Text mining tools for extracting information about microbial biodiversity in food

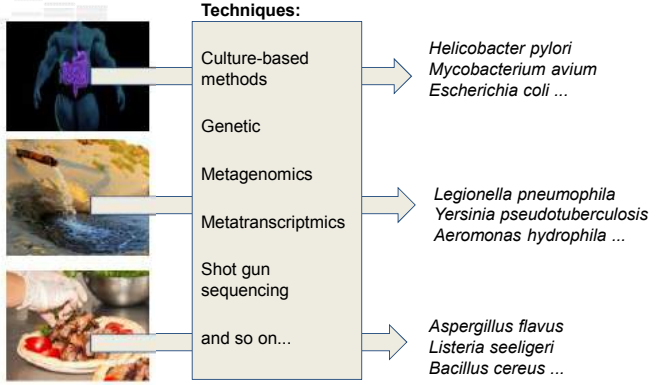
Estelle Chaix, Louise Deléger, Robert Bossy and Claire Nédellec

firstname.lastname@inra.fr

Bibliome Team
MaIAGE, INRA,
Université Paris-Saclay,
78350 Jouy-en-Josas,
France

Microbial ecosystems

Which microbes live in an environment?



Properties of environment? Microbial Interaction?

Crossing between microbial species and habitats

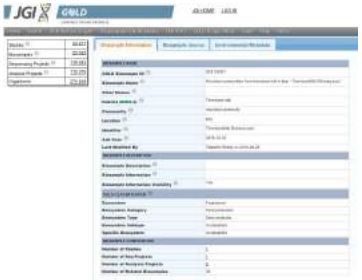
Difficulty : Highly variable forms in text or genomics database (GOLD, SRA, GenBank)


RESEARCH LETTER
Biogenic amine-forming microbial communities in cheese

"Bacteria of the genera *Enterococcus* and *Lactobacillus* and coliform bacteria were isolated from Dutch-type semi-hard cheese"


High incidence of *Listeria monocytogenes* in European red smear cheese

"Out of European red-smear cheese samples of various types [...] 1.2% of the samples were contaminated with *L. seeligeri*"



e.g.
- Artisanal cheeses from Tucuman
- Dairy cheese
- Caciocavallo cheese in Italy

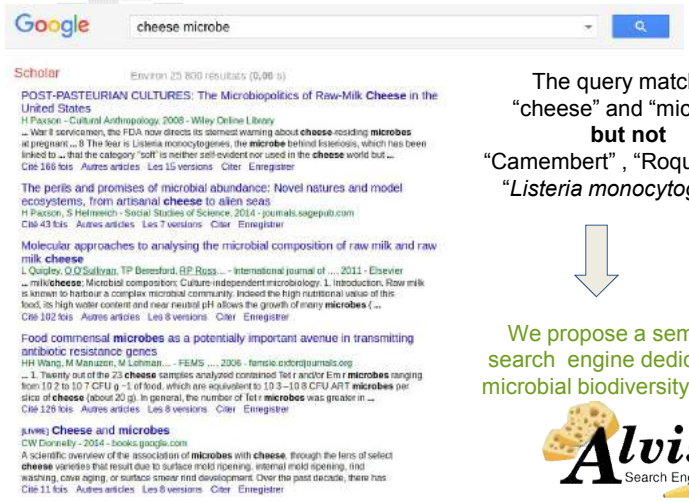
Habitat information is neither queryable nor comparable

Described at different levels of accuracy and not standardized

What is the cheese microflora?

"Geotrichum candidum strains isolated from a traditional Spanish goats' milk cheese."
"Escherichia coli O157:H7 isolated from raw beef, soft cheese and vegetables in Lima"
"Microbial ecology of Gorgonzola rinds and occurrence of different biotypes of *Listeria monocytogenes*."

Classic search engine query



The query matches
“cheese” and “microbe”
but not
“Camembert” , “Roquefort” or
“*Listeria monocytogenes*”

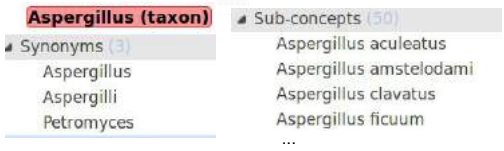
We propose a semantic search engine dedicated to microbial biodiversity in food.

Alvis Search Engine

Semantic search engine of microbial habitat in food

Interpretation of the query

Aspergillus :



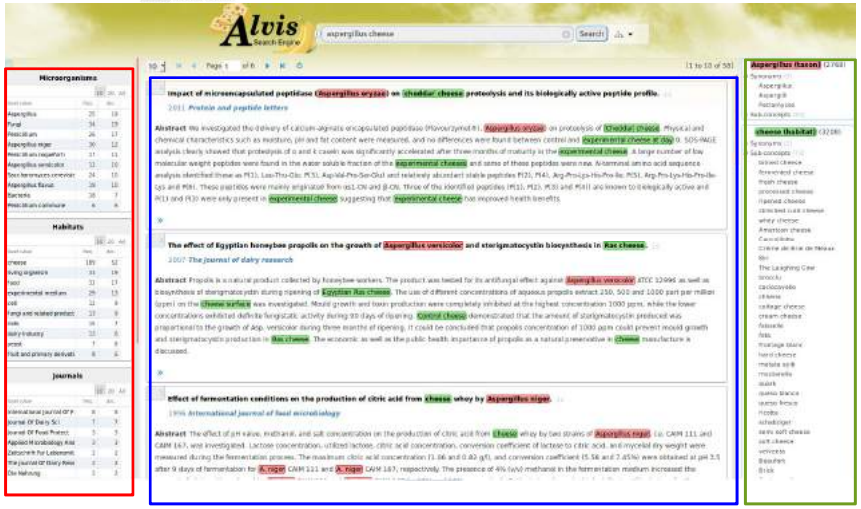
Cheese :

American cheese
Cancoillotte
Crème de Brie de Meaux
Kiri
The Laughing Cow
brocciu
caciocavallo
chhena
cottage cheese
cream cheese

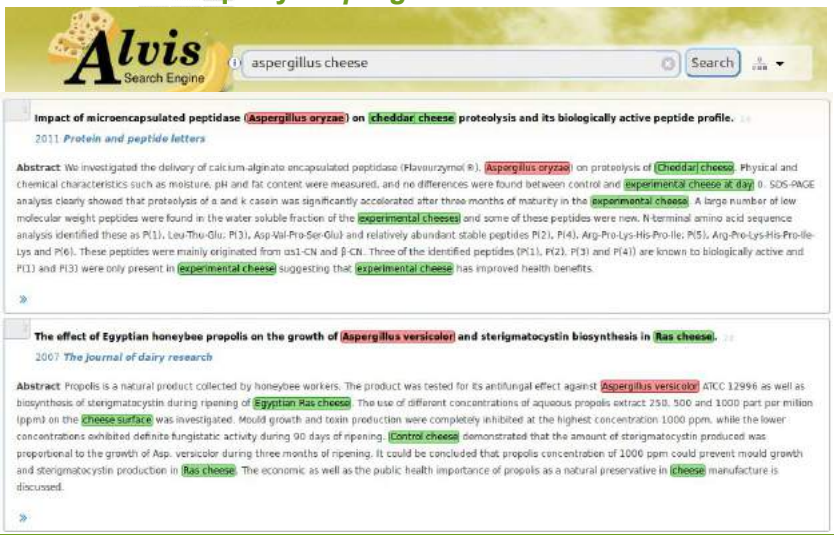
http://bibliome.jouy.inra.fr/demo/food/alvisir/webapi/search
Mini-link: https://frama.link/AlvisFood

Has *Aspergillus* been isolated in cheese?

Results of the query: *aspergillus* cheese



Result of the query: *aspergillus* cheese



Does *Aspergillus* lives in cheese ?

Text mining tools for extracting information about microbial biodiversity in food
Estelle Chaix, Louise Deléger, Robert Bossy and Claire Nédellec
Microbial spoilers in food 2017 - 28th - 30th June - Quimper

Result of the query: aspergillus ~livesin cheese

Text mining tools for extracting information about microbial biodiversity in food
Estelle Chaix, Louise Deléger, Robert Bossy and Claire Nédellec
Microbial spoilers in food 2017 - 28th - 30th June - Quimper

Behind the AlvisFood Search Engine

- Our approach is to extract from text
  - →"Microbe" and "Habitat" concepts
  - →Links between them
- We use
  - →AlvisNLP: Methods and tools for automatic extraction and analysis of biological text (*i.e.* Text Mining and Natural Language Processing)
  - →Machine learning methods trained with examples from microbiological and food domain experts
  - →Internal and external resources
- AlvisFood Search Engine: > 100,000 references from PubMed
  - →Selected by MeSH terms

Text mining tools for extracting information about microbial biodiversity in food
Estelle Chaix, Louise Deléger, Robert Bossy and Claire Nédellec
Microbial spoilers in food 2017 - 28th - 30th June - Quimper

Microbial entity detection

NCBI taxonomy

- Fungi (fungi) *Click on organism name to get more information.*
  - Blastocladiomycota
    - Blastocladiomycetes
      - Blastocladiales
      - Blastocladiales incertae sedis
    - environmental samples
      - uncultured Blastocladiomycota
  - Chytridiomycota
    - Chytridiomycetes
      - Chytridiales
      - Cladochytriales
      - Gromochytriales
      - Lobulomycetales
      - Mesochytriales
      - Polychytriales
      - Rhizophlyctidales
      - Rhizophydiales
      - Spizellomycetales
      - unclassified Chytridiomycetes
      - Chytridiomycetes incertae sedis
      - environmental samples
    - Monoblepharidomycetes
      - Monoblepharidales
      - unclassified Monoblepharidomycetes

Habitat entity detection

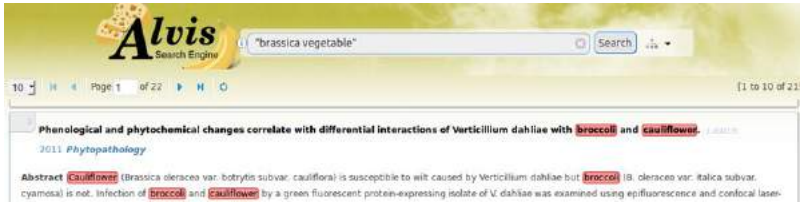
➢Detection in text of nominal or adjectival groups

➢Categorization of these groups with the *Ontobiotope ontology*
→Formal and structured representation of microbial habitats
→Partially reused in AlvisFoodSE

Food sub-categories of Ontobiotope ontology

➢From the EFSA classification
➢Enrichment by microbial and food domains experts

➢Formal indication that “Roquefort” is a “Cheese”
→allows semantic search
➢Our automatic AlvisNLP tools link groups of words from the text to an Ontobiotope category
→achieve normalisation



Relationship between Microbe and Habitat

➢Extraction of ~livesin relationship
➢Hard problems in automatic language processing and artificial intelligence
➢Achieved by machine learning methods trained with annotated examples

What are the taxa living in food?
A query : {taxon}* ~livesin food

Results downloadable as table with occurrence counts.
Displayed as facets

To conclude

http://bibliome.jouy.inra.fr/demo/food/alvisir/webapi/search
Mini-link: https://frama.link/AlvisFood

➢Our tools are pioneers in the field of text-mining for microbial biodiversity

➢Bibliome is a research team so:
→If you use AlvisFoodSE for your research, please cite us
→If you see an error, please send us an email, this will help us to improve our tools

Thank you for your attention

Text mining tools for extracting information about microbial biodiversity in food

Estelle Chaix, Louise Deléger, Robert Bossy and Claire Nédellec

firstname.lastname@inra.fr

Bibliome Team
MaIAGE, INRA,
Université Paris-Saclay,
78350 Jouy-en-Josas, France

http://bibliome.jouy.inra.fr/demo/food/alvisir/webapi/search
Mini-link: https://frama.link/AlvisFood

On going work

➢Ambiguous cases for automatic tools

"Byssochlamys fulva and *Neosartorya fischeri* are heat-resistant fungi which are a concern to food industries"

➢Automatic detection of microbial phenotypes

i.e. halophile, thermophile, phototroph ...

Acknowledgments

INRA Ontobiotope and Florilège working groups
Food Microbiome project

Text mining tools for extracting information about microbial biodiversity in food
Estelle Chaix, Louise Deléger, Robert Bossy and Claire Nédellec
Microbial spoilers in food 2017 - 28th - 30th June - Quimper