



Relevant Facets

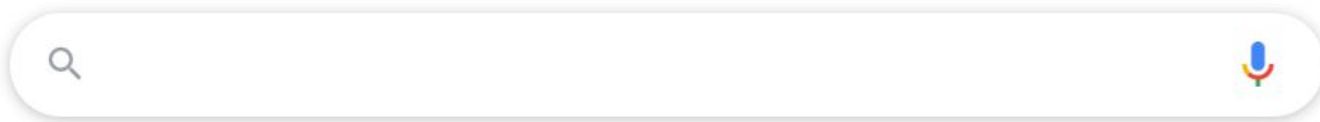
Lucian Precup
Radu Pop
@lucianprecup @a2lean
#haystackconf
Berlin EU 2019

// Poll

- How many of you are using facets with the search engines you implement ?
- Who is doing statistics on facet usage ?
- Who is using Solr ?
- Who is using Elasticsearch ?
- Other search technology ?
- Who speaks French ?



// Why this talk ?



Google Search

I'm Feeling Lucky

Google offered in: [Français](#)

// Facets ?



Used to define filters that refine the initial query



Used for disambiguation



Give a holistic view over the search results



Allow to find the needle in the haystack more quickly



Choisir Drive ou la livraison pour commencer vos courses

Affiner

Affiner par rayon

- ☐ Fruits et Légumes 77
- ☐ Viandes et Poissons 9
- ☐ Crèmerie 1
- ☐ Charcuterie 1
- ☐ Traiteur 65

▼ VOIR PLUS

Marque

🔍 Marques...

- ☐ CARREFOUR 89
- ☐ JARDIN BIO 18
- ☐ PANZANI 14
- ☐ SACLA 14
- ☐ MIX BUFFET 10

« tomate »

469 résultats

Trier par **Pertinence** ▼

DECOUVRIR

Prix imbattable!

Tomates rondes en grappe

le kilo
1.39 € / Kilogramme

FRANCE

1,39€

Tomate ronde en grappe cocktail

la barquette de 350 g
5.71 € / Kilogramme

FRANCE

2,00€

Tomate Cerise rouges rondes

la barquette de 1 kg
3.50 € / Pièce

MAROC

3,50€

Tomate Cerise rouges rondes

REFS.SANS MARQUE

Promotion

Tomates cerises rondes

la barquette de 400g

Tomate ronde noire

le kilo

Tomate côtelée rouge

le kilo

jeudi 13 juin 2019

Did you mean : **water hygiene and sanitation**

Tips : With Diggr you can search from more than 35 different MSF information platforms at the same time



About 153 results in 0.534 seconds

Display : 10 results per page



Order by : Relevance



Gallery



Map



Hide search filters

ALL

OOPS

OOPSMED

INSIDE0CB

MSF WEBSITES

TREASURY

SHERLOG

MAP CENTRE

Selected Filters :

MSF Websites X

eng X

LANGUAGES

☒ eng (153)☐ fra (5)☐ ara (1)☐ zho (1)

DATES OF TOPIC

PRODUCT

SOURCES

ORGANISATION

☐ Médecins Sans Frontières (122)☐ World Health Organization (93)☐ Drugs for Neglected

GO TO THE WATER HYGIENE AND SANITATION PAGE ON SHERLOG

GO TO THE WATER AND SANITATION PAGE ON OOPS MED

**Setting priorities for humanitarian water, sanitation and hygiene research: a meeting report | DMello Guyett et al - 2018 - Setting priorities for humanitarian water, sanitation, and hygiene research.pdf**

Last modified on 18/12/2018

<http://heldresearch.msf.org/msf/bitstream/10144/619251/1>Setting **priorities** for **humanitarian water, sanitation** and **hygiene** research: a meeting report MSF for **humanitarian water, sanitation** and **hygiene** research: a meeting report 2018, 12 (1)**Microsoft Word - The Priorities UK New 2011.doc | the_priorities_uk_2011.pdf**

Last modified on 17/12/2018

<https://evaluation.msf.org/sites/evaluation/files>**Water, hygiene and sanitation.....**9 4. Number, type and location of **water** points **Hygiene and sanitation.....****Using lot quality assurance sampling to assess access to water, sanitation and hygiene services in a refugee camp setting in South Sudan: a feasibility study | Harding E et al - 2017 - Using lot quality assurance sampling**

RECHERCHES

M. Mlle **Dujardin** A CHANCEREL WILLY (A0004 / 000000300)
M. Mlle **Dujardin** PIERRE & (00200 / 000303000)
M. Denis **DUJARDIN** (00020 / 000100000)
Mme Patricia **DUJARDIN** (A0001 / 000090000)
M. Philippe **DUJARDIN** (00500 / 010000000)
M. Matthieu **DUJARDIN** (A1000 / 000000200)
Mme Edith **DUJARDIN** (A0000 / 010100000)
M. Robert **DUJARDIN** (00000 / 000000090)
M. Xavier **DUJARDIN** (A0003 / 010201200)
Mme Sylvie **DUJARDIN** (00090 / 010200202)

2201 RESULTATS TROUVES

Trier les résultats par Pertinence

Essayez avec cette orthographe : DUJARDIN

**M. Denis DUJARDIN** - Prospect Particulier - A1001 / 0101010112 RUE DE PARIS, 75000 PARIS
Intermédiaire : 00000A**Mme Patricia DUJARDIN** - Prospect Particulier - A2001 / 0101010121 RUE DE VERSAILLES, 78000 VERSAILLES
Né(e) le: 31/12/2012 | Intermédiaire : 10001A**M. Philippe DUJARDIN** - Prospect Particulier - A0006 / 01020202025 B RUE DU VILLAGE, 59000 METZ
Né(e) le: 01/01/1970 | Intermédiaire : 10000A**DUJARDIN** - Prospect Entreprise - A0007 / 01201201217 BOULEVARD DU GENERAL DE GAULLE, 75000 Paris
Intermédiaire : 600000**SARL DUJARDIN** - Prospect Entreprise - A1234 / 10101010112 RUE DE LA MONTAGNE, 93000 SAINT DENIS
SIRET : 11008800000010 | Intermédiaire : 15000A

AFFINER LA RECHERCHE

▼ Type client

- ☐ Prospect (1647)
☐ Client (397)
☐ Ancien client (157)

▼ Nature personne

- ☐ Particulier (1939)
☐ Professionnel (116)
☐ Entreprise (114)
☐ Autre (28)
☐ Copropriété (3)
☐ Association (1)

[Effacer les filtres](#)

109 962

Tapez votre recherche ou utilisez directement les filtres ci-dessous

NOM ARTISTE [7 271] NATIONALITÉ [335] ŒUVRES EXPOSÉES [3] SECTEUR COLLECTION [7] TYPE D'ŒUVRE [375] ANNÉE CRÉATION [179] ANNÉE ACQUISITION [115] MODE

Tout déplier

Architecture	[15 227]
• Architecture	[94]
+ Dessin d'architecture	[13 544]
+ Élément d'architecture	[51]
Maquette d'architecture	[1 257]
• Maquette d'étude	[97]
• Maquette de détail	[10]
• Maquette de rendu	[51]
• Maquette de situation	[7]
• Peinture d'architecture	[10]
• Photographie d'architecture	[141]
• Sans domaine déterminé	[12]
Arts plastiques	[89 022]
+ Design	[5 678]
+ _Autre discipline	[1]

Hierarchical facets



All Departments ▾ bike



NextDay delivery

Delivering to 94066

1 - 40 of 1,000+ results

Refine by

Price ▾

Top Brands ▾

Store Availability ▾

Sort by

Best Match



Departments

Sports & Outdoors

Bikes

All Kids' Bikes

Cruiser Bikes

Kids' BMX Bikes

[See all Sports & Outdo...](#)

[See All Departments](#)

Delivery & Pickup

☒ Show all

☐ 2-day delivery

☐ Ship to home

☐ Free pickup + discount

☐ Free pickup

☐ Free pickup today

Bicycle Type

☐ BMX Bikes

☐ Comfort & Hybrid Bikes



Kids bike buying guide

See our guide for ride on toys for kids of all ages.



ROLLBACK

Featured Item

Bonus Helmet and Accessories with Mongoose Excursion Mountain Bike!

★★★★★

· Buy a Mongoose Excursion mountain bike and receive a helmet, lock, and cargo bag!

\$148.00

Free delivery

Free pickup



Men's Bikes



Women's Bikes



Kids Bikes



Rollback



All Bike

BEST SELLER



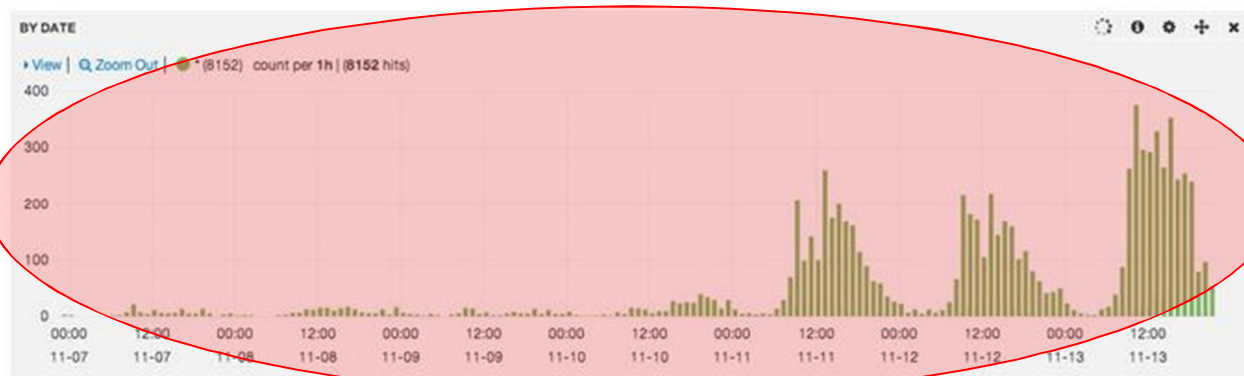
BEST SELLER



BEST SELLER



Other “exotic” facets

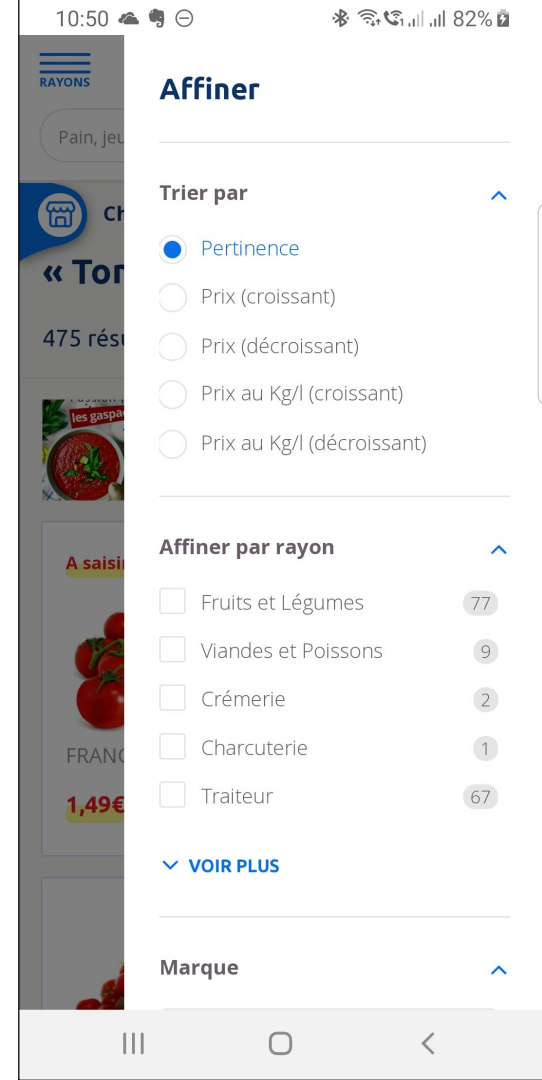
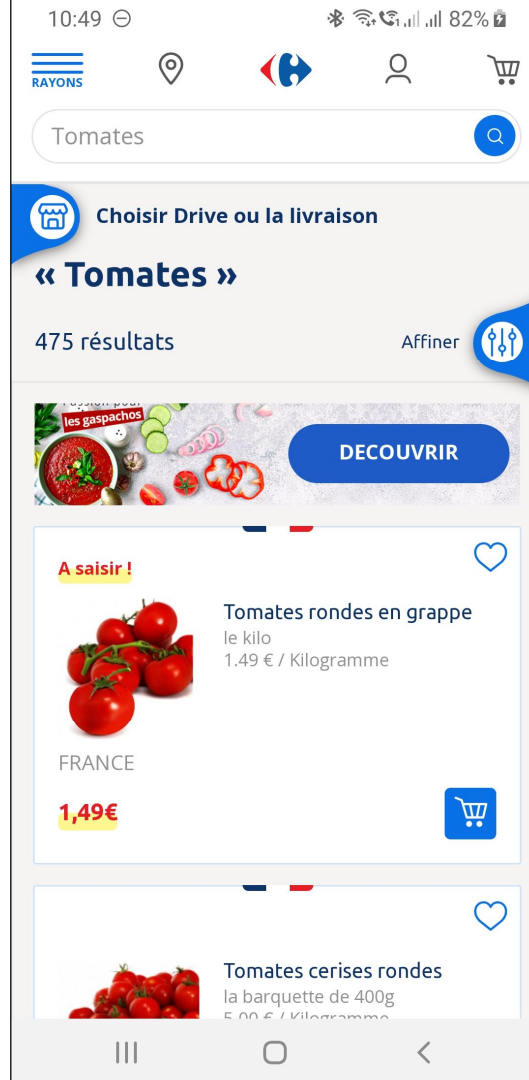


TOP10 MENTION - DEVOXX

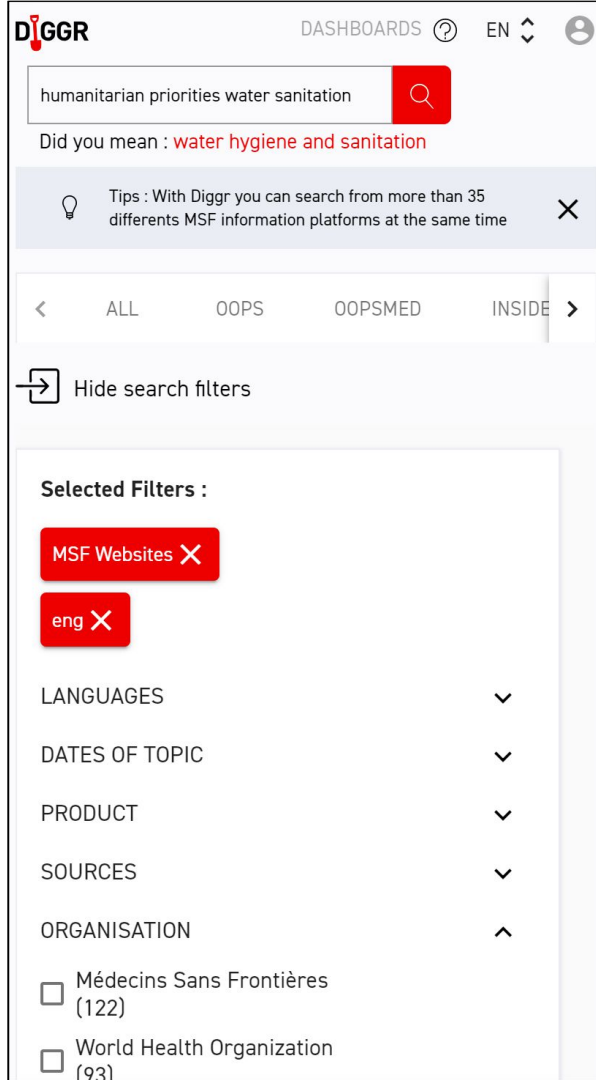
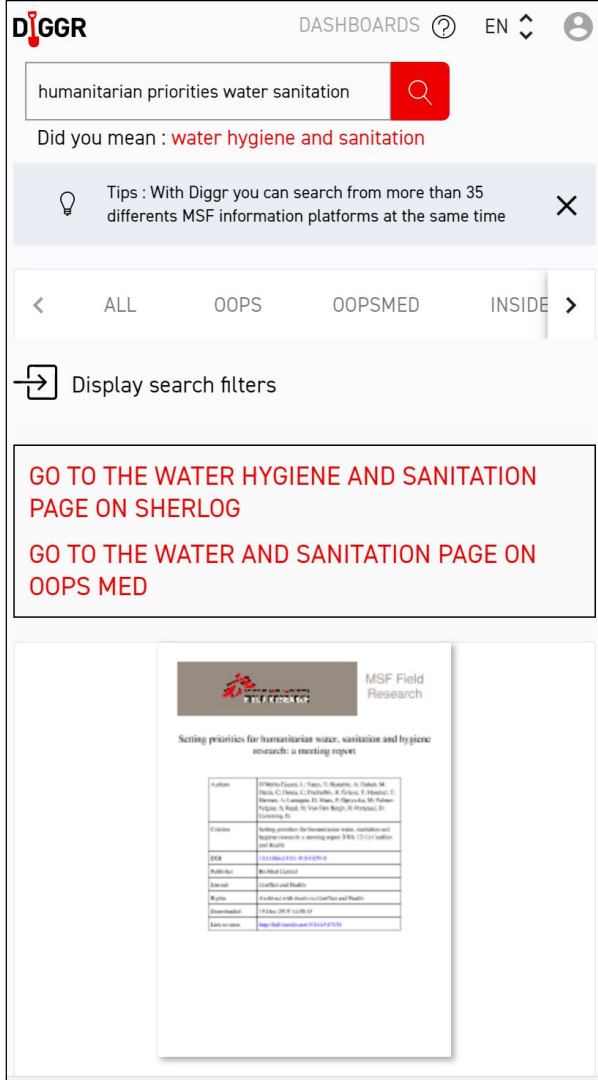
Term	Count	Action
Devoxx	589	Q
java	282	Q
BrianGoetz	179	Q
Stephan007	168	Q
agoncal	163	Q
mraible	159	Q
angularjs	151	Q
arungupta	133	Q
Parleys	116	Q
svenpet	107	Q



Facets on a mobile device



Facets on a mobile device



// Facets and Filters

The screenshot displays the Digger search interface. At the top, the search bar contains 'humanitarian priorities water sanitation' with a suggestion 'Did you mean : water hygiene and sanitation'. A tip box states: 'Tips : With Digger you can search from more than 35 different MSF information platforms at the same time'. Below the search bar, it shows 'About 153 results in 0.534 seconds' and 'Display : 10 results per page'. The 'Order by' is set to 'Relevance'. A navigation bar includes tabs: ALL, OOPS, OOPSMED, INSIDE0CB, MSF WEBSITES, TREASURY, SHERLOG, and MAP CENTRE. The 'MSF WEBSITES' tab is selected. On the left, the 'Selected Filters' section shows 'MSF Websites' and 'eng'. Below this, the 'LANGUAGES' facet is expanded, showing 'eng (153)', 'fra (5)', 'ara (1)', and 'zho (1)'. Other facets like 'DATES OF TOPIC', 'PRODUCT', 'SOURCES', and 'ORGANISATION' are also visible. The main search results area shows several documents, including 'Setting priorities for humanitarian water, sanitation and hygiene research: a meeting report | DMello Guyett et al - 2018', 'Microsoft Word - The Priorities UK New 2011.doc | the_priorities_uk_2011.pdf', and 'Water, hygiene and sanitation...'. Two large red boxes with white text, 'Filters' and 'Facets', are overlaid on the image. Arrows point from the 'Filters' box to the 'Selected Filters' section and the 'LANGUAGES' facet. Arrows point from the 'Facets' box to the 'MSF WEBSITES' tab, the 'LANGUAGES' facet, and the 'ORGANISATION' facet.

Filters

Facets

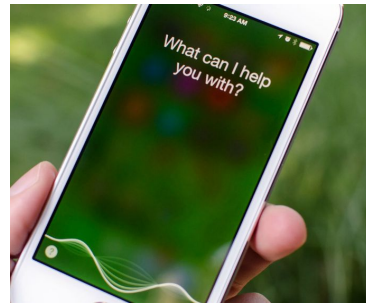
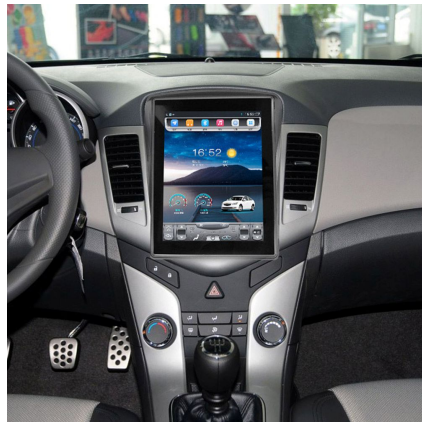
// Why are facets important ?

VOICE + SCREEN
(multimodal)



- More and more data and less and less space to display it
- New ways of searching: voice, assistants, chat bots

VOICE ONLY



// How are facets implemented ?

Facets are a standard feature of modern search engines.

Apache Lucene has great support for everything around facets

- Solr : field value faceting, range faceting, pivot faceting, interval faceting, block join faceting, ...
- Elasticsearch : aggregations, sub-aggregations, top hits aggregation, histogram aggregation, range aggregations, geo aggregations, ...

The User Experience with facets and the way they are "displayed" can be very diverse

// Structure of the talk

- Examples of facet implementations
- Challenges with facets and possible solutions
- Challenges with search in general and how facets can help
- Technical implementation examples are with Elasticsearch
- We are addressing less the "graphical" display of facets and more the technical issues with their relevancy



// Challenge #1: marketplaces

- Issue: the heterogeneity of results and the number of candidate facets

MODE DE LIVRAISON

- ☐ Retrait en magasin
- ☐ Retrait en relais
- ☐ Livraison à domicile

INFORMATIQUE

- ☐ Alimentation (1 654)
- ☐ Hub USB (195)
- ☐ Connectique interne (61)

ACCESSOIRES

- ☐ Câble USB (4 121)
- ☐ Accessoires divers boîtier (531)
- ☐ Connectique informatique (490)

PHOTO, CAMÉRA, DRONE

- ☐ Appareil photo hybride (190)
- ☐ Reflex (69)
- ☐ Drone (31)

3 neufs dès 6,90 €

En stock
Livraison OFFERTE

Ajouter au comparateur

Ajouter au panier

Câble USB XCSOURCE XCSOURCE USB 3.0 À SATA III 22 BROCHES 2.5 POUCES HDD EXTERNE SSD DISQUE DUR CONNECTEUR ADAPTATEUR CÂBLE CÂBLE SATA HAUTE VITESSE AC986

9,99€
6,05€

Vendu par Femency

★★★★★
État : neuf

- Le disque dur d'écriture et de lecture à grande vitesse fait de votre disque dur un disque dur
- Théoriquement, le taux de transfert peut aller jusqu'à 6gb / s (max), avec un disque dur ssd super
- Installation facile, il suffit de brancher et de jouer, aucun pilote nécessaire. Compact et léger, facile à
- Fonction de soutien otg de téléphone portable, le téléphone de connexion a besoin de soutien otg

4 neufs dès 6,05 €

En stock
Livraison OFFERTE

Ajouter au comparateur

Ajouter au panier

Câble USB D2 DIFFUSION CÂBLE D2 DIFFUSION FIBRE OPTIQUE POUR BOX BOUYGUES, SFR ET ORANGE (10M)

18,68€

Vendu par 1foDiscount

★★★★★
État : neuf

En stock
Livraison OFFERTE

Ajouter au panier

Heterogeneity of results: Solution 1

Facets based on top N results:

- Fetch the top N results (first page + a few of the next ones)
- Retain only the facets applicable to these top N results

Implementation details:

- First query: query term
- Fetch the first N document ids (let's say max 1024)
- Second query : terms filter on document ids and aggregations

Heterogeneity of results: Solution 2

- Modeling with a single facet-name / facet-value field tuple and the nested type
- Need to treat differently strings, numbers and booleans

```
1  "string_facets": {
2    "type": "nested",
3    "properties": {
4      "facet-name": {
5        "type": "keyword"
6      },
7      "facet-value": {
8        "type": "keyword"
9      }
10   },
11 },
12 "numeric_facet" : {
13   "type": "nested",
14   "properties": {
15     "facet-name": {
16       "type": "keyword"
17     },
18     "facet-value": {
19       "type": "double"
20     }
21   }
22 }
```


```
1  "string_facets": [
2    {
3      "facet-name": "memory",
4      "facet-value": "64 MB"
5    },
6    {
7      "facet-name": "OS_type",
8      "facet-value": "android"
9    }
10 ],
11 "numeric_facet": [
12   {
13     "facet-name": "price",
14     "facet-value": 599
15   },
16   {
17     "facet-name": "screen_size",
18     "facet-value": 5,7
19   }
20 ]
```

Heterogeneity of results:

Solution 2 – the query

```
1 {
2   "aggs": {
3     "agg_string_facet": {
4       "nested": {
5         "path": "string_facets"
6       },
7       "aggs": {
8         "facet_name": {
9           "terms": {
10            "field": "string_facets.facet-name"
11          },
12          "aggregations": {
13            "facet_value": {
14              "terms": {
15                "field": "string_facets.facet-value"
16              }
17            }
18          }
19        },
20      }
21    },
22    "agg_numeric_facet": {
23      "nested": {
24        "path": "numeric_facet"
25      },
26      "aggs": {
27        "facet_name": {
28          "terms": {
29            "field": "numeric_facet.facet-name"
30          },
31          "aggs": {
32            "facet_value": {
33              "stats": {
34                "field": "numeric_facet.facet-value"
35              }
36            }
37          }
38        },
39      }
40    }
41  }
42 }
```

// Challenge #2: auto-completion

BASK 

BASKETBALL (SPORTWELT)

BASKETBALL > BEKLEIDUNG

BASKETBALLSCHUHE

BASKETBALLTRIKOT

BASKETBALL > SHORTS, HOSEN









BASKETBALL AUSRUSTUNG

PROTEKTOREN, BANDAGEN, TAPES > **BASKETBALL**

BASKETBALL > BALLE

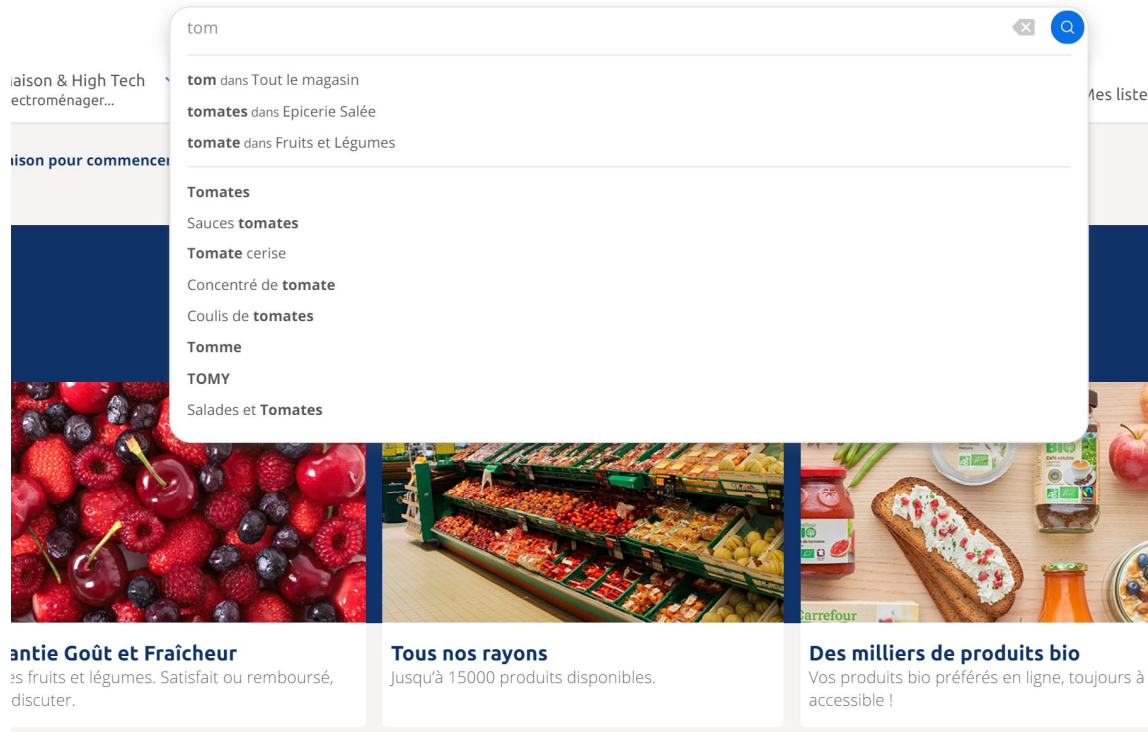
BASKETBALLE, NETBALLE

BASKETBALLSOCKEN

 WANDBEFESTIGUNG FÜR BASKETBALLKORB B300 UND B700. 3 SPIELHOHEN	 DOPPELHUB- BALLPUMPE GELB	 BALLNADELN 3ER- PACK	 SPORTTASCHE KIPOCKET 20LITER BLAU/GELB
 SPORTTASCHE TROLLEY INTENSIVE 30 LITER BLAU	 RUCKSACK INTENSIVE 25 LITER SCHWARZ/KAKI	 SPORTTASCHE INTENSIVE 65 LITER BLAU	 SPORTTASCHE RUCKSACK CLASSIC 25 LITER BLAU

[-Kg](#)

// Challenge #2: auto-completion



tom

- tom dans Tout le magasin
- tomates dans Epicerie Salée
- tomate dans Fruits et Légumes

Tomates

- Sauces **tomates**
- Tomate** cerise
- Concentré de **tomate**
- Coulis de **tomates**

Tomme

TOMY

Salades et **Tomates**

antie Goût et Fraîcheur
s fruits et légumes. Satisfait ou remboursé,
discuter.

Tous nos rayons
Jusqu'à 15000 produits disponibles.

Des milliers de produits bio
Vos produits bio préférés en ligne, toujours à
accessible !

Auto-completion: solution

Products index

```
1 {  
2   "title": "Red tomatoes",  
3   "category": "Tomates",  
4   "department": "Fruits et Légumes",  
5   "price": "3"  
6 }  
7  
8 {  
9   "title": "Yellow tomatoes",  
10  "category": "Tomates",  
11  "department": "Bio et Ecologie",  
12  "price": "30"  
13 }
```

Suggestions index

```
1 {  
2   "text": "Tomates",  
3   "context": {  
4     "department": ["Fruits et Légumes"]  
5   }  
6 }  
7  
8 {  
9   "term": "Tomates",  
10  "context": {  
11    "department": ["Bio et Ecologie"]  
12  }  
13 }
```

Use the Update API here
and also increase the number of occurrences

Auto-completion: solution

The "Suggestions" index

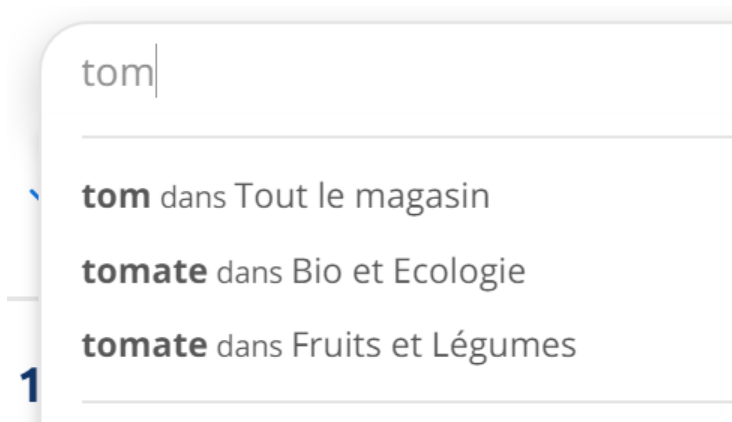
```
1 {  
2   "text": "Tomates",  
3   "context": {  
4     "product_categories_name": [  
5       "Fruits et Légumes",  
6       "Bio et Ecologie"  
7     ]  
8   },  
9   "occurrences": 81,  
10  "popularity": 16  
11 }
```

The query

```
1 {  
2   "query": {  
3     "multi_match": {  
4       "query": "tom",  
5       "fields": ["text.smartComp"]  
6     }  
7   },  
8   "sort": [  
9     {  
10      "occurrences": {  
11        "order": "desc"  
12      }  
13    }  
14  ]  
15 }
```


Auto-completion: solution

The result



Auto-completion: solution

The shortcut

```
1 GET _a2/autocomplete
2 {
3   "text": "tom"
4 }
```

```
1 {
2   "engineTimeInMillis": 1,
3   "totalHits": 1478,
4   "product": [...],
5   "suggestions": [
6     {
7       "text": [
8         "tom"
9       ],
10      "highlight": [
11        "<em>tom</em>"
12      ],
13      "type": "input"
14    },
15    {
16      "text": [
17        "Tomates"
18      ],
19      "highlight": [
20        "<em>Tomates</em>"
21      ]
22    },
23    ...
24  ],
25  "filteredSuggestions": [
26    {
27      "text": "Tomates",
28      "highlight": "<em>Tomates</em>",
29      "facetFilter": {
30        "product.categories.name": "Bio et Ecologie"
31      }
32    },
33    {
34      "text": "Tomates ",
35      "highlight": "<em>Tomates</em>",
36      "facetFilter": {
37        "product.categories.name": "Fruits et Légumes"
38      }
39    },
40    ...
41  ]
42 }
```

// Challenge #3: assistants

- Often the first responses of an assistant are suggestions for additional filters that refine the query.




Simulator


✕


A2

Talk to A2



Alright. Here's the test version of ei two.
Welcome! What are you looking for today?

Chocolate 



I found 920 products for you, what is your favorite brand?

Suggested input

Carrefour

Carrefour Bio



Lindt

Milka







Côte d'Or

cancel

Input

 ▾ Try typing or saying "Talk to A2". 

Change version


Surface

Language

Location

Smart Displ... ▾

English (United States) ▾

Schönwalder Str. 21, 13347 Berlin, Allem... 

DISPLAY

REQUEST





RESPONSE


AUDIO

DEBUG

ERRORS

⋮

 A2

×

I found 920 products for you, what is your favorite brand?

Carrefour

Carrefour Bio

Lindt

Milka

Côte d'Or

How to narrow ?

- Often the first responses of an assistant are suggestions for additional filters that refine the query
- “Quick win” solution :
 - Filters
- Issue :
 - Which facets to choose?
- Prerequisite:
 - Your search engine should already have relevant filters

Brand ^

Q Brand

☐ CARREFOUR 89

☐ JARDIN BIO 18

☐ PANZANI 14

☐ SACLA 14

☐ MIX BUFFET 10

SEE MORE

Price —

\$ to \$ Go

☐ \$0 - \$5

☐ \$5 - \$10

☐ \$10 - \$15

☐ \$15 - \$20

☐ \$20 - \$25

☐ \$25 - \$50

☐ \$50+

Breakfast by Dietary Need —

Gluten-Free Breakfast

Organic Breakfast

Sugar-Free Breakfast

Flavor —

☐ Chocolate

☐ Strawberry

☐ Cinnamon

☐ Peanut Butter

See More Flavors ...

Customer Rating —

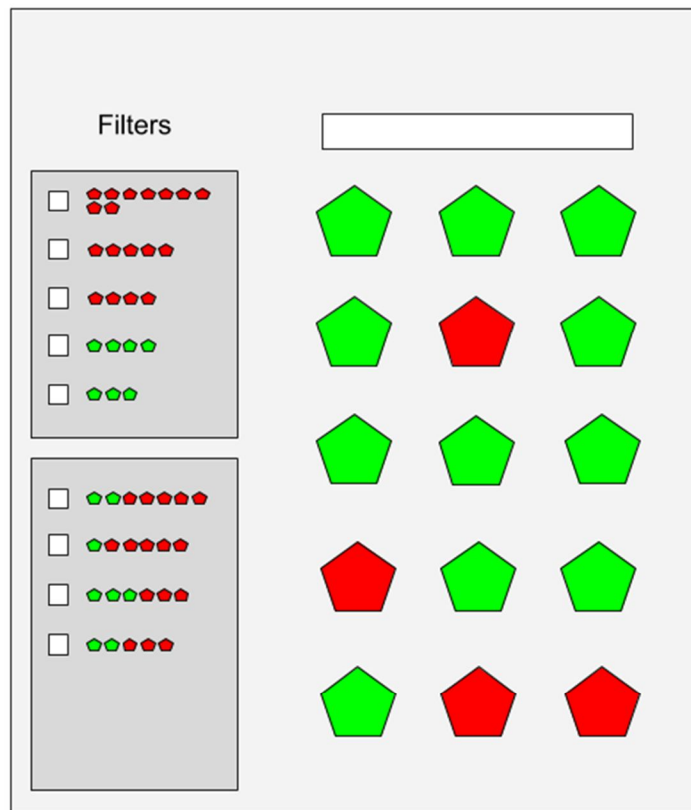
☐ ★★★★★ & Up


☐ ★★★★☆ & Up


☐ ★★★☆☆ & Up


// Challenge #4: relevant facet values

- Issue: how to make facet values relevant in the context of many "less relevant" results ?



page 2 

page 3 

page 4 

- ▼ Rayons
Promo, épicerie...
- ▼ Maison & High Tech
Electroménager...
- ▼ Services
Voyage, spectacle...

📍 Trouver un magasin

📖 Mes listes ▼

👤 Me connecter ▼

Marque



Q Marques...

- ☐ CARREFOUR 106
- ☐ TROPICANA ? 42
- ☐ GERBLE ? 36
- ☐ KARELEA 27
- ☐ JOKER ? 23

▼ VOIR PLUS

Préférences alimentaires



- ☐ Sans sucres ajoutés ? 325
- ☐ Sans sucres ? 100
- ☐ Végétarien 54
- ☐ De saison 5
- ☐ Halal 2



1,28€



0,67€



1,13€



1,25€



Sucre vanillé VAHINE
les 10 sachets de 7,5 g
26.40 € / Kilogramme



Sucre en poudre DADDY
la boîte de 1 kg
1.34 € / Kilogramme



Sucre en poudre DADDY
la boîte de 100 bâchettes - 500 g
4.42 € / Kilogramme



Sucre roux GRAEFFE
le paquet de 1 kg
2.14 € / Kilogramme



Rayons
Promo, épicerie...

Maison & High Tech
Electroménager...

Services
Voyage, spectacle...

📍 Trouver un magasin

👤 Mes listes

👤 Me connecter

Marque

Q Marques...

☐ CARREFOUR



8

☐ BEGHIN SAY

15

☐ DADDY

9

☐ SAINT LOUIS

9

☐ CANDEREL

5

VOIR PLUS

Préférences alimentaires

☐ Végétarien



23

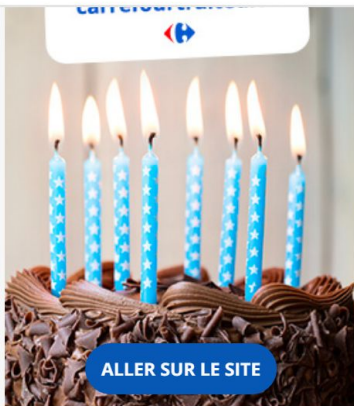
Labels qualité

☐ Bio

14

☐ Commerce équitable

5



0,74€



0,69€



0,99€



Sucre en poudre DADDY

la poche de 900 g
1.34 € / Kilogramme



Sucre Cristal ERSTEIN

le paquet de 1 kg
1.28 € / Kilogramme



Sucre roux morceaux LA

PERRUCHÉ
la boîte de 168 morceaux - 1 kg
1.78 € / Kilogramme



Sucre demi-morceaux DADDY

le paquet de 750 g
1.51 € / Kilogramme



Relevant facet values: the solution

Solutions: work on your search precision



Analytics and data science have clues: for instance, when clients type “tomato”, is there a category which regroup most of the clicks ?



All you must do is prefilter some facets (or even all the results) with this category : 80% of the result set will disappear and your filters will look good !

Examples of prefiltering at Carrefour:

- 11% of results for “tomatos” are in the “Fresh vegetables” category but they represent 86% of products added to basket
- 24% of results for “rice” are in the “Pasta and Rice” category and represent 90% of purchases
- 8% of results for “sugar” are in the “Sugar and sweeteners” category and represent 90% of purchases


// Challenge #5: search in facet values

- Issue: How to bring up facet values beyond the first top N values ?
- Solutions:
 - Pagination
 - Search in Search



Centre Pompidou **LA COLLECTION**
DU MUSÉE NATIONAL D'ART MODERNE


109962

NOM ARTISTE [7272] NATIONALITÉ [335] ŒUVRES EXPOSÉES [3] SECTEUR COLLECTION

☒ un choix ☐ plusieurs choix 

♦ NOM ARTISTE	▼ NB. TOTAL OEUVRE
MAN RAY (RADNITZKY Emmanuel, dit)	[15556]
LOTAR Eli	[11171]
BRANCUSI Constantin	[7749]
MAAR Dora (MARKOVITCH Henriette Théodora, dit)	[2140]
DUFY Raoul	[1765]
KANDINSKY Vassily	[1371]
PROUVÉ Jean	[1364]
MORAL Jean	[1241]
G.R.A.U.	[1187]
ROUAULT Georges	[1089]
STEINER André	[803]
Anonyme (sans précision)	[757]

 1/606 



CATÉGORIE ^

< Image - Son

Téléviseur

TV 8K

TV 4K UHD

TV OLED

TV QLED

Smart TV

TV LED

TV reconditionné

MARQUE ^

Rechercher

☐ Essentielb☐ Grundig☐ Hisense☐ LG☐ Lier

TAILLE DE L'ÉCRAN (DIAGON... ^

☐ 80 pouces et plus (> 203 cm)☐ 75 pouces (189 cm)☐ 65 pouces (163 cm)☐ 55 pouces (140 cm)☐ 49 pouces (123 cm)☐ 40 pouces (100 cm)☐ 32 pouces (80 cm)☐ 28 pouces (70 cm)☐ Comparer

TV LED Essentielb 65UHD-1291-SMART

Jusqu'à 10€ de remise pour l'achat d'une Chromecast*

Automnez-moi !

- 165 cm (65") - 4K UHD
- LxHxP (sans pied) 145.5 x 87.6 x 8.5 cm
- Son 2 x 10 Watts - HDMI x3 - USB x2
- **Le +** : Wifi

699,99€ - 14%

599,99€

AJOUTER AU PANIER

✓ Retrait magasin **dès mardi 29 octobre** à **Velizy 2** et à Les Clayes sous Bois "Alpha" (1) - offert [Autres magasins](#)

✓ Livraison chez vous **dès mardi 29 octobre** - offerte (2) [Détails](#)

☐ Comparer

TV LED Sony KD49XG8096 Android TV

★★★★★ (7 avis)

Automnez-moi !

- 123 cm (49") - 4K UHD
- TV connectée Android TV
- LxHxP (sans pied) 110.1 x 64.5 x 5.7 cm
- Son 2 x 10 Watts - HDMI x4 - USB x3
- **Le +** : Android TV, google assistant, design aux bords fins, passage de câb... intégré au pied

949,00€ - 26%

699,00€

AJOUTER AU PANIER

✓ Retrait magasin **dès demain** à **Velizy 2** et à Les Clayes sous Bois "Alpha" (1) - offert [Autres magasins](#)

✓ Livraison chez vous **dès mardi 29 octobre** - offerte (2) [Détails](#)

MEILLEURE VENTE!



TV OLED Sony KD55AG8 Android TV

★★★★★ (5 avis)

2490,00€ - 20%

1990,00€

Search in facet values: implementation with Elasticsearch

```
1 GET vm-dev_370_2019-07-19_17-18-artworks/_search
2 {
3   "size": 0,
4   "aggregations": {
5     "source.authors.nationality": {
6       "filter": {
7         "bool": {
31       },
32     "aggregations": {
33       "source.authors.nationality.terms": {
34         "terms": {
49       "aggregations": {
50         "details": {
51           "top_hits": {
102     }
103   }
104 },
105 "source.authors.nationality.total": {
106   "cardinality": {
107     "field": "source.authors.nationality.keyword",
108     "precision_threshold": 40000
109   }
110 }
111 }
112 }
113 }
114 }
115 }
```

Search in facet values: details of the filter aggregation

```
6  [
7  [
8  [
9  [
10 [
11 [
12 [
13 [
14 [
15 [
16 [
17 [
18 [
19 [
20 [
21 [
22 [
23 [
24 [
25 [
26 [
27 [
28 [
29 [
30 [
31 [

    "filter": {
      "bool": {
        "should": [
          {
            "multi_match": {
              "query": "fra",
              "fields": [
                "source.authors.nationality.comp^1.0"
              ],
              "type": "phrase",
              "operator": "AND",
              "slop": 2,
              "prefix_length": 0,
              "max_expansions": 50,
              "lenient": false,
              "zero_terms_query": "NONE",
              "boost": 1
            }
          }
        ],
        "disable_coord": false,
        "adjust_pure_negative": true,
        "minimum_should_match": "1",
        "boost": 1
      }
    },
```

Search in facet
values: details of
the terms sub-
aggregation

```
4      "aggregations": {  
5        "source.authors.nationality": {  
6          "filter": {  
32            "aggregations": {  
33              "source.authors.nationality.terms": {  
34                "terms": {  
35                  "field": "source.authors.nationality.keyword",  
36                  "size": 800,  
37                  "min_doc_count": 1,  
38                  "shard_min_doc_count": 0,  
39                  "show_term_doc_count_error": false,  
40                  "order": [  
41                    {  
42                      "_count": "desc"  
43                    },  
44                    {  
45                      "_term": "asc"  
46                    }  
47                  ]  
48                },  
49                "aggregations": {  
50                  "details": {  
51                    "top_hits": {  
102                  }  
103                }  
104              },  
            }  
          }  
        }  
      }
```

Search in facet values: details of the top_hits sub-aggregation and highlighting

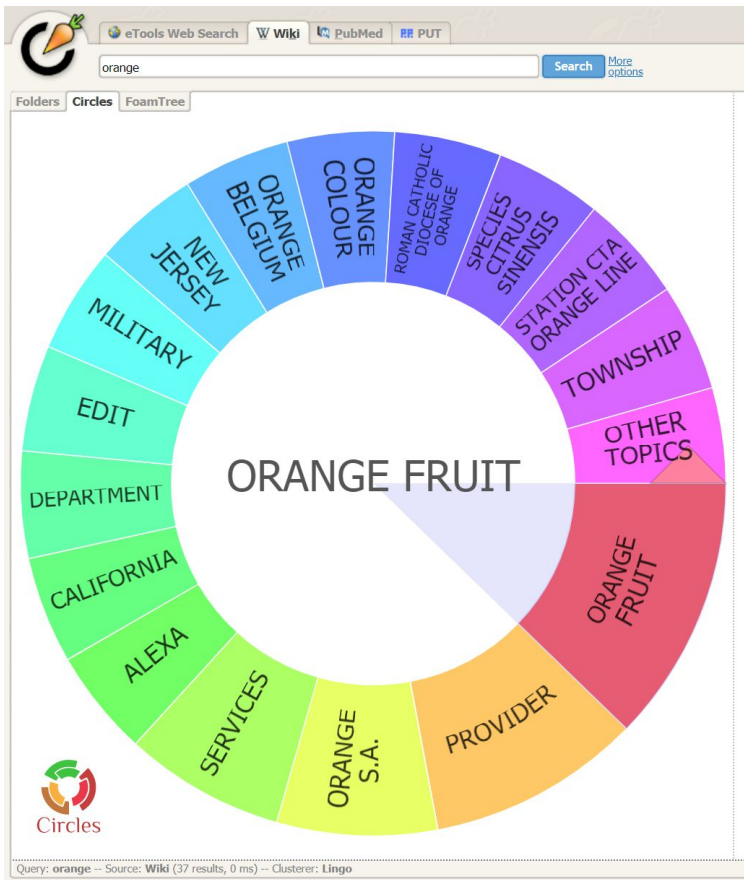
```
51 "top_hits": {  
52   "from": 0,  
53   "size": 1,  
54   "version": false,  
55   "explain": false,  
56   "_source": {  
57     "includes": [  
58       "source.authors.nationality"  
59     ],  
60     "excludes": []  
61   },  
62   "highlight": {  
63     "pre_tags": [  
64       "post_tags": [  
65         "highlight_query": {  
66           "bool": {  
67             "should": [  
68               {  
69                 "multi_match": {  
70                   "query": "fra",  
71                   "fields": [  
72                     "source.authors.nationality.comp^1.0"  
73                   ],  
74                   "type": "phrase",  
75                   "operator": "AND",  
76                   "slop": 2,  
77                   "prefix_length": 0,  
78                   "max_expansions": 50,  
79                   "lenient": false,  
80                   "zero_terms_query": "NONE",  
81                   "boost": 1  
82                 }  
83               }  
84             ],  
85             "disable_coord": false,  
86             "adjust_pure_negative": true,  
87             "minimum_should_match": "1",  
88             "boost": 1  
89           }  
90         ],  
91         "fields": {  
92           "source.authors.nationality.comp": {  
93             "number_of_fragments": 0  
94           }  
95         }  
96       }  
97     }  
98   }  
99 }  
100 }  
101 }
```

// Challenge #6: unstructured data

- Issue: the lack of structure makes difficult to suggest additional query refinements
- Solutions:
 - Clustering (like <http://project.carrot2.org/>)
 - Entity extraction (like <https://www.basistech.com/ext-analytics/rosette/entity-extractor/> or <https://twitter.com/dep4b/status/1121141764503609345>)



Display “facets” with clustering



<http://project.carrot2.org/>

Enrich the data with entity extraction

Haystack is the conference for improving **search** relevance. If you're like us, you work to understand the shiny new tools or dense academic papers out there that promise the moon. Then you puzzle how to apply those insights to your **search** problem, in your **search** stack. But the path isn't always easy, and the promised gains don't always materialize.

Haystack is the conference for organizations where search, matching, and relevance really matters to the bottom line. For **search** managers, developers, relevance engineers & data scientists finding ways to innovate, see past the silver bullets, and share what actually has worked well for their unique problems. Please come share and learn!

<https://haystackconf.com/>

Conference: Haystack

Domain: search

Facets on unstructured text after entity extraction



Categories

msf



DASHBOARDS



EN



Adelean Digger

About 24.368 results in 0.244 seconds

Display : 10 results per page



Order by : Relevance



Gallery



Map



Hide search filters

ALL

OOPS

OOPSMED

MSF WEBSITES

UNIDATA

MAP CENTRE

TREASURY

SHERLOG

NESTOR

MYHELP

ONEDRIVE

Selected Filters :

MSF Websites X

DATES OF TOPIC



PRODUCT



SOURCES



LANGUAGES



ORGANISATION



PERSON



LOCATION



☐ United States of America (12415)

☐ Africa (12402)

☐ France (12257)

☐ Switzerland (11628)

☐ Democratic Republic of the Congo (11487)

☐ United Kingdom (11472)



Microsoft Word - 090810 chagas policy doc FINAL | MSF DNDi Chagas Policy Doc 090810.pdf

Last modified on 17/12/2018

<https://www.dndi.org/wp-content/uploads/2009/10>

population: 8-15 million people • 14,000 Deaths / Year • Largest parasitic killer in the Americas **Medecins Sans Frontieres/Doctors Without Borders (MSF)** and Drugs for Neglected Dise.....



MIM MSF-DNDi SMC symposium at MIM_leaflet | MSF-DNDi _SMCsymposium_MIM.pdf

Last modified on 17/12/2018

<https://www.dndi.org/wp-content/uploads/2013/07>

Médecins Sans Frontières/ **Doctors Without Borders (MSF)** & the Drugs for Neglected 6.30 pm Place: Hall 4B Photo: Patients queue in Moissala District's Bouna health zone, where **MSF** Prof Chris Drakeley, LSHTM, UK SMC: The exper.....



Microsoft Word - 102019 joint msf dndi briefing paper format_v3.doc | msf dndi briefing paper_Feb2010.pdf

Last modified on 17/12/2018

<https://www.dndi.org/wp-content/uploads/2010/02>

Doctors Without Borders/Médecins Sans Frontières (MSF) and the Drugs for Neglected Diseases initiative **Doctors Without Borders/Médecins Sans Frontières (MSF)**.....

// Conclusions and takeaways



More data, less space → Facets are more and more important



In order to be useful → Facets should be relevant



Modern search engines have great support for facets

// Conclusions and takeaways

Marketplaces	When too many possible facets → the relevant ones should be driven by the most relevant results
Auto-completion	Use facet values as suggestions and disambiguation techniques
Assistants	When too many results → chose the facet and filter suggestions that disambiguate most as the first answer
Relevant facet values	When there is a risk of noise in the results → avoid bringing it to facet values
Search in facet values	When too many facet values → bring up those beyond the top N with search (not with JavaScript)
Unstructured data	Use clustering and entity extraction to be able to define facets

Thank You !

- Lucian Precup
- Radu Pop
- @lucianprecup
- @a2lean
- #haystackconf
- @o19s
- Berlin EU 2019



adelean
EXTRACT TRANSFORM SEARCH

HAYSTACK