Chapter 5 - Anatomy of an Analyzer

Example Request - standard_tokenizer

POST http://localhost:9200/_analyze

HEADERS
Content-Type
application/json

BODY
```
{
        "text": "<body>You'll love Elasticsearch 7.0</body>",
        "tokenizer": "standard"
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"<body>You'll love Elasticsearch
7.0</body>\", \"tokenizer\": \"standard\" }"
```

Example Request - standard_tokenizer_lowercase_token_filter

POST http://localhost:9200/_analyze

HEADERS
Content-Type
application/json

BODY
```
{
        "text": "<body>You'll love Elasticsearch 7.0</body>",
        "tokenizer": "standard",
        "filter": ["lowercase"]
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"<body>You'll love Elasticsearch
7.0</body>\", \"tokenizer\": \"standard\", \"filter\": [\"lowercase\"] }"
```

Example Request - html_strip_map_filter_standard_tokenizer

POST http://localhost:9200/_analyze

HEADERS
Content-Type
application/json

BODY
```
{
    "text": "<body>You'll love Elasticsearch 7.0</body>",
    "char_filter": ["html_strip", {"type":"mapping", "mappings": ["You'll => You
will"]}],
    "tokenizer": "standard",
    "filter": ["lowercase"]
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"<body>You'll love Elasticsearch
7.0</body>\", \"char_filter\": [\"html_strip\", {\"type\":\"mapping\", \"mappings\":
[\"You'll => You will\"]}], \"tokenizer\": \"standard\", \"filter\":
[\"lowercase\"] }"
```

Example Request - html_strip_standard_tokenizer_lowercase
POST http://localhost:9200/_analyze

HEADERS
Content-Type
application/json

BODY
```
{
    "text": "<body>You'll love Elasticsearch 7.0</body>",
    "char_filter": ["html_strip"],
    "tokenizer": "standard",
    "filter": ["lowercase"]
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"<body>You'll love Elasticsearch
7.0</body>\", \"char_filter\": [\"html_strip\"], \"tokenizer\": \"standard\",
\"filter\": [\"lowercase\"] }""
```

Example Request - pattern_replace_char_filter
POST http://localhost:9200/_analyze

HEADERS
Content-Type
application/json

BODY
```
{
    "text": "<body>You'll love Elasticsearch 7.0</body>",
    "char_filter": ["html_strip", {"type": "mapping", "mappings": ["You'll=>You
will"]},
                                  {"type":"pattern_replace", "pattern":"(\\d+).(\\d+)",
"replacement":"v$1"}],
```

```
        "tokenizer": "standard",
        "filter": ["lowercase"]
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"<body>You'll love Elasticsearch
7.0</body>\", \"char_filter\": [\"html_strip\", {\"type\": \"mapping\", \"mappings\":
[\"You'll=>You will\"]}, {\"type\":\"pattern_replace\",
\"pattern\":\"(\\\\d+).(\\\\\d+)\", \"replacement\":\"v$1\"}], \"tokenizer\":
\"standard\", \"filter\": [\"lowercase\"] }"
```

Example Request - standard_tokenizer2
POST http://localhost:9200/_analyze

HEADERS
Content-Type
application/json

BODY
```
{
        "text": "POST https://api.iextrading.com/1.0/stock/acwf/company /usr/local",
        "tokenizer": "standard"
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"POST
https://api.iextrading.com/1.0/stock/acwf/company /usr/local\", \"tokenizer\":
\"standard\" }"
```

Example Request - letter_tokenizer
POST http://localhost:9200/_analyze

HEADERS
Content-Type
application/json

BODY
```
{
        "text": "POST https://api.iextrading.com/1.0/stock/acwf/company /user/local",
        "tokenizer": "whitespace"
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"POST
https://api.iextrading.com/1.0/stock/acwf/company /user/local\", \"tokenizer\":
\"whitespace\" }"
```

Example Request - lowercase_tokenizer

POST http://localhost:9200/_analyze

HEADERS
Content-Type
application/json

BODY
```
{
      "text": "POST https://api.iextrading.com/1.0/stock/acwf/company /usr/local",
      "tokenizer": "lowercase"
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"POST
https://api.iextrading.com/1.0/stock/acwf/company /usr/local\", \"tokenizer\":
\"lowercase\" }"
```

Example Request - whitespace_tokenizer

POST http://localhost:9200/_analyze

HEADERS
Content-Type
application/json

BODY
```
{
      "text": "POST https://api.iextrading.com/1.0/stock/acwf/company /usr/local",
      "tokenizer": "whitespace"
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"POST
https://api.iextrading.com/1.0/stock/acwf/company /usr/local\", \"tokenizer\":
\"whitespace\" }"
```

Example Request - uax_url_email_tokenizer

POST http://localhost:9200/_analyze

HEADERS
Content-Type
application/json

BODY
```
{
      "text": "POST https://api.iextrading.com/1.0/stock/acwf/company /usr/local",
      "tokenizer": "uax_url_email"
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"POST
https://api.iextrading.com/1.0/stock/acwf/company /usr/local\", \"tokenizer\":
\"uax_url_email\" }"
```

Example Request - classic-tokenizer

POST http://localhost:9200/_analyze

HEADERS
Content-Type
application/json

BODY
```
{
        "text": "POST https://api.iextrading.com/1.0/stock/acwf/company 192.168.0.1
100-123",
        "tokenizer": "classic"
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"POST
https://api.iextrading.com/1.0/stock/acwf/company 192.168.0.1 100-123\",
\"tokenizer\": \"classic\" }"
```

Example Request - thai_tokenizer

POST http://localhost:9200/_analyze

HEADERS
Content-Type
application/json

BODY
```
{
        "text": "คุณจะรัก Elasticsearch 7.0",
        "tokenizer": "thai"
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"คุณจะรัก Elasticsearch 7.0\",
\"tokenizer\": \"thai\" }"
```

Example Request - ngram_tokenizer

POST http://localhost:9200/_analyze

HEADERS
```

Content-Type
application/json

BODY

```
{
      "text": "Elasticsearch 7.0",
      "tokenizer": {"type":"ngram", "min_gram":2, "max_gram":2, "token_chars":
["punctuation", "digit"]}
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"Elasticsearch 7.0\", \"tokenizer\":
{\"type\":\"ngram\", \"min_gram\":2, \"max_gram\":2, \"token_chars\":
[\"punctuation\", \"digit\"]} }"
```

Example Request - edge_ngram_tokenizer

POST edge_ngram_tokenizer

http://localhost:9200/_analyze

HEADERS
Content-Type
application/json

BODY

```
{
      "text": "Elasticsearch 7.0",
      "tokenizer": {"type":"edge_ngram", "min_gram":2, "max_gram":2, "token_chars":
["punctuation", "digit"]}
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"Elasticsearch 7.0\", \"tokenizer\":
{\"type\":\"edge_ngram\", \"min_gram\":2, \"max_gram\":2, \"token_chars\":
[\"punctuation\", \"digit\"]} }"
```

Example Request - keyword_tokenizer

POST http://localhost:9200/_analyze

HEADERS
Content-Type
application/json

BODY

```
{
      "text": "Elasticsearch 7.0",
      "tokenizer": "keyword"
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"Elasticsearch 7.0\", \"tokenizer\":
\"keyword\" }"
```

Example Request - pattern_tokenizer

POST http://localhost:9200/_analyze

HEADERS
Content-Type
application/json

BODY
```
{
       "text": "Elasticsearch 7.0",
       "tokenizer": "pattern"
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"Elasticsearch 7.0\", \"tokenizer\":
\"pattern\" }"
```

Example Request - char_group_tokenizer

POST http://localhost:9200/_analyze

HEADERS
Content-Type
application/json

BODY
```
{
       "text": "Elasticsearch 7.0",
       "tokenizer": {"type":"char_group", "tokenize_on_chars": ["whitespace",
"punctuation"]}
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"Elasticsearch 7.0\", \"tokenizer\":
{\"type\":\"char_group\", \"tokenize_on_chars\": [\"whitespace\", \"punctuation\"]} }"
```

Example Request - simple_pattern_tokenizer

POST http://localhost:9200/_analyze

HEADERS
Content-Type
application/json

BODY
```

```
{
        "text": "Elasticsearch 7.0",
        "tokenizer": {"type":"simple_pattern", "pattern": "[a-zA-Z]*"}
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"Elasticsearch 7.0\", \"tokenizer\":
{\"type\":\"simple_pattern\", \"pattern\": \"[a-zA-Z]*\"} }"
```

Example Request - simple_pattern_split_tokenizer
POST http://localhost:9200/_analyze

HEADERS
Content-Type
application/json

BODY
```
{
        "text": "Elasticsearch 7.0",
        "tokenizer": {"type":"simple_pattern_split", "pattern": "[a-zA-Z.]*"}
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"Elasticsearch 7.0\", \"tokenizer\":
{\"type\":\"simple_pattern_split\", \"pattern\": \"[a-zA-Z.]*\"} }"
```

Example Request - asciifolding_token_filter
POST http://localhost:9200/_analyze

HEADERS
Content-Type
application/json

BODY
```
{
        "text": "Ÿőű'ľľ ľővė Èľàśťĭćŝėàŕćĥ 7.0",
        "tokenizer": "standard",
        "filter": [{"type":"asciifolding", "preserve_original":true}]

}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"Ÿőű'ľľ ľővė Èľàśťĭćŝėàŕćĥ 7.0\",
\"tokenizer\": \"standard\", \"filter\": [{\"type\":\"asciifolding\",
\"preserve_original\":true}] }"
```

Example Request - ngram_token_filter

http://localhost:9200/_analyze

HEADERS
Content-Type
application/json

BODY

```
{
     "text": "Elasticsearch 7.0",
     "tokenizer": "standard",
     "filter": [{"type":"ngram", "min_gram":10, "max_gram":10}]
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"Elasticsearch 7.0\", \"tokenizer\":
\"standard\", \"filter\": [{\"type\":\"ngram\", \"min_gram\":10, \"max_gram\":10}] }"
```

Example Request - edge_ngram_token_filter

POST http://localhost:9200/_analyze

HEADERS
Content-Type
application/json

BODY

```
{
     "text": "You'll love Elasticsearch 7.0",
     "tokenizer": "standard",
     "filter": [{"type":"edge_ngram", "min_gram":10, "max_gram":10}]
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"You'll love Elasticsearch 7.0\",
\"tokenizer\": \"standard\", \"filter\": [{\"type\":\"edge_ngram\", \"min_gram\":10,
\"max_gram\":10}] }"
```

Example Request - lowercase_token_filter

POST http://localhost:9200/_analyze

HEADERS
Content-Type
application/json

BODY

```
{
     "text": "You'll love Elasticsearch 7.0",
     "tokenizer": "standard",
     "filter": ["lowercase"]
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"You'll love Elasticsearch 7.0\",
\"tokenizer\": \"standard\", \"filter\": [\"lowercase\"] }"
```

Example Request - fingerprint_token_filter

HEADERS
Content-Type
application/json

BODY
```
{
        "text": "You'll love Elasticsearch 7.0",
        "tokenizer": "standard",
        "filter": ["fingerprint"]
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"You'll love Elasticsearch 7.0\",
\"tokenizer\": \"standard\", \"filter\": [\"fingerprint\"] }"
```

Example Request - keep_token_filter

POST http://localhost:9200/_analyze

HEADERS
Content-Type
application/json

BODY
```
{
        "text": "You'll love Elasticsearch 7.0",
        "tokenizer": "standard",
        "filter": [{"type": "keep", "keep_words":["Elasticsearch", "7.0"]}]
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"You'll love Elasticsearch 7.0\",
\"tokenizer\": \"standard\", \"filter\": [{\"type\": \"keep\",
\"keep_words\":[\"Elasticsearch\", \"7.0\"]}] }"
```

Example Request - keep_types_token_filter

POST http://localhost:9200/_analyze

HEADERS
Content-Type
application/json

BODY
```
{
        "text": "You'll love Elasticsearch 7.0",
        "tokenizer": "standard",
        "filter": [{"type": "keep_types", "types":["<NUM>"]}]
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"You'll love Elasticsearch 7.0\",
\"tokenizer\": \"standard\", \"filter\": [{\"type\": \"keep_types\",
\"types\":[\"<NUM>\"]}] }"
```

Example Request - stemmer_token_filter

HEADERS
Content-Type
application/json

BODY
```
{
        "text": "love loves loved loving",
        "tokenizer": "standard",
        "filter": [{"type": "stemmer", "name":"english"}]
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"love loves loved loving\",
\"tokenizer\": \"standard\", \"filter\": [{\"type\": \"stemmer\",
\"name\":\"english\"}] }"
```

Example Request - stop_token_filter

HEADERS
Content-Type
application/json

BODY
```
{
        "text": "A an The and Elasticsearch",
        "tokenizer": "standard",
        "filter": [{"type": "stop", "stopwords":["_english_"], "ignore_case":true}]
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"A an The and Elasticsearch\",
\"tokenizer\": \"standard\", \"filter\": [{\"type\": \"stop\",
\"stopwords\":[\"_english_\"], \"ignore_case\":true}] }"
```

Example Request - unique_token_filter

http://localhost:9200/_analyze

HEADERS
Content-Type
application/json

BODY
```
{
      "text": "love loves loved loving",
      "tokenizer": "standard",
      "filter": [{"type": "stemmer", "name":"english"}, "unique"]
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"love loves loved loving\",
\"tokenizer\": \"standard\", \"filter\": [{\"type\": \"stemmer\",
\"name\":\"english\"}, \"unique\"] }"
```

Example Request - predicate_token_filter

POST http://localhost:9200/_analyze

HEADERS
Content-Type
application/json

BODY
```
{
      "text": "You'll love Elasticsearch 7.0",
      "tokenizer": "standard",
      "filter": [{"type": "predicate_token_filter",
"script":{"source":"token.getType()=='<NUM>'"}}]
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"You'll love Elasticsearch 7.0\",
\"tokenizer\": \"standard\", \"filter\": [{\"type\": \"predicate_token_filter\",
\"script\":{\"source\":\"token.getType()=='<NUM>'\"}}] }"
```

Example Request - conditional_token_filter

POST http://localhost:9200/_analyze

HEADERS
Content-Type
application/json

BODY
```
{
       "text": "You'll love Elasticsearch 7.0",
       "tokenizer": "standard",
       "filter": [{"type": "condition",
"script":{"source":"token.getType()=='<ALPHANUM>'"}, "filter":["reverse"]}]
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"You'll love Elasticsearch 7.0\",
\"tokenizer\": \"standard\", \"filter\": [{\"type\": \"condition\",
\"script\":{\"source\":\"token.getType()=='<ALPHANUM>'\"},
\"filter\":[\"reverse\"]}] }"
```

Example Request - word_delimiter_token_filter_generate_word_parts_true
POST http://localhost:9200/_analyze

HEADERS
Content-Type
application/json

BODY
```
{
       "text": "ElasticSearch 7.0",
       "tokenizer": "standard",
       "filter": [{"type":"word_delimiter", "generate_word_parts":true}]
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"ElasticSearch 7.0\", \"tokenizer\":
\"standard\", \"filter\": [{\"type\":\"word_delimiter\",
\"generate_word_parts\":true}] }"
```

Example Request - word_delimiter_token_filter_generate_word_parts_false
POST http://localhost:9200/_analyze

HEADERS
Content-Type
application/json

BODY
```
{
       "text": "ElasticSearch 7.0",
```

```
        "tokenizer": "standard",
        "filter": [{"type":"word_delimiter", "generate_word_parts":false}]
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"ElasticSearch 7.0\", \"tokenizer\":
\"standard\", \"filter\": [{\"type\":\"word_delimiter\",
\"generate_word_parts\":false}] }"
```

Example Request - word_delimiter_token_filter_generate_number_parts_true

HEADERS
Content-Type
application/json

BODY
```
{
        "text": "ElasticSearch 7.0",
        "tokenizer": "standard",
        "filter": [{"type":"word_delimiter", "generate_number_parts":true}]
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"ElasticSearch 7.0\", \"tokenizer\":
\"standard\", \"filter\": [{\"type\":\"word_delimiter\",
\"generate_number_parts\":true}] }"
```

Example Request - word_delimiter_token_filter_generate_number_parts_false

HEADERS
Content-Type
application/json

BODY
```
{
        "text": "ElasticSearch 7.0",
        "tokenizer": "standard",
        "filter": [{"type":"word_delimiter", "generate_number_parts":false}]
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"ElasticSearch 7.0\", \"tokenizer\":
\"standard\", \"filter\": [{\"type\":\"word_delimiter\",
\"generate_number_parts\":false}] }"
```

Example Request - word_delimiter_token_filter_catenate_words_true

POST http://localhost:9200/_analyze

HEADERS
Content-Type
application/json

BODY
```
{
        "text": "Elastic_Search 7.0",
        "tokenizer": "standard",
        "filter": [{"type":"word_delimiter", "catenate_words": true}]
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"Elastic_Search 7.0\", \"tokenizer\":
\"standard\", \"filter\": [{\"type\":\"word_delimiter\", \"catenate_words\": true}] }"
```

Example Request - word_delimiter_token_filter_catenate_words_false

POST http://localhost:9200/_analyze

HEADERS
Content-Type
application/json

BODY
```
{
        "text": "Elastic_Search 7.0",
        "tokenizer": "standard",
        "filter": [{"type":"word_delimiter", "catenate_words": false}]
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"Elastic_Search 7.0\", \"tokenizer\":
\"standard\", \"filter\": [{\"type\":\"word_delimiter\", \"catenate_words\":
false}] }"
```

Example Request - word_delimiter_token_filter_catenate_numbers_true

POST http://localhost:9200/_analyze

HEADERS
Content-Type
application/json

BODY
```
{
        "text": "Elastic_Search 7.0",
```

```
      "tokenizer": "standard",
      "filter": [{"type":"word_delimiter", "catenate_numbers": true}]
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"Elastic_Search 7.0\", \"tokenizer\":
\"standard\", \"filter\": [{\"type\":\"word_delimiter\", \"catenate_numbers\":
true}] }"
```

Example Request - word_delimiter_token_filter_catenate_number_false
http://localhost:9200/_analyze

HEADERS
Content-Type
application/json

BODY
```
{
      "text": "Elastic_Search 7.0",
      "tokenizer": "standard",
      "filter": [{"type":"word_delimiter", "catenate_numbers": false}]
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"Elastic_Search 7.0\", \"tokenizer\":
\"standard\", \"filter\": [{\"type\":\"word_delimiter\", \"catenate_numbers\":
false}] }"
```

Example Request - word_delimiter_token_filter_catenate_all_true
http://localhost:9200/_analyze

HEADERS
Content-Type
application/json

BODY
```
{
      "text": "Elastic_Search 7.0",
      "tokenizer": "standard",
      "filter": [{"type":"word_delimiter", "catenate_all": true}]
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"Elastic_Search 7.0\", \"tokenizer\":
\"standard\", \"filter\": [{\"type\":\"word_delimiter\", \"catenate_all\": true}] }"
```

Example Request - word_delimiter_token_filter_catenate_all_false

http://localhost:9200/_analyze

HEADERS
Content-Type
application/json

BODY

```
{
    "text": "Elastic_Search 7.0",
    "tokenizer": "standard",
    "filter": [{"type":"word_delimiter", "catenate_all": false}]
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"Elastic_Search 7.0\", \"tokenizer\":
\"standard\", \"filter\": [{\"type\":\"word_delimiter\", \"catenate_all\": false}] }"
```

Example Request - word_delimiter_token_filter_split_on_case_change_true

POST http://localhost:9200/_analyze

HEADERS
Content-Type
application/json

BODY

```
{
    "text": "ElasticSearch 7.0",
    "tokenizer": "standard",
    "filter": [{"type":"word_delimiter", "split_on_case_change": true}]
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"ElasticSearch 7.0\", \"tokenizer\":
\"standard\", \"filter\": [{\"type\":\"word_delimiter\", \"split_on_case_change\":
true}] }"
```

Example Request - word_delimiter_token_filter_split_on_case_change_false

POST http://localhost:9200/_analyze

HEADERS
Content-Type
application/json

BODY

```
{
    "text": "ElasticSearch 7.0",
    "tokenizer": "standard",
    "filter": [{"type":"word_delimiter", "split_on_case_change": false}]
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"ElasticSearch 7.0\", \"tokenizer\":
\"standard\", \"filter\": [{\"type\":\"word_delimiter\", \"split_on_case_change\":
false}] }"
```

Example Request - word_delimiter_token_filter_preserve_original_true

http://localhost:9200/_analyze

HEADERS
Content-Type
application/json

BODY
```
{
        "text": "Elastic_Search 7.0",
        "tokenizer": "standard",
        "filter": [{"type":"word_delimiter", "preserve_original": true}]
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"Elastic_Search 7.0\", \"tokenizer\":
\"standard\", \"filter\": [{\"type\":\"word_delimiter\", \"preserve_original\":
true}] }"
```

Example Request - word_delimiter_token_filter_preserve_original_false

http://localhost:9200/_analyze

HEADERS
Content-Type
application/json

BODY
```
{
        "text": "ElasticSearch 7.0",
        "tokenizer": "standard",
        "filter": [{"type":"word_delimiter", "preserve_original": false}]
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"ElasticSearch 7.0\", \"tokenizer\":
\"standard\", \"filter\": [{\"type\":\"word_delimiter\", \"preserve_original\":
false}] }"
```

Example Request - word_delimiter_token_filter_split_on_numeric_true

POST http://localhost:9200/_analyze

HEADERS
Content-Type
application/json

BODY
```
{
        "text": "SN12X",
        "tokenizer": "standard",
        "filter": [{"type":"word_delimiter", "split_on_numerics":true}]
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"SN12X\", \"tokenizer\": \"standard\",
\"filter\": [{\"type\":\"word_delimiter\", \"split_on_numerics\":true}] }"
```

Example Request - word_delimiter_token_filter_split_on_numeric_false

POST http://localhost:9200/_analyze

HEADERS
Content-Type
application/json

BODY
```
{
        "text": "ElasticSearch 7.0",
        "tokenizer": "standard",
        "filter": [{"type":"word_delimiter", "split_on_numeric": false}]
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"ElasticSearch 7.0\", \"tokenizer\":
\"standard\", \"filter\": [{\"type\":\"word_delimiter\", \"split_on_numeric\":
false}] }"
```

Example Request - word_delimiter_token_filter_stem_english_possesive_true

POST http://localhost:9200/_analyze

HEADERS
Content-Type
application/json

BODY
```
{
        "text": "Elasticsearch's analyer",
        "tokenizer": "standard",
        "filter": [{"type":"word_delimiter", "stem_english_possessive":true}]
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"Elasticsearch's analyer\",
\"tokenizer\": \"standard\", \"filter\": [{\"type\":\"word_delimiter\",
\"stem_english_possessive\":true}] }"
```

Example Request - word_delimiter_token_filter_stem_english_possesive_false

POST http://localhost:9200/_analyze

HEADERS
Content-Type
application/json

BODY
```
{
        "text": "Elasticsearch's analyzer",
        "tokenizer": "standard",
        "filter": [{"type":"word_delimiter", "stem_english_possessive":false}]
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"Elasticsearch's analyzer\",
\"tokenizer\": \"standard\", \"filter\": [{\"type\":\"word_delimiter\",
\"stem_english_possessive\":false}] }"
```

Example Request - standard_analyzer

POST http://localhost:9200/_analyze

HEADERS
Content-Type
application/json

BODY
```
{
        "text": "In Elasticsearch 7.0",
        "analyzer": "standard"
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"In Elasticsearch 7.0\", \"analyzer\":
\"standard\" }"
```

Example Request - simple_analyzer

POST http://localhost:9200/_analyze

Content-Type
application/json

BODY

```
{
        "text": "In Elasticsearch 7.0",
        "analyzer": "simple"
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"In Elasticsearch 7.0\", \"analyzer\":
\"simple\" }"
```

Example Request - whitespace_analyzer
POST http://localhost:9200/_analyze

HEADERS
Content-Type
application/json

BODY

```
{
        "text": "In Elasticsearch 7.0",
        "analyzer": "whitespace"
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"In Elasticsearch 7.0\", \"analyzer\":
\"whitespace\" }"
```

Example Request - stop_analyzer
POST http://localhost:9200/_analyze

HEADERS
Content-Type
application/json

BODY

```
{
        "text": "You'll love Elasticsearch 7.0",
        "analyzer": "whitespace"
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"You'll love Elasticsearch 7.0\",
\"analyzer\": \"whitespace\" }"
```

Example Request - keyword_analyzer

http://localhost:9200/_analyze

HEADERS
Content-Type
application/json

BODY

```
{
      "text": "In Elasticsearch 7.0",
      "analyzer": "keyword"
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"In Elasticsearch 7.0\", \"analyzer\":
\"keyword\" }"
```

Example Request - pattern_analyzer

http://localhost:9200/_analyze

HEADERS
Content-Type
application/json

BODY

```
{
      "text": "In Elasticsearch 7.0",
      "analyzer": "pattern"
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"In Elasticsearch 7.0\", \"analyzer\":
\"pattern\" }"
```

Example Request - fingerprint_analyzer

http://localhost:9200/_analyze

HEADERS
Content-Type
application/json

BODY

```
{
      "text": "In Elasticsearch 7.0",
      "analyzer": "fingerprint"
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"In Elasticsearch 7.0\", \"analyzer\":
\"fingerprint\" }"
```

Example Request - english_analyzer

POST http://localhost:9200/_analyze

HEADERS
Content-Type
application/json

BODY
```
{
     "text": "In Elasticsearch 7.0",
     "analyzer": "english"
}
```

```
curl --location --request POST "http://localhost:9200/_analyze" --header "Content-
Type: application/json" --data "{ \"text\": \"In Elasticsearch 7.0\", \"analyzer\":
\"english\" }"
```

Example Request - define_custom_analyzer

PUT http://localhost:9200/cf_etf

HEADERS
Content-Type
application/json

BODY
```
{
     "mappings": {
          "dynamic": false,
          "properties": {
               "symbol": {"type": "keyword"},
               "fund_name": {"type": "text"},
               "rating": {"type": "byte"},
               "morningstar_category": {"type": "keyword"},
               "category": {"type": "keyword"},
               "family": {"type": "keyword"},
               "market_cap": {"type": "keyword"},
               "description": {"type": "text", "fields": {"raw":
{"type":"keyword"}}, "analyzer":"description_analyzer",
"search_analyzer":"description_analyzer"},
               "exchange": {"type": "keyword"}
          }
     },
     "settings": {
          "analysis": {
               "analyzer" : {
                    "description_analyzer": {
```

```
                                "type": "custom",
                                "tokenizer": "description_tokenizer",
                                "filter": ["lowercase",
"description_pattern_replace_filter", "description_stemmer_filter",
"description_stop_filter", "description_length_filter", "unique"]
                            }
                    },
                    "tokenizer": {
                            "description_tokenizer" : {"type":"char_group",
"tokenize_on_chars": ["whitespace", "digit", "symbol", "\\n", ",", ":", "!", "?", ";",
",", "_", "{", "[", "}", "]", "(", ")", "\\", "\/","\""]}
                    },
                    "filter": {
                            "description_pattern_replace_filter":
{"type":"pattern_replace", "pattern": "(\\w{2,})\\.", "replacement":"$1"},
                            "description_stemmer_filter" : {"type":"stemmer",
"name":"light_english"},
                            "description_stop_filter": {"type":"stop",
"stopwords_path":"stopwords"},
                            "description_length_filter": {"type":"length", "min":2}
                    }
            }
        }
}
```

```
curl --location --request PUT "http://localhost:9200/cf_etf" --header "Content-Type:
application/json" --data $'{ \"mappings\": { \"dynamic\": false, \"properties\":
{ \"symbol\": {\"type\": \"keyword\"}, \"fund_name\": {\"type\": \"text\"},
\"rating\": {\"type\": \"byte\"}, \"morningstar_category\": {\"type\": \"keyword\"},
\"category\": {\"type\": \"keyword\"}, \"family\": {\"type\": \"keyword\"},
\"market_cap\": {\"type\": \"keyword\"}, \"description\": {\"type\": \"text\",
\"fields\": {\"raw\": {\"type\":\"keyword\"}}, \"analyzer\":\"description_analyzer\",
\"search_analyzer\":\"description_analyzer\"}, \"exchange\": {\"type\":
\"keyword\"} } }, \"settings\": { \"analysis\": { \"analyzer\" :
{ \"description_analyzer\": { \"type\": \"custom\", \"tokenizer\":
\"description_tokenizer\", \"filter\": [\"lowercase\",
\"description_pattern_replace_filter\", \"description_stemmer_filter\",
\"description_stop_filter\", \"description_length_filter\", \"unique\"] } },
\"tokenizer\": { \"description_tokenizer\" : {\"type\":\"char_group\",
\"tokenize_on_chars\": [\"whitespace\", \"digit\", \"symbol\", \"\\\\\", \",\", \":\",
\"!\", \"?\", \";\", \",\", \"_\", \"{\", \"[\", \"}\", \"]\", \"(\", \")\", \"\\\\\",
\"\\/\",\"\\\"\"]} }, \"filter\": { \"description_pattern_replace_filter\":
{\"type\":\"pattern_replace\", \"pattern\": \"(\\\\w{2,})\\\\.\",
\"replacement\":\"$1\"}, \"description_stemmer_filter\" : {\"type\":\"stemmer\",
\"name\":\"light_english\"}, \"description_stop_filter\": {\"type\":\"stop\",
\"stopwords_path\":\"stopwords\"}, \"description_length_filter\":
{\"type\":\"length\", \"min\":2} } } } }'
```

Example Request - define_custom_analyzer

GET http://localhost:9200/cf_etf

HEADERS

Content-Type
application/json

```
curl --location --request GET "http://localhost:9200/cf_etf"  --header "Content-Type:
application/json"
```

Example Request - test_custom_analyzer
POST http://localhost:9200/cf_etf/_analyze

HEADERS
Content-Type
application/json

BODY
```
{
     "text":"The investment seeks to track the price and yield performance, before
fees and expenses, of the Bloomberg Barclays U.S. Aggregate Enhanced Yield Index (the
\\\"index\\\").\\n The index is designed to broadly capture the U.S. investment grade,
fixed income securities market while seeking to enhance yield within desired risk
parameters and constraints. The fund is non-diversified.",
     "analyzer": "description_analyzer"
}
```

```
curl --location --request POST "http://localhost:9200/cf_etf/_analyze" --header
"Content-Type: application/json" --data "{ \"text\":\"The investment seeks to track
the price and yield performance, before fees and expenses, of the Bloomberg Barclays
U.S. Aggregate Enhanced Yield Index (the \\\\\\\"index\\\\\\\").\\\ The index is
designed to broadly capture the U.S. investment grade, fixed income securities market
while seeking to enhance yield within desired risk parameters and constraints. The
fund is non-diversified.\", \"analyzer\": \"description_analyzer\" }"
```

Example Request - define_lowercase_normailizer
PUT http://localhost:9200/cf_etf_toy

HEADERS
Content-Type
application/json

BODY
```
{
     "settings": {
           "analysis": {
                 "normalizer" : {
                       "lowercase_normalizer": {
                             "type": "custom",
                             "filter": ["lowercase"]
                       }
                 }
           }
     }
```

```
}
```

```
curl --location --request PUT "http://localhost:9200/cf_etf_toy" --header "Content-
Type: application/json" --data "{ \"settings\": { \"analysis\": { \"normalizer\" :
{ \"lowercase_normalizer\": { \"type\": \"custom\", \"filter\":
[\"lowercase\"] } } } } }"
```

Example Request - test_lowercase_normailizer
POST http://localhost:9200/cf_etf_toy/_analyze

HEADERS
Content-Type
application/json

BODY
```
{
        "text": "You'll love Elasticsearch 7.0",
        "normalizer": "lowercase_normalizer"
}
```

```
curl --location --request POST "http://localhost:9200/cf_etf_toy/_analyze" --header
"Content-Type: application/json" --data "{\"text\": \"You'll love Elasticsearch 7.0\",
\"normalizer\": \"lowercase_normalizer\"}"
```

CHAPTER5