Index definitions

<https://bbopenhack-astpmcrprlfp362.search.windows.net/indexes/webindex?api-version=2019-05-06-Preview>

api-key: C980265C2D26C822195613D8CD6B0412

Content-Type: application/json

Body:

{

"fields": [

{

"name": "id",

"type": "Edm.String",

"key": true,

"searchable": true,

"filterable": true,

"facetable": false,

"sortable": true

},

{

"name": "url",

"type": "Edm.String",

"sortable": true,

"searchable": true,

"filterable": true,

"facetable": false

},

{

"name": "file\_name",

"type": "Edm.String",

"sortable": true,

"searchable": true,

"filterable": true,

"facetable": false

},

{

"name": "content",

"type": "Edm.String",

"sortable": false,

"searchable": true,

"filterable": true,

"facetable": false

},

{

"name": "size",

"type": "Edm.String",

"sortable": true,

"searchable": true,

"filterable": true,

"facetable": false

},

{

"name": "last\_modified",

"type": "Edm.String",

"sortable": true,

"searchable": true,

"filterable": true,

"facetable": false

},

{

"name": "modifieddate",

"type": "Edm.DateTimeOffset",

"sortable": true,

"searchable": false,

"filterable": true,

"facetable": false

},

{

"name": "languageCode",

"type": "Edm.String",

"searchable": true,

"filterable": true,

"facetable": false

},

{

"name": "keyPhrases",

"type": "Collection(Edm.String)",

"searchable": true,

"filterable": true,

"facetable": false

},

{

"name": "organizations",

"type": "Collection(Edm.String)",

"searchable": true,

"sortable": false,

"filterable": false,

"facetable": false

},

{

"name": "score",

"type": "Edm.Double",

"searchable": false,

"filterable": true,

"facetable": false

},

{

"name": "persons",

"type": "Collection(Edm.String)",

"searchable": true,

"sortable": false,

"filterable": true,

"facetable": true

},

{

"name": "locations",

"type": "Collection(Edm.String)",

"searchable": true,

"sortable": false,

"filterable": true,

"facetable": true

},

{

"name": "enriched",

"type": "Edm.String",

"searchable": false,

"filterable": false,

"facetable": false,

"sortable": false

},

{

"name": "merged\_text",

"type": "Edm.String",

"searchable": false,

"filterable": false,

"facetable": false,

"sortable": false

},

{

"name": "name",

"type": "Edm.String",

"sortable": true,

"searchable": true,

"filterable": true,

"facetable": false

},

{

"name": "destinationCity",

"type": "Edm.String",

"sortable": true,

"searchable": true,

"filterable": true,

"facetable": false

},

{

"name": "departureDate",

"type": "Edm.String",

"sortable": true,

"searchable": true,

"filterable": true,

"facetable": false

},

{

"name": "returnDate",

"type": "Edm.String",

"sortable": true,

"searchable": true,

"filterable": true,

"facetable": false

},

{

"name": "dateofBirth",

"type": "Edm.String",

"sortable": true,

"searchable": true,

"filterable": true,

"facetable": false

},

{

"name": "categories",

"type": "Collection(Edm.ComplexType)",

"fields": [

{

"name": "name",

"type": "Edm.String",

"searchable": true,

"filterable": false,

"facetable": false

},

{

"name": "score",

"type": "Edm.Double",

"searchable": false,

"filterable": false,

"facetable": false

},

{

"name": "detail",

"type": "Edm.ComplexType",

"fields": [

{

"name": "celebrities",

"type": "Collection(Edm.ComplexType)",

"fields": [

{

"name": "name",

"type": "Edm.String",

"searchable": true,

"filterable": false,

"facetable": false

},

{

"name": "faceBoundingBox",

"type": "Collection(Edm.ComplexType)",

"fields": [

{

"name": "x",

"type": "Edm.Int32",

"searchable": false,

"filterable": false,

"facetable": false

},

{

"name": "y",

"type": "Edm.Int32",

"searchable": false,

"filterable": false,

"facetable": false

}

]

},

{

"name": "confidence",

"type": "Edm.Double",

"searchable": false,

"filterable": false,

"facetable": false

}

]

},

{

"name": "landmarks",

"type": "Collection(Edm.ComplexType)",

"fields": [

{

"name": "name",

"type": "Edm.String",

"searchable": true,

"filterable": false,

"facetable": false

},

{

"name": "confidence",

"type": "Edm.Double",

"searchable": false,

"filterable": false,

"facetable": false

}

]

}

]

}

]

},

{

"name": "description",

"type": "Collection(Edm.ComplexType)",

"fields": [

{

"name": "tags",

"type": "Collection(Edm.String)",

"searchable": true,

"filterable": false,

"facetable": false

},

{

"name": "captions",

"type": "Collection(Edm.ComplexType)",

"fields": [

{

"name": "text",

"type": "Edm.String",

"searchable": true,

"filterable": false,

"facetable": false

},

{

"name": "confidence",

"type": "Edm.Double",

"searchable": false,

"filterable": false,

"facetable": false

}

]

}

]

},

{

"name": "faces",

"type": "Collection(Edm.ComplexType)",

"fields": [

{

"name": "age",

"type": "Edm.Int32",

"searchable": false,

"filterable": false,

"facetable": false

},

{

"name": "gender",

"type": "Edm.String",

"searchable": false,

"filterable": false,

"facetable": false

},

{

"name": "faceBoundingBox",

"type": "Collection(Edm.ComplexType)",

"fields": [

{

"name": "x",

"type": "Edm.Int32",

"searchable": false,

"filterable": false,

"facetable": false

},

{

"name": "y",

"type": "Edm.Int32",

"searchable": false,

"filterable": false,

"facetable": false

}

]

}

]

},

{

"name": "tags",

"type": "Collection(Edm.ComplexType)",

"fields": [

{

"name": "name",

"type": "Edm.String",

"searchable": true,

"filterable": false,

"facetable": false

},

{

"name": "confidence",

"type": "Edm.Double",

"searchable": false,

"filterable": false,

"facetable": false

}

]

},

{

"name": "topWords",

"type": "Collection(Edm.String)",

"searchable": true,

"sortable": false,

"filterable": true,

"facetable": true

}

],

"suggesters": [

{

"name": "sg",

"searchMode": "analyzingInfixMatching",

"sourceFields": ["keyPhrases", "locations"]

}

],

"scoringProfiles": [

{

"name":"geo",

"text": {

"weights": {

"keyPhrases": 1,

"locations": 4,

"size": 2,

"last\_modified": 3

}

}

},

{

"name": "newAndHighlyMod",

"functions": [

{

"type": "freshness",

"fieldName": "modifieddate",

"boost": 10,

"interpolation": "quadratic",

"freshness": {

"boostingDuration": "P365D"

}

}

]

}

]

}

Skill index:

<https://bbopenhack-astpmcrprlfp362.search.windows.net/skillsets/webindexskillset?api-version=2019-05-06-Preview>

api-key: C980265C2D26C822195613D8CD6B0412

Content-Type: application/json

Body:

{

"description":

"Extract entities, detect language and extract key-phrases",

"skills":

[

{

"@odata.type": "#Microsoft.Skills.Text.EntityRecognitionSkill",

"context": "/document",

"categories": [ "Organization", "Person", "Location" ],

"defaultLanguageCode": "en",

"inputs": [

{

"name": "text",

"source": "/document/content"

}

],

"outputs": [

{

"name": "organizations",

"targetName": "organizations"

},

{ "name": "locations", "targetName": "locations" },

{ "name": "persons", "targetName": "persons" }

]

},

{

"@odata.type": "#Microsoft.Skills.Text.SentimentSkill",

"inputs": [

{

"name": "text",

"source": "/document/content"

}

],

"outputs": [

{

"name": "score",

"targetName": "score"

}

]

},

{

"@odata.type": "#Microsoft.Skills.Text.KeyPhraseExtractionSkill",

"inputs": [

{

"name": "text",

"source": "/document/content"

}

],

"outputs": [

{

"name": "keyPhrases",

"targetName": "keyPhrases"

}

]

},

{

"description": "Extract text (plain and structured) from image.",

"@odata.type": "#Microsoft.Skills.Vision.OcrSkill",

"context": "/document/normalized\_images/\*",

"defaultLanguageCode": "en",

"detectOrientation": true,

"inputs": [

{

"name": "image",

"source": "/document/normalized\_images/\*"

}

],

"outputs": [

{

"name": "text"

}

]

},

{

"@odata.type": "#Microsoft.Skills.Text.MergeSkill",

"description": "Create merged\_text, which includes all the textual representation of each image inserted at the right location in the content field.",

"context": "/document",

"insertPreTag": " ",

"insertPostTag": " ",

"inputs": [

{

"name":"text", "source": "/document/content"

},

{

"name": "itemsToInsert", "source": "/document/normalized\_images/\*/text"

},

{

"name":"offsets", "source": "/document/normalized\_images/\*/contentOffset"

}

],

"outputs": [

{

"name": "mergedText", "targetName" : "merged\_text"

}

]

},

{

"description": "Extract image analysis.",

"@odata.type": "#Microsoft.Skills.Vision.ImageAnalysisSkill",

"context": "/document/normalized\_images/\*",

"defaultLanguageCode": "en",

"visualFeatures": [

"Tags",

"Categories",

"Description",

"Faces"

],

"inputs": [

{

"name": "image",

"source": "/document/normalized\_images/\*"

}

],

"outputs": [

{

"name": "categories"

},

{

"name": "tags"

},

{

"name": "description"

},

{

"name": "faces"

}

]

},

{

"@odata.type": "#Microsoft.Skills.Custom.WebApiSkill",

"description": "word count list",

"uri": "https://bbcustomfunc.azurewebsites.net/api/Function1?code=FDRmLY0RyuxU1t6IuL6h12KJsV36tx6TCk79f3aMjXUy0oGH8nLQ8A==",

"httpHeaders": {

"Ocp-Apim-Subscription-Key": "FDRmLY0RyuxU1t6IuL6h12KJsV36tx6TCk79f3aMjXUy0oGH8nLQ8A=="

},

"context": "/document",

"inputs": [

{

"name": "mergedText",

"source": "/document/merged\_text"

}

],

"outputs": [

{

"name": "topWords",

"targetName": "topWords"

}

]

},

{

"@odata.type": "#Microsoft.Skills.Util.ShaperSkill",

"name": "Projection Shaper",

"description": "Shape projections",

"context": "/document",

"inputs": [

{

"name": "file\_id",

"source": "/document/id"

},

{

"name": "file\_name",

"source": "/document/file\_name"

},

{

"name": "url",

"source": "/document/url"

},

{

"name": "top\_words",

"source": "/document/topWords/\*"

},

{

"name": "sentiment\_score",

"source": "/document/score"

},

{

"name": "locations",

"source": "/document/locations/\*"

}

],

"outputs": [

{

"name": "output",

"targetName": "projection"

}

]

}

],

"knowledgeStore": {

"storageConnectionString": "DefaultEndpointsProtocol=https;AccountName=bbopenhackstore;AccountKey=CpE7U+UfKepApBdUyVxHLb4Ijdz3xg4uZpOnqv73T+khI3VgQbJWtttIQufuHn6+ULH2QSb3xGhsGWqpvQfUsA==;EndpointSuffix=core.windows.net",

"projections": [

{

"tables": [],

"objects": [

{

"storageContainer": "kmstore1",

"source": "/document/projection"

}

]

},

{

"tables": [

{

"tableName": "WebDocs",

"generatedKeyName": "document\_id",

"source": "/document/projection"

},

{

"tableName": "TopWords",

"generatedKeyName": "topword\_id",

"sourceContext": "/document/projection/top\_words/\*",

"inputs": [

{

"name": "word",

"source": "/document/projection/top\_words/\*"

}

]

},

{

"tableName": "SentimentScores",

"generatedKeyName": "sentiment\_id",

"sourceContext": "/document/projection/sentiment\_score",

"inputs": [

{

"name": "sentiment",

"source": "/document/projection/sentiment\_score"

}

]

},

{

"tableName": "Locations",

"generatedKeyName": "location\_id",

"sourceContext": "/document/projection/locations/\*",

"inputs": [

{

"name": "location",

"source": "/document/projection/locations/\*"

}

]

}

],

"objects": []

}

]

},

"cognitiveServices": {

"@odata.type": "#Microsoft.Azure.Search.CognitiveServicesByKey",

"description": "bbopenhackcog",

"key": "4e7edb3052ec438b8043569df4f722fc"

}

}

---

Indexer:

<https://bbopenhack-astpmcrprlfp362.search.windows.net/indexers/webindexer?api-version=2019-05-06-Preview>

api-key: C980265C2D26C822195613D8CD6B0412

Content-Type: application/json

Body:

{

"name":"webindexer",

"dataSourceName" : "webdata",

"targetIndexName" : "webindex",

"skillsetName" : "webindexskillset",

"fieldMappings" : [

{

"sourceFieldName" : "metadata\_storage\_path",

"targetFieldName" : "id",

"mappingFunction" :

{ "name" : "base64Encode" }

},

{

"sourceFieldName" : "metadata\_storage\_path",

"targetFieldName" : "url"

},

{

"sourceFieldName" : "metadata\_storage\_name",

"targetFieldName" : "file\_name"

},

{

"sourceFieldName" : "content",

"targetFieldName" : "content"

},

{

"sourceFieldName" : "metadata\_storage\_size",

"targetFieldName" : "size"

},

{

"sourceFieldName" : "metadata\_storage\_last\_modified",

"targetFieldName" : "modifieddate"

},

{

"sourceFieldName" : "metadata\_storage\_last\_modified",

"targetFieldName" : "last\_modified"

},

{

"sourceFieldName" : "metadata\_storage\_path",

"targetFieldName" : "metadata\_storage\_path"

}

],

"outputFieldMappings" :

[

{

"sourceFieldName" : "/document/organizations",

"targetFieldName" : "organizations"

},

{

"sourceFieldName" : "/document/keyPhrases",

"targetFieldName" : "keyPhrases"

},

{

"sourceFieldName": "/document/score",

"targetFieldName": "score"

},

{

"sourceFieldName" : "/document/persons",

"targetFieldName" : "persons"

},

{

"sourceFieldName" : "/document/locations",

"targetFieldName" : "locations"

},

{

"sourceFieldName": "/document/normalized\_images/\*/categories/\*",

"targetFieldName": "categories"

},

{

"sourceFieldName": "/document/normalized\_images/\*/tags/\*",

"targetFieldName": "tags"

},

{

"sourceFieldName": "/document/normalized\_images/\*/description",

"targetFieldName": "description"

},

{

"sourceFieldName": "/document/normalized\_images/\*/faces/\*",

"targetFieldName": "faces"

},

{

"sourceFieldName" : "/document/merged\_text",

"targetFieldName" : "merged\_text"

},

{

"sourceFieldName" : "/document/topWords",

"targetFieldName" : "topWords"

}

],

"parameters":

{

"maxFailedItems":-1,

"maxFailedItemsPerBatch":-1,

"configuration":

{

"dataToExtract": "contentAndMetadata",

"imageAction":"generateNormalizedImages"

}

}

}

<https://bbopenhack-astpmcrprlfp362.search.windows.net/indexes/insuranceindex?api-version=2019-05-06-Preview>

api-key: C980265C2D26C822195613D8CD6B0412

Content-Type: application/json

Body:

{

"fields": [

{

"name": "id",

"type": "Edm.String",

"key": true,

"searchable": true,

"filterable": true,

"facetable": false,

"sortable": true

},

{

"name": "url",

"type": "Edm.String",

"sortable": true,

"searchable": true,

"filterable": true,

"facetable": false

},

{

"name": "file\_name",

"type": "Edm.String",

"sortable": true,

"searchable": true,

"filterable": true,

"facetable": false

},

{

"name": "content",

"type": "Edm.String",

"sortable": false,

"searchable": true,

"filterable": true,

"facetable": false

},

{

"name": "size",

"type": "Edm.String",

"sortable": true,

"searchable": true,

"filterable": true,

"facetable": false

},

{

"name": "last\_modified",

"type": "Edm.String",

"sortable": true,

"searchable": true,

"filterable": true,

"facetable": false

},

{

"name": "modifieddate",

"type": "Edm.DateTimeOffset",

"sortable": true,

"searchable": false,

"filterable": true,

"facetable": false

},

{

"name": "languageCode",

"type": "Edm.String",

"searchable": true,

"filterable": true,

"facetable": false

},

{

"name": "keyPhrases",

"type": "Collection(Edm.String)",

"searchable": true,

"filterable": true,

"facetable": false

},

{

"name": "organizations",

"type": "Collection(Edm.String)",

"searchable": true,

"sortable": false,

"filterable": false,

"facetable": false

},

{

"name": "score",

"type": "Edm.Double",

"searchable": false,

"filterable": true,

"facetable": false

},

{

"name": "persons",

"type": "Collection(Edm.String)",

"searchable": true,

"sortable": false,

"filterable": true,

"facetable": true

},

{

"name": "locations",

"type": "Collection(Edm.String)",

"searchable": true,

"sortable": false,

"filterable": true,

"facetable": true

},

{

"name": "enriched",

"type": "Edm.String",

"searchable": false,

"filterable": false,

"facetable": false,

"sortable": false

},

{

"name": "merged\_text",

"type": "Edm.String",

"searchable": false,

"filterable": false,

"facetable": false,

"sortable": false

},

{

"name": "name",

"type": "Edm.String",

"sortable": true,

"searchable": true,

"filterable": true,

"facetable": false

},

{

"name": "destinationCity",

"type": "Edm.String",

"sortable": true,

"searchable": true,

"filterable": true,

"facetable": false

},

{

"name": "departureDate",

"type": "Edm.String",

"sortable": true,

"searchable": true,

"filterable": true,

"facetable": false

},

{

"name": "returnDate",

"type": "Edm.String",

"sortable": true,

"searchable": true,

"filterable": true,

"facetable": false

},

{

"name": "dateofBirth",

"type": "Edm.String",

"sortable": true,

"searchable": true,

"filterable": true,

"facetable": false

}

]

}

Skillset:

<https://bbopenhack-astpmcrprlfp362.search.windows.net/skillsets/insuranceindexskillset?api-version=2019-05-06-Preview>

api-key: C980265C2D26C822195613D8CD6B0412

Content-Type: application/json

Body:

{

"description":

"Extract entities, detect language and extract key-phrases",

"skills":

[

{

"@odata.type": "#Microsoft.Skills.Text.EntityRecognitionSkill",

"context": "/document",

"categories": [ "Organization", "Person", "Location" ],

"defaultLanguageCode": "en",

"inputs": [

{

"name": "text",

"source": "/document/content"

}

],

"outputs": [

{

"name": "organizations",

"targetName": "organizations"

},

{ "name": "locations", "targetName": "locations" },

{ "name": "persons", "targetName": "persons" }

]

},

{

"@odata.type": "#Microsoft.Skills.Text.SentimentSkill",

"inputs": [

{

"name": "text",

"source": "/document/content"

}

],

"outputs": [

{

"name": "score",

"targetName": "score"

}

]

},

{

"@odata.type": "#Microsoft.Skills.Text.KeyPhraseExtractionSkill",

"inputs": [

{

"name": "text",

"source": "/document/content"

}

],

"outputs": [

{

"name": "keyPhrases",

"targetName": "keyPhrases"

}

]

},

{

"description": "Extract text (plain and structured) from image.",

"@odata.type": "#Microsoft.Skills.Vision.OcrSkill",

"context": "/document/normalized\_images/\*",

"defaultLanguageCode": "en",

"detectOrientation": true,

"inputs": [

{

"name": "image",

"source": "/document/normalized\_images/\*"

}

],

"outputs": [

{

"name": "text"

}

]

},

{

"@odata.type": "#Microsoft.Skills.Text.MergeSkill",

"description": "Create merged\_text, which includes all the textual representation of each image inserted at the right location in the content field.",

"context": "/document",

"insertPreTag": " ",

"insertPostTag": " ",

"inputs": [

{

"name":"text", "source": "/document/content"

},

{

"name": "itemsToInsert", "source": "/document/normalized\_images/\*/text"

},

{

"name":"offsets", "source": "/document/normalized\_images/\*/contentOffset"

}

],

"outputs": [

{

"name": "mergedText", "targetName" : "merged\_text"

}

]

},

{

"description": "Extract image analysis.",

"@odata.type": "#Microsoft.Skills.Vision.ImageAnalysisSkill",

"context": "/document/normalized\_images/\*",

"defaultLanguageCode": "en",

"visualFeatures": [

"Tags",

"Categories",

"Description",

"Faces"

],

"inputs": [

{

"name": "image",

"source": "/document/normalized\_images/\*"

}

],

"outputs": [

{

"name": "categories"

},

{

"name": "tags"

},

{

"name": "description"

},

{

"name": "faces"

}

]

},

{

"@odata.type": "#Microsoft.Skills.Custom.WebApiSkill",

"name": "formrecognizer",

"description": "Extracts fields from a form using a pre-trained form recognition model",

"uri": "https://bbanalyzeform.azurewebsites.net/api/Function1?code=KQrpcv/TNmGbJwz3FgqvsdUqdwEUDtceuGnasW21Eq0VtB8gxBxv/A==",

"context": "/document",

"batchSize": 1,

"inputs": [

{

"name": "formUrl",

"source": "/document/metadata\_storage\_path"

},

{

"name": "formSasToken",

"source": "/document/metadata\_storage\_sas\_token"

}

],

"outputs": [

{

"name": "name",

"targetName": "name"

},

{

"name": "destinationCity",

"targetName": "destinationCity"

},

{

"name": "departureDate",

"targetName": "departureDate"

},

{

"name": "returnDate",

"targetName": "returnDate"

},

{

"name": "dateofBirth",

"targetName": "dateofBirth"

}

]

}

],

"cognitiveServices": {

"@odata.type": "#Microsoft.Azure.Search.CognitiveServicesByKey",

"description": "bbopenhackcog",

"key": "4e7edb3052ec438b8043569df4f722fc"

}

}

Indexer:

<https://bbopenhack-astpmcrprlfp362.search.windows.net/indexers/insuranceindexer?api-version=2019-05-06-Preview>

api-key: C980265C2D26C822195613D8CD6B0412

Content-Type: application/json

{

"name":"insuranceindexer",

"dataSourceName" : "insurancedata",

"targetIndexName" : "insuranceindex",

"skillsetName" : "insuranceindexskillset",

"fieldMappings" : [

{

"sourceFieldName" : "metadata\_storage\_path",

"targetFieldName" : "id",

"mappingFunction" :

{ "name" : "base64Encode" }

},

{

"sourceFieldName" : "metadata\_storage\_path",

"targetFieldName" : "url"

},

{

"sourceFieldName" : "metadata\_storage\_name",

"targetFieldName" : "file\_name"

},

{

"sourceFieldName" : "content",

"targetFieldName" : "content"

},

{

"sourceFieldName" : "metadata\_storage\_size",

"targetFieldName" : "size"

},

{

"sourceFieldName" : "metadata\_storage\_last\_modified",

"targetFieldName" : "modifieddate"

},

{

"sourceFieldName" : "metadata\_storage\_last\_modified",

"targetFieldName" : "last\_modified"

},

{

"sourceFieldName" : "metadata\_storage\_path",

"targetFieldName" : "metadata\_storage\_path"

}

],

"outputFieldMappings" :

[

{

"sourceFieldName" : "/document/organizations",

"targetFieldName" : "organizations"

},

{

"sourceFieldName" : "/document/keyPhrases",

"targetFieldName" : "keyPhrases"

},

{

"sourceFieldName": "/document/score",

"targetFieldName": "score"

},

{

"sourceFieldName" : "/document/persons",

"targetFieldName" : "persons"

},

{

"sourceFieldName" : "/document/locations",

"targetFieldName" : "locations"

},

{

"sourceFieldName" : "/document/merged\_text",

"targetFieldName" : "merged\_text"

},

{

"sourceFieldName" : "/document/name",

"targetFieldName" : "name"

},

{

"sourceFieldName" : "/document/destinationCity",

"targetFieldName" : "destinationCity"

},

{

"sourceFieldName" : "/document/departureDate",

"targetFieldName" : "departureDate"

},

{

"sourceFieldName" : "/document/returnDate",

"targetFieldName" : "returnDate"

},

{

"sourceFieldName" : "/document/dateofBirth",

"targetFieldName" : "dateofBirth"

}

],

"parameters":

{

"maxFailedItems":-1,

"maxFailedItemsPerBatch":-1,

"configuration":

{

"dataToExtract": "contentAndMetadata",

"imageAction":"generateNormalizedImages"

}

}

}