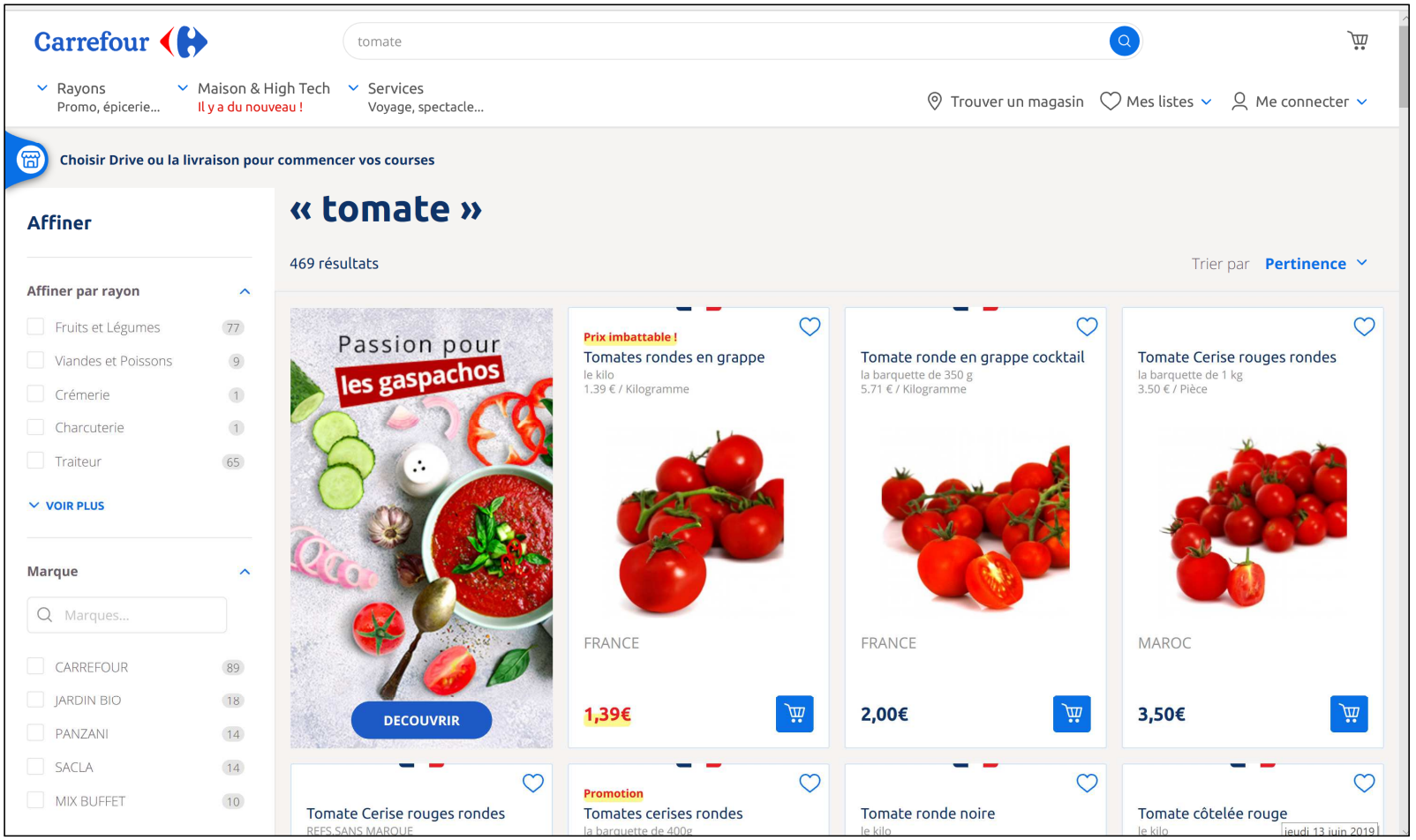
adelean
EXTRACT TRANSFORM SEARCH

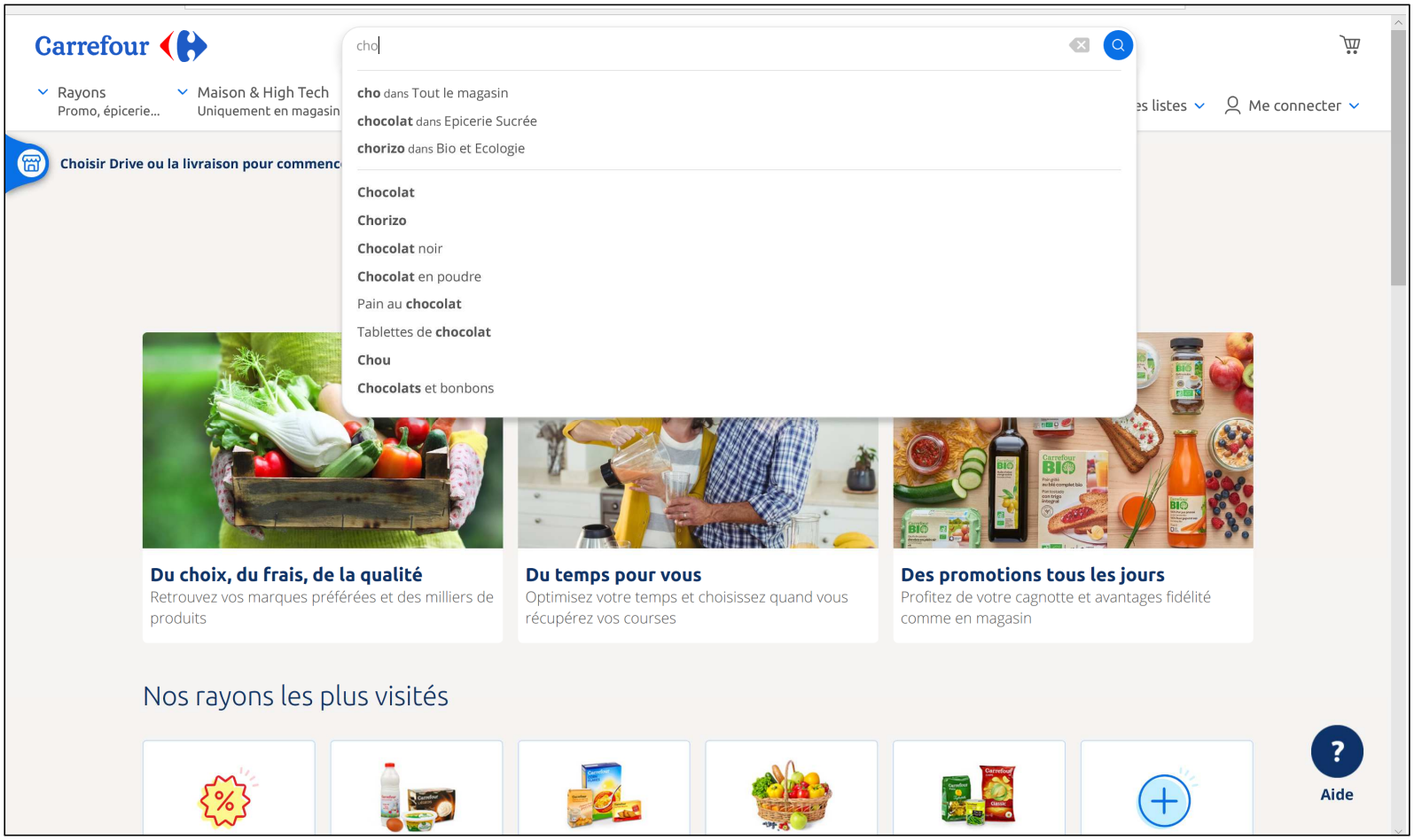
- @lucianprecup
- @margauxwagner
- @a2lean

- #bbuzz

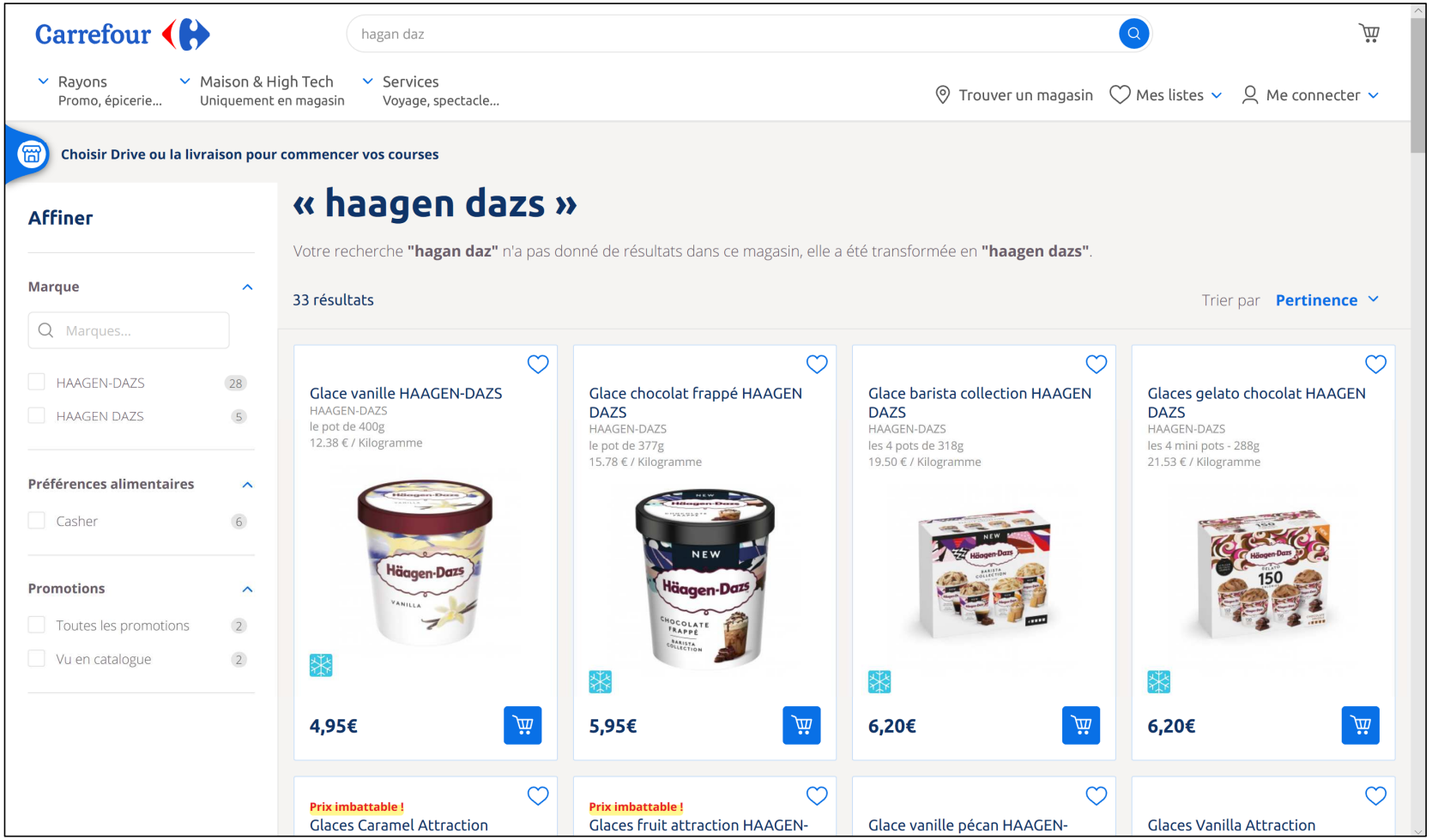
The journey of a search engine to voice search



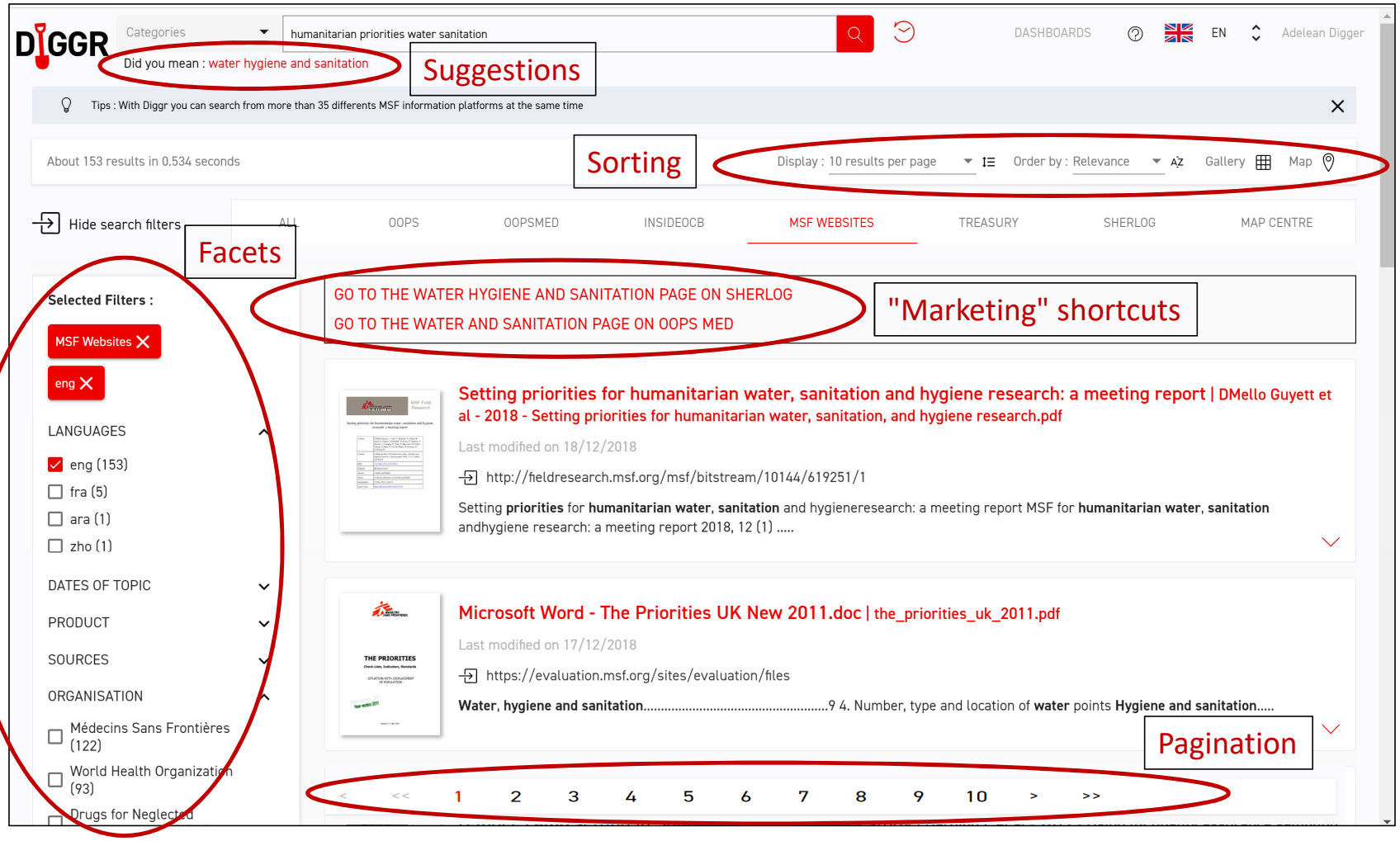
A search engine on “large screen”



A search engine on “large screen”

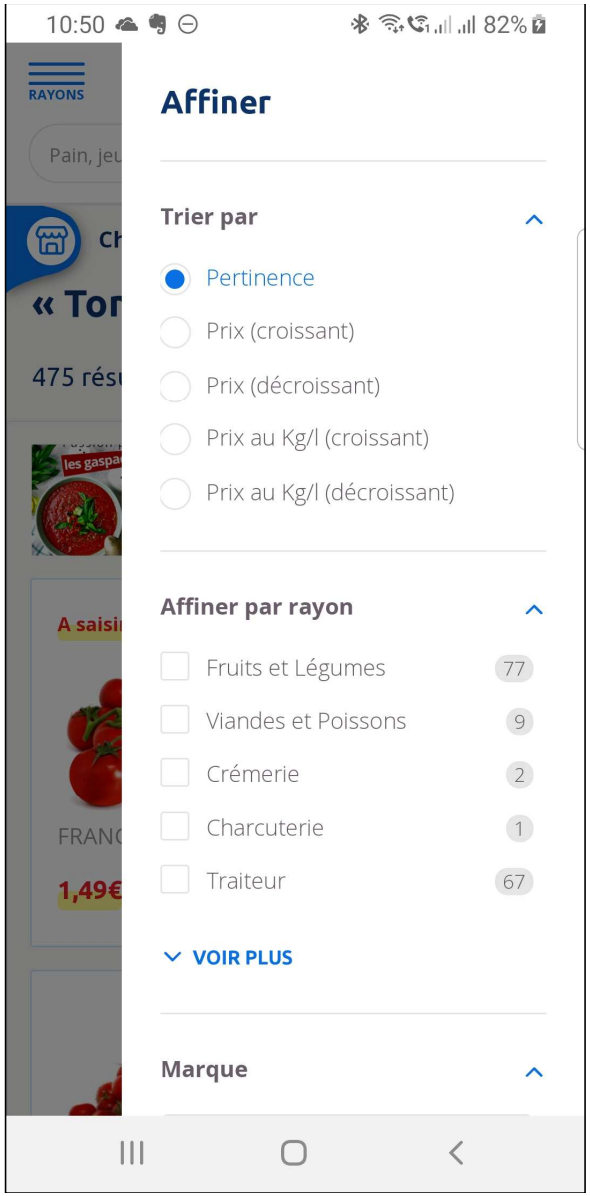


A search engine on “large screen”

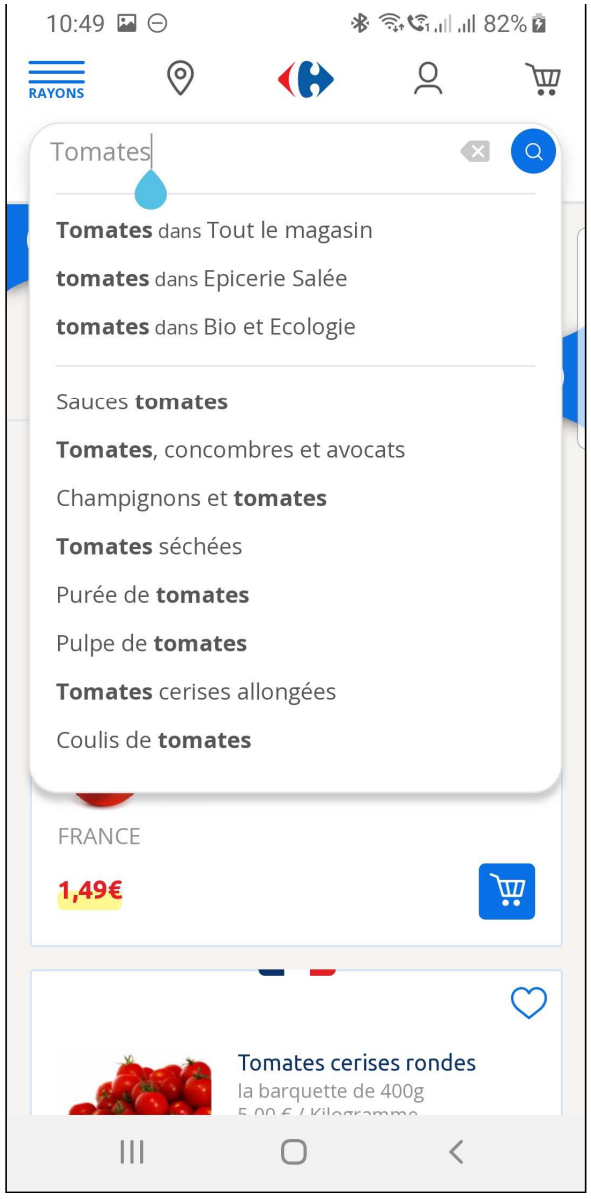


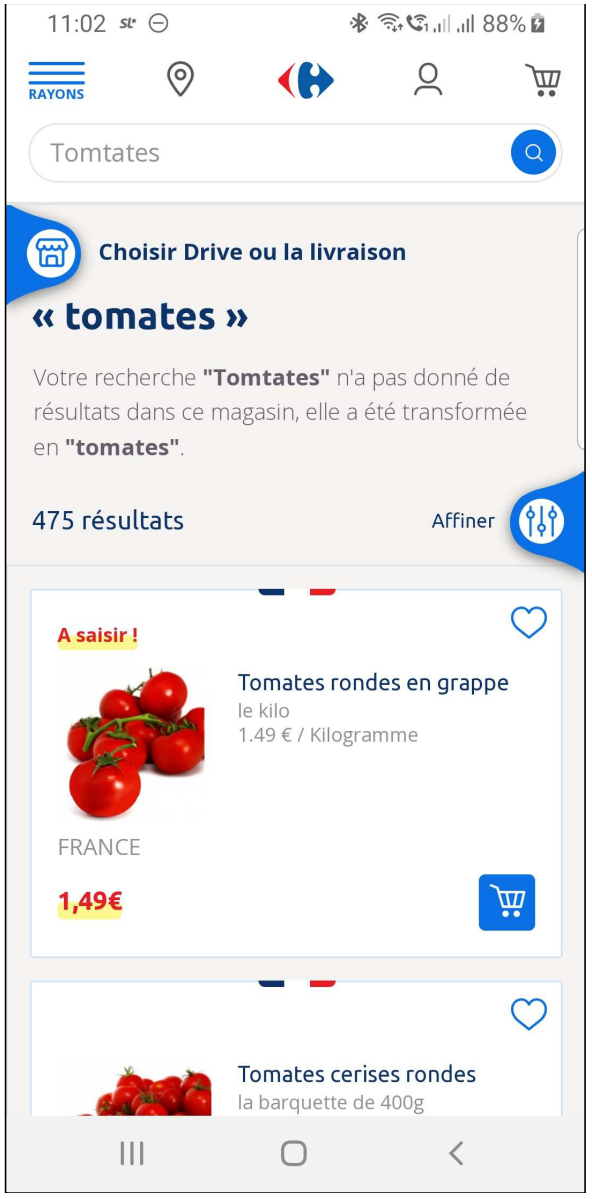
A search engine on “large screen”

Search on a mobile device

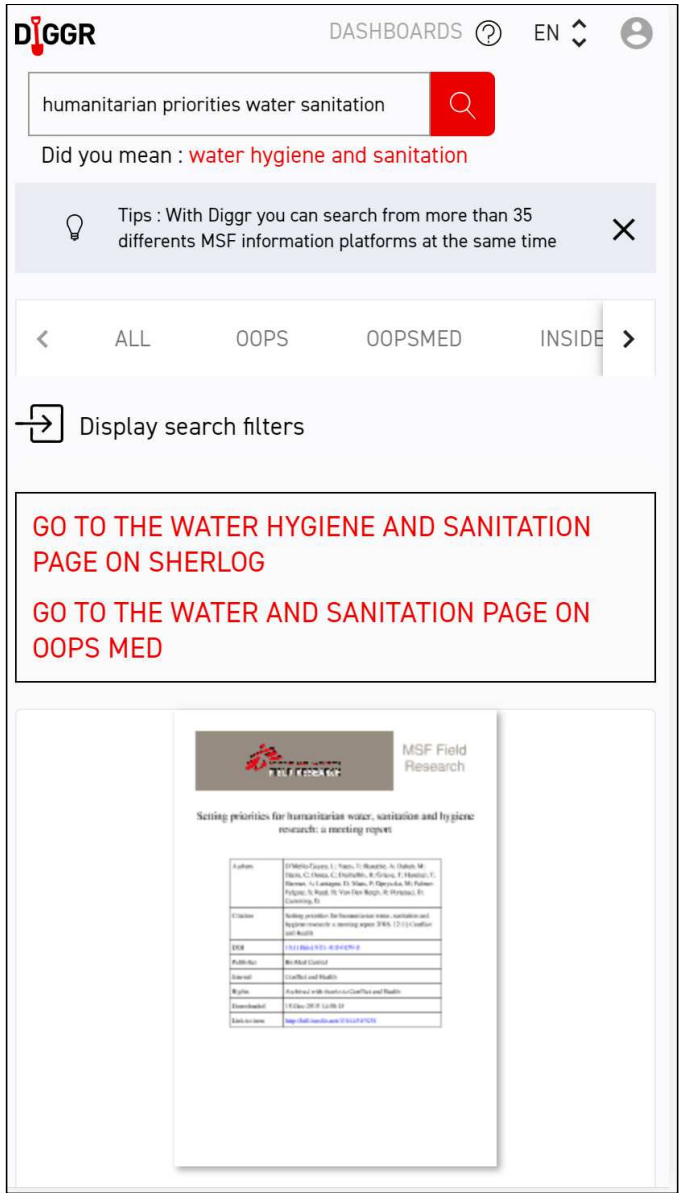


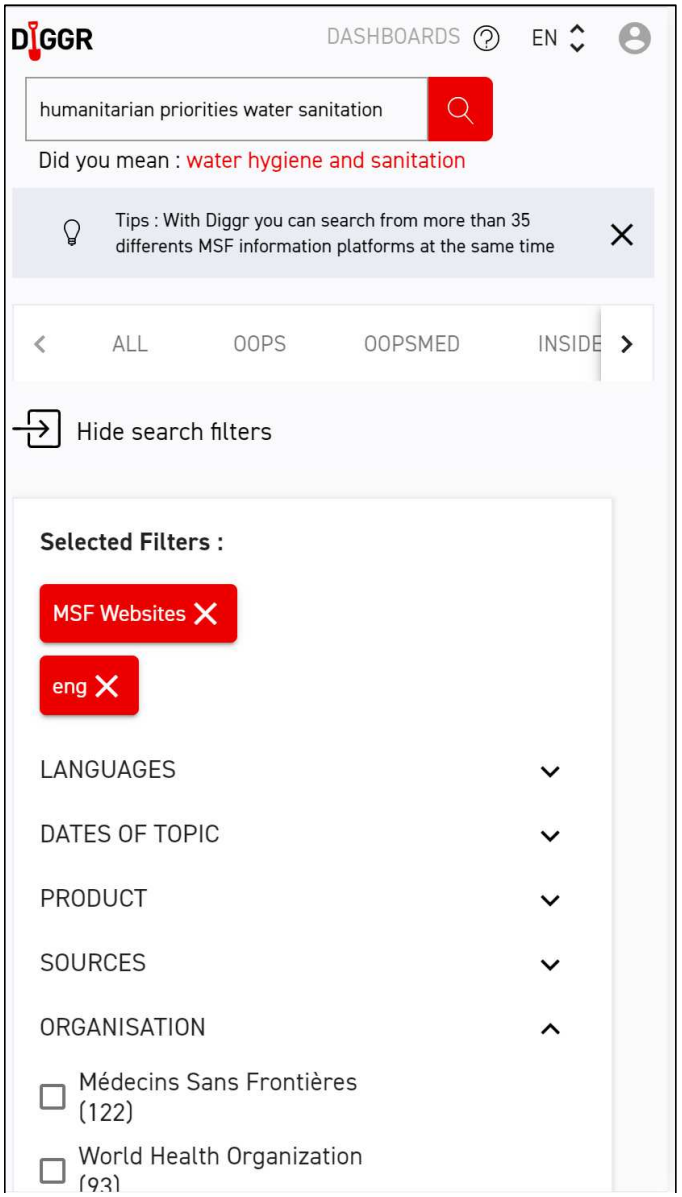
Search on a mobile device





Search on a mobile device





Issues and challenges with voice search

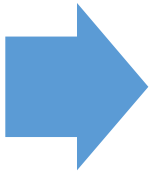
- Lack of space to display information
- Large amount of results : how to narrow?
- Unclear search: how to clarify?
- Stop words or “noise” in the query
- Filters expressed as search keywords

Description of a search API

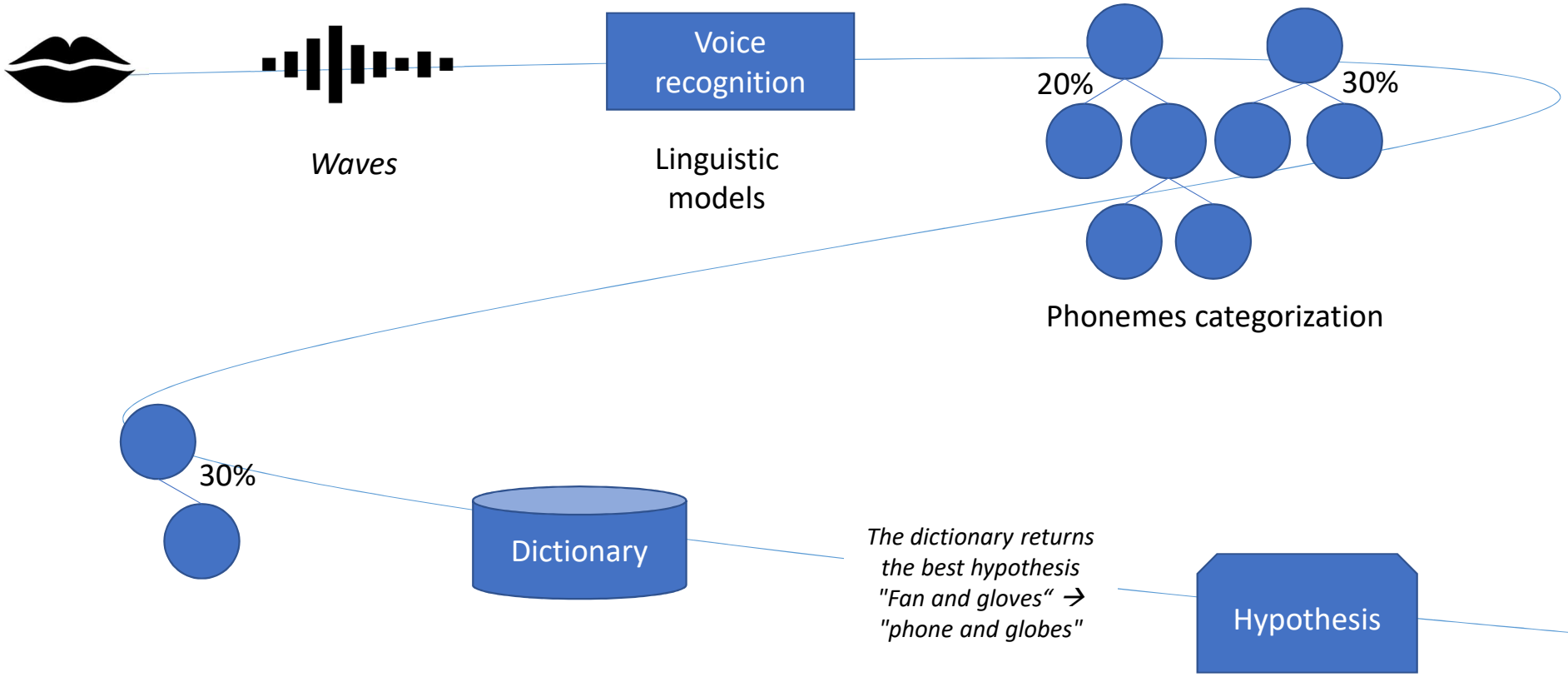
- Example of a search API:
  - Input :
    - Text
    - Filters
    - Sorting
  - Output :
    - Suggestions
    - Results
    - Marketing shortcuts
    - Facets
- Challenge :
  - We developed a classic search engine and we need to adapt it for a voice

First steps from voice to API calls

1) Speech to Text



2) Knowing the service to call - link a sentence to a service


Voice recognition
Waves
Linguistic models
20%
30%
Phonemes categorization
30%
Dictionary
The dictionary returns the best hypothesis "Fan and gloves“ → "phone and globes"
Hypothesis

INTENT / UTTERANCES/ SLOTS

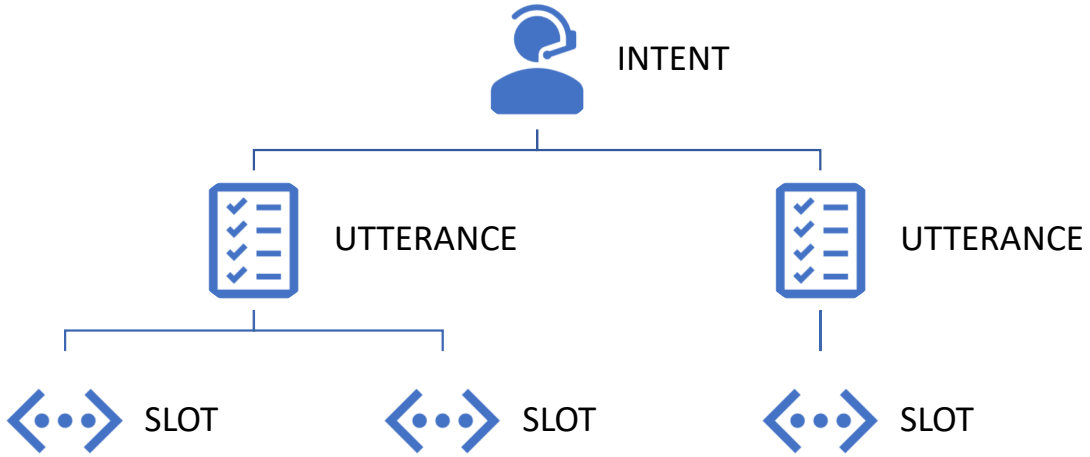
Intent = the requested service .

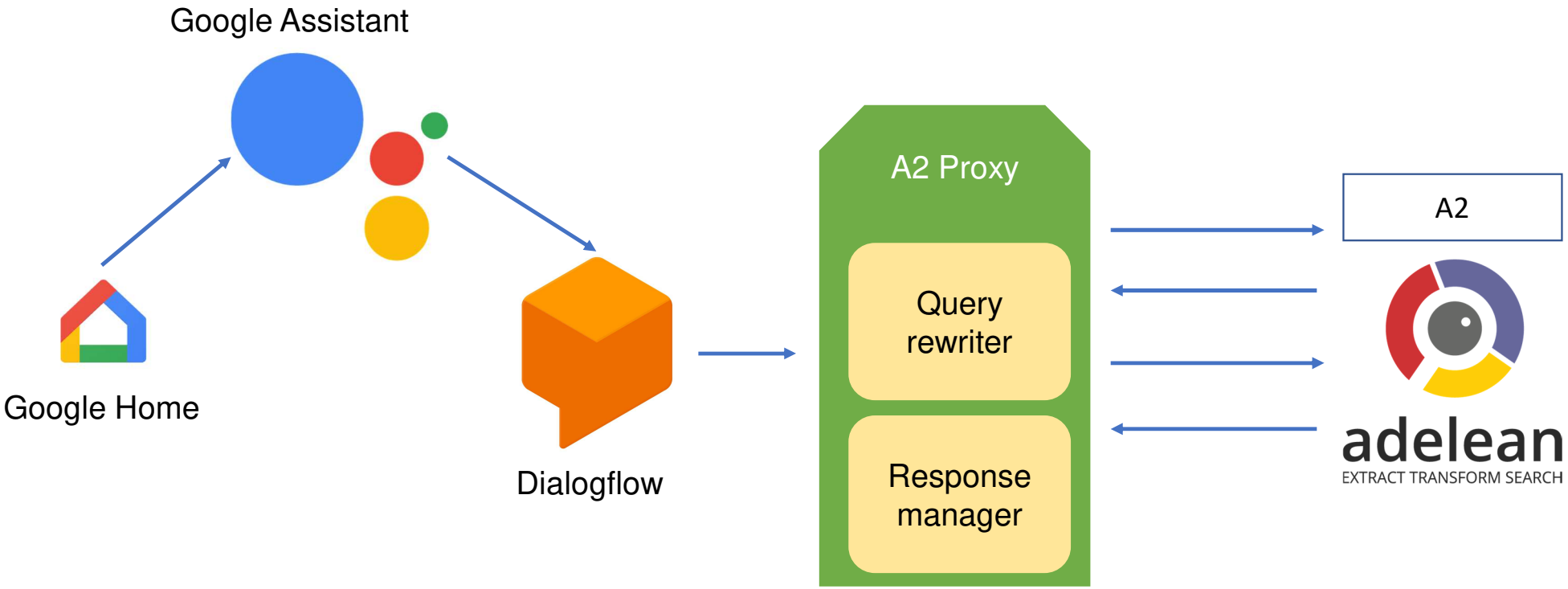
Utterances = several means of calling it, different phrases that correspond to the same intention
Each Intent has several Utterances.

Slot = represents variable information within the Intent.

"I am looking for **tomatoes**"
"I need some **chocolate** in the **frozen foods** department«

"Where is the **Kulturbrauerei** entrance ?"




Google Assistant
Google Home
Dialogflow
A2 Proxy
Query rewriter
Response manager
A2
adelean
EXTRACT TRANSFORM SEARCH

Steps from voice to search API

The user speaks to his/her **voice** assistant

The voice is transformed to **text**

The system recognize a **utterance** related to an **intent** and extract keywords from the **slots**

Now you can connect your **search engine**

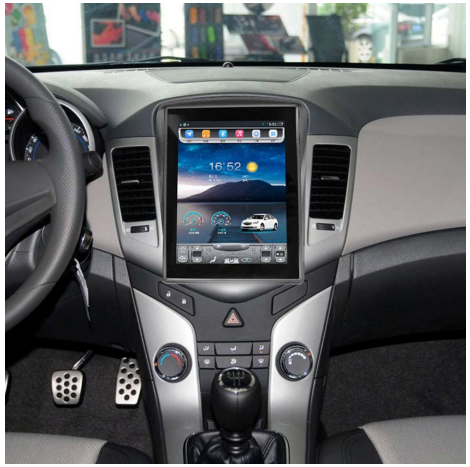
Issues and challenges with voice search

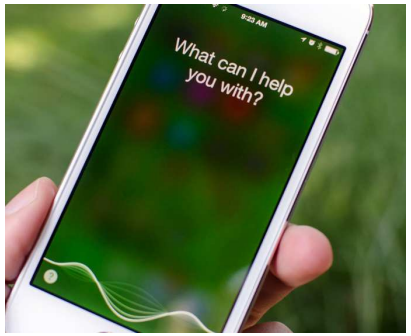
- Lack of space to display information
- Large amount of results : how to narrow?
- Unclear search: how to clarify?
- Stop words or “noise” in the query
- Filters expressed as search keywords

VOICE + SCREEN
(multimodal)

Additional Workflow
To Display less

VOICE ONLY



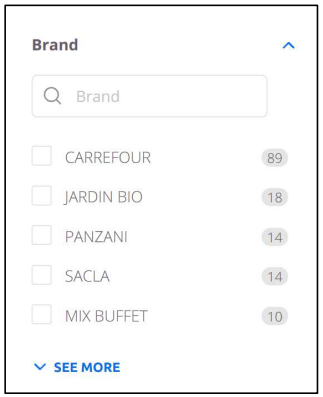


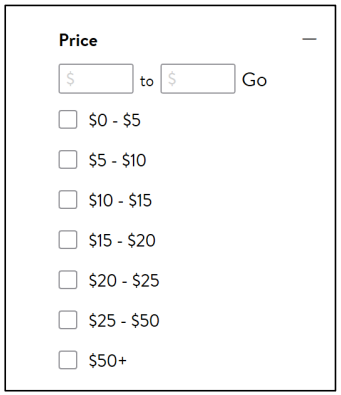
Issues and challenges with voice search

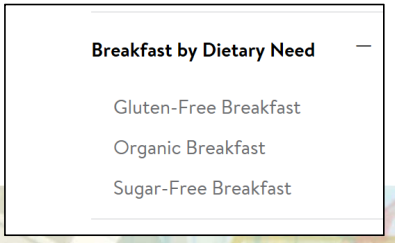
- Lack of space to display information
- Large amount of results : how to narrow?
- Unclear search: how to clarify?
- Stop words or “noise” in the query
- Filters expressed as search keywords

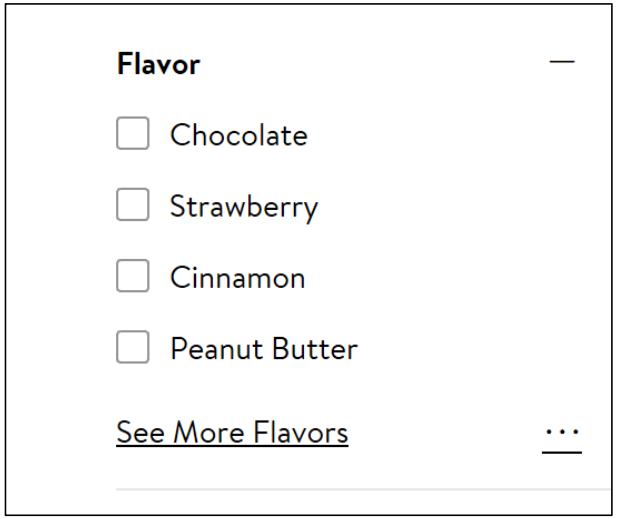
How to narrow

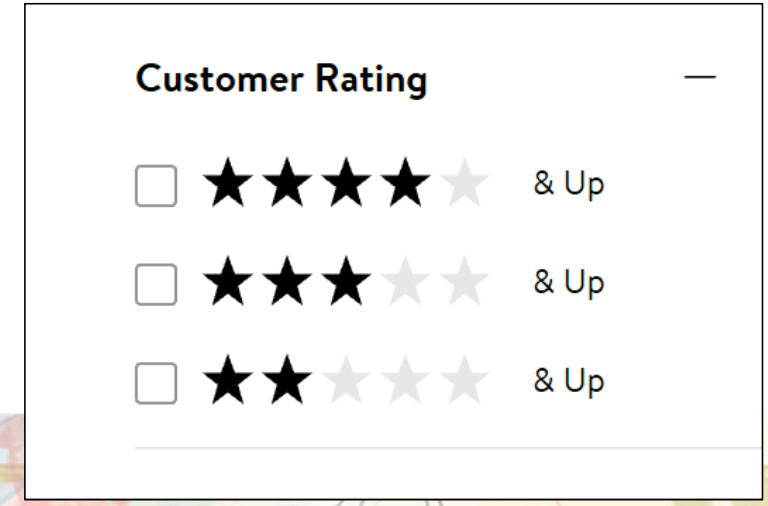
- “Quick win” solution :
  - Filters
- Issue :
  - Which filters to choose?
- Prerequisite:
  - Your search engine should already have relevant filters
- Go further:
  - Clustering
- Go even further:
  - Personalization





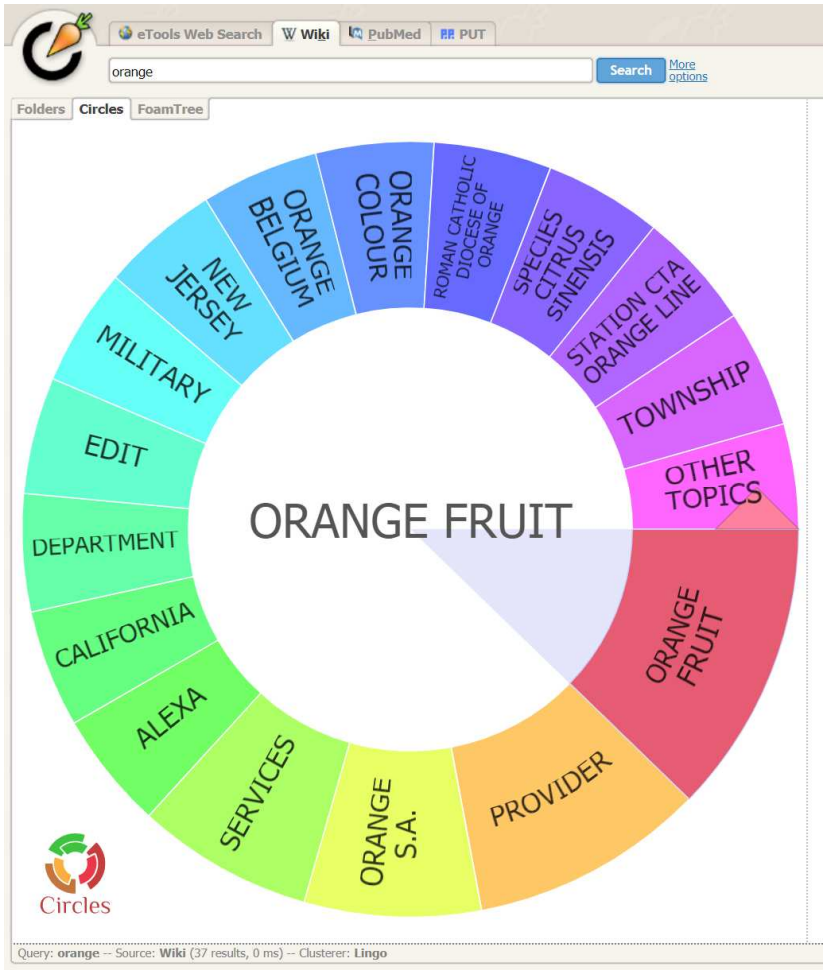


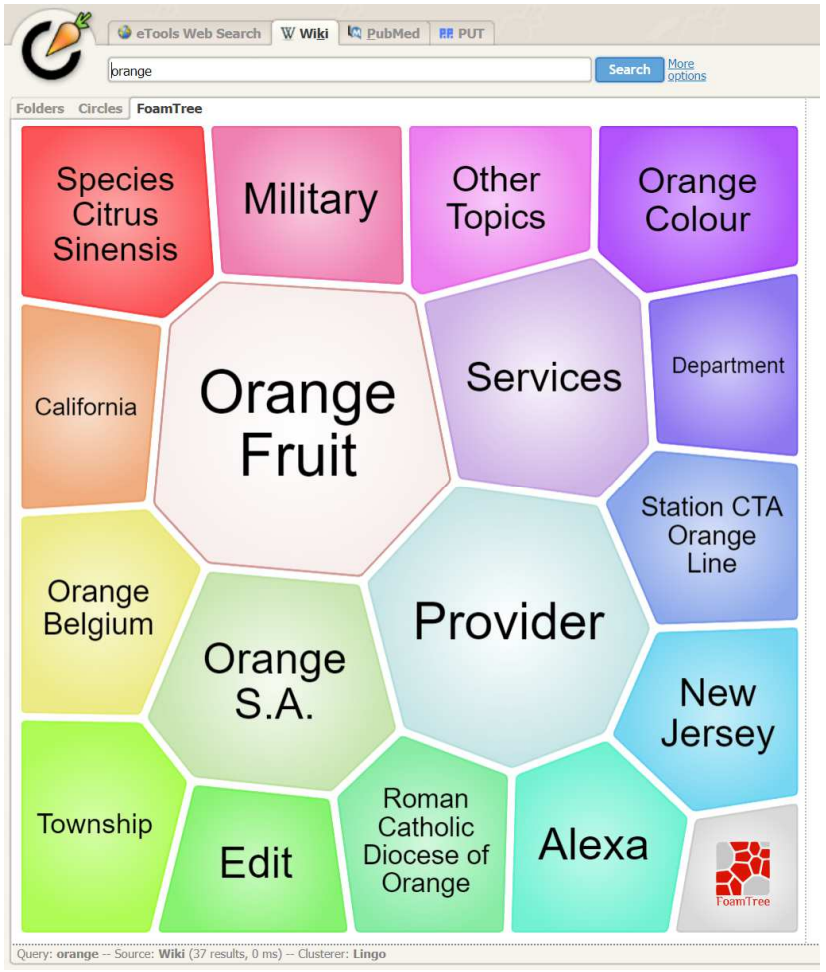




Issues and challenges with voice search

- Lack of space to display information
- Large amount of results : how to narrow?
- **Unclear search: how to clarify?**
- Stop words or “noise” in the query
- Filters expressed as search keywords


eTools Web Search
Wiki
PubMed
PUT
orange
Search
More options
Folders
Circles
FoamTree
ORANGE FRUIT
ORANGE COLOUR
ROMAN CATHOLIC DIOCESE OF ORANGE
SPECIES CITRUS SINENSIS
STATION CTA ORANGE LINE
TOWNSHIP
OTHER TOPICS
ORANGE FRUIT
PROVIDER
ORANGE S.A.
SERVICES
ALEXA
CALIFORNIA
DEPARTMENT
EDIT
MILITARY
NEW JERSEY
ORANGE BELGIUM
Circles
Query: orange -- Source: Wiki (37 results, 0 ms) -- Clusterer: Lingo


eTools Web Search
Wiki
PubMed
PUT
orange
Search
More options
Folders
Circles
FoamTree
Species Citrus Sinensis
Military
Other Topics
Orange Colour
California
Orange Fruit
Services
Department
Station CTA Orange Line
Orange Belgium
Orange S.A.
Provider
New Jersey
Township
Edit
Roman Catholic Diocese of Orange
Alexa
FoamTree
Query: orange -- Source: Wiki (37 results, 0 ms) -- Clusterer: Lingo

Issues and challenges with voice search

- Lack of space to display information
- Large amount of results : how to narrow?
- Unclear search: how to clarify?
- Stop words or “noise” in the query
- Filters expressed as search keywords

What if we get off-road?

- How to manage search queries such as :

  - I am searching for Simon Willnauer conferences on elasticsearch.
  - What is the best way to plug this specific cable?
  - How to win the FIFA Women's World Cup in France this year?
  - …

Stop words or “noise” in the query

“I am looking for **chocolate** please. Could you help me with that?”

- Issue: The query is (very) verbose
- Challenge: You have to rewrite the query
- Pitfall: We are not talking about stop words in the data ! The “Common Terms Query” will not help !
- Solution: index all the user queries and figure out the stop words

Dashboard / Editing [voice] text_en (unsaved) Save Cancel Add Options Share Reporting Auto-refresh Last 15 minutes

Search... (e.g. status:200 AND extension:PHP) Uses lucene query syntax

Add a filter +

[voice] text_en

text: Descending	Count
i	7
for	6
am	6
the	5
search	4
to	3
look	3
wai	2
thi	2
some	2

[voice] text_en standard_stopwords

standard stopwords	Count
i	7
am	6
search	4
look	3
wai	2
some	2
chocol	2
best	2
get	1
frozen	1

[voice] text_en extra_stopwords

extra stopwords	Count
wai	2
some	2
chocol	2
best	2
need	1
machin	1
kulturbrauerei	1
how	1
get	1
frozen	1

[voice] text_en shingles

shingles	Count
i am searching	3
i am looking	3
am searching for	3
am looking for	3
the best way	2
best way to	2
get to the	1
frozen foods department	1
france this year	1
for the best	1

Issues and challenges with voice search

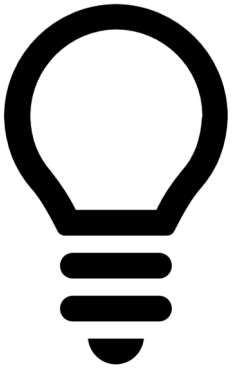
- Lack of space to display information
- Large amount of results : how to narrow?
- Unclear search: how to clarify?
- Stop words or “noise” in the query
- Filters expressed as search keywords

Filters expressed as search keywords

Filters can be expressed as a search keyword

Search for « organic tomatos » → "organic" can be a filter

Search for "Simon Willnauer's talks" on elasticsearch →
"Simon Willnauer" can be a filter



Solution 1/ Entities extraction and query rewriting with entities as a filter

Organic tomato →
text:tomato
organic_flag:true

Simon Willnauer talk on elasticsearch →
text:elasticsearch
speaker:Simon Willnauer
type:talk

Solution 2/ Filter fields are also "searchable" and the scoring algorithms will put the most relevant results first

To summarize

- Lack of space to display information → **ask questions**
- Large amount of results : how to narrow? → **add filters**
- Unclear search: how to clarify? → **Use facets, clustering, personalization**
- Stop words or “noise” in the query → **index queries and build your own list of stop words**
- Filters expressed as search keywords → **rewrite the query or make sure filters are also “searchable”**

THANK YOU FOR YOUR ATTENTION

VIELEN DANK

- @lucianprecup
- @margauxwagner
- @a2lean
- #bbuzz

adelean
EXTRACT TRANSFORM SEARCH