Harpocrates API

This is the Rest API specification for the Harpocrates application
More information: https://harpocrates-app.gitlab.io/harpocrates/
Version: 9.2.0
MIT
https://gitlab.com/harpocrates-app/harpocrates/-/raw/master/LICENSE.md

Access

Methods

[Jump to Models]

Table of Contents

Document

POST /documentSet/{setId}

Up

Add a new document to the set (**createDocument**)

Contents of the document in the body of the request. This should be in plain text. The Content-Type header should be appropriately set to text/plain.

Path parameters

setId (required)
Path Parameter — ID of a set

Consumes

This API call consumes the following media types via the Content-Type request header:

- `text/plain`

Request body

body string (required)
Body Parameter —

Return type

document

Example data

Content-Type: application/json

```
{
  "textSplitGranularity" : "paragraph",
  "textContents" : [ {
    "sensitiveSections" : {
      "sensitiveSections" : [ "", "" ]
    },
```

```
    "content" : "content"
  }, {
    "sensitiveSections" : {
      "sensitiveSections" : [ "", "" ]
    },
    "content" : "content"
  } ],
  "predictedClassification" : {
    "classifier" : "SVM",
    "explanations" : [ {
      "features" : [ "", "" ],
      "explainer" : "Lime"
    }, {
      "features" : [ "", "" ],
      "explainer" : "Lime"
    } ],
    "sensitive" : true,
    "sensitivity" : 78.3
  },
  "name" : "",
  "documentId" : ""
}
```


Produces
This API call produces the following media types according to the Accept request header; the media type will be conveyed by the Content-Type response header.

- `application/json`

Responses
200
Successfully created document in set document
400
The request was invalid JSON or missing some attributes httpStatus
404
The specified resource was not found httpStatus
409
Conflicting element already present httpStatus

DELETE /documentSet/{setId}/{docId}

Up

delete document from set (**deleteDocument**)

Path parameters

setId (required)
Path Parameter — ID of a set

docId (required)
Path Parameter — ID of a document

Return type
document

Example data
Content-Type: application/json


```
{
  "textSplitGranularity" : "paragraph",
  "textContents" : [ {
    "sensitiveSections" : {
      "sensitiveSections" : [ "", "" ]
    },
    "content" : "content"
  }, {
    "sensitiveSections" : {
      "sensitiveSections" : [ "", "" ]
    },
    "content" : "content"
  } ],
  "predictedClassification" : {
    "classifier" : "SVM",
    "explanations" : [ {
      "features" : [ "", "" ],
      "explainer" : "Lime"
    }, {
```

```
      "features" : [ "", "" ],
      "explainer" : "Lime"
    } ],
    "sensitive" : true,
    "sensitivity" : 78.3
  },
  "name" : "",
  "documentId" : ""
}
```


Produces
This API call produces the following media types according to the Accept request header; the media type will be conveyed by the Content-Type response header.

- `application/json`

Responses
200
Successfully deleted document in the set document
400
The request was invalid JSON or missing some attributes httpStatus
404
The specified resource was not found httpStatus

GET /documentSet/{setId}/{docId}

get document from set (**getDocument**)

Path parameters

setId (required)
Path Parameter — ID of a set

docId (required)
Path Parameter — ID of a document

Return type
document

Example data
Content-Type: application/json


```
{
  "textSplitGranularity" : "paragraph",
  "textContents" : [ {
    "sensitiveSections" : {
      "sensitiveSections" : [ "", "" ]
    },
    "content" : "content"
  }, {
    "sensitiveSections" : {
      "sensitiveSections" : [ "", "" ]
    },
    "content" : "content"
  } ],
  "predictedClassification" : {
    "classifier" : "SVM",
    "explanations" : [ {
      "features" : [ "", "" ],
      "explainer" : "Lime"
    }, {
      "features" : [ "", "" ],
      "explainer" : "Lime"
    } ],
    "sensitive" : true,
    "sensitivity" : 78.3
  },
  "name" : "",
  "documentId" : ""
}
```


Produces
This API call produces the following media types according to the Accept request header; the media type will be conveyed by the Content-Type response header.

- `application/json`

Responses
200
Successfully got document from set document
400
The request was invalid JSON or missing some attributes httpStatus
404
The specified resource was not found httpStatus

GET /documentSet/{setId}/{docId}/original

get original document content as plain text (**getOriginalContent**)

Path parameters

setId (required)
Path Parameter — ID of a set

docId (required)
Path Parameter — ID of a document

Return type
String

Example data
Content-Type: application/json

```
"The source was E. Macron a renowed banker working in the French Presidency"
```

Produces
This API call produces the following media types according to the Accept request header; the media type will be conveyed by the Content-Type response header.

- `text/plain`
- `application/json`

Responses
200
Successfully got the original document content as plain text String
400
The request was invalid JSON or missing some attributes httpStatus
404
The specified resource was not found httpStatus

GET /documentSet/{setId}/{docId}/predictedClassification

Get the predicted classification for the document (**getPredictedClassification**)

Path parameters

setId (required)
Path Parameter — ID of a set

docId (required)
Path Parameter — ID of a document

Return type
predictedClassification

Example data
Content-Type: application/json

```
{
  "classifier" : "SVM",
  "explanations" : [ {
    "features" : [ "", "" ],
    "explainer" : "Lime"
  }, {
    "features" : [ "", "" ],
    "explainer" : "Lime"
  } ],
  "sensitive" : true,
  "sensitivity" : 78.3
}
```

Produces
This API call produces the following media types according to the Accept request header; the media type will be conveyed

by the Content-Type response header.

- `application/json`

Responses
200
Successfully got the predicted classification predictedClassification
400
The request was invalid JSON or missing some attributes httpStatus
404
The specified resource was not found httpStatus

GET /documentSet/{setId}/{docId}/redacted

Up

get redacted document content as plain text (**getRedactedContent**)

Path parameters

setId (required)
Path Parameter — ID of a set

docId (required)
Path Parameter — ID of a document

Return type
String

Example data
Content-Type: application/json

```
"The source was ████████  a renowned ████ working in ████████████"
```

Produces
This API call produces the following media types according to the Accept request header; the media type will be conveyed by the Content-Type response header.

- `text/plain`
- `application/json`

Responses
200
Successfully got the redacted document content as plain text String
400
The request was invalid JSON or missing some attributes httpStatus
404
The specified resource was not found httpStatus

Set

POST /documentSet

Up

Add a new document set set to the engine (**createSet**)

Consumes
This API call consumes the following media types via the Content-Type request header:

- `application/json`

Request body

body documentSet (required)
Body Parameter — documentSet descriptor that needs to be added to the engine

Return type
documentSet

Example data
Content-Type: application/json

```
{
  "documentCount" : 42,
  "name" : "",
  "setId" : ""
}
```

Produces
This API call produces the following media types according to the Accept request header; the media type will be conveyed by the Content-Type response header.

- `application/json`

Responses
200
The created document set Object documentSet
400
The request was invalid JSON or missing some attributes httpStatus
409
Conflicting element already present httpStatus

DELETE /documentSet/{setId}

Up

delete the set (**deleteSet**)

Path parameters

setId (required)
Path Parameter — ID of a set

Return type
documentSet

Example data
Content-Type: application/json

```
{
  "documentCount" : 42,
  "name" : "",
  "setId" : ""
}
```

Produces
This API call produces the following media types according to the Accept request header; the media type will be conveyed by the Content-Type response header.

- `application/json`

Responses
200
Successfully deleted the document set documentSet
400
The request was invalid JSON or missing some attributes httpStatus
404
The specified resource was not found httpStatus

GET /documentSet/{setId}

Up

lists all documents in the set (**getSet**)

Path parameters

setId (required)
Path Parameter — ID of a set

Return type
documents

Example data
Content-Type: application/json

```
{
  "documents" : [ {
    "textSplitGranularity" : "paragraph",
    "textContents" : [ {
      "sensitiveSections" : {
        "sensitiveSections" : [ "", "" ]
      },
      "content" : "content"
    }, {
      "sensitiveSections" : {
```

```
        "sensitiveSections" : [ "", "" ]
      },
      "content" : "content"
    } ],
    "predictedClassification" : {
      "classifier" : "SVM",
      "explanations" : [ {
        "features" : [ "", "" ],
        "explainer" : "Lime"
      }, {
        "features" : [ "", "" ],
        "explainer" : "Lime"
      } ],
      "sensitive" : true,
      "sensitivity" : 78.3
    },
    "name" : "",
    "documentId" : ""
  }, {
    "textSplitGranularity" : "paragraph",
    "textContents" : [ {
      "sensitiveSections" : {
        "sensitiveSections" : [ "", "" ]
      },
      "content" : "content"
    }, {
      "sensitiveSections" : {
        "sensitiveSections" : [ "", "" ]
      },
      "content" : "content"
    } ],
    "predictedClassification" : {
      "classifier" : "SVM",
      "explanations" : [ {
        "features" : [ "", "" ],
        "explainer" : "Lime"
      }, {
        "features" : [ "", "" ],
        "explainer" : "Lime"
      } ],
      "sensitive" : true,
      "sensitivity" : 78.3
    },
    "name" : "",
    "documentId" : ""
  } ]
}
```


Produces
This API call produces the following media types according to the Accept request header; the media type will be conveyed by the Content-Type response header.

- `application/json`

Responses
200
Successfully got list of documents in set documents
400
The request was invalid JSON or missing some attributes httpStatus
404
The specified resource was not found httpStatus

GET /documentSet

List all document sets known by the engine (**getSets**)

Return type
documentSets

Example data
Content-Type: application/json


```
{
  "documentSets" : [ {
    "documentCount" : 42,
    "name" : "",
```

```
    "setId" : ""
  }, {
    "documentCount" : 42,
    "name" : "",
    "setId" : ""
  } ]
}
```

Produces
This API call produces the following media types according to the Accept request header; the media type will be conveyed by the Content-Type response header.

- `application/json`

Responses
200
Listing of all known document sets documentSets

TextContent

POST /documentSet/{setId}/{docId}/{textContentIndex}/sensitiveSection

Up

add a sensitive section to the document (**addSensitiveSection**)

add a sensitive section to the document

Path parameters

setId (required)
Path Parameter — ID of a set

docId (required)
Path Parameter — ID of a document

textContentIndex (required)
Path Parameter — Index of the position of a TextContent object in the list of TextContent objects making up the document format: int32

Consumes
This API call consumes the following media types via the Content-Type request header:

- `application/json`

Request body

body sensitiveSection (optional)
Body Parameter —

Return type
sensitiveSection

Example data
Content-Type: application/json

```
""
```

Produces
This API call produces the following media types according to the Accept request header; the media type will be conveyed by the Content-Type response header.

- `application/json`

Responses
201
Successfully added the sensitive section sensitiveSection
400
The request was invalid JSON or missing some attributes httpStatus
404
The specified resource was not found httpStatus

PUT /documentSet/{setId}/{docId}/{textContentIndex}/sensitiveSections

Up

Overwrite the sensitive sections of a TextContent object (**addSensitiveSections**)

Overwrite the sensitive sections of a TextContent object

Path parameters

setId (required)
Path Parameter — ID of a set

docId (required)
Path Parameter — ID of a document

textContentIndex (required)
Path Parameter — Index of the position of a TextContent object in the list of TextContent objects making up the document format: int32

Consumes
This API call consumes the following media types via the Content-Type request header:

- `application/json`

Request body

body sensitiveSections (optional)
Body Parameter —

Return type
sensitiveSections

Example data
Content-Type: application/json

```
{
  "sensitiveSections" : [ "", "" ]
}
```

Produces
This API call produces the following media types according to the Accept request header; the media type will be conveyed by the Content-Type response header.

- `application/json`

Responses
200
Successfully overwrote the sensitive sections sensitiveSections
400
The request was invalid JSON or missing some attributes httpStatus
404
The specified resource was not found httpStatus

GET /documentSet/{setId}/{docId}/{textContentIndex}/sensitiveSections

get sensitive sections of a TextContent object (**getSensitiveSections**)

get sensitive sections of a TextContent object

Path parameters

setId (required)
Path Parameter — ID of a set

docId (required)
Path Parameter — ID of a document

textContentIndex (required)
Path Parameter — Index of the position of a TextContent object in the list of TextContent objects making up the document format: int32

Return type
sensitiveSections

Example data
Content-Type: application/json

```
{
  "sensitiveSections" : [ "", "" ]
}
```

Produces
This API call produces the following media types according to the Accept request header; the media type will be conveyed by the Content-Type response header.

- `application/json`

Responses
200
Successfully got the sensitive sections for the TextContent object sensitiveSections
400
The request was invalid JSON or missing some attributes httpStatus
404
The specified resource was not found httpStatus

Models

[Jump to Methods]

Table of Contents

document Up

A document to review

name (optional)

documentId

textSplitGranularity (optional)
String Granularity of the split of the document's content
Enum:
document
paragraph
line
example: paragraph

predictedClassification (optional)
predictedClassification

textContents
array[textContent] list of textContent object representing the content of the document

documentSet Up

schema for a set of Documents

name

setId (optional)

documentCount (optional)
Integer number of documents in set format: int32
example: 42

documentSets Up

an object representing multiple document sets

documentSets
array[documentSet] array of document sets

documents Up

an object representing multiple documents

documents
array[document] array of documents

feature Up

startOffset
Integer start offset of a section in characters from the beginning of the text format: int32
example: 4

endOffset
Integer end offset of a section in characters from the beginning of the text format: int32
example: 20

text (optional)
String textual representation of the section
example: NSA

weight (optional)
Float The contribution of that feature to the classification, if positive weight, this feature contributes to a document's sensitivity and conversely format: float

`httpStatus` Up

This object describes an HTTP status

code
Integer HTTP status code format: int32

message
String HTTP status message

`idBase` Up

Base for the id schema of an object.

`nameBase` Up

The name of the object.

`predictedClassification` Up

The predicted Sensitivity (Sensitive or Not) of a text along with feature explanation for that classification

sensitive
Boolean true if the document is predicted to be sensitive, false otherwise
example: true

sensitivity (optional)
Double Document sensitivity format: double
example: 78.3

classifier (optional)
String Classifier algorithm used for this classification
example: SVM

explanations (optional)
array[predictedClassificationExplanation] list of explanations for a predicted classification

`predictedClassificationExplanation` Up

Explanation for a predicted classification

features
array[feature] List of features with weights explaining a text's classification

explainer
String Explainer used to obtain explanation
example: Lime

`section` Up

Section of characters within a text

startOffset
Integer start offset of a section in characters from the beginning of the text format: int32
example: 4

endOffset
Integer end offset of a section in characters from the beginning of the text format: int32
example: 20

text (optional)
String textual representation of the section
example: NSA

sensitiveSection [Up]

startOffset
Integer start offset of a section in characters from the beginning of the text format: int32
example: 4

endOffset
Integer end offset of a section in characters from the beginning of the text format: int32
example: 20

text (optional)
String textual representation of the section
example: NSA

name
String what was the exemption that was found (mandatory)
example: S27

description (optional)
String This is an optional textual description of what the predicted sensitivity is, suitable for presentation to the user
example: This matches the Royal Family checklist

sensitiveSections [Up]

an object representing multiple sensitive sections

sensitiveSections
array[sensitiveSection] array of sensitive sections

textContent [Up]

A textContent from a document

predictedClassification (optional)
predictedClassification

sensitiveSections (optional)
sensitiveSections

content
String content of the textContent