FACTSET NER (NAMED ENTITY RECOGNITION) API

FACTSET SEE THE ADVANTAGE

FactSet NER API

Contents

1. Overview

The FactSet NER (Named Entity Recognition) service identifies companies, people, locations, health conditions, drug names, numbers, monetary values, and dates from unstructured or semi-structured documents. In addition to providing the text and type of the entity names, along with their start and end offsets in the document, this service also provides the best matching FactSet identifiers for companies and people found in the text. This unique FactSet identifier allows you to link any document with other FactSet content sets, such as historical prices or fundamental data.

The FactSet NER service accepts user queries through a Swagger-based restful API (details below).

a. Methodology

The FactSet NER machine learning model was trained on business and financial documents such as EDGAR filings, news, and press releases. It uses a combination of proprietary and third party NER engines. By merging FactSet's best-in-class data with cutting-edge AI technologies, NER identifies a range of entities from company tickers to health conditions and drugs found in the StreetAccount drug database. The service also provides the best matching FactSet identifiers for companies and people found in the text.

b. Sample NER Functionality

Given a sample text:

As demand for the coronavirus vaccine plateaus, Ohio Governor Mike DeWine is giving state residents a shot to win $1 million. Starting May 26, the state will award $1 million each week to an adult who has received at least the first dose of the COVID-19 vaccine, DeWine said. There will be five total drawings. This will give anyone who has not been vaccinated time to get the first dose of Pfizer or Moderna and be well on the way to full immunity, DeWine said Wednesday. West Virginia Governor Jim Justice wanted to give every vaccinated resident between the ages of 16 and 35 a $100 savings bond. States and localities have been given wide discretion by the Treasury Department in spending federal aid.

The FactSet NER service can identify companies, people, locations, health conditions, drug names, dates, monetary values, and numbers from the document along with their associated FactSet identifiers. Below is an image of a sample output taken from an internal user interface for NER. The image is for illustrative purposes only.


As demand for the coronavirus vaccine plateaus, Ohio Governor Mike DeWine is giving state residents a shot to win $1 million. Starting May 26, the state will award $1 million each week to an adult who has received at least the first dose of the COVID-19 vaccine, DeWine said. There will be five total drawings. "This will give anyone who has not been vaccinated time to get the first dose of Pfizer or Moderna and be well on the way to full immunity," DeWine said Wednesday. West Virginia Governor Jim Justice wanted to give every vaccinated resident between the ages of 16 and 35 a $100 savings bond. States and localities have been given wide discretion by the Treasury Department in spending federal aid.

A government-run study of Gilead's remdesivir showed that the medicine is effective against four variants of Covid-19, the disease caused by the novel coronavirus. A trial at a Chicago hospital suggested that over 400 patients were doing better than expected on remdesivir. Days later, a summary of results from a December study conducted in China showed that patients on the drug did not improve more than those in a control group. Scott Gottlieb, the former commissioner of the Federal Drug Administration, said he expected there was enough evidence for the agency to issue an emergency use authorization for remdesivir. Effectiveness of the drug is expected to save an estimated 10 million dollars in United States vaccine purchases in 2021.

In the sample output image above, items in orange are company or organization names, items in blue are person names, items in red are locations, items in purple are drug names and health conditions, items in brown are dates, and items in green are numbers and monetary values.

c. Accessing the NER API

The FactSet NER service is accessible to authorized FactSet clients only. When accessing the service, please make sure that your account has been enabled for the NER API. If not, calls to the NER service will fail. Please contact your FactSet representative if you encounter authentication issues while calling the NER service.

d. Document Size and Other Specifications

Currently, this service identifies names from English documents only. It expects input text with characters in their usual cases (e.g., a company or person name must begin with a capital letter).

The NER service will process the first 10,000 characters of any document. If you need to process additional characters, please split the document into sections, each containing 10,000 characters or fewer before sending it to NER. The service might also impose a request rate limit as needed.

e. Expected Results

While we are constantly working to improve the performance of the service, no entity recognition service will be 100% correct. It is possible that the service will miss an entity, mis-identify an entity as the wrong type, or incorrectly match an entity name to the wrong FactSet entity ID. Please send us your feedback about any of these issues. This will be a great help in improving the service over time.

2. Submitting a Request to the NER API

The FactSet NER service supports a restful HTTP-based API.

a. FactSet NER API Endpoint

Production URL for the service:

- https://api.factset.com/cognitive/nlp/v1/ner/entities

The service supports HTTP POST only

b. Building a Request-Payload for the NER API

A sample JSON payload to the FactSet NER service follows the template below:

text*	**String (required)** *example: As demand for the coronavirus vaccine plateaus, Ohio Governor Mike DeWine is giving state residents a shot to win $1 million. Starting May 26, the state will award $1 million each week to an adult who has received at least the first dose of the COVID-19 vaccine, DeWine said. There will be five total drawings. This will give anyone who has not been vaccinated time to get the first dose of Pfizer or Moderna and be well on the way to full immunity, DeWine said Wednesday. West Virginia Governor Jim Justice wanted to give every vaccinated resident between the ages of 16 and 35 a $100 savings bond. States and localities have been given wide discretion by the Treasury Department in spending federal aid.* **English plain text to extract named entities from. Maximum of 10,000 characters.**
filterEntities	**boolean (optional)** *default: true* **Drop lower probability entities using machine learning filtering rules.**
enableIDLookup	**boolean (optional)** *default: true* **Retrieve FactSet IDs for entities**

Sample Payload

Below is a sample document from which we need to identify named entities:

As demand for the coronavirus vaccine plateaus, Ohio Governor Mike DeWine is giving state residents a shot to win $1 million. Starting May 26, the state will award $1 million each week to an adult who has received at least the first dose of the COVID-19 vaccine, DeWine said. There will be five total drawings. This will give anyone who has not been vaccinated time to get the first dose of Pfizer or Moderna and be well on the way to full immunity, DeWine said Wednesday. West Virginia Governor Jim Justice wanted to give every vaccinated

resident between the ages of 16 and 35 a $100 savings bond. States and localities have been given wide discretion by the Treasury Department in spending federal aid.

A request payload for this sample document is constructed as a JSON-formatted query which can be saved as a file (ner.json):

```
{
  "data": {
    "text": "As demand for the coronavirus vaccine plateaus, Ohio Governor Mike DeWine is giving state
residents a shot to win $1 million. Starting May 26, the state will award $1 million each week to an adult
who has received at least the first dose of the COVID-19 vaccine, DeWine said. There will be five total
drawings. This will give anyone who has not been vaccinated time to get the first dose of Pfizer or Moderna
and be well on the way to full immunity, DeWine said Wednesday. West Virginia Governor Jim Justice wanted to
give every vaccinated resident between the ages of 16 and 35 a $100 savings bond. States and localities have
been given wide discretion by the Treasury Department in spending federal aid.",
    "filterEntities": true,
    "enableIDLookup": true
  }
}
```

c. Sending a Sample Request to the NER API

Below is a command using cURL to send a POST request to the service using the sample input payload from a JSON file (ner.json).

```
curl -s https://api.factset.com/cognitive/nlp/v1/ner/entities -d@ner.json -H "Content-Type:
application/json"
```

d. Output Response from the NER API

In response to a valid request, the NER service returns a response with a list of named entities extracted from the document in JSON format:

data: {

entities: [

{

type string

example: NE_GPE

Type of the extracted named entity (NE).

text string

```
                                    Text for the named entity, as
                                    found in the source text.

                startChar           integer
                                    minimum: 0

                                    Starting character position of
                                    the entity text.

                endChar             integer
                                    minimum: 0

                                    Ending character position of
                                    the entity text.

                lookupText          string

                                    Name of the FactSet entity
                                    linked to the text (if
                                    applicable/found)

                entityId            string

                                    FactSet Entity ID for the
                                    extracted entity (if
                                    applicable/found)

                lookupURL           string

                                    URL to the extracted entity in
                                    FactSet (if applicable/found)

                associatedOrgs      array

                                    (Coming Soon) Organizations
                                    associated with this entity.
        }
        ]
}
```

Please Note: The associatedOrgs response parameter will be available at a later date and will return organizations associated with the identified entity.

NER Entity Types

The following is a list of the current entity types (“type”) returned by the API.

```
NE_GPE – Geopolitical Entities (Locations)
NE_ORG – Companies and Organizations
NE_PERSON - People
NE_NUMBER - Numbers
NE_MONEY – Monetary Values
NE_DATETIME - Dates
NE_DRUG – Drug Names
NE_HEALTH_INDICATION – Health Conditions
NE_PRODUCT - Products
```

Sample Response from NER API

For the given sample payload, the NER service will return a response in the following format (full response truncated on this document for space):


```
{
"data": {
  "entities": [
    {
      "type": "NE_HEALTH_INDICATION",
      "text": "coronavirus",
      "startChar": 18,
      "endChar": 29,
      "lookupText": "Coronavirus",
      "entityId": "1638",
      "lookupUrl": "https://my.apps.factset.com/drugs/indication?kw_id=1638",
      "associatedOrgs": []
    },
    {
      "type": "NE_GPE",
      "text": "Ohio",
      "startChar": 48,
      "endChar": 52,
      "lookupText": "",
      "entityId": "0",
      "lookupUrl": null,
      "associatedOrgs": []
    },
    {
      "type": "NE_PERSON",
      "text": "Mike DeWine",
      "startChar": 62,
      "endChar": 73,
      "lookupText": "Richard Michael DeWine",
      "entityId": "0FCYK1-E",
      "lookupUrl": "https://my.apps.factset.com/navigator/people/people-snapshot/0FCYK1-E",
      "associatedOrgs": []
    },
```

```
    {
      "type": "NE_MONEY",
      "text": "$1 million",
      "startChar": 114,
      "endChar": 124,
      "lookupText": null,
      "entityId": null,
      "lookupUrl": null,
      "associatedOrgs": []
    },
    {
      "type": "NE_DATETIME",
      "text": "May 26",
      "startChar": 135,
      "endChar": 141,
      "lookupText": null,
      "entityId": null,
      "lookupUrl": null,
      "associatedOrgs": []
    },
    {
      "type": "NE_MONEY",
      "text": "$1 million",
      "startChar": 164,
      "endChar": 174,
      "lookupText": null,
      "entityId": null,
      "lookupUrl": null,
      "associatedOrgs": []
    },
  (etc…)]
  }
}
```

3. A Sample Client for the NER API

Sample Python code for a simple client utilizing the FactSet NER service

```
import requests

from requests.auth import HTTPBasicAuth

def extract_from_text(doc_text,
                      min_doc_text_size= 10,
                      fds_ner_svc_url= 'https://api.factset.com/cognitive/nlp/v1/ner/entities'):
    if (not doc_text or not doc_text.strip() or min_doc_text_size> len(doc_text)):
        return None

    payload = {
                'data':{
                'text': doc_text
                }
    }
    response_json = None
    try:
        resp = requests.post(anna_ner_svc_url, json=payload, auth=HTTPBasicAuth('username-serial', 'api-key'))
        if not resp:
            status_code = resp.status_code if (resp is not None) else -1
            raise ValueError(f'Received unexpected response from service: status_code: {status_code}')
        response_json = resp.json()
    except Exception as ex:
        #print(str(ex))   # debug only
        raise ex

    if (not response_json or not isinstance(response_json, dict)
        or ('errors' in response_json) or ('data' not in response_json)):
        return None
    return response_json['data']['entities']
```