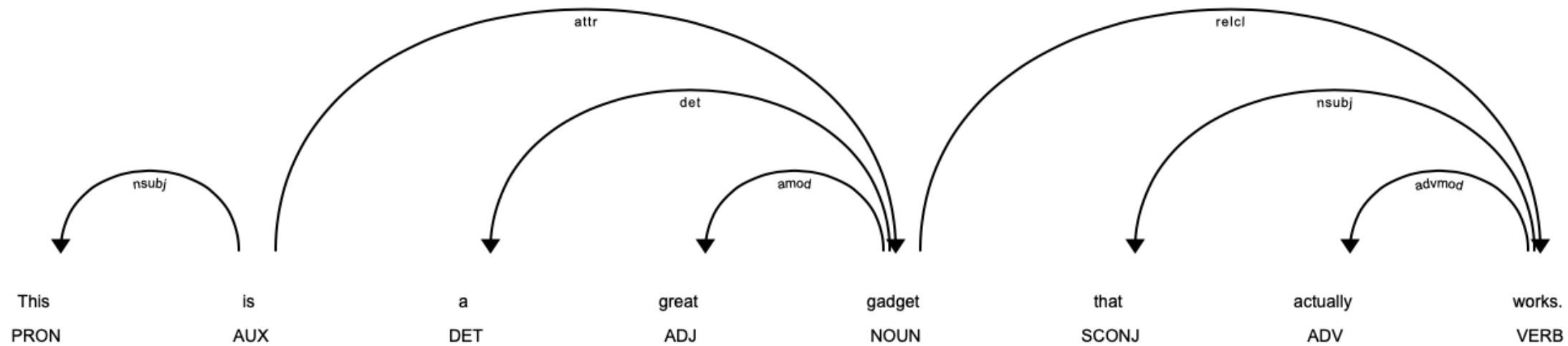Паттерны:

- (not) + adjective + noun/compound noun

- noun/compound noun+ be + (not) + adjective

spaCy

Библиотека **spaCy**, модель **en_core_web_lg**

Поиск словосочетаний в текстах отзывов
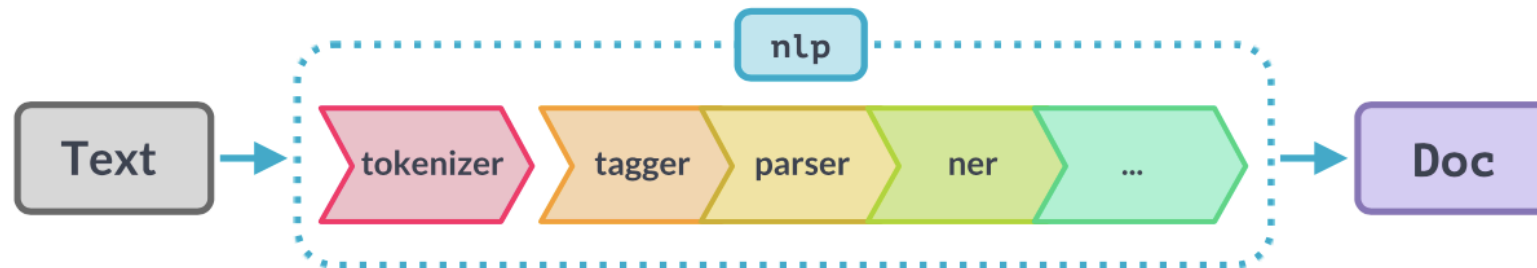
Группировка схожих словосочетаний

«great value», *«low price»*, *«right price»*, *«great price»*   —>   *«great price»*

- Косинусная мера близости векторов **.similarity()**

- Threshold = 0.8 (0.7 для потенциальных)

Выходные данные

	review	aspect	key_aspect	count	product
0	i was impressed by this very inexpensive playe...	inexpensive player	best player	16	Apex AD2600 Progressive-scan DVD player
1	i looked into buying an inexpensive dvd player...	inexpensive player	best player	16	Apex AD2600 Progressive-scan DVD player
2	this is the best dvd player i have purchased	best player	best player	16	Apex AD2600 Progressive-scan DVD player
3	i think apex is the best dvd player you can ge...	best player	best player	16	Apex AD2600 Progressive-scan DVD player
4	for the first few weeks this player was everyt...	affordable player	best player	16	Apex AD2600 Progressive-scan DVD player
...	...	...	...	...	...
162	i give it 4stars only because the installation...	average user	average user	1	norton
163	this program would not let me send outlook ema...	audible books	audible books	1	norton
164	i am an advanced computer user so i cannot eve...	advanced user	advanced user	1	norton
165	i do not know what is happened to norton but 2...	absolute garbage	absolute garbage	1	norton
166	the ms scandisk would appear due to manual shu...	2nd bar	2nd bar	1	norton

3359 rows × 5 columns

Выходные данные: для анализа тональности

	review	aspect	key_aspect	count	product
0	i was impressed by this very inexpensive playe...	inexpensive player	best player	16	Apex AD2600 Progressive-scan DVD player
1	i looked into buying an inexpensive dvd player...	inexpensive player	best player	16	Apex AD2600 Progressive-scan DVD player
2	this is the best dvd player i have purchased	best player	best player	16	Apex AD2600 Progressive-scan DVD player
3	i think apex is the best dvd player you can ge...	best player	best player	16	Apex AD2600 Progressive-scan DVD player
4	for the first few weeks this player was everyt...	affordable player	best player	16	Apex AD2600 Progressive-scan DVD player
...	...	...	...	...	...
162	i give it 4stars only because the installation...	average user	average user	1	norton
163	this program would not let me send outlook ema...	audible books	audible books	1	norton
164	i am an advanced computer user so i cannot eve...	advanced user	advanced user	1	norton
165	i do not know what is happened to norton but 2...	absolute garbage	absolute garbage	1	norton
166	the ms scandisk would appear due to manual shu...	2nd bar	2nd bar	1	norton

3359 rows × 5 columns

Выходные данные: для отзыва-резюме

	review	aspect	key_aspect	count	product
0	i was impressed by this very inexpensive playe...	inexpensive player	best player	16	Apex AD2600 Progressive-scan DVD player
1	i looked into buying an inexpensive dvd player...	inexpensive player	best player	16	Apex AD2600 Progressive-scan DVD player
2	this is the best dvd player i have purchased	best player	best player	16	Apex AD2600 Progressive-scan DVD player
3	i think apex is the best dvd player you can ge...	best player	best player	16	Apex AD2600 Progressive-scan DVD player
4	for the first few weeks this player was everyt...	affordable player	best player	16	Apex AD2600 Progressive-scan DVD player
...	...	...	...	...	...
162	i give it 4stars only because the installation...	average user	average user	1	norton
163	this program would not let me send outlook ema...	audible books	audible books	1	norton
164	i am an advanced computer user so i cannot eve...	advanced user	advanced user	1	norton
165	i do not know what is happened to norton but 2...	absolute garbage	absolute garbage	1	norton
166	the ms scandisk would appear due to manual shu...	2nd bar	2nd bar	1	norton

3359 rows × 5 columns

Результат применения
для генерации отзыва-резюме

Выделение аспектов товара

Apex AD2600 Progressive-scan DVD player:

- 206 уникальных пар отзыв-аспект
- 142 уникальных аспекта

100 пар проанализировано:

- 32 содержат нерелевантные аспекты
- 8 аспектов недостаточно близки по смыслу

Выделение аспектов товара (ошибки)

- «*apex player*», «*apex product*»

- «*many times*», «*several occasions*»

- «*good reviews*»

- «*higher quality»:* only owned it about two weeks so i expect were looking at getting another progressive scan dvd player of <u>higher quality</u> very soon

- «*not worth player»:* this <u>player</u> is <u>not worth</u> any price and i recommend that you do not purchase it

Классификация тональности аспектов (ошибки)

из 100 пар:

- 12 аспектов ложноположительного класса

«*reliable brand*»	i read in consumer reports that apex is a <u>reliable brand</u> but with my experience and reading the other reviews i think i will be doing a little more homework before considering another apex product
«*strong smell*»	it did also seem to have a <u>strong</u> new electronics <u>smell</u> for some time

Отзыв-резюме: данные

	review	aspect	key_aspect	count	product	sentiment
0	i was impressed by this very inexpensive playe...	inexpensive player	best player	16	Apex AD2600 Progressive-scan DVD player	1
1	i looked into buying an inexpensive dvd player...	inexpensive player	best player	16	Apex AD2600 Progressive-scan DVD player	1
2	this is the best dvd player i have purchased	best player	best player	16	Apex AD2600 Progressive-scan DVD player	1
3	i think apex is the best dvd player you can ge...	best player	best player	16	Apex AD2600 Progressive-scan DVD player	1
4	for the first few weeks this player was everyt...	affordable player	best player	16	Apex AD2600 Progressive-scan DVD player	1
...	...	...	...	...	...	...
201	i use this with a home theater system and its ...	amazing system	amazing system	1	Apex AD2600 Progressive-scan DVD player	1
202	it plays alternate video formats vcds svcds cv...	alternate formats	alternate formats	1	Apex AD2600 Progressive-scan DVD player	1
203	imagine my shock and absolute instant disdain ...	absolute disdain	absolute disdain	1	Apex AD2600 Progressive-scan DVD player	0
204	the 2nd dvd player had a faulty power supply w...	2nd player	2nd player	1	Apex AD2600 Progressive-scan DVD player	0
205	this unit was working fine for the 1st month a...	1st month	1st month	1	Apex AD2600 Progressive-scan DVD player	0

206 rows × 6 columns

Отзыв-резюме

Reviews summary for
Apex AD2600 Progressive-scan DVD player
(top 20 aspects)

+ best player   + sleek design
+ extra features   + sharp jpeg pictures
+ apex player   + scan players
+ right price   + regular disk
+ progressive player   + new smell
+ nice machines   - remote control
+ great picture   - several times
+ cheap brand   - bad quality
+ apex product   - poor video
+ small buttons   - normal size

- ошибки в аспектах/тональности

Отзыв-резюме

Reviews summary for
Apex AD2600 Progressive-scan DVD player
(top 20 aspects)

+ best player   + sleek design
+ extra features   + sharp jpeg pictures
+ apex player   + scan players
+ right price   + regular disk
+ progressive player   + new smell
+ nice machines   - remote control
+ great picture   - several times
+ cheap brand   - bad quality
+ apex product   - poor video
+ small buttons   - normal size

Reviews summary for
Apex AD2600 Progressive-scan DVD player
(top 40 aspects)

+ best player   + wide range
+ extra features   + visual artist
+ apex player   + unmatched picture clarity
+ right price   + universal remotes
+ progressive player   + top notch
+ nice machines   + third product
+ great picture   + stylish case
+ cheap brand   - remote control
+ apex product   - several times
+ small buttons   - bad quality
+ sleek design   - poor video
+ sharp jpeg pictures   - normal size
+ scan players   - great deal
+ regular disk   - busy service line
+ new smell   - white noise
+ good sound   - useless codes
+ good reviews   - universal control
+ free shipping   - technical support
+ basic usage   - svideo connections
+ apex tv   - substandard guide

■ ошибки в аспектах/тональности

Заключение:

- BERT для анализа тональности аспектов даёт хороший результат (accuracy 0.88 на выделенных аспектах);

- алгоритм выделения аспектов товаров необходимо доработать/ пересмотреть;

- протестировать на актуальных отзывах и сравнить с резюме у Amazon.

Анализ тональности аспектов товаров на основе отзывов покупателей

Киреева Алена Николаевна