

open MIN 7ED

INRA  
SCIENCE & IMPACT

MaiAGE

## Text mining tools for extracting information about microbial biodiversity in food

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

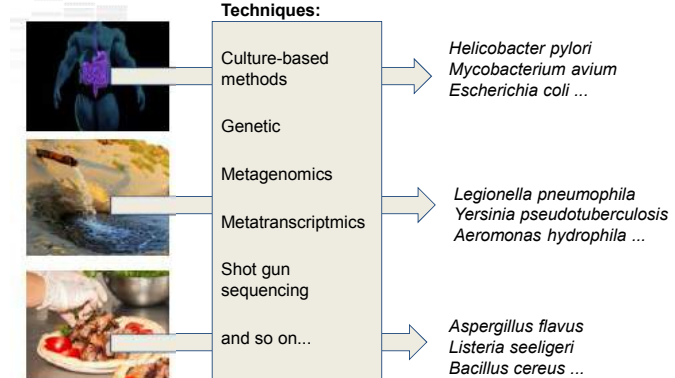
Estelle Chaix, Louise Deléger, Robert Bossy  
and Claire Nédellec

*firstname.lastname@inra.fr*

CC BY

## Microbial ecosystems

Which microbes live in an environment?



Properties of environment? Microbial Interaction?

INRA  
SCIENCE & IMPACT

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 - 28<sup>th</sup> - 30<sup>th</sup> June - Quimper

CC BY

## 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**  
Radha Sundharam & Thomas Corns  
Department of Food Technology, Imperial College of Science, Technology and Innovation, London, United Kingdom

"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

Project	Accession	Accession	Accession
Project 1	Accession 1	Accession 2	Accession 3
Project 2	Accession 4	Accession 5	Accession 6
Project 3	Accession 7	Accession 8	Accession 9
Project 4	Accession 10	Accession 11	Accession 12
Project 5	Accession 13	Accession 14	Accession 15
Project 6	Accession 16	Accession 17	Accession 18
Project 7	Accession 19	Accession 20	Accession 21
Project 8	Accession 22	Accession 23	Accession 24
Project 9	Accession 25	Accession 26	Accession 27
Project 10	Accession 28	Accession 29	Accession 30

## 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*."

INRA  
SCIENCE & IMPACT

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 - 28<sup>th</sup> - 30<sup>th</sup> June - Quimper

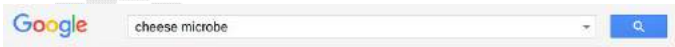
CC BY

INRA  
SCIENCE & IMPACT

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 - 28<sup>th</sup> - 30<sup>th</sup> June - Quimper

CC BY

## Classic search engine query



**Scholar** Environ 25 800 résultats (0,96 s)

**POST-PASTEURIAN CULTURES: The Microbiopolitics of Raw-Milk Cheese in the United States**  
H Paxson - *Cultural Anthropology*, 2008 - Wiley Online Library  
... View 4 services, the FDA now directs its sternest warnings about **cheese** residing **microbes** at pregnant ... 8 This fear is *Listeria monocytogenes*, the **microbe** behind listeriosis, which has been linked to ... that the category "soft" is neither self-evident nor used in the **cheese** world but ...  
Cite 166 fois - Autres articles - Les 15 versions - Citer - Enregistrer

**The perils and promises of microbial abundance: Novel natures and model ecosystems, from artisanal cheese to alien seas**  
H Pearson, S Helmreich - *Social Studies of Science*, 2014 - journals.sagepub.com  
Cite 43 fois - Autres articles - Les 7 versions - Citer - Enregistrer

**Molecular approaches to analysing the microbial composition of raw milk and raw milk cheese**  
L Quirley, O O'Sullivan, TP Beresford, BP Ross ... - *International journal of ...*, 2011 - Elsevier  
... milk**cheese**: Microbial composition; Culture-independent microbiology; 1. Introduction. Raw milk is known to harbour a complex microbial community. Indeed the high nutritional value of this food, its high water content and near neutral pH allows the growth of many **microbes** (...  
Cite 102 fois - Autres articles - Les 8 versions - Citer - Enregistrer

**Food commensal microbes as a potentially important avenue in transmitting antibiotic resistance genes**  
HH Wang, M Manuazz, M Lehman ... - *FEMS* ... 2006 - femsic.oxfordjournals.org  
... 1. Twenty out of the 23 **cheese** samples analyzed contained *Tet r* and/or *Em r* **microbes** ranging from 10.2 to 10.7 CFU/g ± 1 of food, which are equivalent to 10.3–10.9 CFU/ART **microbes** per slice of **cheese** (about 20 g). In general, the number of *Tet r* **microbes** was greater in ...  
Cite 126 fois - Autres articles - Les 8 versions - Citer - Enregistrer

**juvies Cheese and microbes**  
CW Donnelly - 2014 - books.google.com  
A scientific overview of the association of **microbes** with **cheese**, through the lens of select **cheese** varieties that result due to surface mold ripening, internal mold ripening, and washing, cave aging, or surface smear rind development. Over the past decade, there has  
Cite 11 fois - Autres articles - Les 8 versions - Citer - Enregistrer

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.



## Semantic search engine of microbial habitat in food

### Interpretation of the query

**Aspergillus :**

**Aspergillus (taxon)**  
Synonyms (3)  
Aspergillus  
Aspergilli  
Petrymyces

Sub-concepts (50)  
Aspergillus aculeatus  
Aspergillus amstelodami  
Aspergillus clavatus  
Aspergillus ficuum  
...

**Cheese :**

American cheese  
Cancailotte  
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>

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



## Has *Aspergillus* been isolated in cheese?



## Results of the query: *aspergillus* cheese

**Microorganisms**

Microorganism	Count
Aspergillus	25
Penicillium	14
Trichoderma	17
Aspergillus niger	30
Aspergillus fumigatus	11
Aspergillus oryzae	10
Aspergillus nidulans	10
Aspergillus terreus	10
Aspergillus nidulans	10
Aspergillus nidulans	10

**Habitats**

Habitat	Count
Aspergillus	10
Aspergillus	10
Aspergillus	10
Aspergillus	10
Aspergillus	10
Aspergillus	10
Aspergillus	10
Aspergillus	10
Aspergillus	10
Aspergillus	10

**Journals**

Journal	Count
International journal of food microbiology	1
International journal of food microbiology	1
International journal of food microbiology	1
International journal of food microbiology	1
International journal of food microbiology	1
International journal of food microbiology	1
International journal of food microbiology	1
International journal of food microbiology	1
International journal of food microbiology	1
International journal of food microbiology	1

**Impact of microencapsulated peptide (*Aspergillus niger*) on *Staphylococcus aureus* proteolysis and its biologically active peptide profile.**  
2011: *Protein and peptide letters*

**Abstract:** We investigated the delivery of calcium alginate encapsulated peptides (Pheocystinyl-His-Arg-Asp-Val) on proteolysis of *Staphylococcus aureus* physical and chemical characteristics such as moisture, pH and fat content were measured, and no differences were found between control and *Aspergillus niger* (ATCC 9642) or *Aspergillus niger* (ATCC 9642) analysis. Results showed that proteolysis of *S. aureus* was significantly accelerated after three months of maturity in the *Aspergillus niger* medium. A large number of low molecular weight peptides were found in the agar media as a result of the *Aspergillus niger* proteolysis, and some of these peptides were new. Amino acid analysis identified these as P13, Lys-Thr-Glu, P13, Arg-Thr-Pro-Ser-Glu and relatively abundant stable peptides P13, P14, Arg-Thr-Pro-Ser-Glu, P13, Arg-Thr-Pro-Ser-Glu, P13, Arg-Thr-Pro-Ser-Glu and P14. These peptides were mainly composed from Lys, Thr and Glu. Those of the identified peptides P13, P14, P13 and P14 are known to be biologically active and P13 and P14 were only present in *Aspergillus niger* suggesting that *Aspergillus niger* has expressed health benefits.

**The effect of Egyptian honey bee propolis on the growth of *Aspergillus niger* and *Stenotrophomonas maltophilia* in *Penicillium*.**  
2010: *The Journal of dairy research*

**Abstract:** Propolis is a natural product collected by honey bees. The product was tested for its antifungal effect against *Aspergillus niger* ATCC 10942 as well as its inhibitory effect on *Stenotrophomonas maltophilia* during ripening of *Aspergillus niger*. The use of different concentrations of aqueous propolis extract (200, 500 and 1000 ppm per ml) against the *Aspergillus niger* was investigated. Mold growth and toxin production were completely inhibited at the highest concentration (1000 ppm), while the lower concentrations exhibited definite fungistatic activity during 30 days of ripening. *Aspergillus niger* demonstrated that the amount of *Stenotrophomonas maltophilia* produced was proportional to the growth of *Aspergillus niger* during three months of ripening. It could be concluded that propolis concentration of 1000 ppm could prevent mold growth and *Stenotrophomonas maltophilia* production. *Aspergillus niger*. The economic as well as the public health importance of propolis as a natural preservative in *Aspergillus niger* is discussed.

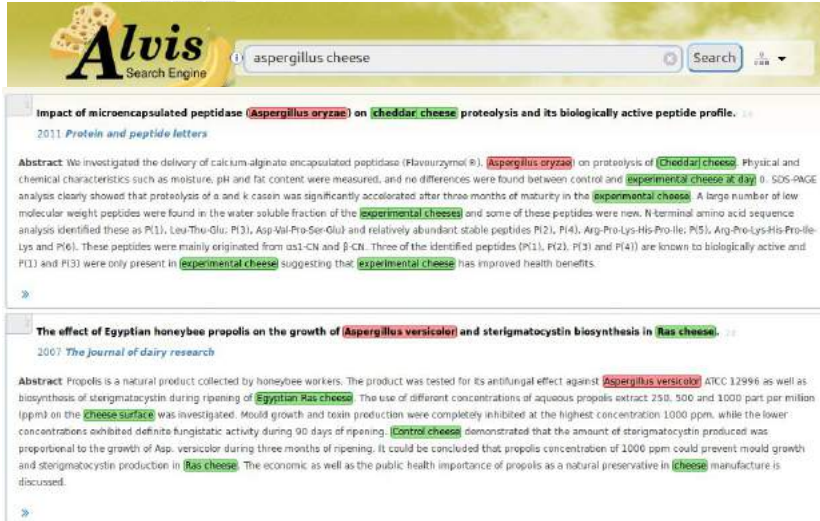
**Effect of fermentation conditions on the production of citric acid from cheese whey by *Aspergillus niger*.**  
1990: *International journal of food microbiology*

**Abstract:** The effect of pH value, substrate, and salt concentration on the production of citric acid from cheese whey by two strains of *Aspergillus niger* (i.e. CERN 111 and CERN 143) was investigated. Lactose concentration, citric acid concentration, conversion coefficient of lactose to citric acid, and microbial dry weight were measured during the fermentation process. The maximum citric acid concentration (1.36 and 0.80 g/L) and conversion coefficient (15.56 and 7.45%) were obtained at pH 3.5 after 9 days of fermentation by CERN 111 and CERN 143, respectively. The presence of 4% (v/v) methanol in the fermentation medium increased the

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



## Result of the query: *aspergillus* cheese



**Impact of microencapsulated peptidase (*Aspergillus oryzae*) on cheddar cheese proteolysis and its biologically active peptide profile.** 2011 *Protein and peptide letters*

**Abstract:** We investigated the delivery of calcium alginate encapsulated peptidase (Havazymol®), *Aspergillus oryzae* on proteolysis of *Cheddar cheese*. Physical and chemical characteristics such as moisture, pH and fat content were measured, and no differences were found between control and *experimental cheese at day 0*. SDS-PAGE analysis clearly showed that proteolysis of  $\alpha$  and  $\kappa$  casein was significantly accelerated after three months of maturity in the *experimental cheese*. A large number of low molecular weight peptides were found in the water soluble fraction of the *experimental cheese* and some of these peptides were new. N-terminal amino acid sequence analysis identified these as P(1), Leu-Thr-Glu; P(3), Asp-Val-Pro-Ser-Glu and relatively abundant stable peptides P(2), Arg-Pro-Lys-His-Pro-Ile; P(5), Arg-Pro-Lys-His-Pro-Ile-Lys and P(6). These peptides were mainly originated from  $\alpha$ s1-CN and  $\beta$ -CN. Three of the identified peptides (P(1), P(2), P(3) and P(4)) are known to be biologically active and P(1) and P(3) were only present in *experimental cheese* suggesting that *experimental cheese* has improved health benefits.

**The effect of Egyptian honeybee propolis on the growth of (*Aspergillus versicolor*) and sterigmatocystin biosynthesis in *Ras cheese*.** 2007 *The Journal of dairy research*

**Abstract:** Propolis is a natural product collected by honeybee workers. The product was tested for its antifungal effect against *Aspergillus versicolor* ATCC 12996 as well as biosynthesis of sterigmatocystin during ripening of *Egyptian Ras cheese*. The use of different concentrations of aqueous propolis extract 250, 500 and 1000 part per million (ppm) on the *cheese surface* was investigated. Mould growth and toxin production were completely inhibited at the highest concentration 1000 ppm, while the lower concentrations exhibited definite fungistatic activity during 90 days of ripening. *Control cheese* demonstrated that the amount of sterigmatocystin produced was proportional to the growth of *Asp. versicolor* during three months of ripening. It could be concluded that propolis concentration of 1000 ppm could prevent mould growth and sterigmatocystin production in *Ras cheese*. The economic as well as the public health importance of propolis as a natural preservative in *cheese* manufacture is discussed.

## Does *Aspergillus* lives in cheese ?



**The effect of Egyptian honeybee propolis on the growth of (*Aspergillus versicolor*) and sterigmatocystin biosynthesis in *Ras cheese*.** 2007 *The Journal of dairy research*

**Abstract:** Propolis is a natural product collected by honeybee workers. The product was tested for its antifungal effect against *Aspergillus versicolor* ATCC 12996 as well as biosynthesis of sterigmatocystin during ripening of *Egyptian Ras cheese*. The use of different concentrations of aqueous propolis extract 250, 500 and 1000 part per million (ppm) on the *cheese surface* was investigated. Mould growth and toxin production were completely inhibited at the highest concentration 1000 ppm, while the lower concentrations exhibited definite fungistatic activity during 90 days of ripening. *Control cheese* demonstrated that the amount of sterigmatocystin produced was proportional to the growth of *Asp. versicolor* during three months of ripening. It could be concluded that propolis concentration of 1000 ppm could prevent mould growth and sterigmatocystin production in *Ras cheese*. The economic as well as the public health importance of propolis as a natural preservative in *cheese* manufacture is discussed.

**Effect of Zataria multiflora Boiss. essential oil on growth and aflatoxin formation by *Aspergillus flavus* in culture media and *cheese*.** 2009 *Food and chemical toxicology : an international journal published for the British Industrial Biological Research Association*

**Abstract:** The effect of Zataria multiflora Boiss. essential oil (EO) against growth, spore production and aflatoxin formation by *Aspergillus flavus* ATCC 15946 was investigated in synthetic media as well as in *Iranian ultra-filtered white cheese* in brine. EO effectively inhibited radial growth and spore production on potato dextrose agar (PDA) in a dose-dependent manner. At 200 ppm, the radial growth and sporulation reduced by 79.4% and 92.5%, respectively. The growth was completely prevented at 400 ppm on PDA, and minimum fungicidal concentration (MFC) of the oil was estimated at 1000 ppm. The oil also significantly suppressed mycelial growth and aflatoxin synthesis in broth medium at all concentrations tested ( $P < 0.05$ ). At 150 ppm of EO, the mycelial growth and aflatoxin accumulation reduced by 90% and 99.4%, respectively. The EO at all concentrations tested, had an inhibitory effect against radial fungal growth and aflatoxin production by *A. flavus* in *cheese*. However, no concentration of EO examined was able to completely inhibit the growth and aflatoxin production in *cheese*. The results suggested the potential substitution of the antifungal chemicals by this EO as a natural inhibitor to control the growth of molds in foods such as *cheese*.

## Result of the query: *aspergillus* ~lives in cheese



**The effect of Egyptian honeybee propolis on the growth of (*Aspergillus versicolor*) and sterigmatocystin biosynthesis in *Ras cheese*.** 2007 *The Journal of dairy research*

**Abstract:** Propolis is a natural product collected by honeybee workers. The product was tested for its antifungal effect against *Aspergillus versicolor* ATCC 12996 as well as biosynthesis of sterigmatocystin during ripening of *Egyptian Ras cheese*. The use of different concentrations of aqueous propolis extract 250, 500 and 1000 part per million (ppm) on the *cheese surface* was investigated. Mould growth and toxin production were completely inhibited at the highest concentration 1000 ppm, while the lower concentrations exhibited definite fungistatic activity during 90 days of ripening. *Control cheese* demonstrated that the amount of sterigmatocystin produced was proportional to the growth of *Asp. versicolor* during three months of ripening. It could be concluded that propolis concentration of 1000 ppm could prevent mould growth and sterigmatocystin production in *Ras cheese*. The economic as well as the public health importance of propolis as a natural preservative in *cheese* manufacture is discussed.

**Effect of Zataria multiflora Boiss. essential oil on growth and aflatoxin formation by *Aspergillus flavus* in culture media and *cheese*.** 2009 *Food and chemical toxicology : an international journal published for the British Industrial Biological Research Association*

**Abstract:** The effect of Zataria multiflora Boiss. essential oil (EO) against growth, spore production and aflatoxin formation by *Aspergillus flavus* ATCC 15946 was investigated in synthetic media as well as in *Iranian ultra-filtered white cheese* in brine. EO effectively inhibited radial growth and spore production on potato dextrose agar (PDA) in a dose-dependent manner. At 200 ppm, the radial growth and sporulation reduced by 79.4% and 92.5%, respectively. The growth was completely prevented at 400 ppm on PDA, and minimum fungicidal concentration (MFC) of the oil was estimated at 1000 ppm. The oil also significantly suppressed mycelial growth and aflatoxin synthesis in broth medium at all concentrations tested ( $P < 0.05$ ). At 150 ppm of EO, the mycelial growth and aflatoxin accumulation reduced by 90% and 99.4%, respectively. The EO at all concentrations tested, had an inhibitory effect against radial fungal growth and aflatoxin production by *A. flavus* in *cheese*. However, no concentration of EO examined was able to completely inhibit the growth and aflatoxin production in *cheese*. The results suggested the potential substitution of the antifungal chemicals by this EO as a natural inhibitor to control the growth of molds in foods such as *cheese*.

## 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



## Microbial entity detection

### NCBI taxonomy

- **Fungi** (fungi) Click on organism name to get more information.
  - **Blastocladiomycota**
    - **Blastocladiomycetes**
      - **Blastocladales**
      - **Blastocladales incertae sedis**
    - **environmental samples**
      - **uncultured Blastocladiomycota**
  - **Chytridiomycota**
    - **Chytridiomycetes**
      - **Chytridiales**
      - **Cladochytriales**
      - **Gromochytriales**
      - **Lobulomycetales**
      - **Mesochytriales**
      - **Polychytriales**
      - **Rhizophyctidiales**
      - **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

## On going work

- Ambiguous cases for automatic tools

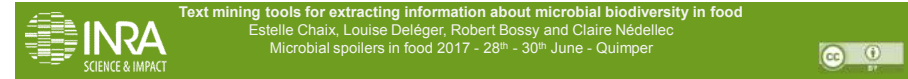
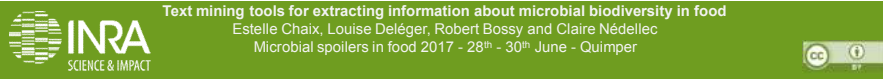
*"Byssoschlamys fulva and Neosartorya fischeri are heat-resistant fungi which are a concern to food industries"*

- Automatic detection of microbial phenotypes

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

## Acknowledgments

INRA Ontobiotope and Florilège working groups  
Food Microbiome project





# 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](mailto: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>



