

Cognitive Insights: Themes API

Version 1.1



1 OVERVIEW

This REST API is the interface designed to provide access to our Content Analysis Engine. It is intended for anyone interested in exploring the features of our API.

The objective of the Content Analysis Engine is to understand topics and main ideas of text input. Our program utilizes a unique machine learning algorithm, the Content Analysis Engine (CAE) to parse text and extract the most important concepts from what is provided. For any text that is given to the CAE, the algorithm returns a list of concepts sorted by their importance vectors – the more important the concept this higher its vector value, and the higher up on the list it is.

On top of the base algorithm of the CAE, our API includes functionality for retrieving and manipulating content. That functionality is described below.

1.1 Date Format

All dates stored in the API, or passed to the API (as described below), should/will have the following format:

```
"Thu Mar 21 2013 13:51:22 GMT+0000"
```

In code format, for languages that support strftime or strptime, it looks as follows:

```
"%a %b %d %Y %H:%M:%S %Z%z"
```

1.2 Errors

Errors are returned to the user in the standard HTTP error code syntax. If there is any additional information for the specific error, it will be included in the body of the return, which will be JSON-formatted. Each REST API method below specifies its possible error returns.

1.3 Authentication

The authentication methodology for this API is form-based. When users create new accounts, they will set up a unique username/password combination, which they will use within their API calls to authenticate themselves once their account is set up.

1.4 SSL Only

All requests must be done over SSL only.

1.5 Data Upload Limits

In order to accommodate multiple users, there are some limits to how much data a user is allowed to upload for processing at one time. The maximum amount of data allowed for any one author is 300 kB of text, and the maximum amount of data allowed for any one POST request is 1 MB of text.

2 REST API METHODS

2.1 Setting Up An Account

This section introduces methods for setting up an account on our REST API.

/createUser

Description

Creates a new account from provided username and password information. This account can be used for all of our APIs, once it is created.

URL Structure

<https://bpc.bmedianet.com/themes-api/1/createUser.json>

Parameters

user

The string that you would like to use as your username for logging into the API repository

pswd

The string that you would like to use as your password for logging into the API repository

Method

POST

Return

The return will either return:

"User Created successfully"

if the requested username and password are available, or it will return an error.

Errors

400 – User not created in database

2.2 Making New Predictions

This section introduced the calls to our API associated with making a new prediction on previously unseen data.

/add_authors

Description

Requests to run the Content Analysis Engine on inputted text.

URL Structure:

https://bpc.bmedianet.com/themes-api/1/add_authors.json

Parameters

text

This must contain JSON-formatted content containing the text to be analyzed. For each author, there must be a separate object, labeled by the author's ID. Within each author object, each separate piece of text to be analyzed should be labeled by its date, as specified in the date format. This request assumes that, no matter if the authors exist in the database yet, that all pieces of text per job request call contain are new, and do not exist in the database yet.

Sample JSON Format:

```
{
  "AuthorID1" : {
    "Thu Mar 21 2013 13:51:22 GMT+0000" : "text",
    "Thu Mar 21 2013 13:58:40 GMT+0000" : "text"
  },
  "AuthorID2" : {
    "Thu Mar 21 2013 11:22:22 GMT+0000" : "text",
    "Thu Mar 21 2013 23:47:35 GMT+0000" : "text",
    "Fri Mar 22 2013 00:05:03 GMT+0000" : "text",
    "Fri Mar 22 2013 00:06:40 GMT+0000" : "text"
  }
}
```

Input Value Definitions:

AuthorID1

sample ID for an author of inputted text

Thu Mar 21 2013 13:51:22 GMT+0000

sample date formatting, used to mark each separate piece of text being inputted.

Method:

POST

Return:

The status of the job, as well as a link to use to check the status of the job, in JSON format. All job requests for inputted data are processed asynchronously.

Sample JSON return:

```
{
  "returnLink":
    "/content/personality/2013/6/21/job/results.json"
  "statusLink": "/content/personality/2013/6/21/job/status.json"
}
```

Response Value Definitions:

statusLink

String that indicates the link that should be used as a request to see if the data has been processed. The status link will contain a json-formatted string that gives the status of the analysis process. When

the status link says “finished”, that means the analysis is complete. In order to see the status, make a GET request to the location provided with “https://bpc.bmedianet.com”.

returnLink

String that indicates the link that should be used as a request to get the results of the analysis. This link will return a 404 response until the analysis is finished, after which it will contain the prediction data. In order to see the results, make a GET request to the location provided with “https://bpc.bmedianet.com”.

Sample JSON return for processed data:

```
{
  "AuthorID1": {
    "Concept1": 0.53,
    "Concept2": 0.64,
    "Concept3": 0.45
  }
  "AuthorID2": {
    "Concept1": 0.9,
    "Concept2": 1.43,
    "Concept3": 0.89
  }
}
```

Response Value Definitions:

AuthorID1

sample author that has been analyzed

Concept[1,2,3,...]

A concept found in the text, along with its importance vector. The higher the vector value, the more important it is.

Errors:

400 - Bad Input Parameter
405 - Request Method Not Expected

2.3 Retrieving and Manipulating Existing Author Information

This section describes all of the calls relating to getting, changing and removing content that already exists in your personality database.

/all_authors

Description

Returns the contents of your database.

URL Structure

https://bpc.bmedianet.com/themes-api/1/all_authors.json

Parameters

count

Optional value, determines # of most recent authors to retrieve. Must be an integer number, i.e. 1,2,3,... Default value that is used if this parameter is unspecified, returns all authors in the database.

timeline

Optional value, determines whether to return theme vectors for an author by day for that author has stored in the database. Value must be either "True" or "False". "True" means that each authors' object of theme vectors contains values by month for the most recent year, on top of their overall values. "False", the default value for when this parameter is unspecified, means that only overall concept vectors will be returned.

max_concepts

Optional value, an integer that determines how many of the top concepts to return for each author (and day, year, etc. should you choose the timeline option). Default value for this is 10.

Method

GET

Return

JSON-formatted object of all the authors in the database (or of the amount specified). Concepts are contained in a JSON-list, ranked by importance.

Sample HTTP Request and Response:

https://bpc.bmedianet.com/themes-api/1/all_authors.json?timeline=True

Sample JSON return for processed data:

```
{
  "AuthorID1": {
    "total" : [
      "Concept1": 2.53,
      "Concept2": 1.64,
      "Concept3": 0.4
    ],
    "2013" : {
      "total" : [
        "Concept1": 2.53,
        "Concept2": 1.64,
        "Concept3": 0.4
      ],
      "January" : {
        "total" : [
          "Concept1": 2.53,
          "Concept2": 1.64,
          "Concept3": 0.4
        ],
        "25" : [
          "Concept1": 2.53,
          "Concept2": 1.64,
          "Concept3": 0.4
        ]
      ]
    }
  }
}
```

Errors:

400 - Bad Input Parameter
405 - Request Method Not Expected

Notes:

Not all of the URL parameters specified above need to be present. If they are not present, default values will be assigned. For the JSON response, not all days will be present uniformly for each author - they will differ depending on the data. Each month's key is the name of that month, starting with the most recent one, and the overall scores key is "Total". If no authors are in the database, and empty JSON list will be returned.

/authors_by_concept

Description

Returns a list of authors within your themes database that have all of the requested theme(s). The user may specify as many themes as desired.

URL Structure

https://bpc.bmedianet.com/themes-api/1/authors_by_concept.json

Parameters

conceptN

In this instance, 'N' stands for any natural number 1,2,... This requests is designed to accept any number of potential concepts to be searched for. For each concept, it should be labeled as Concept1, Concept2, Concept3, etc. Each concept may be the string of a word, or of a two-word phrase – if the string is a two-word phrase, the phrase must be separated by an underscore. For example, “a lot” would become “a_lot” as a parameter here.

Method

GET

Return

JSON-formatted object of all the authors that contain the desired concept(s) from the parameter(s). Each author object contains an object of occurrences, separated by date. The date is formatted as listed in the introduction, but object will only be separated by day. The time of day is irrelevant, as all instances for a given data are put together into one importance vector. Each date object contains a concept-vector mapping, where the importance vector is for the date of the date object.

Sample HTTP Request and Response:

https://bpc.bmedianet.com/themes-api/1/authors_by_concept.json?concept1=document&concept2=have_fun

Sample JSON return:

```
{
  "AuthorID1": {
    "Thu Mar 21 2013 13:51:22 GMT+0000" : {
      "documents" : 0.43,
      "a lot" : 1.75
    }
    "Fri Mar 29 2013 13:51:22 GMT+0000" : {
      "documents" : 0.93,
      "a lot" : 0.78
    }
  }
}
```

Errors

400 - Bad Input Parameter
 405 - Request Method Not Expected

/author

Description

Returns an author by their ID

URL Structure

<https://bpc.bmedianet.com/themes-api/1/author.json>

Parameters

ID

Author's ID that is being searched for.

Method

GET

Returns

JSON-Formatted object containing the author's overall concept use, as well as scores for each day for which there is data (as if the timeline parameter were true)

Sample HTTP Request and Response:

<https://bpc.bmedianet.com/themes-api/1/author.json?ID=AuthorID1>

```
{
  "AuthorID1": {
    "total" : [
      "Concept1": 2.53,
      "Concept2": 1.64,
      "Concept3": 0.4
    ],
    "2013" : {
      "total" : [
        "Concept1": 2.53,
        "Concept2": 1.64,
        "Concept3": 0.4
      ],
      "January" : {
        "total" : [
          "Concept1": 2.53,
          "Concept2": 1.64,
          "Concept3": 0.4
        ],
        "25" : [
          "Concept1": 2.53,
          "Concept2": 1.64,
          "Concept3": 0.4
        ]
      ]
    }
  }
}
```

Errors
404 - Author requested was not found in the database

Notes
None

/remove_author

Description
Deletes an author from your themes database

URL Structure
https://bpc.bmedianet.com/themes-api/1/remove_author.json

Parameters
ID
Author's ID that is to be removed

Method
POST

Returns
JSON-formatted string saying that the author has been successfully removed.

Sample HTTP Request and Response:

https://bpc.bmedianet.com/themes-api/1/remove_author.json?ID=AuthorID1

response: "author successfully removed"

Errors
404 - Author requested was not found in the database

Notes
None