



Extraction d'information à partir de textes

Fouille de texte, un enjeu scientifique majeur

Plus de 60 millions d'articles, 2.5 par an
160 millions de documents scientifiques indexés

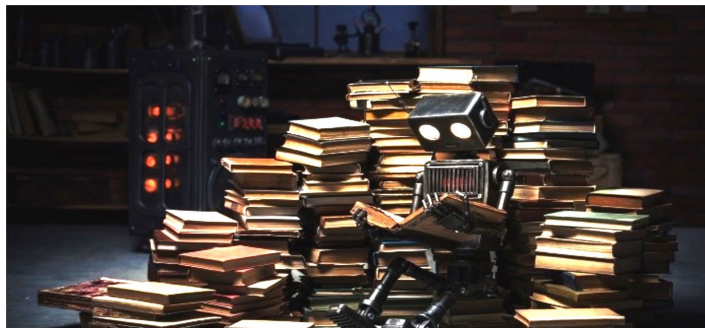
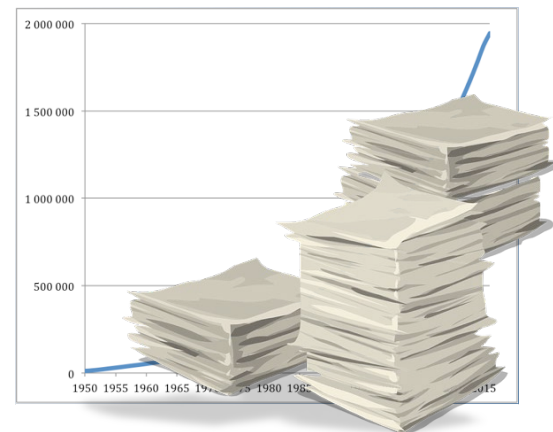
50% des articles ne sont jamais lus
90% des articles ne sont pas cités
80% des articles cités ne sont pas lus

*The STM report,
2015*

*Orduña-Malea et al.,
Scientometrics 2014*

*Lokman I. Meho, the
rise and rise of citation
analysis, 2007.*

*Simkin & Roychowdhury.
Read before you cite!, 2002*



Exploiter les données de la recherche

Donner du sens aux données textuelles

Transformer une donnée non structurée
en donnée structurée, manipulable par un ordinateur

Intégrer la Fouille de Texte
au cœur de l'activité de l'utilisateur
non spécialiste

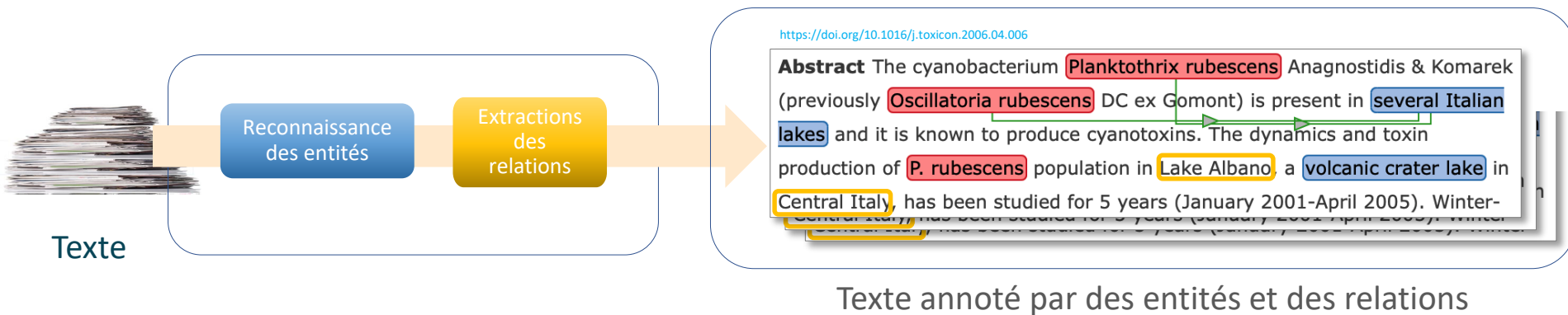


Extraction d'information et de
connaissance - partage et réutilisation

Notion de référentiel commun



Extraction automatique d'information



Exemple

- Les entités de type **microbe**, **habitat** et **lieu géographique**
- Reliés par des relations orientées **vit-dans**, **localisée-à**

Textes de documents

... *P. rubescens* population in Lake Albano, a volcanic crater lake in Central Italy,...

Extraction
d'information

<u>Microbe</u>	<u>Habitat</u>	<u>Lieu géographique</u>
<i>P. rubescens</i>	volcanic crater lake	Lake Albano
...	...	...

Vit-dans

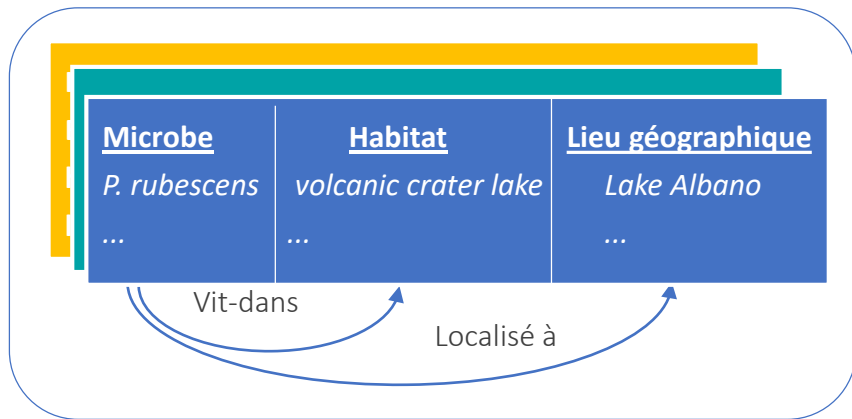
Localisé à

Représentation formelle de l'information

Extraction automatique d'information

Transforme une donnée non
structurée, du texte
en donnée structurée manipulable
par un ordinateur

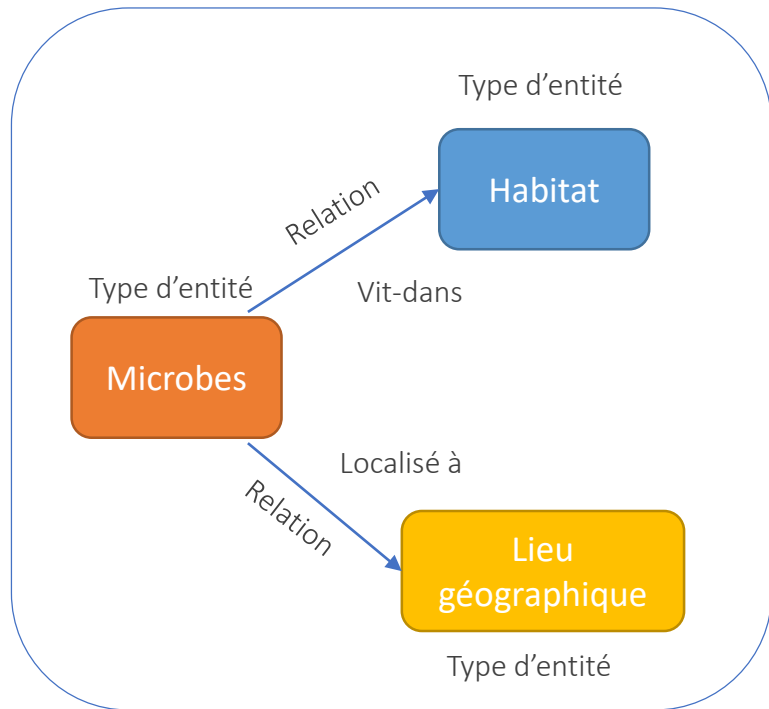
Schéma d'annotation des informations



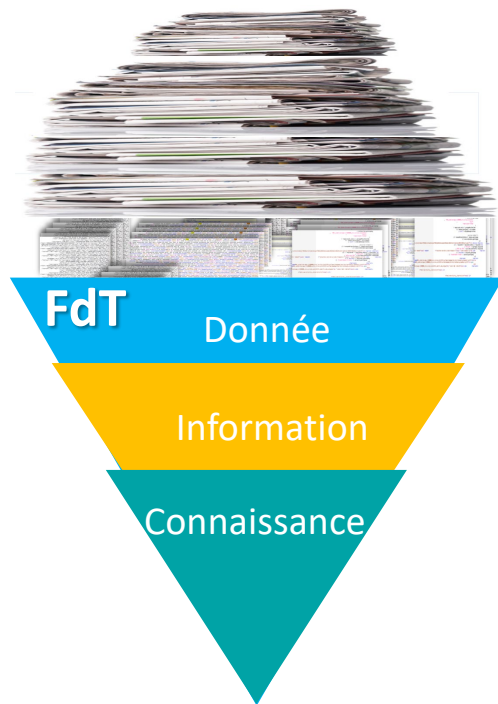
Les champs du formulaire correspondent au schéma des données

Schéma des données, entité - relation

- En général, des relations binaires entre deux entités
- Orientées
- Les arguments de la relation sont typés



Extraction automatique d'information et de connaissances *Partage et réutilisation*



L'extraction d'information transforme
des données en information

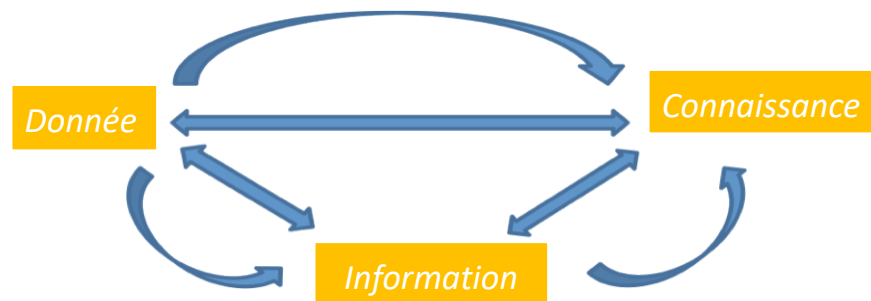
Partager l'information nécessite

un référentiel de connaissance commun

et de relier l'information aux connaissances connues

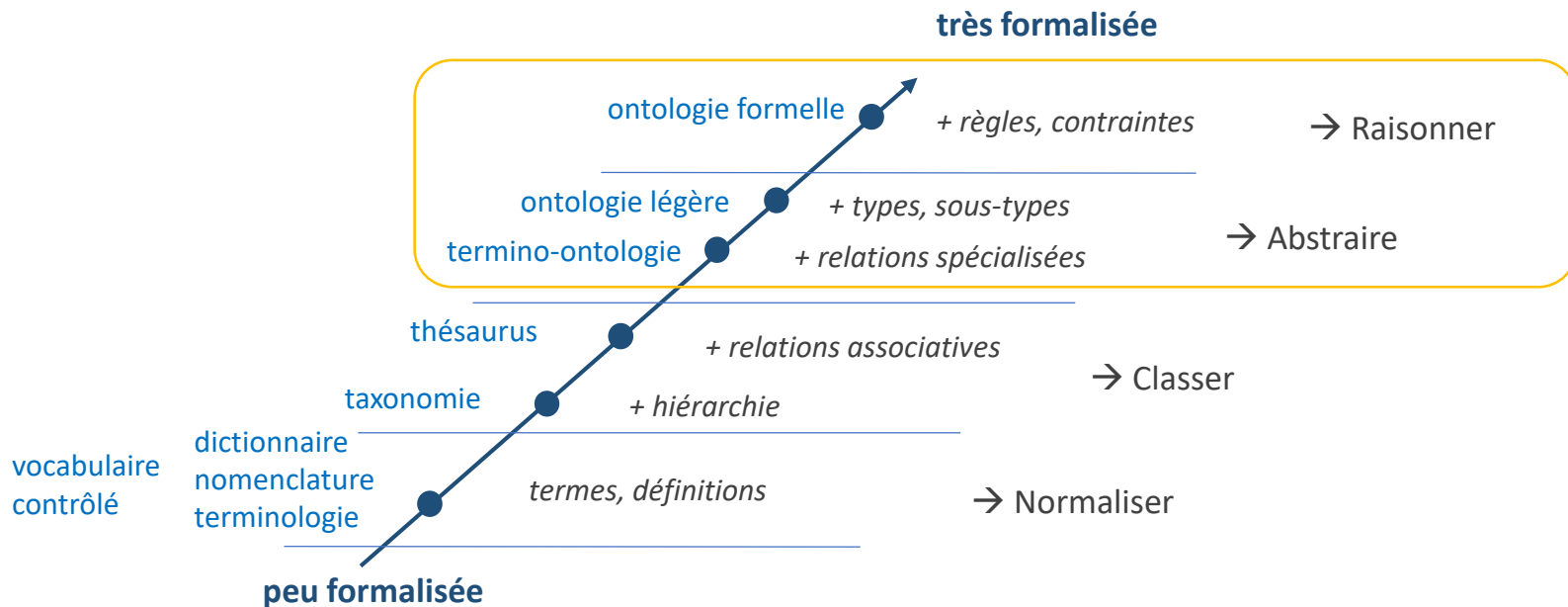
Données, informations, connaissances

- Données.** Éléments discrets, nombres, bases de données, texte
- Informations.** Données filtrées, contextualisées, interprétées et reliées localement à l'aide de la connaissance de façon à être réutilisables
- Connaissances.** Informations organisées, synthétisées et reliées à un modèle de connaissance : un système unifié de définitions formelles et de principes fondamentaux, pour la compréhension partagée et des traitements fiables (*sémantique formelle*)



Référentiels

D'après Sophie Aubin



Termino-ontologie

Chaque concept est

Relié à ses **parents** (concepts plus généraux)

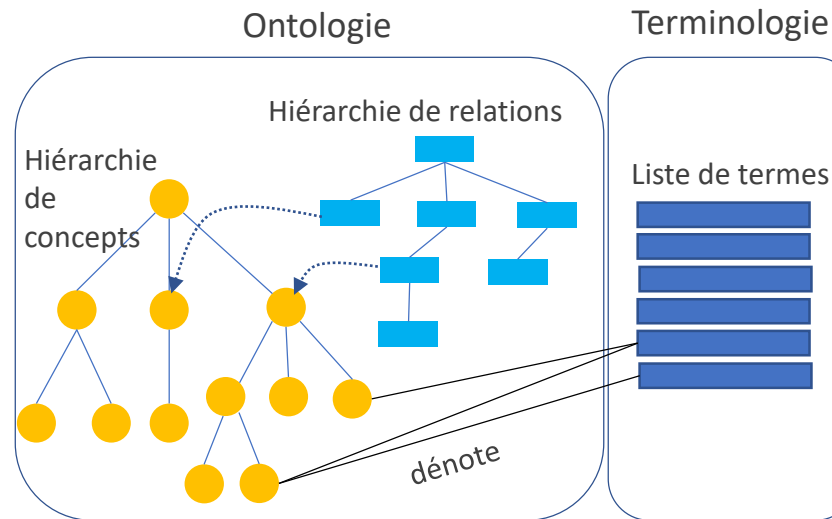
Relié à ses **descendants**
(concepts plus spécifiques)

Relié à d'autres concepts par des **relations**
spécialisées

Il est associé à

des **informations terminologiques**, label, synonyme

Éventuellement, les sens et usages du terme



D'après Anis Tissaoui

Extraction d'information *pour des bases de connaissance intégrées*

Objectif

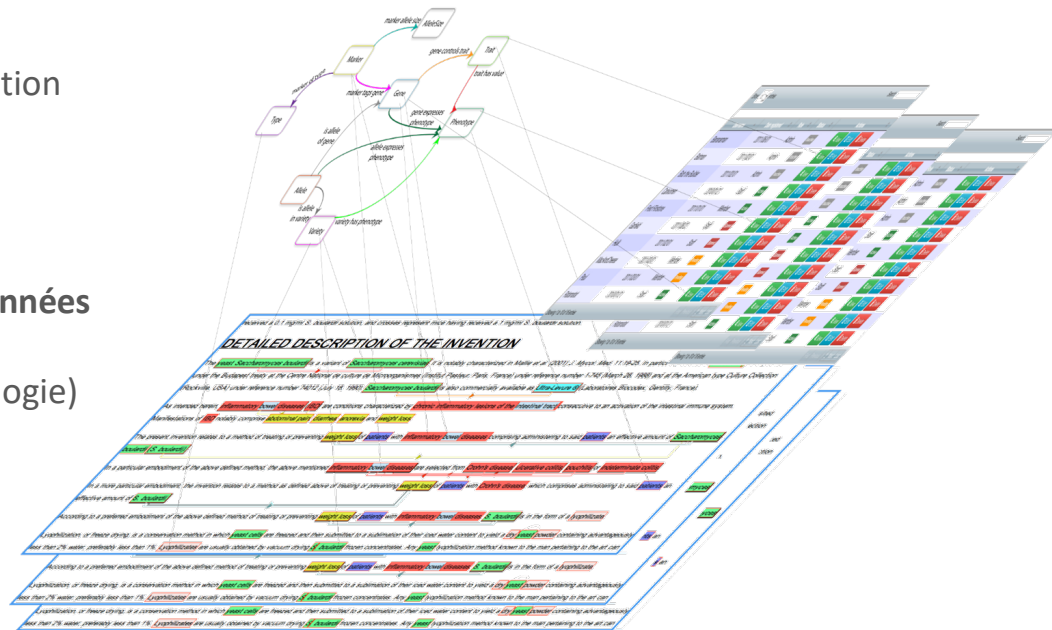
Centraliser, structurer et standardiser l'information
Pour en faciliter l'accès et l'utilisation

Données en **texte libre nombreuses**

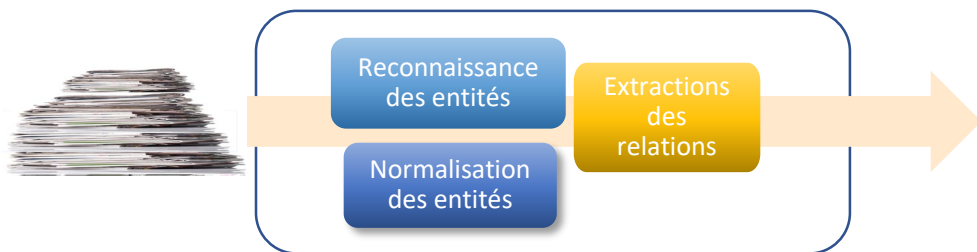
Les structurer pour les combiner à d'autres données

- Détecter les informations
- Les associer à une référence partagée (ontologie)
- Les relier entre elles
- Les mettre à disposition

Par des méthodes de *text mining*



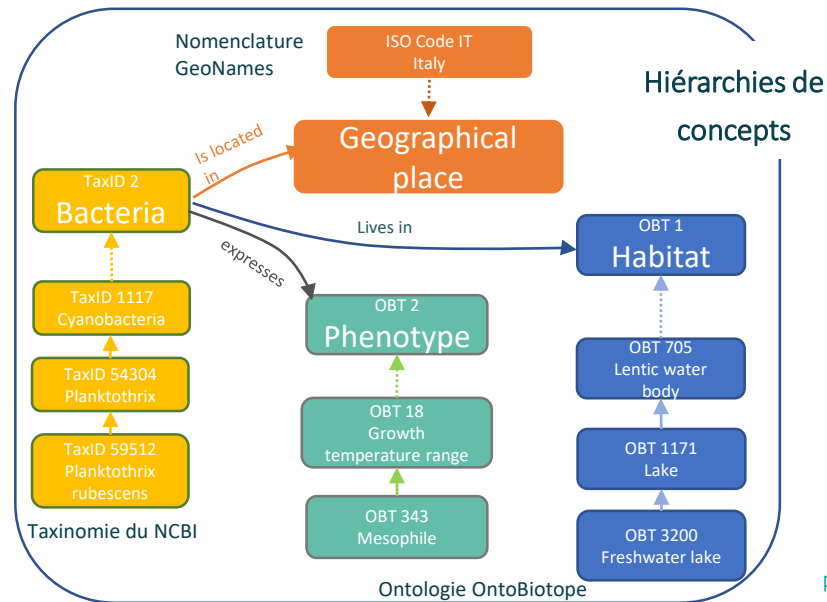
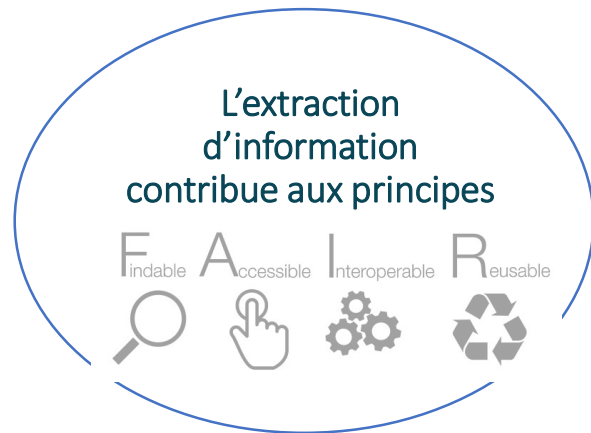
Formaliser l'association entre données du texte et référentiels



Extraction automatique d'information basée sur
des ontologies

Abstract The cyanobacterium *Planktothrix rubescens* Anagnostidis & Komarek (previously *Oscillatoria rubescens* DC ex Gomont) is present in several Italian lakes and it is known to produce cyanotoxins. The dynamics and toxin production of *P. rubescens* population in Lake Albano, a volcanic crater lake in Central Italy, has been studied for 5 years (January 2001-April 2005). Winter-

Texte



Extraction d'information : identifier et *typer* les entités du texte

Microcystin diversity in a *Planktothrix rubescens* population from Lake Albano (Central Italy)

The cyanobacterium *Planktothrix rubescens* Anagnostidis & Komarek (previously *Oscillatoria rubescens* DC ex Gomont) is present in several Italian lakes and it is known to produce cyanotoxins. The dynamics and toxin production of *P. rubescens* population in Lake Albano, a volcanic crater lake in Central Italy, has been studied for 5 years (January 2001–April 2005).

<https://doi.org/10.1016/j.toxicon.2006.04.006>

Types d'entité

Microbe



Habitat



Lieu géographique



Types et classes d'entités en extraction d'information

1. La reconnaissance d'entité nommée associe un type général aux entités reconnues

Date, lieu, personne

Microbe, habitat, maladie, symptôme, organe

Entreprise, fonction

2. La normalisation associe aux entités un nom standard provenant d'un référentiel

Microbe: nom scientifique

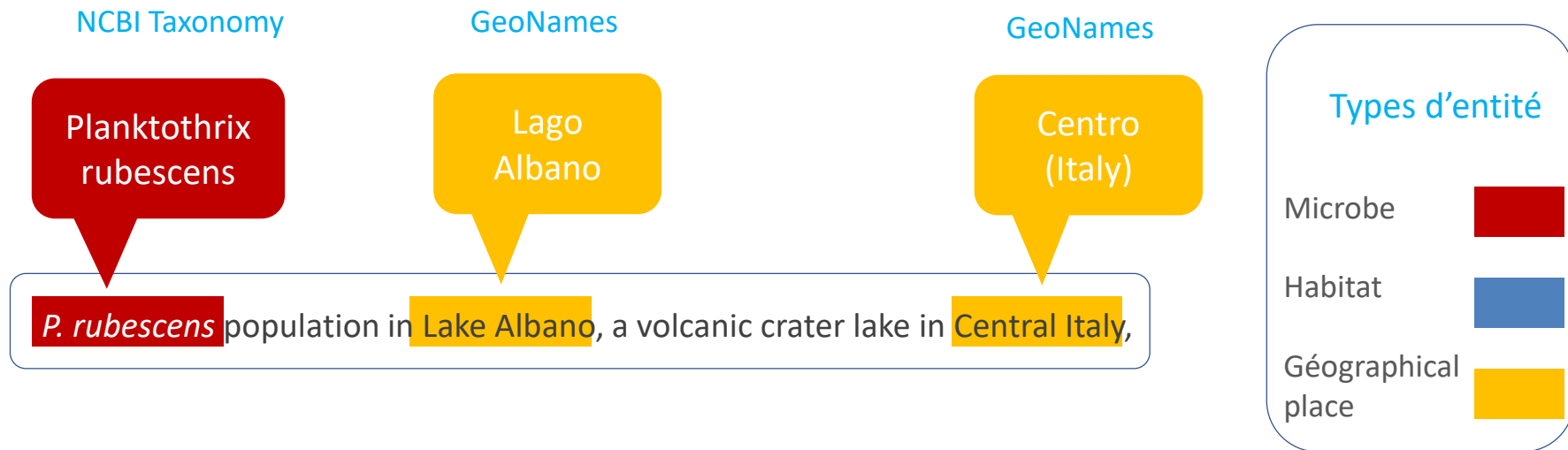
Molécule : nom de la molécule dans le référentiel (ex CheBI)

3. L'annotation sémantique associe aux entités une catégorie d'un référentiel, un concept précis de l'ontologie

Microbe: taxon dans la taxinomie (ex NCBI)

Maladie : nom de la maladie dans la classification (ex CIM-10)

2. Normalisation et référentiel : associer un nom standard



Annotation sémantique : relier entité et connaissance du domaine

associe aux entités *une catégorie précise du référentiel*



Classe de l'entité
OBT 3200
Freshwater lake

Exemple de texte annoté

P. rubescens population in Lake Albano, a volcanic crater lake in Central Italy,

TaxID 59512
Planktothrix
rubescens

ID 3183358
Lago Albano

ID 12089032
Centro (Italy)

Différents référentiels pour différents types d'entités et de relations

Le schéma d'annotation des données du texte peut faire référence à des **référentiels distincts** :

nomenclatures, dictionnaires, thesaurus, ontologies

Dans **l'exemple**

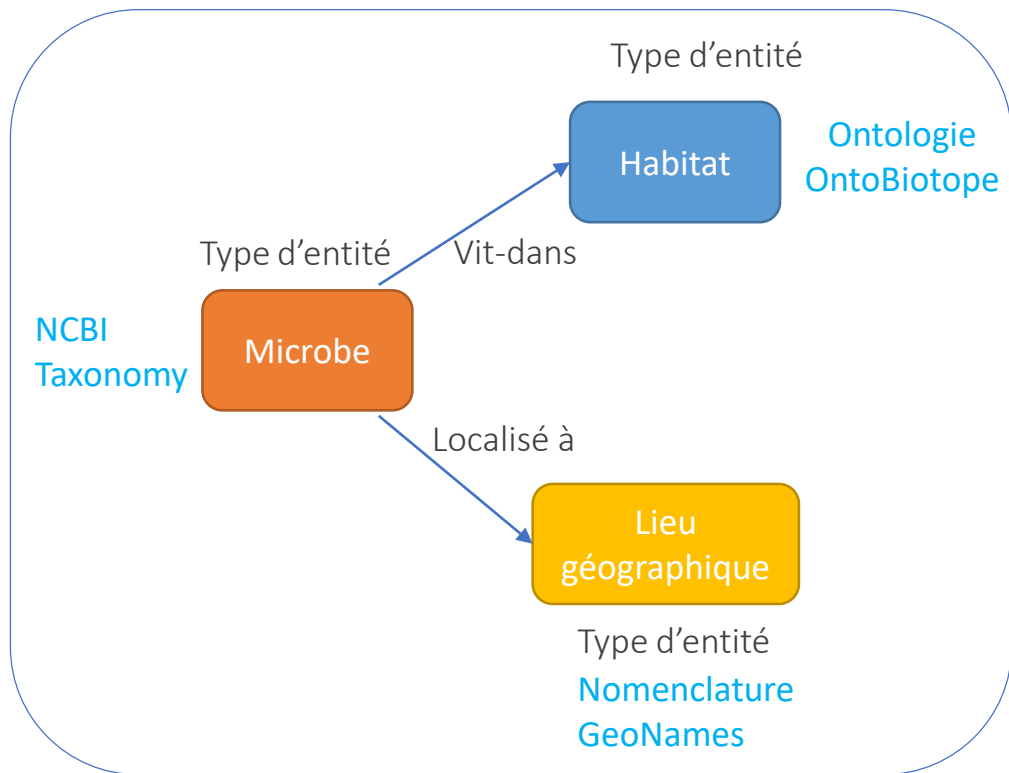
Une taxinomie : *NCBI Taxonomy*

Une nomenclature : *GeoNames*

Une ontologie : *OntoBiotope*

Des relations externes

Différents types de référentiels pour différents niveaux de standardisation et d'expressivité



Exemple de schéma d'annotation entité – relation et référentiels

p. 17

Que faire en recherche documentaire en fonction du référentiel

1. La reconnaissance d'entité nommée associe un type général aux entités reconnues

Retrouver tous les documents qui mentionnent un type d'entité

Produire la liste des entités dans un ensemble de documents

Ex. Trouver tous les documents qui mentionnent une maladie,

Ex. Identifier toutes les expressions du texte qui dénotent des maladies

2. La normalisation associe aux entités un nom standard provenant d'un référentiel

Retrouver précisément les données ou documents contenant une entité étant donné son nom,
Indépendamment des variations

Ex. Trouver tous les documents qui mentionnent la bactérie *Planktothrix rubescens*

3. L'annotation sémantique associe aux entités une catégorie précise de l'ontologie

Retrouver des données ou documents contenant une entité, à différents niveaux de généralité
Sans connaître son nom

Ex. Trouver tous les documents qui mentionnent des types d'habitat, par exemple, des milieux aquatiques



Aperçu des méthodes d'extraction d'information

Projection de lexiques

Rechercher dans le texte des fragments identiques à des entrées dans une liste préétablie de formes connues.

Recovery of Gram-Negative Bacteria from Aerobic Blood Culture Bottles Containing Antibiotic Binding Resins after Exposure to β -Lactam and Fluoroquinolone Concentrations.

1-4142125

Authors: Iris H Chen David P Nicolau Joseph L Kuti

2019 *Journal of clinical microbiology*

Abstract Blood culture bottles containing antibiotic binding resins are routinely used to minimize artificial sterilization in the presence of antibiotics. However, the resin binding kinetics can differ between antibiotics and concentrations. This study assessed the impact of clinically meaningful peak, midpoint, and trough concentrations of meropenem, imipenem, cefepime, ceftazidime, levofloxacin, and piperacillin-tazobactam on the recovery of *Pseudomonas aeruginosa*, *Escherichia coli*, and *Klebsiella pneumoniae* from resin-containing BacT/Alert FA Plus and Bactec Aerobic/F blood culture bottles. *P. aeruginosa*-inoculated bottles alarmed positive in 4/20 (20%), 16/20 (80%), and 20/20 (100%) of those with peak, midpoint, and trough concentrations of antipseudomonal agents, respectively ($P \leq 0.001$). *E. coli* was recovered from 8/24 (33%), 11/24 (46%), and 14/24 (58%) of bottles with peak, midpoint, and trough concentrations, respectively ($P = 0.221$). *K. pneumoniae* was recovered from 8/16 (50%) at all concentrations of the studied antibiotics ($P = 1.0$). BacT/Alert and Bactec bottles inoculated with antibiotics and *P. aeruginosa* had similar times to detection (TTD) ($P = 0.352$); however, antibiotic-containing BacT/Alert bottles had a shorter TTD compared with antibiotic-containing Bactec bottles for *E. coli* ($P = 0.026$) and *K. pneumoniae* ($P \leq 0.001$). Pathogen recovery in BacT/Alert FA Plus and Bactec Aerobic/F blood culture bottles containing antibiotic binding resins was greatly reduced in the presence of antibiotics, especially at higher concentrations. These data support the practice of drawing blood cultures immediately before an antibiotic dose to maximize the chances of pathogen recovery.

- Exploitation de ressources existantes :
gazetteers, glossaires, annuaires,
terminologies, ontologies, etc.
- Approche simple pour la reconnaissance
d'entités, voire l'annotation sémantique si les
étiquettes sont associées à un identifiant.

Limites

- Silence sur les entités hors-nomenclature.
- Peu robuste aux variations typographiques ou
aux variations de langage.
- Entrées ambiguës.

Capture des variations

Saturation automatique des lexiques pour les abréviations.

Escherichia coli	→	E. coli
Klebsiella pneumoniae	→	K. pneumoniae
Pseudomonas aeruginosa	→	P. aeruginosa
Streptococcus aureus	→	S. aureus
Listeria monocytogenes	→	L. monocytogenes

Recovery of Gram-Negative **Bacteria from Aerobic Blood Culture Bottles Containing Antibiotic Binding Resins after Exposure to β -Lactam and Fluoroquinolone Concentrations.**

1:4142135

Authors: Iris H Chen David P Nicolau Joseph L Kuti

2019 *Journal of clinical microbiology*

Abstract Blood culture bottles containing antibiotic binding resins are routinely used to minimize artificial sterilization in the presence of antibiotics. However, the resin binding kinetics can differ between antibiotics and concentrations. This study assessed the impact of clinically meaningful peak, midpoint, and trough concentrations of meropenem, imipenem, cefepime, cefazolin, levofloxacin, and piperacillin-tazobactam on the recovery of **Pseudomonas aeruginosa**, **Escherichia coli**, and **Klebsiella pneumoniae** from resin-containing BacT/Alert FA Plus and Bactec Aerobic/F blood culture bottles. **P. aeruginosa** -inoculated bottles alarmed positive in 4/20 (20%), 16/20 (80%), and 20/20 (100%) of those with peak, midpoint, and trough concentrations of antipseudomonal agents, respectively ($P \leq 0.001$). **E. coli** was recovered from 8/24 (33%), 11/24 (46%), and 14/24 (58%) of bottles with peak, midpoint, and trough concentrations, respectively ($P = 0.221$). **K. pneumoniae** was recovered from 8/16 (50%) at all concentrations of the studied antibiotics ($P = 1.0$). BacT/Alert and Bactec bottles inoculated with antibiotics and **P. aeruginosa** had similar times to detection (TTD) ($P = 0.352$); however, antibiotic-containing BacT/Alert bottles had a shorter TTD compared with antibiotic-containing Bactec bottles for **E. coli** ($P = 0.026$) and **K. pneumoniae** ($P \leq 0.001$). Pathogen recovery in BacT/Alert FA Plus and Bactec Aerobic/F blood culture bottles containing antibiotic binding resins was greatly reduced in the presence of antibiotics, especially at higher concentrations. These data support the practice of drawing blood cultures immediately before an antibiotic dose to maximize the chances of pathogen recovery.

Lemmatisation pour les flexions.

plants	→	plant
plant	→	plant
Plant	→	plant
PLANT	→	plant

Comparison of microbial and transient expression (tobacco plants and **plant-cell packs) for the production and purification of the anticancer mistletoe lectin viscum.**

1:4142135

Authors: Benjamin B Gengenbach Linda L Keil Patrick Opdensteinen Catherine R Müschen Georg Melmer Hans Lentzen Jens Bührmann Johannes F Buyel

2019 *Biotechnology and bioengineering*

Abstract Cancer is the leading cause of death in industrialized countries. Cancer therapy often involves monoclonal antibodies or small-molecule drugs, but carbohydrate-binding lectins such as mistletoe (*Viscum album*) viscum offer a potential alternative treatment strategy. Viscum is toxic in mammalian cells, ruling them out as an efficient production system, and it forms inclusion bodies in *Escherichia coli* such that purification requires complex and lengthy refolding steps. We therefore investigated the transient expression of viscum in intact *Nicotiana benthamiana* **plants** and *Nicotiana tabacum* Bright Yellow 2 **plant**-cell packs (PCPs), comparing a full-length viscum gene construct to separate constructs for the A and B chains. As determined by capillary electrophoresis the maximum yield of purified heterodimeric viscum in *N. benthamiana* was ~7 mg/kg fresh biomass with the full-length construct. The yield was about 50% higher in PCPs but reduced 10-fold when coexpressing A and B chains as individual polypeptides. Using a single-step lactosyl-Sepharose affinity resin, we purified viscum to ~54%. The absence of refolding steps resulted in estimated cost savings of more than 80% when transient expression in tobacco was compared with *E. coli*. Furthermore, the **plant**-derived product was ~3-fold more toxic than the bacterially produced counterpart. We conclude that **plants** offer a suitable alternative for the production of complex biopharmaceutical proteins that are toxic to mammalian cells and that form inclusion bodies in bacteria.



INRAE

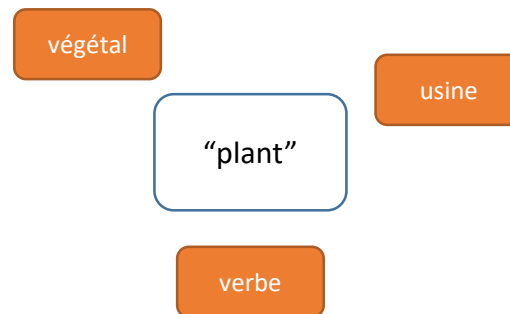
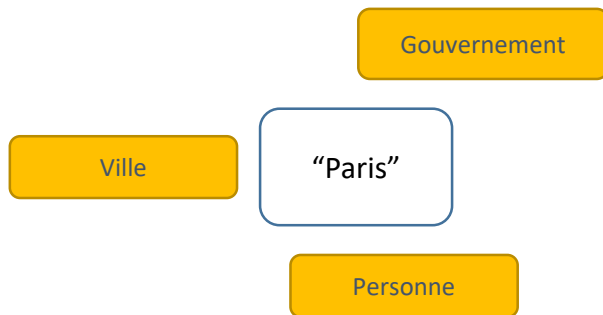
Extraction d'information, ANF 2021

Robert Bossy & Claire Nédélec

Formes ambiguës

Sources d'ambiguïté

- homonymie
- métonymie
- polysémie



Désambiguïisation par le contexte

Identifier des mots dans le contexte qui permettront de discriminer entre plusieurs sens.

Par exemple : *plant/factory* et *plant/organisme* ?

new strains can be rapidly implemented into existing infrastructures such as <u>bioethanol production</u>	plants	
Evaluation of methanogenic activity of <u>biogas</u>	plant	slurry
two reference points (indoor and outdoor) that are assumed not to be <u>contaminated</u> by the	plant	's activities
a phage can survive in a <u>cheese</u>	plant	for more than a year
The first large-scale <u>biogas</u>	plant	was put into operation

a problem that is present in the transformation of all	plant	<u>species</u>
The tendency of closely related	plant	<u>species</u> to share natural enemies
Rhizosphere is the complex place of numerous interactions between	plant	<u>roots</u> , microbes and soil fauna
an appropriate method of transgene introduction into a	plant	<u>cell</u>
common mechanisms for	plant	<u>cell</u> reprogramming during endosymbiosis

Exploitation du contexte par des règles d'extraction

Il existe des outils qui permettent d'exprimer des règles de désambiguïsation selon les mots du contexte.

("biogas" ou <nourriture>) "plant" → étiqueter "plant" comme une usine.

"plant" ("species" ou "cell") → étiqueter "plant" comme une plante.

Les mêmes types de règles peuvent servir aussi à extraire des relations.

<microorganism> "isolated from" <NOM>

... mycobacteria isolated from seals ...

<microorganism> "isolated from" <DET> <ADJ> <NOM>

... Mycobacterium malmoense isolated from soil ...

... Nocardia africana isolated from a feline mycetoma ...

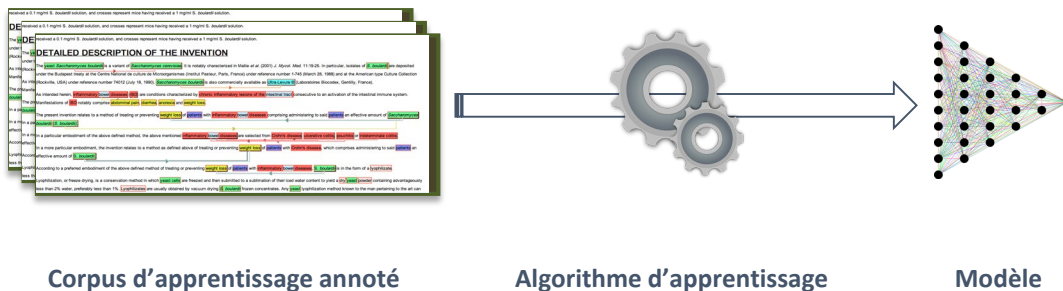
Limites

- La performance des règles dépend des types d'entités ou de relation, des domaines, et du genre des documents.
- La qualité dépend de la variabilité d'expression dans les documents.
- La maintenance d'un système basé sur des règles écrites à la main peut être coûteuse à maintenir.



Apprentissage supervisé

Les algorithmes d'apprentissage supervisé induisent des règles à partir d'exemples étiquetés.



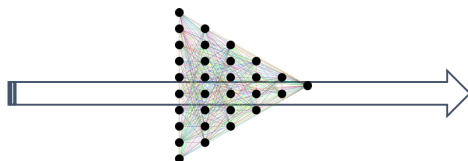
1. Phase d'apprentissage

- Lors de la phase d'apprentissage, l'algorithme induit un modèle à partir d'exemples annotés par des experts.
- La nature du modèle dépend de l'algorithme utilisé. Ce peuvent être des droites de régression, des arbres de décision, etc.

Apprentissage supervisé

Les algorithmes d'apprentissage supervisé induisent des règles à partir d'exemples étiquetés.

Abstract The cyanobacterium *Planktothrix rubescens* Anagnostidis & Komarek (previously *Oscillatoria rubescens* DC ex Gomont) is present in several Italian lakes and it is known to produce cyanotoxins. The dynamics and toxin production of *P. rubescens* population in Lake Albano, a volcanic crater lake in Central Italy, has been studied for 5 years (January 2001-April 2005). Winter-



Abstract The cyanobacterium *Planktothrix rubescens* Anagnostidis & Komarek (previously *Oscillatoria rubescens* DC ex Gomont) is present in several Italian lakes and it is known to produce cyanotoxins. The dynamics and toxin production of *P. rubescens* population in Lake Albano, a volcanic crater lake in Central Italy, has been studied for 5 years (January 2001-April 2005). Winter-

Corpus non-annoté

Modèle

Corpus annoté
automatiquement

Phase de production

- Lors de la phase d'étiquetage, le modèle sert à annoter automatiquement de nouveaux documents.
- Les erreurs peuvent être quantifiées en appliquant le modèle sur le corpus d'apprentissage.

À savoir

- L'élaboration du corpus d'apprentissage représente un effort initial considérable.
- Il existe des corpus d'apprentissage et même des modèles pré-appris pour un certain nombre de types d'annotation.

Conclusion

Les différentes méthodes ont différentes qualités.

Projection de
lexiques

Application de
règles

Apprentissage
supervisé

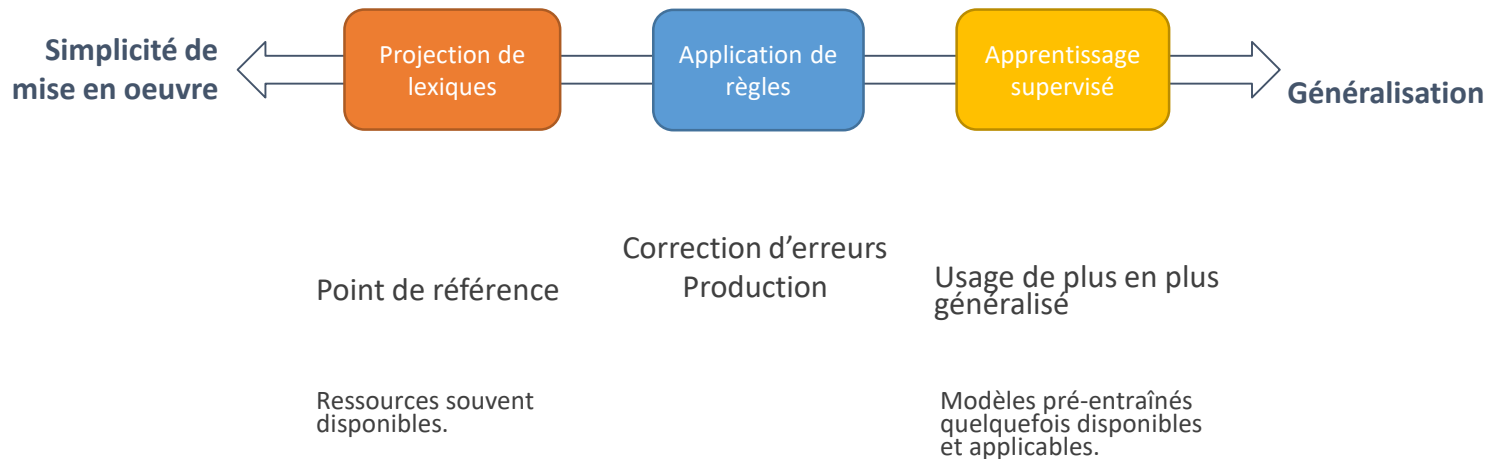
Conclusion

Les différentes méthodes ont différentes qualités.



Conclusion

Les différentes méthodes ont différents rôles.





Exemples d'utilisation

Moteurs de recherche sémantique

Intégration de données

Exemples d'applications

Recherche d'information

- Recherche sémantique de documents
- Intégration de données de sources diverses : bibliographie, données expérimentales, données de référence, etc.

Aide à la décision

- Cartographie thématique (synthèse de corpus, tendances)
- Profilage d'experts (par ex. sélection de relecteurs)

Extension de référentiel

- Recherche de nouveaux synonymes ou concepts

Extraction d'information et applications

Ressources
sémantiques
Ontologies

Bases
documentaires

Information
extraction

Données et
métadonnées
des textes

Système
d'information

Services

Source	Type	Métadonnées	Valeur	Actions
PROJET TEST	Document	Document	Document	Document
PROJET TEST2	Document	Document	Document	Document
PROJET TEST3	Document	Document	Document	Document
PROJET TEST4	Document	Document	Document	Document
PROJET TEST5	Document	Document	Document	Document
PROJET TEST6	Document	Document	Document	Document
PROJET TEST7	Document	Document	Document	Document
PROJET TEST8	Document	Document	Document	Document
PROJET TEST9	Document	Document	Document	Document
PROJET TEST10	Document	Document	Document	Document
PROJET TEST11	Document	Document	Document	Document
PROJET TEST12	Document	Document	Document	Document
PROJET TEST13	Document	Document	Document	Document
PROJET TEST14	Document	Document	Document	Document
PROJET TEST15	Document	Document	Document	Document
PROJET TEST16	Document	Document	Document	Document
PROJET TEST17	Document	Document	Document	Document
PROJET TEST18	Document	Document	Document	Document
PROJET TEST19	Document	Document	Document	Document
PROJET TEST20	Document	Document	Document	Document

Deux services exploitant l'Extraction d'Information

Deux démonstrations

- AlvisIR : un moteur de recherche sémantique.
- Florilège: une base de données intégrative de référence.

Thème : biodiversité microbienne

- Motivations : recherche biomédicale, sécurité et innovation alimentaires, bioréacteurs...
- Nous allons examiner les entités de type Taxon (bactéries, levures, virus) et Habitats.
- Les deux services exploitent les mêmes annotations.



Un moteur de recherche sémantique,
AlvisIR



<http://bibliome.jouy.inra.fr/demo/ontobiotope/alvisir2/webapi/search?>



Un exemple d'intégration de données,
Florilege



<http://migale.jouy.inra.fr/Florilege/#&welcome>

Extraction d'information et base de données



Recherche documentaire / recherche d'information

- La **recherche documentaire** identifie les documents qui répondent à tous les critères de la requête
- La **recherche d'information** identifie les informations provenant potentiellement de différents documents
Ou d'autres sources non documentaires

Extraction d'information et applications



Combiner information structurée et information textuelle

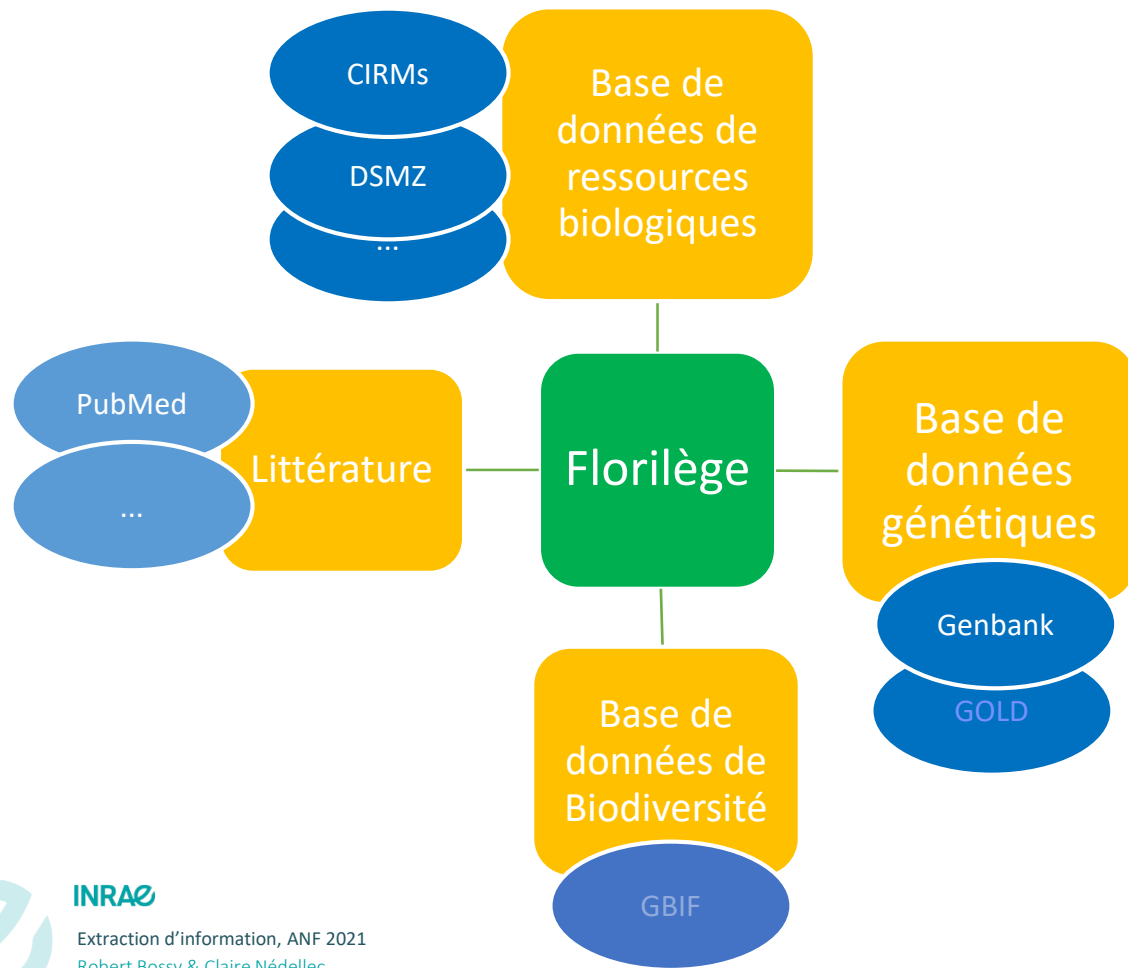
L'extraction d'information ne relève pas seulement de la fouille de texte

Une fois l'information textuelle *structurée*
elle peut être intégrée avec tout ensemble de données structurées,
qui serait indexé suivant les mêmes référentiels



39

Base de Florilege de connaissance publique en ligne

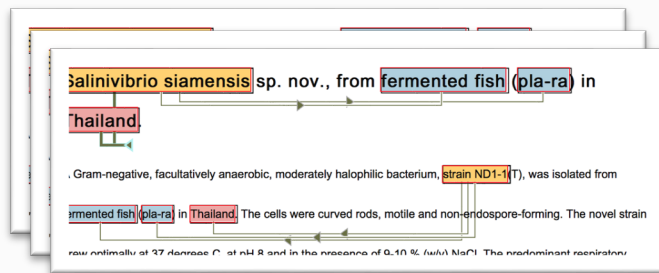


Agrège l'information sur les microbes

Permet une interrogation transparente de données représentées de différentes façon

<http://migale.jouy.inra.fr/florilege/>

Document



Base de données

Base de données

Base de données

	Taxon	Géo	Habitat
CLIB number	Species name	§ Country	Environment
CLIB 1	Saccharomyces cerevisiae	Switzerland	Environment
CLIB 3	Saccharomyces fibuliger	Belgium	Food
CLIB 5	Saccharomyces cerevisiae	USA	Plant
CLIB 6	Saccharomyces cerevisiae	USA	Plant
CLIB 7	Saccharomyces viti	Chile	Oenology
CLIB 8	Saccharomyces viti	Brazil	Oenology
CLIB 10	Saccharomyces malanga	Indonesia	Food
CLIB 11	Saccharomyces cerevisiae	South Africa	Plant
CLIB 12	Saccharomyces cerevisiae	n/a	n/a
CLIB 13	Saccharomyces viti	Brazil	Oenology
CLIB 14	Saccharomyces viti	Chile	Oenology
CLIB 15	Saccharomyces fibuliger	Germany	Food
CLIB 16	Saccharomyces fibuliger	n/a	n/a
CLIB 17	Saccharomyces fibuliger	n/a	n/a
CLIB 18	Saccharomyces fibuliger	n/a	n/a
CLIB 19	Saccharomyces fibuliger	n/a	n/a
CLIB 20	Saccharomyces fibuliger	n/a	n/a

Lives in

Fouille de texte

SOURCE TEXT	TAXON	RELATION TYPE	HABITAT	QPS	SOURCE
21478365	Lactobacillus acidophilus 30SC	Lives in	intestine		PubMed
21478365	Lactobacillus acidophilus 30SC	Lives in	milk and milk product		PubMed
22366464	Lactobacillus acidophilus 30SC	Lives in	rainwater treatment utility		PubMed
21478365	Lactobacillus acidophilus 30SC	Lives in	swine		PubMed
29154473	Lactobacillus acidophilus 30SC	Lives in	cell		PubMed
	Lactobacillus acidophilus CIRM-BIA 442	Lives in	fermented milk		CIRM
	Lactobacillus acidophilus CIRM-BIA 445	Lives in	fermented milk		CIRM
	Lactobacillus acidophilus DSM 20079 = JCM 1132 = NBRC 13951	Lives in	biofilm		PubMed
21342140	Lactobacillus acidophilus DSM 20079 = JCM 1132 = NBRC 13951	Lives in	dinner plate		PubMed
21342140	Lactobacillus acidophilus DSM 20079 = JCM 1132 = NBRC 13951	Lives in	mouth		PubMed
23608032	Lactobacillus acidophilus DSM 20079 = JCM 1132 = NBRC 13951	Lives in	fish		PubMed
23608032	Lactobacillus acidophilus DSM 20079 = JCM 1132 = NBRC 13951	Lives in	intestine		PubMed
23608032	Lactobacillus acidophilus DSM 20079 = JCM 1132 = NBRC 13951	Lives in	kidney		PubMed
23608032	Lactobacillus acidophilus DSM 20079 = JCM 1132 = NBRC 13951	Lives in	spleen		PubMed
23608032	Lactobacillus acidophilus DSM 20079 = JCM 1132 = NBRC 13951	Lives in	gut		PubMed

Taxon

Habitat



Florilege, démonstration sur
l'intégration de données par les
ontologies



Discussion et conclusion

Valeur ajoutée de l'annotation sémantique

Abstraction

- Facilite l'**accès** à l'information
- Permet l'expression de **requêtes puissantes** et expressives
- Aide à l'**interprétation** et l'explication d'observations
- Permet la **découverte** de nouvelles connaissances (ex. par analogie)

Standardisation

- Permet l'**intégration** de données hétérogènes
- Permet la **confrontation** des données, la validation, la curation
- Permet le **partage**, la mutualisation et la réutilisation
- Facilite l'**identification** des données par des nouveaux utilisateurs
- Facilite la **diffusion** de la connaissance, la critique et la **reproductibilité**

Points d'attention

Sources ... Applications

Accès aux sources documentaires

Accès technique, accès légal, format

Anticiper les moyens de mise à jour

Application Client

Quel usage, quelle intégration au poste de travail

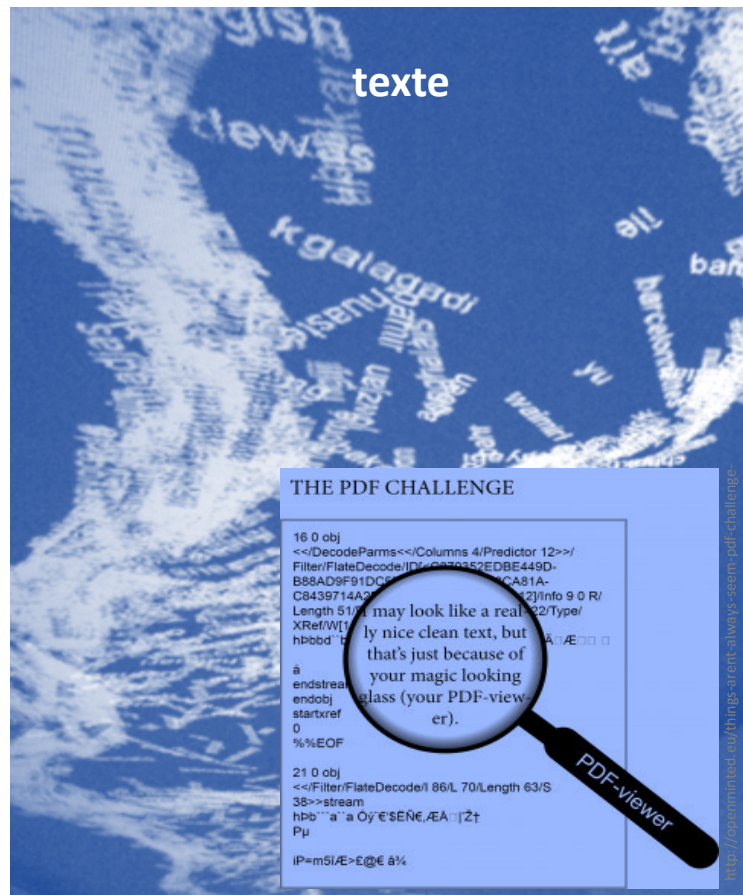
Pour une interface homme machine adaptée aux différents profils

Anticiper les questions de corpus, d'information recherchée, de qualité requise,

Quels coûts, quelle expertise nécessaire

Démarche itérative et adaptative





Accès aux contenus

- **Cadre légal**

Code de la Propriété intellectuelle
Loi de 2016 *Pour une République Numérique*.

Directive européenne (2019) *si accès licite, alors FdT permise pour la recherche*
transposition en cours (2021)

- **Verrous techniques** d'accès aux documents : identifier, accéder, formater

- **Perspective.** Accès et format transparent, centralisation des références et des documents pris en charge par des services

INRAE

