

Unstructured Text Analysis using HPE Haven OnDemand (HOD)

(A Juvenile Justice App developed as part of AngelHack DC Hackathon)


Pragyansmita Nayak, PhD
@SorishaPragyan

Juvenile Justice Application

```
C:\Users\pnayak\havenondemand-python\examples>python vacode.py
-----
----- INPUT -----
-----
Which STATE of US are you in? --> VA
What are you planning to COMMIT? Think Twice! --> Robbery
-----
-----
Searching for: Robbery in state (US) VA
-----
-----
From collection: ['Code of Virginia']
Title: b'\xc2\xa7 18.2-90. Entering dwelling house, etc., with intent to commit murder, rape, robbery or arson; penal
Refer: http://law.lis.virginia.gov/vacode/title18.2/chapter5/section18.2-90/
Detailed information (read carefully please - Think Twice!): b'Code of Virginia If any person in the nighttime enters
-----
...
...
...
-----
Total number of matching codes found: 8
JAMJustice - Think Twice!
-----
Want to continue searching (yes/no)? -->
```

HOD - Online Search Query

Text Indexes /vacodeindex

**vacodeindex** vacodeindex
No description

MetadataIndex DocumentsQuery IndexAPIs

Sort by: Relevance

Showing 3 results (out of a total of 3 results)

Document Reference: <http://law.lis.virginia.gov/vacode/title18.2/chapter4/section18.2-58.1/>

A. Any person who commits **carjacking**, as herein defined, shall be guilty of a felony punishable by imprisonment for life or a term not less than fifteen years.

Document Reference: <http://law.lis.virginia.gov/vacode/title18.2/chapter4/section18.2-58/>

Carjacking: Penalty • Print page Print • Download PDF PDF • email Creating a Report: Check the sections you'd like to appear in the report, then use the "Create Report" button at the bottom of the page to generate your report.

Document Reference: <http://law.lis.virginia.gov/vacode/title18.2/chapter4/section18.2-53.1/>

in a threatening manner while committing or attempting to commit murder, rape, forcible sodomy, inanimate or animate object sexual penetration as defined in § 18.2-67.2, robbery, **carjacking**, burglary, malicious wounding as defined in § 18.2-51, malicious

<https://dev.havenondemand.com/account/text-indexes.html#q=Carjacking&p=1&index=vacodeindex>


HOD 101

<https://www.havenondemand.com/>


- Cloud-based
- Big Data Platform
- API Orchestration
- Client library support
 - Python, Java, Node.js, Angular.js, Ruby, PHP, R, Go

Categories

Audio-Video Analytics
Connectors
Format Conversion
Geo Analysis
Graph Analysis
Image Analysis
Labs Sandbox
Policy
Prediction
Query Profile and Manipulation
Search
Text Analysis
Unstructured Text Indexing

Unstructured Text Indexing			⬆
	Add to Text Index	Indexes a document.	Try
	Create Text Index	Creates a text index.	Try
	Delete Text Index	Deletes a text index.	Try
	Delete from Text Index	Deletes a document from the index.	Try
	Index Status	Returns index status information.	Try
	List Resources	Lists your dynamic resources.	Try
	Restore Text Index	Restores a text index at a previous state.	Try

HOD – Create Index vacodeindex



Create Text Index

Creates a text index.

[Overview](#) [Request](#) [Response](#) [Try It](#)

API Keys a5582336-3213-4ecf-919e-56f90ea2405b

Parameters

Required		
Name	Type	Value
index	string	<input type="text" value="vacodeindex"/>
flavor	string	<div>explorer</div> <div><div>explorer</div><div>standard</div><div>categorization</div><div>custom_fields</div><div>jumbo</div><div>querymanipulation</div></div>

Optional		
Name	Type	Value
description	string	<input type="text"/>
index_fields	array	<div>Value for index_fields</div> <div>+ Add another value</div>
parametric_fields	array	<div>Value for parametric_fields</div> <div>+ Add another value</div>
expire_time	integer	<input type="text" value="Value for expire_time"/>
expire_date_fields	array	<div>Value for expire_date_fields</div> <div>+ Add another value</div>
numeric_fields	array	<div>Value for numeric_fields</div> <div>+ Add another value</div>
display_name	string	<input type="text" value="Value for display_name"/>

HOD – Create Web Connector

Create web_cloud connector
web_cloud_crime

```
config =  
{ "url" :  
  "http://law.lis.virginia.gov/vacode/title18.2/chapter4/" }  
destination =  
{ "action" : "addtotextindex", "index" : "vacodeindex" }
```

Create Connector
Creates a connector.

Overview Request Response Try It

API Keys a5582336-3213-4ecf-919e-56f90ea2405b

Parameters

Name	Type	Value
flavor	enum	web_cloud
connector	string	web_cloud_crime
config	json	{ "url" : "http://law.lis.virginia.gov/vacode/title18.2/chapter4/" }
destination	json	{ "action" : "addtotextindex", "index" : "vacodeindex" }

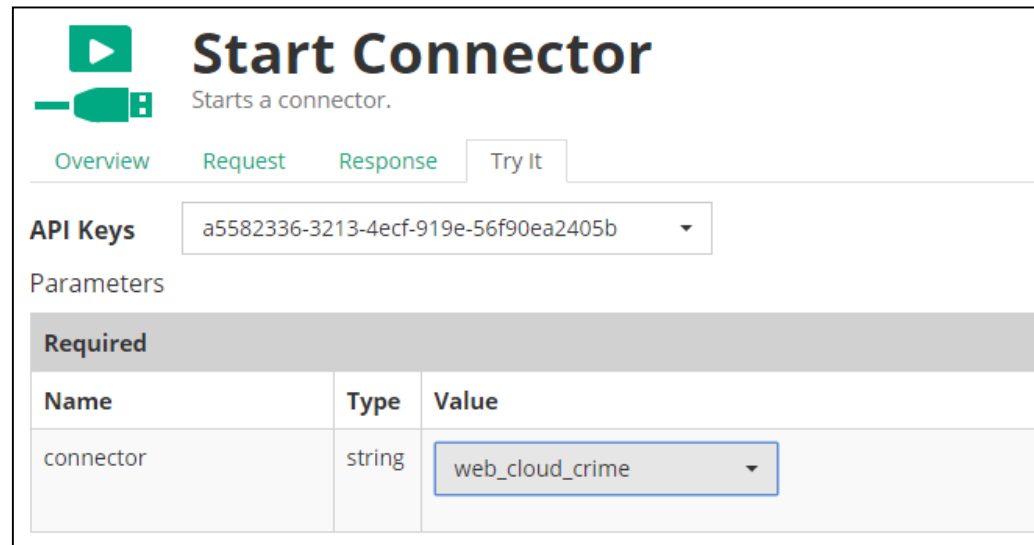
```
curl  
"https://api.havenondemand.com/1/api/sync/createconnector/v1?flavor=web_cloud&connector=web_cloud_c  
rime&config=%7B%22url%22+%3A%22http%3A%2F%2Flaw.lis.virginia.gov%2Fvacode%2Ftitle18.2%2Fchapter4%  
2F%22%7D&destination=%7B%22action%22+%3A%22addtotextindex%22%2C%22index%22+%3A%22vacodeindex%22  
+%7D&credentials=&credentials_policy=&schedule=&apikey=a5582336-3213-4ecf-919e-56f90ea2405b"  
  
curl -X POST --form "flavor=web_cloud" --form "connector=web_cloud_crime" --form  
"config={\"url\":\"http://law.lis.virginia.gov/vacode/title18.2/chapter4/\"}" --form  
"destination={\"action\":\"addtotextindex\",\"index\":\"vacodeindex\"}" --form "apikey=a5582336-3213-4ecf-  
919e-56f90ea2405b" https://api.havenondemand.com/1/api/sync/createconnector/v1
```

HOD – Process the indexing

- Start the connector web_cloud_crime
- Data from the provided URL is indexed in the cloud

- Connector Status:

- QUEUED
- PROCESSING
- DONE



The screenshot shows a web interface titled "Start Connector" with a play button icon. Below the title is a description "Starts a connector." and four tabs: "Overview", "Request", "Response", and "Try It". The "API Keys" section shows a dropdown menu with the value "a5582336-3213-4ecf-919e-56f90ea2405b". The "Parameters" section has a "Required" header and a table with columns "Name", "Type", and "Value". The table contains one row with "connector" as the name, "string" as the type, and a dropdown menu showing "web_cloud_crime" as the value.

Required		
Name	Type	Value
connector	string	web_cloud_crime

Python Client Library-based Search Query Application

```
def formulateRequest(iSearchState, iSearchTerm):
    # curl -X POST --form "text=cariack" --form "absolute_max_results=10" --form "highlight=terms" --form "ignore_operators=false" --form
    "indexes=vacodeindex" --form "print=all" --form "promotion=false" --form "total_results=true" --form
    "apikey=a5582336-3213-4ecf-919e-56f90ea2405b" https://api.havenondemand.com/1/api/sync/querytextindex/v1

    params = dict()
    params["text"] = iSearchTerm
    print("-----")
    print("Searching for:", iSearchTerm, "in state (US)", iSearchState, sep=" ")
    print("-----")

    params["absolute_max_results"] = 10
    params["highlight"] = "terms"
    params["ignore_operators"] = "false"
    params["indexes"] = [iSearchState.lower()+"codeindex"]
    params["print"] = "all"
    params["promotion"] = "false"
    params["total_results"] = "true"

    oResults=postrequests('querytextindex',params)

    for document in oResults["documents"]:
        print("-----")
        print("From collection:", document["collection"])
        print("Title:", document["title"].encode("utf-8"))
        print("Refer:", document["reference"])

        detail = BeautifulSoup(document["content"],"html.parser")

        print("Detailed information (read carefully please - Think Twice!):\n", detail.encode("utf-8"))
    #return oResults
    print("-----")
    print("Total number of matching codes found:", oResults["totalhits"])
    print("JAMJustice - Think Twice!")
    print("-----")
```


References

- GitHub Link of the code used in the presentation : <https://github.com/pragyansmita/AngelHack-May2016>
- Getting Started with Haven OnDemand : https://dev.havenondemand.com/docs/HowTo_GettingStarted.html
- HPE Haven OnDemand APIs : <https://dev.havenondemand.com/apis>
- HPE Haven OnDemand Developer Documentation : <https://dev.havenondemand.com/docs>
- Client library for various programming languages (for example, Go, Python, iOS, Android, PHP, R, Windows Universal 8.1, Ruby et al.) to call Haven OnDemand APIs : <https://github.com/HPE-Haven-OnDemand>
- Client library for Python to call Haven OnDemand APIs : <https://github.com/HPE-Haven-OnDemand/havenondemand-python>
- Synchronous and Asynchronous API : <https://dev.havenondemand.com/docs/AsynchronousAPI.htm>
- Web Cloud Connector : https://dev.havenondemand.com/docs/Connectors_Web.html
- Advanced Haven OnDemand Unstructured Text Indexing : https://dev.havenondemand.com/docs/HowTo_IndexAdvanced.html
- Use Haven OnDemand Search Functionality : https://dev.havenondemand.com/docs/HowTo_Search.html
- Review available text indexes in the Haven OnDemand account and execute search queries : <https://dev.havenondemand.com/account/text-indexes.html>