# AI Themes API

# AI Themes API

## Contents

# 1. Overview

The AI Themes service extracts significant concepts in the form of phrases from unstructured text.  Document Themes can be used to quickly understand the contents of a document, as the first step step in determining the sentiment of a document, or when compared with themes from similar documents, to discover when new ideas emerge. Every theme is paired with a numeric value to indicate its significance within the document and can additionally be paired with a sentiment score.

The FactSet Workstation displays themes from earnings call transcripts as a word cloud in the document viewer.  However, the service can be used on any unstructured text, including news stories, text sections of company filings, text from web pages, and research reports.

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

## a. Methodology

Unsupervised key-phrase extraction techniques are used to extract themes that are representative of the document. Third-party APIs, open-source tools, and internal FactSet research have been used to ensure we arrive at the best output.

## b. Sample AI Themes Functionality

Given a selection of unstructured text:

*Studio Entertainment: At Studio Entertainment, operating income decreased in the quarter due to lower theatrical distribution and home entertainment results. Worldwide theatrical results continued to be adversely impacted by COVID-19, as theaters were closed in many key markets both domestically and internationally. With no significant worldwide theatrical releases in the quarter, we faced a difficult comparison against the strong performance of The Lion King and Toy Story 4 in the prior-year quarter. Operating Results: On our last earnings call, we said that we expected Q4 operating results of our DTC businesses to improve by approximately $100mm relative to the prior-year quarter. Our results came in better than that guidance, with operating income at our DTC businesses improving by approximately $300mm vs. the prior year due to better-than-expected performance across all three of our streaming services. I will note that we do not plan to further update any of our subscriber numbers until our Investor Day on December 10 at our International Channels, lower results were due to lower affiliate and advertising revenues, partially offset by a decrease in cost.*

The AI Themes service will return a list of themes and associated relevancy scores. The service can also return the sentiment of the theme and an associated sentiment score. Below are themes, sentiment, and scores shown in a table. This is for illustrative purposes only.

| Theme | Relevance Score | Sentiment | Sentiment Score |
|---|---|---|---|
| prior-year quarter | 100% | Positive | 98.76% |
| worldwide theatrical results | 99.88% | Negative | 99.10% |

3   FactSet Research Systems Inc. | www.factset.com

| home entertainment results | 97.98% | Negative | 99.13% |
| --- | --- | --- | --- |
| studio entertainment | 88.27% | Negative | 99.13% |
| last earnings call | 88.18% | Positive | 98.87% |
| subscriber numbers | 85.55% | Negative | 99.01% |
| strong performance | 83.02% | Negative | 97.66% |

# c. Accessing the AI Themes API

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

# d. Document Size and Other Specifications

Currently, this service only supports English language documents in plain text format. The minimum input length is 100 characters. The maximum input length is 15,000 characters per request. Any text beyond the maximum length will not be considered for output.

# e. Expected Results

It is possible that the service will miss themes you find significant or will return themes that you were not expecting.  Please send us your feedback.  This will be a great help in improving the service over time.

# 2. Submitting a Request to the AI Themes API

The AI Themes service supports a restful HTTP-based API.

## a. AI Themes Endpoint

Production URLs for the service:

https://api.factset.com/cognitive/nlp/v1/themes

https://api.factset.com/cognitive/nlp/v1/themes/{id}/status

https://api.factset.com/cognitive/nlp/v1/themes/{id}

The service supports HTTP POST for the **themes** endpoint and HTTP GET for the **ID** and **status** endpoints.

## b. Building a Request-Payload for the AI Themes API

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

**Data**{
    `includeSentiments`**\***

              **Boolean (required)**
              *default: false*

              **Retrieve sentiment for themes**

    `input`**\***

              **String (required)**
              *Example: At Studio Entertainment, operating income decreased in the quarter due to lower theatrical distribution and home entertainment results. Worldwide theatrical results continued to be adversely impacted by COVID-19, as theaters were closed in many key markets both domestically and internationally. With no significant worldwide theatrical releases in the quarter, we faced a difficult comparison against the strong performance of The Lion King and Toy Story 4 in the prior-year quarter. Operating Results: On our last earnings call, we said that we expected Q4 operating results of our DTC businesses to improve by approximately $100mm relative to the prior-year quarter. Our results came in better than that guidance, with operating income at our DTC businesses improving by approximately $300mm vs. the prior year due to better-than-expected performance across all three of our streaming services. I will note that we do not plan to further update any of our subscriber numbers until our Investor Day on December 10 at our International Channels, lower*

```
            results were due to lower affiliate and advertising revenues,
            partially offset by a decrease in cost.

            English plain text to extract themes from. Maximum of 1,024
            words.

    }
```

## Sample Payload

Below is a sample input of unstructured text for the AI Themes service:

*Studio Entertainment: At Studio Entertainment, operating income decreased in the quarter due to lower theatrical distribution and home entertainment results. Worldwide theatrical results continued to be adversely impacted by COVID-19, as theaters were closed in many key markets both domestically and internationally. With no significant worldwide theatrical releases in the quarter, we faced a difficult comparison against the strong performance of The Lion King and Toy Story 4 in the prior-year quarter. Operating Results: On our last earnings call, we said that we expected Q4 operating results of our DTC businesses to improve by approximately $100mm relative to the prior-year quarter. Our results came in better than that guidance, with operating income at our DTC businesses improving by approximately $300mm vs. the prior year due to better-than-expected performance across all three of our streaming services. I will note that we do not plan to further update any of our subscriber numbers until our Investor Day on December 10 at our International Channels, lower results were due to lower affiliate and advertising revenues, partially offset by a decrease in cost.*

A request payload for the sample text is constructed as a JSON-formatted query that can be saved as a file (themes.json):

```json
{
  "data": {
    "includeSentiments": true,
    "input": "Studio Entertainment: At Studio Entertainment, operating income decreased in the quarter due
to lower theatrical distribution and home entertainment results. Worldwide theatrical results continued to
be adversely impacted by COVID-19, as theaters were closed in many key markets both domestically and
internationally. With no significant worldwide theatrical releases in the quarter, we faced a difficult
comparison against the strong performance of The Lion King and Toy Story 4 in the prior-year quarter.
Operating Results: On our last earnings call, we said that we expected Q4 operating results of our DTC
businesses to improve by approximately $100mm relative to the prior-year quarter. Our results came in better
than that guidance, with operating income at our DTC businesses improving by approximately $300mm vs. the
prior year due to better-than-expected performance across all three of our streaming services. I will note
that we do not plan to further update any of our subscriber numbers until our Investor Day on December 10 at
our International Channels, lower results were due to lower affiliate and advertising revenues, partially
offset by a decrease in cost."
  },
  "meta": {}
}
```

# c. Sending a Sample Request to the AI Themes API

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

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

# d. Output Response from the AI Themes API

In response to a valid request, the AI Themes **status** endpoint returns a unique ID {id} from the request in JSON format:

```
Response
{
    id                      string <uuid>
                            example: ecdf284b-2db8-4499-898e-8d6b59f28d2a
                            Unique ID to be used to retrieve the themes
                            extracted from the text.

    started_at              string <date-time>
                            example: 2022-10-31T16:31:30.423086
                            The date/time when the task started in UTC.

    status                  string
                            example: queued
                            Status of the task.
}
```

## Sample Response from the AI Themes API

For the given sample payload, the AI Themes service will return a unique ID in the following format:

```
{
  "data": {
    "id": "954d024e-47af-4aa0-9354-bba3603e1615",
    "startedAt": null,
    "status": "queued"
  }
}
```

# e. Using the Unique ID to Determine the Status of the Request

Most documents will be processed immediately, but longer documents may take a few seconds to process. The average processing time is about one second for every 1,000 words sent to the service. To determine whether or not the job has finished, use cURL to send a GET request to the **status** endpoint using the unique ID.

```
curl --request GET \
  --url https://api.factset.com/cognitive/nlp/v1/themes/ecdf284b-2db8-4499-898e-8d6b59f28d2a/status \
  --header 'Content-Type: application/json'
```

## Sample Response from the Status Endpoint

For the given unique ID, the **status** endpoint will return the current status of the job. When the job reaches a status of "created", the themes are ready to be retrieved using the **ID** endpoint.

**Please Note:** For security reasons, results expire within 500 seconds of being generated. Please retrieve your results using the **ID** endpoint as soon as possible.

```
{
  "data": {
    "id": "954d024e-47af-4aa0-9354-bba3603e1615",
    "startedAt": "2023-07-21T17:38:53.172081",
    "status": "created"
  }
}
```

# f. Using the Unique ID to Retrieve Themes

Once the status of the job is "finished", use cURL to send a GET request to the **ID** endpoint using the unique ID.

```
curl --request GET \
  --url https://api.factset.com/cognitive/nlp/v1/themes/ecdf284b-2db8-4499-898e-8d6b59f28d2a \
  --header 'Content-Type: application/json'
```

## Sample Response from the ID Endpoint

For the given unique ID, the **ID** endpoint will return the themes, confidence score, and sentiment (if requested):

```json
{
  "data": [
    {
      "sentiment": "NEGATIVE",
      "sentimentScore": 0.9910721183,
      "themeScore": 1,
      "themeText": "Worldwide theatrical results"
    },
    {
      "sentiment": "POSITIVE",
      "sentimentScore": 0.9887331724,
      "themeScore": 0.84,
      "themeText": "last earnings call"
    },
    {
      "sentiment": "POSITIVE",
      "sentimentScore": 0.9887331724,
      "themeScore": 0.97,
      "themeText": "Operating Results"
    },
    {
      "sentiment": "NEGATIVE",
      "sentimentScore": 0.9913682342,
      "themeScore": 0.92,
      "themeText": "home entertainment results"
    },
    {
      "sentiment": "POSITIVE",
      "sentimentScore": 0.97756019235,
      "themeScore": 0.72,
      "themeText": "DTC businesses"
    },
    {
      "sentiment": "NEGATIVE",
      "sentimentScore": 0.9901952147,
      "themeScore": 0.77,
      "themeText": "subscriber numbers"
    },
    {
      "sentiment": "NEGATIVE",
      "sentimentScore": 0.9766709208,
      "themeScore": 0.81,
      "themeText": "strong performance"
    },
    {
      "sentiment": "POSITIVE",
      "sentimentScore": 0.9663872123,
      "themeScore": 0.74,
      "themeText": "streaming services"
    },
```

```json
    {
      "sentiment": "NEGATIVE",
      "sentimentScore": 0.9901952147,
      "themeScore": 0.76,
      "themeText": "advertising revenues"
    },
    {
      "sentiment": "NEGATIVE",
      "sentimentScore": 0.9913682342,
      "themeScore": 0.77,
      "themeText": "theatrical distribution"
    },
    {
      "sentiment": "NEGATIVE",
      "sentimentScore": 0.9910721183,
      "themeScore": 0.66,
      "themeText": "many key markets"
    },
    {
      "sentiment": "NEGATIVE",
      "sentimentScore": 0.9766709208,
      "themeScore": 0.54,
      "themeText": "Toy Story"
    },
    {
      "sentiment": "NEGATIVE",
      "sentimentScore": 0.9766709208,
      "themeScore": 0.73,
      "themeText": "theatrical releases"
    },
    {
      "sentiment": "NEGATIVE",
      "sentimentScore": 0.9910721183,
      "themeScore": 0.48,
      "themeText": "COVID-19"
    },
    {
      "sentiment": "NEGATIVE",
      "sentimentScore": 0.9913682342,
      "themeScore": 0.73,
      "themeText": "Studio Entertainment"
    }
  ]
}
```

FactSet Research Systems Inc.  |  www.factset.com

# 3. A Sample Client for the AI Themes API

Sample Python code for a simple client utilizing the FactSet AI Themes service

```python
import requests

from requests.auth import HTTPBasicAuth

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

    payload = {"data": {'text': doc_text}}
    try:
        resp = requests.post(fds_themes_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}')
        result_id = resp.json()["data"]["id"]

        while(True):
            resp = requests.get(
                f"{fds_themes_svc_url}/{result_id}", auth=auth)
            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}')
            if resp.json() != 'Processing':
                return resp.json()
            time.sleep(1)
    except Exception as ex:
        #print(str(ex))   # debug only
        raise ex
```

FactSet Research Systems Inc. | www.factset.com