



# Reddit Jokes

...

Ensemble de publications sur la section “ Jokes ” du site  
web Reddit

Goulley Matthias, Apollon Vieira  
22/10/2018

# Les données (Kaggle)

- Fichier CSV de plus de 130 000 entrées
- Jokes entre 2010 et février 2018

*What do you call a soldier who survives Mustard gas and pepper spray?*

*A seasoned veteran.*

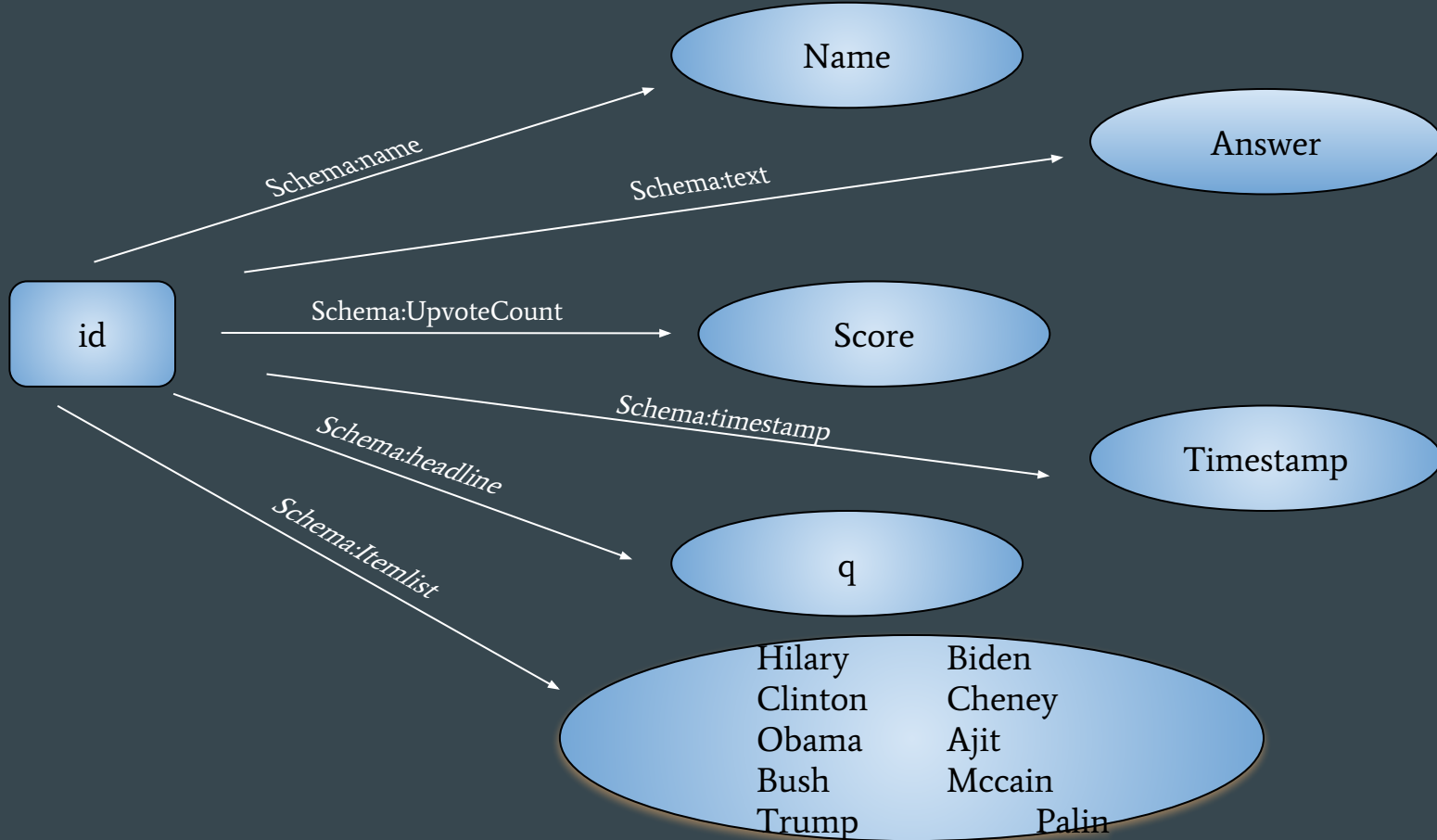
- *Runciblespoon77*

Une Joke est représentée par:

- Une référence vers le site reddit
- Un “Titre” de la blague
- La blague (Texte)
- Nom de l’auteur (Pseudonyme)
- Le timestamp (date + heure)
- Informations politiques (10)

Hilary, Clinton, Obama, Bush, Trump, Biden, Cheney, Ajit, McCain, Palin

# Transformation des données en rdf



# Requête n°1 : Sélectionner la meilleure et la pire blague du dataset

```
SELECT * WHERE {  
  {  
    SELECT * WHERE {  
      ?link <http://schema.org/upvoteCount> ?score;  
      <http://schema.org/headline> ?question;  
      <http://schema.org/text> ?reponse  
    } ORDER BY ASC(?score) LIMIT 1  
  }  
  UNION  
  {  
    SELECT * WHERE {  
      ?link <http://schema.org/upvoteCount> ?score;  
      <http://schema.org/headline> ?question;  
      <http://schema.org/text> ?reponse  
    } ORDER BY DESC(?score) LIMIT 1  
  }  
}
```

link	score	question	reponse
<https://www.reddit.com/r/Jokes/comments/2l024h>	"10"	"CNN reporter"	"In Jerusalem, a female CNN journalist heard about a very old Jewish"...
<https://www.reddit.com/r/Jokes/comments/5b9nap>	"9993"	"A joke my grandma told me before she passed."	"So a classroom teacher was giving candy to all the students. "...

# Requête n°2 : Sélectionner l'auteur ayant reçu le plus d' upvotes

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
SELECT ?name (SUM(STRDT(?score, xsd:integer)) AS ?sum) (COUNT(DISTINCT ?score) AS ?count)
WHERE {
  ?x <http://schema.org/upvoteCount> ?score;
      <http://schema.org/additionalName> ?name
} GROUP BY ?name
ORDER BY DESC(?sum) LIMIT 5
```

name	sum	count
"madazzahatter"	1657823	542
"wackoclown"	492586	234
"raydeep"	339241	165
"iRyaaanM"	205155	66
"supercoooldudewithab"	204598	148

# Utilisation du dataset

## *“Volcano Eruptions Global Distribution”*

Bouvet Henri, Coutand Maxence

Requête n3 : Est-ce qu’il y a des blagues qui contiennent le mot “volcano” qui ont été postées dans un délai d’une semaine après une éruption volcanique

```

SELECT ?headline ?blague ?joketimestamp ?nomvolcan ?volcanotimestamp ?delta
WHERE{
  ?x schema:headline ?headline;
     schema:text ?blague;
     schema:dateCreated ?time;
     schema:text ?answer.
  ?y dbo:year ?annee;
     dbo:month ?mois;
     dbo:day ?jour;
     dbo:volcano ?nomvolcan.
  FILTER(regex(?headline, "volcano"))
  BIND(xsd:integer (strbefore(?time, ".")) AS ?joketimestamp)
  BIND(xsd:integer (?annee) AS ?Annee)
  BIND(xsd:integer (?mois) AS ?Mois)
  BIND(xsd:integer (?jour) AS ?Jour)
  BIND( (?Annee - 1970) * 31557600 + (?Mois - 1) * 2629743 + (?Jour - 1) * 86400 AS ?volcanotimestamp)
  BIND( ?joketimestamp - ?volcanotimestamp AS ?delta)
  FILTER(?joketimestamp > ?volcanotimestamp - 86400 && ?joketimestamp < ?volcanotimestamp + 604800)
}

```

```

PREFIX schema: <http://schema.org/>
PREFIX dbo: <http://dbpedia.org/ontology/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

```

headline	blague	joketimestamp	nomvolcan	volcanotimestamp	delta
"What do you get when you throw a chicken in a volcano?"	"Baklava"	1513864956	"Merapi"	1513603173	261783
"Why are volcanoes so jolly?"	"Whenever they hear a joke, they erupt with laughter"	1489931198	"Etna"	1489676286	254912
"The Calbuco volcano is very hot right now... "	"But everything around it is Chile. "	1429807197	"Calbuco"	1429795629	11568

# Utilisation du dataset

## *“World happiness report”*

Maréchal Roxane, Pagot Stella

Requête n°4 : Sélectionner la meilleure blague pour chaque pays et compter le nombre de blagues sur ce pays



```

SELECT ?country ?HappinessRank ?count ?maxscore MIN(?headline) AS ?titre)
(MIN(?blague) AS ?reponse)
WHERE {
    ?x schema:text ?blague;
        schema:headline ?headline;
        schema:upvoteCount ?maxscore.

    ?y geof:country ?country;
        rank:happinessRank ?HappinessRank.
    {
        SELECT DISTINCT ?country (MAX(?score) AS ?maxscore) (COUNT(?x) AS ?count)
        WHERE {
            ?x schema:headline ?headline;
                schema:text ?blague;
                schema:upvoteCount ?score.

            ?y geof:country ?country.

            FILTER(regex(?headline, ?country) || regex(?blague, ?country))
        } GROUP BY ?country
    }
    FILTER(regex(?headline, ?country) || regex(?blague, ?country))
} GROUP BY ?country ?HappinessRank ?count ?maxscore
ORDER BY DESC(?count)
LIMIT 5

```

```

| PREFIX schema:
| <http://schema.org/>
| PREFIX dbo:
| <http://dbpedia.org/ontology/>
| PREFIX geof:
| <http://www.mindswap.org/2003/owl/geo/
| geoFeatures20040307.owl#>
| PREFIX rank:
| <http://dbpedia.org/ontology/rank#>
| PREFIX xsd:
| <http://www.w3.org/2001/XMLSchema#>
| PREFIX rdf:
| <http://www.w3.org/2000/01/rdf-schema#
| >
| PREFIX head:
| <http://dbpedia.org/ontology/head#>

```





# Top 6 des meilleures blagues sur Trump

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
```

```
PREFIX schema: <http://schema.org/>
```

```
SELECT ?q ?answer WHERE {  
  ?x schema:text ?answer;  
      schema:headline ?q;  
      schema:upvoteCount ?score;  
      schema:ItemList "trump"^^<http://schema.org/itemListElement>.  
}  
ORDER BY DESC(?score) LIMIT 6
```

intitulé	réponse
"So Trump is working with Putin on cybersecurity... "	"In other news, the principal at my school is working with the boys to install a surveillance system to insure privacy in the girls' locker room."
"Trump has left the historical Paris Climate Accord in which countries around the world agreed to fight global warming... "	"It was the first time he pulled out of a working model."
"What is Donald Trump's favorite vegetable?"	"Leeks"
"Donald Trump was once asked if he could quote any Bible verses."	"Give a man a fish and he will eat for a day. Deport him and you will never have to feed him again. - Donald 20:17"
"Whether you love him or hate him, Donald Trump has given us something we haven't had in decades."	"A First lady we can masturbate to."
"How does Donald Trump plan on deporting 12 million illegal immigrants?"	"Juan by Juan. "

# Description du Dataset avec le vocabulaire VoID

Voici comment nous avons choisi de définir notre dataset avec le vocabulaire VoID. Ceci permettra aux autres contributeurs du web sémantique d'avoir des informations sur la manière dont nous avons produit notre dataset.

```
:RedditJokes      a      void:Dataset;
                  dcterms:title      "Reddit Jokes";
                  dcterms:description "Blagues issues de la section jokes du site
Reddit";
                  dcterms:creator      :CreatorPen;
                  dcterms:publisher     :Goulley_Matthias;
                  dcterms:publisher     :Vieira_Apollon;
                  dcterms:contributor   :Univ_Nantes;
                  dcterms:source        :source;
                  dcterms:created       "2018-02-01"^^xsd:date ;
                  dcterms:issued        "2018-10-14"^^xsd:date ;
                  dcterms:modified      "2018-11-28"^^xsd:date ;
                  dcterms:license       <http://www.opendatacommons.org/licenses/pddl/> ;
                  void:feature          <http://www.w3.org/ns/formats/Turtle>;
                  void:exampleResource <https://www.reddit.com/r/Jokes/comments/2nwsagg>;
                  void:exampleResource <https://www.reddit.com/r/Jokes/comments/2nwc7k>;
                  void:entities         133329;
                  void:vocabulary       <http://schema.org/>.
```

# Conclusion

Ce projet nous a permis de découvrir le web sémantique et ses problématiques.

A travers ce projet, nous avons produit des données standardisées permettant d'être partagées et réutilisées par les autres utilisateurs.

L'un des avantages de cette standardisation c'est qu'elle permet de lier facilement des jeux de données entre eux. Nous avons pu étudier cela en liant nos données avec celles du "World happiness report" ainsi que les données "Volcano Eruptions Global Distribution" et avons pu obtenir des résultats intéressants.

# Merci de votre attention

*What do you call a moose with no name?*

*Anonymoose*

- *Chelcheld2*

*A wife is like a hand grenade*

*Take off the ring and say goodbye to your house*

- *levji\_kralj*