Watson Language Understanding API

IBM API 사용 가이드 먼저

Analyze Text

text, HTML, webpage를 입력으로 넣어서

category
Concepts
Emotion
Entities
Keywords
Metadata
Relations
Semantic Roles
Sentiment
Syntax 와 같은 정보를 추출 할 수 있습니다.

1. 요청 파라미터 이해하기 - Features

Features

Analysis features and options.

Features

concepts

Returns high-level concepts in the content. For example, a research paper about deep learning might return the concept, "Artificial Intelligence" although the term is not mentioned.

Supported languages: English, French, German, Italian, Japanese, Korean, Portuguese, Spanish.

Example: View

ConceptsOptions

limit
int

Maximum number of concepts to return.

Constraints: value ≤ 50

emotion

Detects anger, disgust, fear, joy, or sadness that is conveyed in the content or by the context around target phrases specified in the targets parameter. You can analyze emotion for detected entities with `entities.emotion` and for keywords with `keywords.emotion`.

Supported languages: English.

Example: View

EmotionOptions

document
bool

Set this to `false` to hide document-level emotion results.

Default: `true`

targets
List[str]

Emotion results will be returned for each target string that is found in the document.

entities

Identifies people, cities, organizations, and other entities in the content. For more information, see Entity types and subtypes.

Supported languages: English, French, German, Italian, Japanese, Korean, Portuguese, Russian, Spanish, Swedish. Arabic, Chinese, and Dutch are supported only through custom models.

Example: View

EntitiesOptions

limit
int

Maximum number of entities to return.

Constraints: value ≤ 250

mentions
bool

Set this to `true` to return locations of entity mentions.

Default: `false`

model
str

Enter a custom model ID to override the standard entity detection model.

sentiment
bool

Set this to `true` to return sentiment information for detected entities.

Default: `false`

emotion
bool

Set this to `true` to analyze emotion for detected keywords.

Default: `false`

keywords

Returns important keywords in the content.

Supported languages: English, French, German, Italian, Japanese, Korean, Portuguese, Russian, Spanish, Swedish.

Example: View

KeywordsOptions

limit
int

Maximum number of keywords to return.

Constraints: value ≤ 250

sentiment
bool

Set this to `true` to return sentiment information for detected keywords.

Default: `false`

emotion
bool

Set this to `true` to analyze emotion for detected keywords.

Default: `false`

1. 요청 파라미터 이해하기 - Features

metadata
MetadataOptions

Returns information from the document, including author name, title, RSS/ATOM feeds, prominent page image, and publication date. Supports URL and HTML input types only.

relations

Recognizes when two entities are related and identifies the type of relation. For example, an `awardedTo` relation might connect the entities "Nobel Prize" and "Albert Einstein". For more information, see Relation types.

Supported languages: Arabic, English, German, Japanese, Korean, Spanish. Chinese, Dutch, French, Italian, and Portuguese custom models are also supported.

RelationsOptions

model
str

Enter a custom model ID to override the default model.

semantic_roles

Parses sentences into subject, action, and object form.

Supported languages: English, German, Japanese, Korean, Spanish.

Example: View

SemanticRolesOptions

limit
int

Maximum number of semantic_roles results to return.

keywords
bool

Set this to `true` to return keyword information for subjects and objects.

Default: `false`

entities
bool

Set this to `true` to return entity information for subjects and objects.

Default: `false`

sentiment

Analyzes the general sentiment of your content or the sentiment toward specific target phrases. You can analyze sentiment for detected entities with `entities.sentiment` and for keywords with `keywords.sentiment`.

Supported languages: Arabic, English, French, German, Italian, Japanese, Korean, Portuguese, Russian, Spanish.

Example: View

SentimentOptions

document
bool

Set this to `false` to hide document-level sentiment results.

Default: `true`

targets
List[str]

Sentiment results will be returned for each target string that is found in the document.

categories

Returns a five-level taxonomy of the content. The top three categories are returned.

Supported languages: Arabic, English, French, German, Italian, Japanese, Korean, Portuguese, Spanish.

CategoriesOptions

explanation
bool

Set this to `true` to return explanations for each categorization. **This is available only for English categories.**.

Default: `false`

limit
int

Maximum number of categories to return.

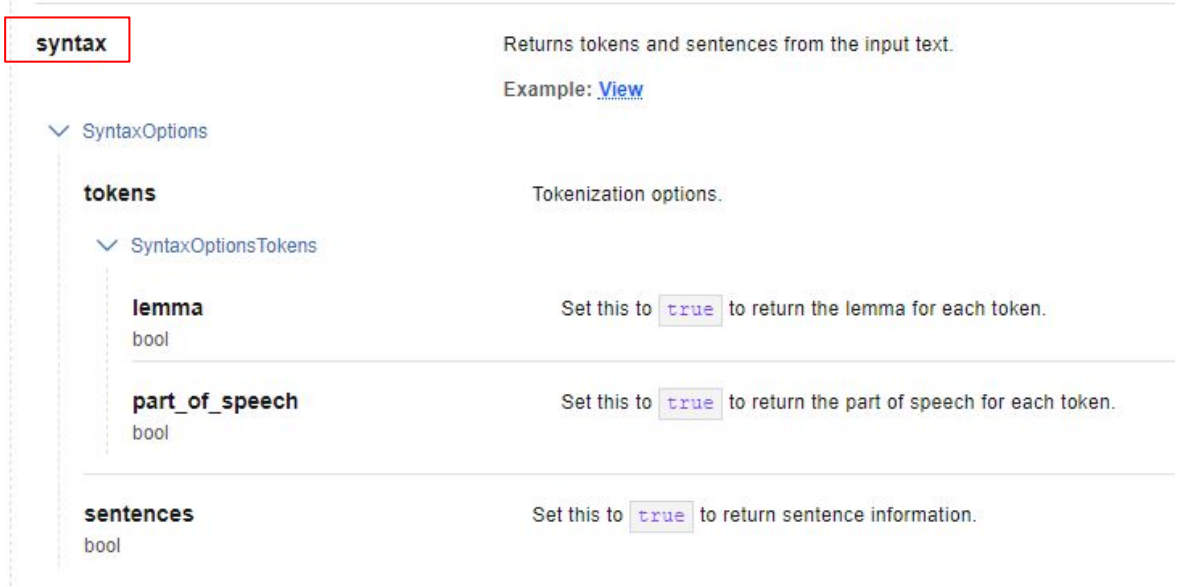
Constraints: value ≤ 10

model
str

Enter a custom model ID to override the standard categories model.

The custom categories experimental feature will be retired on 19 December 2019. On that date, deployed custom categories models will no longer be accessible in Natural Language Understanding. The feature will be removed from Knowledge Studio on an earlier date. Custom categories models will no longer be accessible in Knowledge Studio on 17 December 2019.

1. 요청 파라미터 이해하기 - Features




```
features=Features(
    entities=EntitiesOptions(emotion=True, sentiment=True, limit=2),
    categories=CategoriesOptions(limit=3)
    keywords=KeywordsOptions(emotion=True, sentiment=True, limit=2))
```

처럼 Featrues라는 함수 안에 빨간색으로 테두리 쳤던 속성들 = 각 속성Options (각 속성별 하위 파라미터)
와 같이 사용한다.

1. 요청 파라미터 이해하기 - 나머지

text
str
The plain text to analyze. One of the `text`, `html`, or `url` parameters is required.

html
str
The HTML file to analyze. One of the `text`, `html`, or `url` parameters is required.

url
str
The webpage to analyze. One of the `text`, `html`, or `url` parameters is required.

clean
bool
Set this to `false` to disable webpage cleaning. For more information about webpage cleaning, see Analyzing webpages.
Default: `true`

xpath
str
An XPath query to perform on `html` or `url` input. Results of the query will be appended to the cleaned webpage text before it is analyzed. To analyze only the results of the XPath query, set the `clean` parameter to `false`.

fallback_to_raw
bool
Whether to use raw HTML content if text cleaning fails.
Default: `true`

return_analyzed_text
bool
Whether or not to return the analyzed text.
Default: `false`

language
str
ISO 639-1 code that specifies the language of your text. This overrides automatic language detection. Language support differs depending on the features you include in your analysis. For more information, see Language support.

limit_text_characters
int
Sets the maximum number of characters that are processed by the service.

```
response = natural_language_understanding.analyze(
    text= 'Four score and seven years ago our fathers brought forth on
this continent a new nation, conceived in liberty, and dedicated to the
proposition that all men are created equal.',

    features=Features(
        entities=EntitiesOptions(emotion=True, sentiment=True,
limit=2),
        keywords=KeywordsOptions(emotion=True, sentiment=True,
                                 limit=2))).get_result()
```

natural_language_understanding.analyze() 안에 속성으로
파란색 테두리 = 값
의 형태로 사용한다.

2. Analyze Text 전체 코드

```
import json
from ibm_watson import NaturalLanguageUnderstandingV1
from ibm_cloud_sdk_core.authenticators import IAMAuthenticator
from ibm_watson.natural_language_understanding_v1 import Features, EntitiesOptions, KeywordsOptions

authenticator = IAMAuthenticator('발급받은 API키 입력')   # 발급 받은 키
natural_language_understanding = NaturalLanguageUnderstandingV1(                    # 사용할 클래스를 받아옴
    version='2019-07-12',
    authenticator=authenticator)

natural_language_understanding.set_service_url('발급받은 URL입력')      #사용자 URL 설정

response = natural_language_understanding.analyze(
    text=   'Four score and seven years ago our fathers brought forth on this continent a new nation, conceived in liberty, and dedicated to the
proposition that all men are created equal.',
    features=Features(                # Analysis 메소드
        entities=EntitiesOptions(emotion=True, sentiment=True, limit=2), #이 단어가 무엇을 나타내는 단어인가? ex) 수량 / 사람 / 회사 ….
        keywords=KeywordsOptions(emotion=True, sentiment=True,      #중요 키워드를 리턴함
                                 limit=2))).get_result()

print(json.dumps(response, indent=2))
```

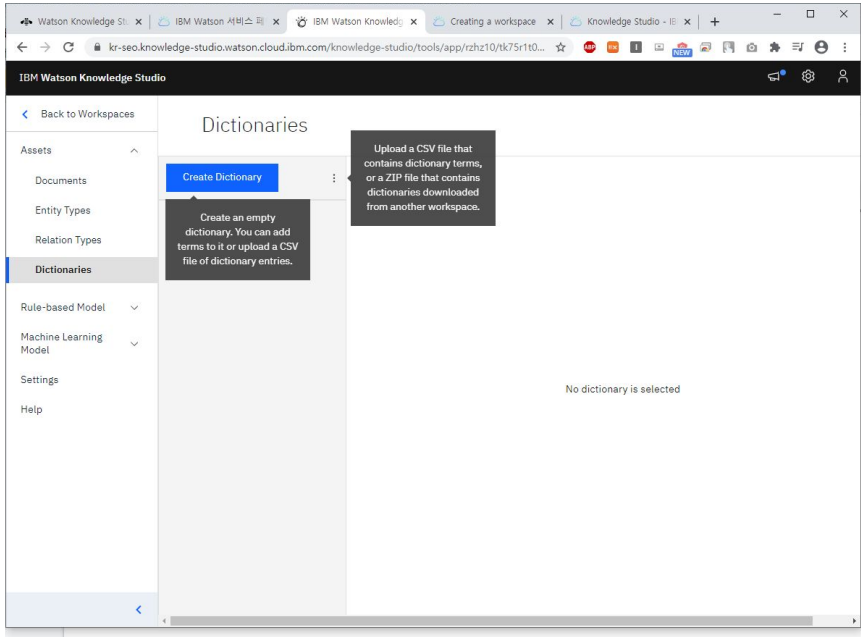
3. 실행 결과

keywords limit=2에 의해서 키워드는 2개만 결과가 나왔다.

entities의 경우 limit은 2이지만, 명사형(사람 / 회사 / 수량)단어가 본문에 1개 뿐이라 1개만 나왔다.


```
{
  "usage": {
    "text_units": 1,
    "text_characters": 175,
    "features": 2
  },
  "language": "en",
  "keywords": [
    {
      "text": "new nation",
      "sentiment": {
        "score": 0.867754,
        "label": "positive"
      },
      "relevance": 0.89316,
      "emotion": {
        "sadness": 0.273896,
        "joy": 0.644126,
        "fear": 0.024682,
        "disgust": 0.043153,
        "anger": 0.019833
      },
      "count": 1
    },
    {
      "text": "score",
      "sentiment": {
        "score": 0.867754,
        "label": "positive"
      },
      "relevance": 0.769836,
      "emotion": {
        "sadness": 0.273896,
        "joy": 0.644126,
        "fear": 0.024682,
        "disgust": 0.043153,
        "anger": 0.019833
      },
      "count": 1
    }
  ],
  "entities": [
    {
      "type": "Quantity",
      "text": "seven years",
      "sentiment": {
        "score": 0.867754,
        "label": "positive"
      },
      "relevance": 0.978348,
      "emotion": {
        "sadness": 0.273896,
        "joy": 0.644126,
        "fear": 0.024682,
        "disgust": 0.043153,
        "anger": 0.019833
      },
      "count": 1,
      "confidence": 0.8
    }
  ]
}
```

4. Knowledge stduio를 활용하여 커스텀 모델 제작



knowledge studio로 이동하여 커스텀 모델을 만든 후
response = natural_language_understanding.list_models().get_result()
로 커스텀 모델을 확인할 수 있고
response = natural_language_understanding.delete_model(model_id='model_id').get_result()
로 커스텀 모델을 삭제할 수 있다.
자세히는 다루지 않겠다. knowlege stduio에 대해 궁금하면 아래 링크를 참고
https://cloud.ibm.com/docs/watson-knowledge-studio?topic=watson-knowledge-studio-wks_overview_full

참고 자료 : https://cloud.ibm.com/apidocs/natural-language-understanding?code=python#text-analytics-features