Filtering and Scoring Search Results

Chad Campbell
INDEPENDENT SOFTWARE ENGINEER
@chadcampbell www.ecofic.com

Overview

Filtering search results

Measuring document relevancy

Filtering Search Results

OData Expression Syntax for Azure Search

Provides operators for filtering

Geospatial functions

Comparison

Logical

“Any” and “All”

OData Operators

Comparison Operators

Equal

Not equal

Greater than

Less than

Greater than or equal to

Less than or equal to

Comparison Operators

OData	C# / Java / JavaScript
ne	!=
gt	>
ge	>=
lt	<
le	<=

Logical Operators

And

Or

Not

Logical Operators

OData	C# / Java / JavaScript
and	&&
or	||
not	!

“Any” and “All” Operators

Lambda operators for collection filtering

Used on `Collection(Edm.String)` **fields**

Any and All Scenarios

Find all beers that have a chocolatey flavor

```
"filter":"flavors/any(f: f eq ‘chocolate')"
```

Find all beers that don’t have a smokey flavor

```
"filter":"flavors/all(f: f ne ‘smokey')"
```

“Any” and “All” Operators

“Any” returns true when *any* string in the collection is matched

“All” returns true when *every* string in the collection is matched

Searches by Distance

geo.distance function

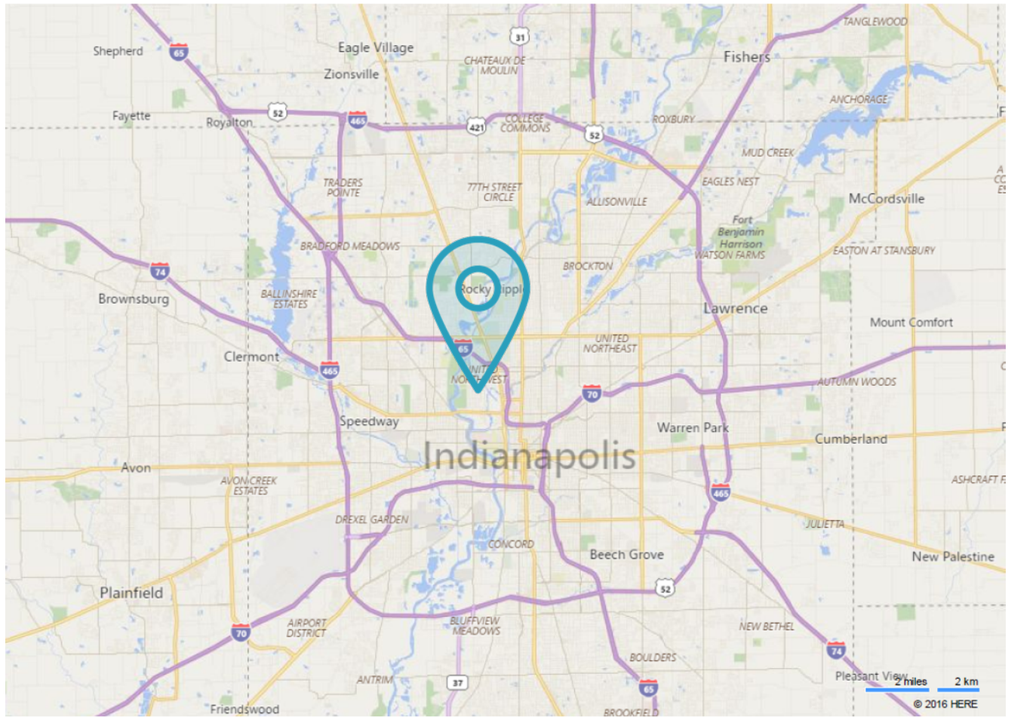
returns distance between two points

distance in kilometers

Getting the Distance

Pass each point as an

`Edm.GeographyPoint`



Microsoft product screen shot(s) reprinted with permission from Microsoft Corporation.

geo.distance Function

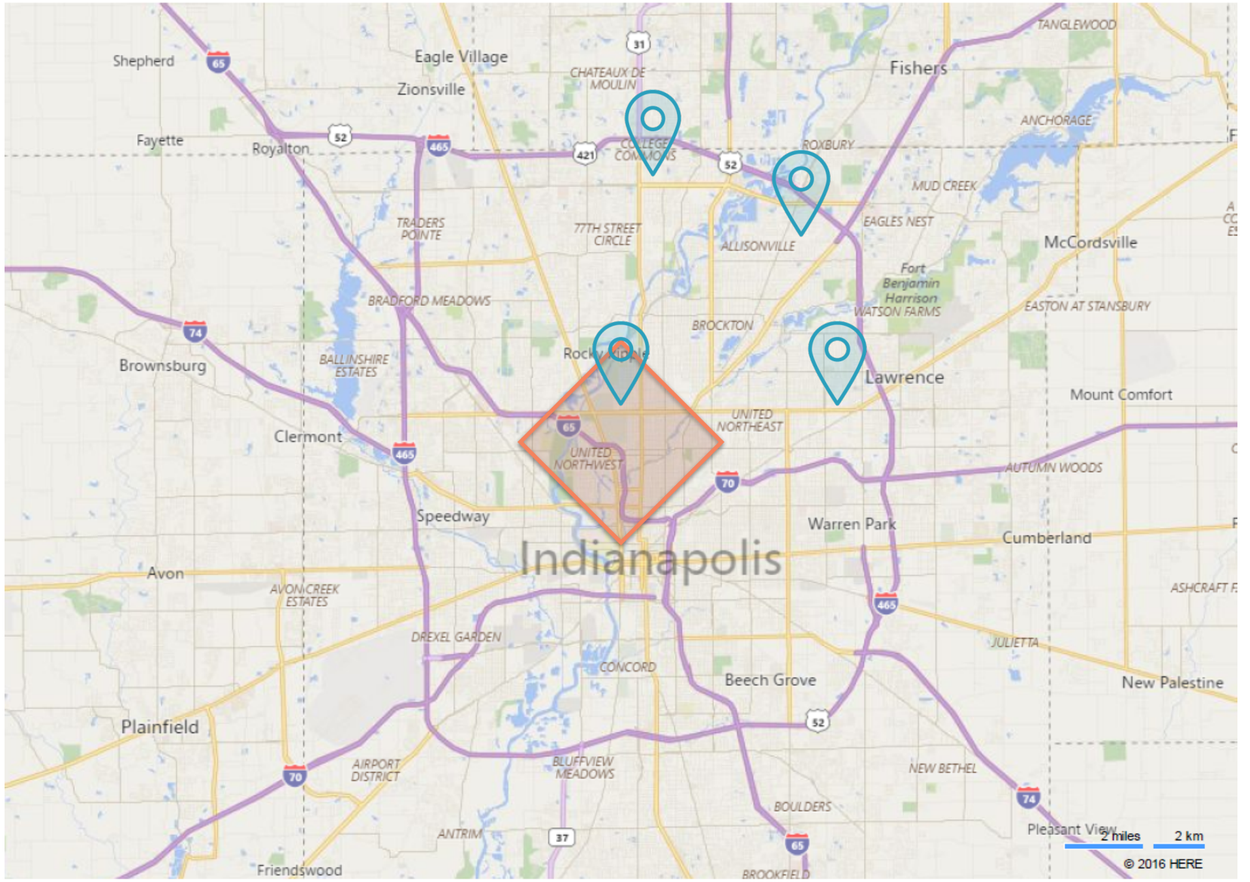
Cannot use “eq” or “ne” operators

Only “gt”, “lt”, “ge”, and “le”

geo.intersects Function

Determines if a point is in the bounds of a polygon

Defining the Geofence



Microsoft product screen shot(s) reprinted with permission from Microsoft Corporation.

Measuring Document Relevancy

@search.score

Document relevancy against a query

Higher the score, the more relevant the document

More relevancy, means higher ranking

TF-IDF

Term frequency-inverse document frequency

TF-IDF

Looks at how often a term appears

Looks at how often it appears in other docs

The more frequent a term appears, the more relevant the document

Scoring profiles let you influence how results are ranked

Scoring Profiles

Let you promote one or more fields

“Boost”

Connect search results with business goals

Creating a Scoring Profile

Part of a search index's definition

Get the current index definition

Update the index definition

Search Results

W/Out Scoring Profile

Rank	Name	Brewery
1	North Adjule Lager	Southern Hemisphere Brewco
2	Hyote Chocolate Stout	North American Brewco
3	Igopogo Pilsner	North American Brewco
4	Jackalobe Lager	North American Brewco
5	Pope Lick Porter	North American Brewco
6	Chocolate Pukwudgie Stout	North American Brewco
7	Sharlie Pilsner	North American Brewco
8	Snallygaster Pale Ale	North American Brewco

With Scoring Profile

Rank	Name	Brewery
1	Hyote Chocolate Stout	North American Brewco
2	Igopogo Pilsner	North American Brewco
3	Jackalobe Lager	North American Brewco
4	Pope Lick Porter	North American Brewco
5	Chocolate Pukwudgie Stout	North American Brewco
6	Sharlie Pilsner	North American Brewco
7	Snallygaster Pale Ale	North American Brewco
8	North Adjule Lager	Southern Hemisphere Brewco

Customizing a Scoring Profile

Text

Numeric values

Data freshness

Proximity

Collections of strings

```
"scoringProfiles": [
  {
   "name":"breweryName",
   "text": {
      "weights": {
        "breweryName":5
      }
    }
  }
]
```

- Name of the scoring profile

- Specifies which fields to boost
- Boost this field by the magnitude

```
"scoringProfiles": [
  {
   "name":"beerName",
   "text": {
     "weights": {
       "name":5,
       "breweryName":2.5
     }
   }
  }
]
```

◀2.5x more important

Weighting Numbers

Edm.Int32

Edm.Int64

Edm.Double

`magnitude` Function

Boost numeric values

Based on how high or how low

Breaks ties between values

Search for “Ale”

Rank	Score	Name	ABV
1	0.28196415	Ahool Ale	5.4
2	0.28196415	Agogwe Ale	3.9
3	0.28196415	Aswang Ale	4.2
4	0.22557132	Megalodon Pale Ale	5.7
5	0.22557132	Pale Popobawa Ale	4.4
6	0.22557132	Snallygaster Pale Ale	9.7

```
"scoringProfiles": [
 {
  "name":"abv",
  "functions": [
   {
    "type":"magnitude",
    "fieldName":"abv",
    "boost": 2.5,
    "magnitude": {
     "boostingRangeStart":0,
     "boostingRangeEnd":100
    }
   }
  ]
 }
]
```

abv Scoring Profile

Search for “Ale”

Rank	Score	Name	ABV
1	0.28196415	Ahool Ale	5.4
2	0.28196415	Agogwe Ale	3.9
3	0.28196415	Aswang Ale	4.2
4	0.22557132	Megalodon Pale Ale	5.7
5	0.22557132	Pale Popobawa Ale	4.4
6	0.22557132	Snallygaster Pale Ale	9.7

Rank	Score	Name	ABV
1	0.30480325	Ahool Ale	5.4
2	0.2997279	Aswang Ale	4.2
3	0.29845905	Agogwe Ale	3.9
4	0.25839195	Snallygaster Pale Ale	9.7
5	0.24485767	Megalodon Pale Ale	5.7
6	0.24045902	Pale Popobawa Ale	4.4

```
"scoringProfiles": [
 {
  "name":"abv",
  "functions": [
   {
    "type":"magnitude",
    "fieldName":"abv",
    "boost": 2.5,
    "magnitude": {
     "boostingRangeStart":0,
     "boostingRangeEnd":100
    }
   }
  ]
 }
]
```

abv Scoring Profile

`boostingRange`

Edm.Int32

Edm.Int64

Edm.Double

```
"magnitude": {
  "boostingRangeStart":0,
  "boostingRangeEnd":100
}
```

Magnitude Function

Higher values are preferred

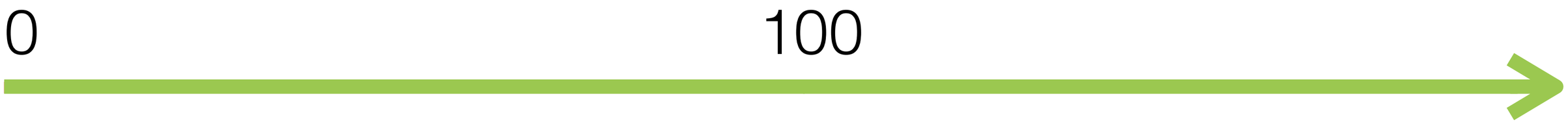
```
"magnitude": {
  "boostingRangeStart":100,
  "boostingRangeEnd":0
}
```

Magnitude Function

Lower values are preferred

constantBoostBeyondRange

Whether to limit boosting to the range defined by the `boostingRangeStart` and `boostingRangeEnd` values


0
100
boostingRangeStart
boostingRangeEnd

constantBoostBeyondRange

false (default)	true
Range	Vector

freshness Function

The most recent data is the most relevant data

```
{
 “name":"freshBeers",
 "functions": [
  {
   "type": "freshness",
   “fieldName": "lastTappedOn",
   "boost": 10,
   "freshness": {
    "boostingDuration": "P90D"
   }
  }
 ]
}
```

- Defines the time span in which values are boosted

`+boostingDuration (P90D)`

query execution time

`boostingDuration` Property

`boostingDuration` Format

Span	Example
1 Day	P1D
12 Hours	PT12H
30 Minutes	PT30M
10.123 Seconds	PT10.123S

Date Boosting

Boost past dates
P90D

Boost future dates
-P90D

Boost dates at the edges
boost: 0.5

`distance` Function

Boost data based on proximity

Works with `Edm.GeographyPoint` data

`tag`
Function

Boosts tag data

Works with `Collection(Edm.String)` **data**

```
{
  "name":"preferredFlavors",
  "functions": [
    {
      "type": "tag",
      "fieldName": "flavors",
      "boost": 10,
      "tag": {
        “tagsParameter":
          "myBelovedFlavors"
      }
    }
  ]
}
```

Scoring Function by Type

	Weighted Text	magnitude	freshness	distance	tag
Collection(Edm.String)					✓
Edm.Boolean					
Edm.DateTimeOffset			✓		
Edm.Double		✓			
Edm.GeographyPoint				✓	
Edm.Int32		✓			
Edm.Int64		✓			
Edm.String	✓				✓

Scoring Profiles

Can use multiple functions at once

Identify what is most relevant to a search

Help you connect results with business goals

Choosing a Default Scoring Profile

defaultScoringProfile

Sets the default search behavior when querying an index

Summary

Scoring profiles

Filter search results