

---

# **Taobao Crawler Documentation**

**rwv, GuoYuanfang**

**2017 12 26**



<b>1</b>	<b>Crawler</b>	<b>1</b>
<b>2</b>	<b>Item Crawler</b>	<b>3</b>
<b>3</b>	<b>Rate Crawler</b>	<b>5</b>
<b>4</b>	<b>Size Info Analyzer</b>	<b>7</b>
<b>5</b>	<b>Size Info Analyzer with Time</b>	<b>9</b>
<b>6</b>	<b>Sold Number Analyzer</b>	<b>11</b>
<b>7</b>	<b>Sentiment Analyzer</b>	<b>13</b>
<b>8</b>	<b>Comment Analyzer</b>	<b>15</b>
<b>9</b>	<b>Final Score Analyzer</b>	<b>17</b>
<b>10</b>	<b>DB</b>	<b>19</b>
<b>11</b>		<b>21</b>



# CHAPTER 1

---

## Crawler

---

### 1.1

,:

```
keywords = ['', 'Phone']  
from taobao_crawler.crawler import Crawler  
crawler = Crawler(keywords, db)
```

Crawler(keywords, db) db *DB*

#### 1.1.1

```
crawler.item.run()
```

#### 1.1.2

```
crawler.rate.run()
```

#### 1.1.3

```
crawler.run()
```

mongodb *Item Crawler, Rate Crawler*

### 1.2



## 2.1

,

### 2.1.1

```
keywords = ['', 'Phone']  
from taobao_crawler.crawler.item import ItemCrawler  
crawler = ItemCrawler(keywords, db)
```

ItemCrawler(keywords, db) db *DB*

### 2.1.2

```
crawler.run()
```

## 2.2

```
{  
    "is_crawled" : true,  
    "seller_id" : "360622108",  
    "sellerLoc" : " ",  
    "location" : " ",  
    "title" : "4+64G!4G5.5",  
    "item_id" : "561319321061",  
    "price" : "529.00",  
    "area" : "",  
}
```

```
"sold" : "0"  
}
```

## 2.3



## CHAPTER 3

---

### Rate Crawler

---

#### 3.1

,

##### 3.1.1

```
from taobao_crawler.crawler.rate import RateCrawler
rate_crawler = RateCrawler(db)
```

RateCrawler(db) db *DB*

##### 3.1.2

```
rate_crawler.run()
```

#### 3.2

```
{
  "_id" : ObjectId("5a1d600db0d7ee38b9b0c942"),
  "buyCount" : 0,
  "useful" : true,
  "item_id" : "560697135358",
  "rate_id" : NumberLong("331495062062"),
  "rateDate" : "2017-11-23 23:16:40",
  "rate_content" : " ",
  "auctionSku" : ";;;:64GB;;",
  "anony" : true,
```

```
"size_info" : ";;;64GB;:"  
}
```

## 3.3

---

### Size Info Analyzer

---

#### 4.1

##### 4.1.1

```
keywords = {'': ['', 'mi'], '': ['', 'apple', 'iphone']}  
classifiers = ['16G', '32G', '64G', '128G', '256G']  
from taobao_crawler.analyzer.size_info_analyzer import SizeInfoAnalyzer  
analyzer = SizeInfoAnalyzer(keywords, classifiers, db)
```

```
SizeInfoAnalyzer(keywords, classifiers, db) db DB
```

##### 4.1.2

```
analyzer.run()
```

#### 4.2



---

### Size Info Analyzer with Time

---

## 5.1

### 5.1.1

```
keywords = {'': ['', 'mi'], '': ['', 'apple', 'iphone']}
classifiers = ['16G', '32G', '64G', '128G', '256G']
from taobao_crawler.analyzer.size_info_analyzer_with_time import _
↪SizeInfoAnalyzerWithTime
analyzer = SizeInfoAnalyzerWithTime(keywords, classifiers, db)
```

```
SizeInfoAnalyzerWithTime(keywords, classifiers, db) db DB
```

### 5.1.2

```
analyzer.run()
```

## 5.2



---

### Sold Number Analyzer

---

## 6.1

### 6.1.1

```
keywords = {'': ['', 'mi'], '': ['', 'apple', 'iphone']}  
from taobao_crawler.analyzer.sold_number_analyzer import SoldNumberAnalyzer  
analyzer = SoldNumberAnalyzer(keywords, db)
```

```
SoldNumberAnalyzer(keywords, db) db DB
```

### 6.1.2

```
analyzer.run()
```

## 6.2





## CHAPTER 7

---

### Sentiment Analyzer

---

rates collection, rates\_sentiments collection

## 7.1

### 7.1.1

```
from taobao_crawler.analyzer.sentiment_analyzer import SentimentAnalyzer
analyzer = SentimentAnalyzer(db)
```

SentimentAnalyzer(db) db *DB*

### 7.1.2

```
analyzer.run()
```

## 7.2



---

## Comment Analyzer

---

### 8.1

#### 8.1.1

```
brands = ['vivo', 'oppo', 'mi', 'huawei', 'apple', 'SMARTISAN']
from taobao_crawler.analyzer.comment_analyzer import CommentAnalyzer
comm_analyzer = CommentAnalyzer(db)
comm_analyzer.set_brand(brands)
```

CommentAnalyzer(db) db *DB*

#### 8.1.2

```
comm_analyzer.frequency_run()
```

#### 8.1.3

```
comm_analyzer.times_run()
```

### 8.2

```
[(' ', 20.856220399929835),
 (' ', 20.856220399929835),
 (' ', 20.856220399929835),
 (' ', 20.856220399929835),
 (' ', 20.856220399929835),
```

```
( ' ', 20.856220399929835) ,  
( ' ', 20.856220399929835) ,  
( ' ', 20.856220399929835) ,  
( ' ', 20.856220399929835) ,  
( ' ', 20.856220399929835) ]
```

## 8.3

---

### Final Score Analyzer

---

## 9.1

### 9.1.1

```
brands = ['vivo', 'oppo', 'mi', 'huawei', 'apple', 'SMARTISAN']
from taobao_crawler.analyzer.final_score_analyzer import FinalScoreAnalyzer
score_analyzer = FinalScoreAnalyzer(db)
score_analyzer.set_brand(brands)
```

FinalScoreAnalyzer(db) db *DB*

### 9.1.2

```
item_id = ''
score_analyzer.score_distribution_run(item_id)
```

### 9.1.3

```
item_id = ''
score_analyzer.score_distribution_run(item_id)
```

### 9.1.4

```
score_analyzer.compared_score_run()
```

## 9.2

```
final_scores = self.three_score_to_final_score(self.three_scores, [0.43, 0.25, 0.32])
self.draw_final_score(final_scores)
```

## 9.3

*DB*

## 10.1

### 10.1.1

```
db_config = {'db_user': '', #
             'db_pass': '', #
             'db_host': 'localhost', #
             'db_port': 27017, #
             'db_name': 'taobao'} # collection
mongo = DB(db_config)
```

### 10.1.2

```
rate_crawler = RateCrawler(mongo.db)
```

### 10.1.3

```
mongo.close()
```

## 10.2





## CHAPTER 11

---

---

- `genindex`
- `modindex`
- `search`