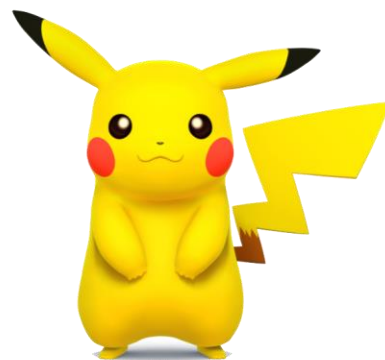


# Pokemon Go

MongoDB – projekat

Milica Todorović RA17-2015  
Fakultet tehničkih nauka, Novi Sad  
Predmet: Sistemi baza podataka  
Datum: 17.6.2019.



## Sadržaj

Opis postupka

Finalna šema

Analiza agregacija

# Opis postupka

## Struktiranje podataka

- Iskorištena je Python skripta da bi se promenila struktura podataka
- Atributi poput Type\_1 i Type\_2, te cooc\_1-cooc\_151 pretvoreni u nizove objekata
- Pojedini atributi su grupisani
- Od atributa longitude i latitude kreiran je GeoJSON objekat
- Kod susreta je id pokemona zamenjen proširenom referencom.
- Susreti su baktirani prema kontinentu, gradu i datumu. Ukoliko za neki grad ima više od 150 susreta taj dan, oni su podeljeni na više dokumenata.

## Unos podataka u bazu

- Iskorištena je Python biblioteka PyMongo

# Konačna šema - kolekcija pokemon

Number	Name	Type_1	Type_2	Total	HP	Attack	Defense	Sp_Atk	Sp_Def	Speed	Generat	isLegen	Color	hasGen	Pr_Mal	Egg_Gr	Egg_Gr	hasMeg	Height	Weight	Catch_F	Body_S
16	Pidgey	Normal	Flying	251	40	45	40	35	35	56	1	False	Brown	True		0,5	Flying		False	0,3	1,8	255 two_wings

▼ (1) { \_id : 5d04d2987cdf7ccaab495020 }

Id

Id

Name

▼ Type

0

1

> Stats

Generation

Legendary

> Gender

> EggGroup

HasMegaEvolution

> Body

CatchRate

{ 12 fields }

5d04d2987cdf7ccaab495020

16

Pidgey

[ 2 elements ]

Normal

Flying

{ 7 fields }

1

true

{ 2 fields }

[ 1 elements ]

true

{ 4 fields }

255

Document

ObjectId

Int32

String

Array

String

String

Object

Int32

Bool

Object

Array

Bool

Object

Int32

# Konačna šema - kolekcija encounter

pokemon	latitude	longitude	terrainType	closeToWater	city	continent	weather	temperature	windSpeed	windBearing	pressure	weatherIcon	urban	rural	gymDistanceKm	gymIn100m	gymIn5000m	pokestopDistanceKm	pokestopIn100m	pokestopIn250m	pokestopIn5000m	cooc_1	cooc_151
16	20,525745	-97,460829	14	false	Mexico_City	America	Foggy	25,5	4,79	269	1018,02	fog	true	false	0,049869	true	true	0,081776	true	true	true	true	false
16	38,90359	-77,19978	13	false	New_York	America	Clear	24,2	4,29	218	1015,29	clear-night	false	false	0,489886	false	true	0,338602	false	false	true	false	false
16	26,235552	-98,197249	13	false	Chicago	America	Clear	28	11,26	142	1016,69	clear-night	true	false	0,782757	false	true	0,921315	false	false	true	false	false

▼ (1) (1) {id : 5d04d2997cdf7ccaab4952e3}

id

▼ (1) CommonInfo

Continent

City

Day

Month

Year

DayOfWeek

▼ (1) Encounters

▼ (1) 0

▼ (1) Pokemon

Id

Name

> (1) Type

Total

> (1) Location

> (1) Time

> (1) Weather

> (1) NearbyObjects

▼ (1) Coencounters

> (1) 0

> (1) 1

> (1) 2

{ 3 fields }

5d04d2997cdf7ccaab4952e3

{ 6 fields }

America

New York

8

8

2016

dummy\_day

[ 151 elements ]

{ 6 fields }

{ 4 fields }

16

Pidgey

[ 2 elements ]

251

{ 5 fields }

{ 6 fields }

{ 6 fields }

{ 4 fields }

[ 3 elements ]

{ 4 fields }

{ 4 fields }

{ 4 fields }

Document

ObjectId

Object

String

String

Int32

Int32

Int32

String

Array

Object

Object

Int32

String

Array

Int32

Object

Object

Object

Object

Array

Object

Object

Object

# Agregacije

U koje doba dana se desilo najviše susreta sa pokemonima?

Odrediti kontinent sa najmanje i kontinent sa najviše susreta?

Za svakog pokemona odrediti ukupan broj susreta i prosečnu temperaturu prilikom susreta?

Odrediti prosečnu udaljenost teretane i pkestopa za pokemone kojima je barem jedan od tipova vodeni.

Za svako vreme odrediti 3 tipa koja su se najviše puta pojavila.

Kojih 10 pokemona se najčešće pojavilo kada se pojavio Bulbasaur?

Kojih 5 tipova se najčešće pojavilo kada se pojavio pokemon čiji barem jedan tip je vodeni?

Koja kombinacija vremena i tipa ima najmanji broj pojava. Uzeti u obzir samo tipove koji imaju više od 10000 pojava.

Prikazati za svaki kontinent pokemone koji su se pojavili samo na njemu.

Pronaći pokemona primećenog najbliže Novom Sadu (koordinate:[45.267136, 19.833549]).



# Promena agregacija u odnosu na predlog projekta

Stara agregacija	Nova agracija	Razlog
Po kakvom vremenu se pojavljuje najviše pokemona?	U koje doba dana se desilo najviše susreta sa pokemonima?	Zbog promene agracije vezane za tipove i vreme dana u kojem se pojavljuju.
Za svako doba dana odrediti pokemona koji se najčešće pojavljuje.	Za svako vreme odrediti 3 tipa koja su se najviše puta pojavila.	U trenutku beleženja susreta nije bilo dovoljno tipova čije pojave zavise od doba dana.
Po kakvom vremenu se najčešće pojavljuje Bulbasaur?	Odrediti kontinent sa najmanje i kontinent sa najviše susreta?	Upit previše jednostavan i sličan drugim.
Za svaki tip terena pronaći broj pojava svakog tipa pokemona.	Pronaći pokemona primećenog najbliže Novom Sadu (koordinate:[45.267136, 19.833549]).	Upit sličan upitu vezanom za doba dana i broj pojava tipova.

# U koje doba dana se desilo najviše susreta sa pokemonima?

Rešenje:

```
db.encounters.aggregate([
  {$project: {"_id": 0, "TimeOfDay": "$Encounters.Time.TimeOfDay"}},
  {$unwind: "$TimeOfDay"},
  {$sortByCount: "$TimeOfDay" },
  {$limit: 1},
  {$project: {"_id": 0, "Time of the day": "$_id", "Number of encountered pokemon": "$count" } }
])
```

Rezultat:

```
{
  "Time of the day" : "night",
  "Number of encountered pokemon" : NumberInt(147990)
}
```

## Analiza

- Vreme kada su igrači najaktivniji je pogodno za organizovanje događaja u igrici.
- Podaci su skupljeni u toku avgusta, većinski u SAD, tako da je igračima bilo lakše da šetaju i igraju igricu noću, kada je temperatura manja.



# Odrediti kontinent sa najmanje i kontinent sa najviše susreta?

Rešenje:

```
db.encounters.aggregate([
  {$group: {
    "_id": "$CommonInfo.Continent",
    "Count": {$sum: { $size:"$Encounters" }}
  }},
  {$sort: {"Count": -1 }},
  {$project: {"_id": 0, "Encounter.Continent": "$_id", "Encounter.Count": "$Count"}},
  {$group: {"_id": null, "EncountersPerContinent":{$push: "$Encounter" } }},
  {$project: {
    "_id": 0,
    "Continent with most encounters": { $arrayElemAt: [ "$EncountersPerContinent", 0 ] },
    "Continent with least encounters": { $arrayElemAt: [ "$EncountersPerContinent", -1 ] }
  }}
])
```

Rezultat:

```
{
  "Continent with most encounters" : {
    "Continent" : "America",
    "Count" : NumberInt(152711)
  },
  "Continent with least encounters" : {
    "Continent" : "Indian",
    "Count" : NumberInt(3)
  }
}
```

## Analiza

- U periodu kada su skupljeni podaci, igrice je legalno bila dostupna isključivo u Australiji, SAD i Novom Zelandu.
- Podaci mogu prikazati da li su igrice ipak igrali i ljudi u Evropi itd.

# Za svakog pokemona odrediti ukupan broj susreta i prosečnu temperaturu prilikom susreta?

Rešenje:

```
db.encounters.aggregate([
  {$project: {
    "Encounters.Pokemon.Id": 1,
    "Encounters.Pokemon.Name": 1,
    "Encounters.Weather.Temperature": 1
  }},
  {$unwind: "$Encounters"},
  {$group: {
    "_id": {"Number": "$Encounters.Pokemon.Id", "Name": "$Encounters.Pokemon.Name"},
    "Count": {$sum: 1},
    "Temperature": {$avg: "$Encounters.Weather.Temperature"}
  }},
  {$project: {
    "_id": 0, "Pokedex Number": "$_id.Number",
    "Name": "$_id.Name",
    "Number of encounters": "$Count",
    "Average Temperature": "$Temperature"
  }},
  {$sort: {"Number of encounters": -1}}
])
```

## Analiza

- Provera da li se ledeni pokemoni pojavlju tamo gde je hladno, a vatreni gde je toplo.
- Za pojedine regionalne, moguća provera da li su se samo tamo pojavili (npr. Kangaskhan značajno odstupa, jer je se pojavljivao samo u Australiji)

# Odrediti prosečnu udaljenost teretane i pokestopa za pokemone kojima je barem jedan od tipova vodeni?

Rešenje:

```
db.encounters.aggregate([
  {$project: {
    "Encounters.Pokemon.Type": 1,
    "Encounters.NearbyObjects.ClosestGym": 1,
    "Encounters.NearbyObjects.ClosestPokestop": 1
  }},
  {$unwind: "$Encounters"},
  {$match: {"Encounters.Pokemon.Type": "Water"} },
  {$group: {
    "_id": null,
    "Average Closest Gym Distance": {$avg: "$Encounters.NearbyObjects.ClosestGym"},
    "Average Closest Pokestop Distance": {$avg: "$Encounters.NearbyObjects.ClosestPokestop"}
  }},
  {$project: {"_id": 0} }
])
```

Rezultat:

```
{
  "Average Closest Gym Distance" : 1.9188675028379507,
  "Average Closest Pokestop Distance" : 0.4771367950166632
}
```

## Analiza

- Testiranje teorije da pokemoni većinski izlaze samo gde ima dosta teretana i pokestopova (koji su obično istorijski spomenici, značajne zgrade itd)
- Da li je igricu bolje igrati u gradu gde ima više objekata, nego u prirodi?
- Dobijene udaljenosti su u metrima.

# Za svako vreme odrediti 3 tipa koja su se najviše puta pojavila?

Rešenje:

```
db.encounters.aggregate([
  {$project: {"Encounters.Pokemon.Type": 1, "Encounters.Weather.Weather": 1}},
  {$unwind: "$Encounters"},
  {$project: {"Type": "$Encounters.Pokemon.Type", "Weather": "$Encounters.Weather.Weather"}},
  {$unwind: "$Type"},
  {$group: {"_id": {"Weather": "$Weather", "Type": "$Type"}, "Count": {$sum: 1}} },
  {$project: {"_id": "$_id.Weather", "Encounters.Type": "$_id.Type", "Encounters.Count": "$Count"} },
  {$sort: {"Encounters.Count": -1} },
  {$group: { "_id": "$_id", "EncountersByType": { $push: "$Encounters" } } },
  {$project: {
    "_id": 0,
    "Weather": "$_id",
    "Most Encountered Type": { $slice: [ "$EncountersByType", 3 ] }
  }}
])
```

Rezultat:

```
...
{
  "Weather" : "BreezyandOvercast",
  "Most Encountered Type" : [
    {
      "Type" : "Water",
      "Count" : 66.0
    },
  ],
}
```

## Analiza

- Da li vodeni tipovi izlaze kada je kiša, a travnati kada je sunce?
- Dobijeni su rezultati u kojima normal tip najviše izlazi po gotovo svakom vremenu. Razlog tome je neravnomerna zastupljenost tipova, kao i vreme kada su zabeleženi susreti.

# Kojih 10 pokemona se najčešće pojavilo kada se pojavio Bulbasaur?

Rešenje:

```
db.encounters.aggregate([
  {$unwind: "$Encounters"},
  {$project: {"Name": "$Encounters.Pokemon.Name", "Coencounters": "$Encounters.Coencounters"} },
  {$match: {"Name": "Bulbasaur"} },
  {$unwind: "$Coencounters"},
  {$project: {"_id": "$Coencounters.Name"}},
  {$group: {"_id": "$_id", "Count": {$sum: 1} }},
  {$sort: {"Count": -1}},
  {$limit: 10}
])
```

Rezultat:

```
{
  "_id" : "Nidoranâ",
  "Count" : 273.0
}
// -----
```

## Analiza

- Koji pokemoni su slični jedan drugom?
- Najčešće su se pojavljivali pokemoni istog tipa, ali i pokemoni koji su jednostavno jako česti.



# Kojih 5 tipova se najčešće pojavilo kada se pojavio pokemon čiji barem jedan tip je vodeni?

Rešenje:

```
db.encounters.aggregate([
  {$unwind: "$Encounters"},
  {$project: {"Type": "$Encounters.Pokemon.Type", "Coencounters": "$Encounters.Coencounters"} },
  {$match: {"Type": "Water"} },
  {$unwind: "$Coencounters"},
  {$project: {"Type": "$Coencounters.Type"} },
  {$unwind: "$Type"},
  {$group: {"_id": "$Type", "Count": {$sum: 1}}},
  {$sort: {"Count": -1}},
  {$limit: 5}
])
```

Rezultat:

```
{
  "_id" : "Water",
  "Count" : 17391.0
}
// -----
{
  "_id" : "Poison",
  "Count" : 16492.0
}
// -----
```

## Analiza

- Da li se isti tipovi često pojavljuju zajedno? Koje kombinacije tipova su najčešće?
- Dobijeni su većinski drugi tipovi za vodene pokemone, ali i tipovi za koje se smatra da bi trebali da se pojavljuju po sličnim vremenskim uslovima.

Koja kombinacija vremena i tipa ima najmanji broj pojava. Uzeti u obzir samo tipove koji imaju više od 10000 pojava.

Rešenje:

```
db.encounters.aggregate([
  {$unwind: "$Encounters"},
  {$project: {"Type": "$Encounters.Pokemon.Type", "Weather": "$Encounters.Weather.WeatherIcon"}},
  {$unwind: "$Type"},
  {$group: {"_id": "$Type", "TotalEncounters": {$sum: 1}, "EncountersWeather": {$push: "$Weather"} }},
  {$match: {"TotalEncounters": {$gt: 10000} }},
  {$unwind: "$EncountersWeather"},
  {$group: {"_id": {"Type": "$_id", "Weather": "$EncountersWeather"}, "Count": {$sum: 1}}},
  {$sort: {"Count": 1}},
  {$limit: 1},
  {$project: {"_id": 0, "Type": "$_id.Type", "Weather": "$_id.Weather", "Encounters count": "$Count"}}
])
```

Rezultat:

```
{
  "Type" : "Psychic",
  "Weather" : "fog",
  "Encounters count" : 93.0
}
```

## Analiza

- Granica od 10000 pojava je dodana, kako bi se ignorisali retki tipovi ili tipovi za koje je u tom trenutku bilo značajno manje pokemona.
- Agragcija treba da da odgovor na pitanje kada ne hvatati neki tip pokemona, npr. vatrene po kisi.

# Prikazati za svaki kontinent pokemone koji su se pojavili samo na njemu.

Rešenje:

```
db.encounters.aggregate([
  {$unwind: "$Encounters"},
  {$project: {"Continent": "$CommonInfo.Continent", "Pokemon": "$Encounters.Pokemon.Name"}},
  {$group: {"_id": "$Pokemon", "Continents": {$addToSet: "$Continent"} }},
  {$match: {"Continents": { $size: 1 } }},
  {$unwind: "$Continents"},
  {$group: {"_id": "$Continents", "Pokemon": {$push: "$_id"} }},
  {$project: {"_id": 0, "Pokemon": 1, "Continent": "$_id"}}
])
```

Rešenje:

```
{
  "Pokemon" : [
    "Farfetch'd"
  ],
  "Continent" : "Asia"
}
// -----
```

## Analiza

- Koji pokemoni su regionalni?
- Rezultati su pokazali neke zaista regionalne pokemone, ali i 2 evolucije, koje se samo zbog broja podataka nisu pojavile na drugim kontinentima (pojava evolucija je značajno ređa)

# Pronaći pokemona primećenog najbliže Novom Sadu (koordinate:[45.267136, 19.833549]).

Rešenje:

```
db.encounters.ensureIndex({'Encounters.Location.GeoLocation':'2dsphere'})

db.encounters.aggregate([
  { $geoNear: {
    near: {"type": "Point", "coordinates": [19.833549, 45.267136]},
    distanceField: "distance",
    includeLocs: "location",
    spherical: true
  }},
  { $redact: {
    $cond: {
      if: { $eq: [ { $ifNull: [ "$Location.GeoLocation", "$$ROOT.location" ] }, "$$ROOT.location" ] },
      then: "$$DESCEND",
      else: "$$PRUNE"
    }
  }},
  {$unwind: "$Encounters"},
  {$project: {
    "Pokemon": "$Encounters.Pokemon",
    "Distance": "$distance",
    "Continent": "$CommonInfo.Continent",
    "City": "$CommonInfo.City"
  }},
  {$sort: {"Distance": 1}},
  {$limit: 1}
])
```

## Analiza

- S obzirom da se za igrače na istoj lokaciji pojavljuju isti pokemoni, veliki značaj imaju susreti koji su se dogodili relativno skoro blizu trenutne lokacije igrača.
- Ukoliko bi se upit proširio sa vremensim intervalom, te skup susreta ažurirao u realnom vremenu, igrači bi mogli lakše pronaći pokemone koje žele.

# Pronaći pokemona primećenog najbliže Novom Sadu (koordinate:[45.267136, 19.833549]).

Rezultat:

```
{
  "_id" : ObjectId("5d04d3037cdf7ccaab495b95"),
  "Pokemon" : {
    "Id" : NumberInt(16),
    "Name" : "Pidghey",
    "Type" : [
      "Normal",
      "Flying"
    ],
    "Total" : NumberInt(251)
  },
  "Distance" : 193046.47106034038,
  "Continent" : "Europe",
  "City" : "Sarajevo"
}
```

## Analiza

- U toku skupljanja podataka, igrice nije bila leglno dostupna u Evropi.



# Reference

1. [Set podataka o pokemonima](#)
2. [Set podataka o susretima](#)
3. Materijali sa vežbi i predavanja
4. [MongoDB dokumentacija](#)