Relations and entities extraction from full texts, and their use in an end-user platform. The case of the epidemiosurveillance VESPA platform.

Nicolas Turenne

Thursday 27 November 2014

Université Paris Est Marne la Vallée – UMR LISIS

Availability of Information

-500 BC 1500

1500 2000

2000 2015

+ 50 % of all information
informational deluge

Text and Corpus processing

a needle in a haystac

- Existing Research Communities since 1950
  - Natural Language Processing (Syntax Analysis, Logic)
  - Lexical Statistics
  - Information retrieval

- Different Tasks (basic usages)
  - Text Classification
  - Question/Answering
  - Automatic Translation
  - Ontology and Thesaurus Acquisition
  - Information Extraction
  - Open Survey Analytics
  - Opinion & Sentiment Analysis

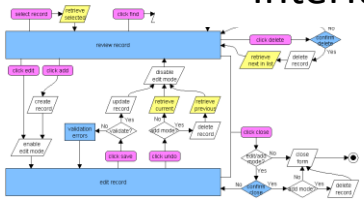
Information Extraction

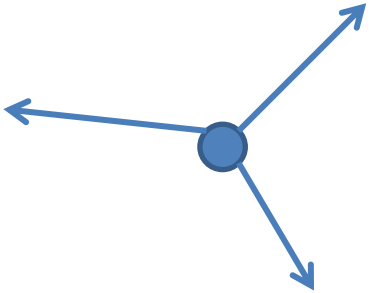
- IE = extracting information from text
- Extract entities
  - People, organizations, locations, times, dates, prices, ...
  - Or sometimes: genes, proteins, diseases, medicines, ...
- Extract the relations between entities
  - Located in, employed by, part of, married to, ...
- Extract scenario of relations
  - network reconstruction, figure out set of events

3 cornerstones

Users
Usages

Algorithms
Interfaces





Datasets
Databases

Our dataset:
Agricultural Warning Report Newsletter

Inform weekly **farmers** about agression on **crops** by **pathogens** i.e microbes (diseases) and insects

Goal of the newspaper*: invite farmers to use Treatment against Pests.*

1946 : first issue, on typewritter

2001: all issues are published in PDF

France is divided in 22 regions in Europe, and overseas regions, each region publishes its own newsletter (own graphical layout).
In future this number will be reduced ~13 regions

Our dataset:

Agricultural Warning Report Newsletter

Service Régional de la Protection des Végétaux
Z.I. de Thise - BP 199
25023 - BESANCON Cedex
Tél : 03 81 88 10 71 - Fax : 03 81 88 27 27

Avertissements Agricoles

Bourgogne et Franche-Comté

Bulletin n° 27/97 - 28 octobre 1997

COLZA

Stades : 6 à 10 feuilles

Ravageurs

Les **pucerons** sont encore très présents. L'effectif est soit en stabilisation pour des populations importantes soit en progression en situations d'attaques moyennes ou de recolonisation après traitement. L'arrivée de conditions plus froides devrait enfin faire régresser les populations.

Essais pucerons d'automne sur colza : voir premiers résultats page suivante.

Quelques captures de **charançon du bourgeon terminal** sont encore observées mais le vol s'est essentiellement regroupé sur la période du 16 au 22 octobre. Les larves d'**altises** sont rares sur pétioles. Des comptages larvaires faisant un bilan d'avant-hivernage du bourgeon terminal seront réalisés ultérieurement.

Au cours des dernières semaines, des larves de **mouche du chou** (adultes blancs aux collets et sur les pivots des colzas) ont été signalées en quelques endroits : Sepenoux, Delain (70), Minchun, Soucy (21)... Ces dégâts sont liés essentiellement à la précocité des dates de semis et/ou de levée. Aucune lutte n'est possible en post-levée de la culture.

Préconisations : Les interventions contre le charançon et/ou les pucerons ont dû être réalisées. Compte-tenu de la baisse des températures, peu favorable à l'activité des insectes, ne plus envisager d'intervention insecticide sur colza.

Désherbage de rattrapage sur sanves et ravenelles

Après les premières gelées hivernales qui provoquent la mise en place de la cuticule cireuse du colza, l'utilisation de Cent 7 est possible en rattrapage contre les sanves et les ravenelles (les autres crucifères sont peu sensibles au produit). L'efficacité, optimale lorsque le traitement est effectué au stade rosette des crucifères adventices, décroît à partir du stade montaison de celles-ci (sans avoir atteint cet automne) mais le produit sensibilise néanmoins les plantes au gel hivernal.

Dose d'emploi du Cent 7 : 0,4 l/ha

Précautions : proscrire toute association avec un autre produit phytosanitaire ; traiter hors d'un ralentissement.

CEREALES

Stades : Selon les dates de semis, les cultures peuvent attendre le début tallage, voire 2-3 talles pour les situations les plus avancées. Les semis de début octobre sont à leur 3ème feuille.

Ravageurs

Le vol de **pucerons** se ralentit sensiblement avec l'abaissement des températures (voir ci-après le graphe des captures à la tour d'Auxerre).

Les comptages réalisés en cultures donnent des résultats variables selon les secteurs et le stade des cultures. Le grand Val de Saône présente les populations les plus importantes (Saône-et-Loire, Jura, secteur Val de Saône et début plaine Dijonnaise en Côte-d'Or). Des niveaux atteignant 10 à 15 % de pieds porteurs sur cultures à 2 feuilles et au-delà peuvent être observés. Les autres secteurs présentent des niveaux variables mais généralement inférieurs, ne dépassant pas 3 à 5 % de pieds porteurs pour les mêmes stades.

Les **cicadelles** restent présentes dans les secteurs favorables de l'Yonne et de la Nièvre où les captures dans le réseau de cuvettes n'ont pas diminué.

Préconisations :

➢Pour les parcelles atteignant 3 feuilles-début tallage, en tous secteurs, les niveaux de population observés et la durée de présence des pucerons justifient une intervention rapide.

➢Pour les parcelles de Saône-et-Loire, Jura et Sud Est Côte-d'Or actuellement à 2 feuilles-3ème feuille pointante, l'intervention est aussi à réaliser très prochainement.

➢Pour les parcelles des autres secteurs au même stade, les populations observées dépassent rarement 3 à 5 % de pieds porteurs. L'intervention peut être encore différée.

➢Pour les parcelles n'ayant pas atteint 2 feuilles, toute intervention est actuellement prématurée. La surveillance doit cependant se maintenir dans le grand Val de Saône.

Pour les situations de l'Yonne et de la Nièvre, exposées au risque cicadelles, il convient de rester attentif aux fortes populations sur les jeunes levées et intervenir le cas échéant, même sur des stades jeunes.

COLZA

Plus d'intervention contre les ravageurs.

CEREALES

Protection anti-pucerons selon les secteurs et les stades.

P21

Essais pucerons d'automne sur colza : Premiers résultats

Lieu	Stade de la culture	Pucerons lors du traitement		Date de traitement	Conditions lors du traitement	Efficacité 7 jours après traitement			
		Fréquence de pieds avec pucerons	Nombre de pucerons par plante			Sur fréquence de pieds porteurs de pucerons		Sur nombre de pucerons par plante	
						Karaté Vert 0,15 l/ha	Karaté K 1,25 l/ha	Karaté Vert 0,15 l/ha	Karaté K 1,25 l/ha
21 Vignoles	B8	35 %	5,5	16/10	T = 15,9°C HR = 75 %	23 %	84 %	53 %	96 %
58 Authume	B8-B10	84 %	11,9	20/10	T = 22°C HR = 73 %	25 %	91 %	69 %	99 %

Deux essais d'opportunité ont été mis en place cet automne compte tenu des infestations importantes en pucerons. Les premiers résultats confirment l'efficacité insecticide des pyréthrinoïdes seules, appliquées sur populations déjà bien installées, et la nécessité de recourir dans ce cas à un aphicide spécifique. Ces essais seront conduits jusqu'au rendement pour des mesures de nuisibilité.

Céréales (suite)

Mouche jaune

Le vol vient de débuter. Compte-tenu de sa tardivité et des conditions météorologiques actuelles, le niveau d'attaque sera certainement faible. En effet les mouches jaunes sont surtout dangereuses quand les pontes se concentrent sur les toutes premières levées, ce qui ne sera pas le cas cette année.

En semis précoces, ne pas utiliser des produits agressifs si le risque de gelées persiste (en variétés plutôt sensibles au froid l'hiver dernier a montré qu'il fallait choisir des produits très sélectifs ou ne pas désherber avant l'hiver).

Les parcelles n'ayant pas atteint le stade 2-3 feuilles il convient d'attendre ; En général les levées d'adventices sont peu nombreuses et un retour des pluies en novembre peut entraîner de nouvelles germinations.

Désherbage

La baisse actuelle et les risques de gel matinaux conduisent à la prudence vis-à-vis des applications herbicides.

Deux dates à retenir

Le **samedi 15 novembre 97**, les S.R.P.V. organisent avec les Services Vétérinaires une journée d'information du public sur la qualité et la sécurité de l'alimentation.

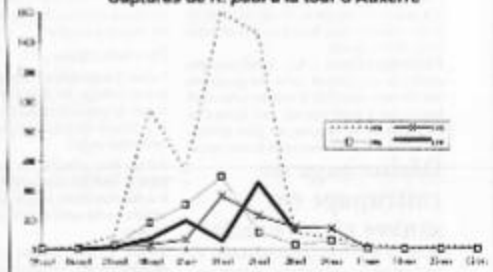
En *Bourgogne* cette journée se tiendra à Dijon - Centre Commercial Dauphine de 10 h à 19 h.

En *Franche-Comté*, au siège de la DRAF, 191, Rue de Belfort - Besançon - de 10 h à 17 h.

Le **jeudi 11 décembre 97**, aura lieu la réunion de synthèse Grandes Cultures Franche-Comté.

La journée se déroulera au Lycée Agricole de Dannemarie/Crête (LEGTA GRANVELLE).

Captures de *R. padi* à la tour d'Auxerre



Résultats JNO - Pots piège - Reulée (21)

Semaine	% de pots attribués	% de plantes infestées	Nbre de pucerons par pot	Nbre de pucerons par plante	% de pots infestés par le virus
Du 23/09 au 26/09	21	6,4	0,6	0,1	0
Du 26/09 au 02/10	66	28,0	3,2	0,7	4
Du 02/10 au 09/10	88	46,5	6,4	1,3	1
Du 9/10 au 17/10	66	25,0	3,0	0,7	4
Du 17/10 au 24/10	100	63,0	10,0	2,4	Analyse en cours

Our dataset:
Agricultural Warning Report Newsletter

- Potential Dataset is **50000** issues.
- 20000 are under paper format.
- We need to scan them. They are shared at the BNF (Bibliotheque Francois-Mitterrand). And need to make an OCR (optical character recognition) (with Jouve Corp.). The price is 50000 €
- At moment we work with an OCR sample of two regions (Bourgogne and Midi-Pyrénées). Means 1800 files (1963-2001). Plus a sample of 523 PDF files (2004-2011).
- Our actual database is **2323** files.

JOUVE

Users/usages

- Project is sponsored by French Ministry of Agriculture and French Ministry of Research

- Project includes specialists in biology and ecology of key pathogens: epidemiology and environment sciences (pest forecasts) around a network called PIC (Protection Intégrée des Cultures / Integrated Crop Protection)

- PIC was created in 2004, with 400 members
- 4 specialists (potato, wheat) from PIC are active with us.

Algorithm & Model

- **Named Entity Recognition (NER)**
  - **find names in text**
  - **classify them by type, for instance {organization:*org*, person:*per*, localisation:*loc*, protein:*prot*, disease:*dis*, ...}**

"We performed exome sequencing in a family with <Crohn's disease/DIS> (CD) and severe autoimmunity, analysed immune cell phenotype and function in affected and non-affected individuals, and performed in silico and in vitro analyses of <cytotoxic T lymphocyte-associated protein 4/PROT> (CTLA-4) structure and function."

Algorithm & Model

- **Our goal is mainly to extract relations between crops and pathogens with high level of relevance.**

- **Information Extraction = good tool in NLP.**
  - Step 1 named entity recognition
  - Step 2 relation extraction
  - Step 3 extraction of context information (Pest significance, development stage, climat, location).

Information Extraction

- Named entity recognition (NER) is considered as a solved problem
- Pattern-based algorithms (expert rules)
- Statistical Learning algorithms (HMM, MAXENT, CRF)

- First conferences MUC in 1987 (message understanding conferences)
- Lasts conferences BioCreativ in bioinformatics (2008), CONLL for news (2003)
- Score > 90% for NER , human ~97%

- Relation Extraction is more difficult (generally F-score < 0.6)

Information Extraction: extern ontology database

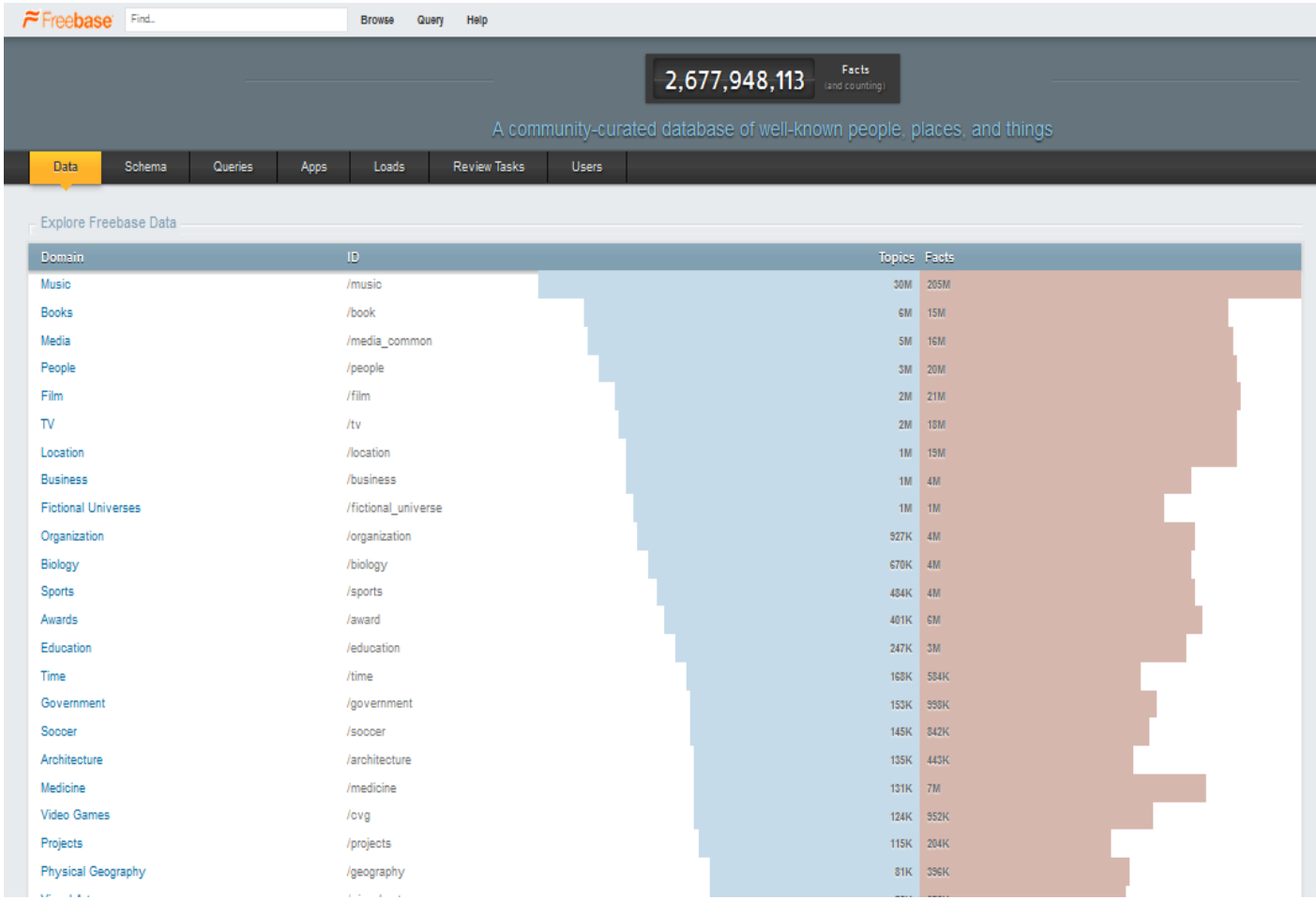
Generic ones

FREEBASE
3 billions entities, with links
Yesterday ←→ John Lennon, Paul McCartney
YAGO
10 millions entities/120 millions facts
WORDNET
WIKIPEDIA

But generally not sufficiantly specific for a specialized domain.

Specific ones

GENE ONTOLOGY
AGROVOC (UNO)



Information Extraction: NER systems

- OpenNLP nameFinder
- LingPipe NER system
- IllinoisNER system
  - 90.6 F1 (CoNLL03 NER shared task data)
- SNER (Stanford NER)
  - 86.86 F1 (CoNLL03 NER shared task data)
- LAITOR ***L**iterature **A**ssistant for **I**dentification of **T**erms co-**O**ccurrences and **R**elationships*
- HIT LTP NER system

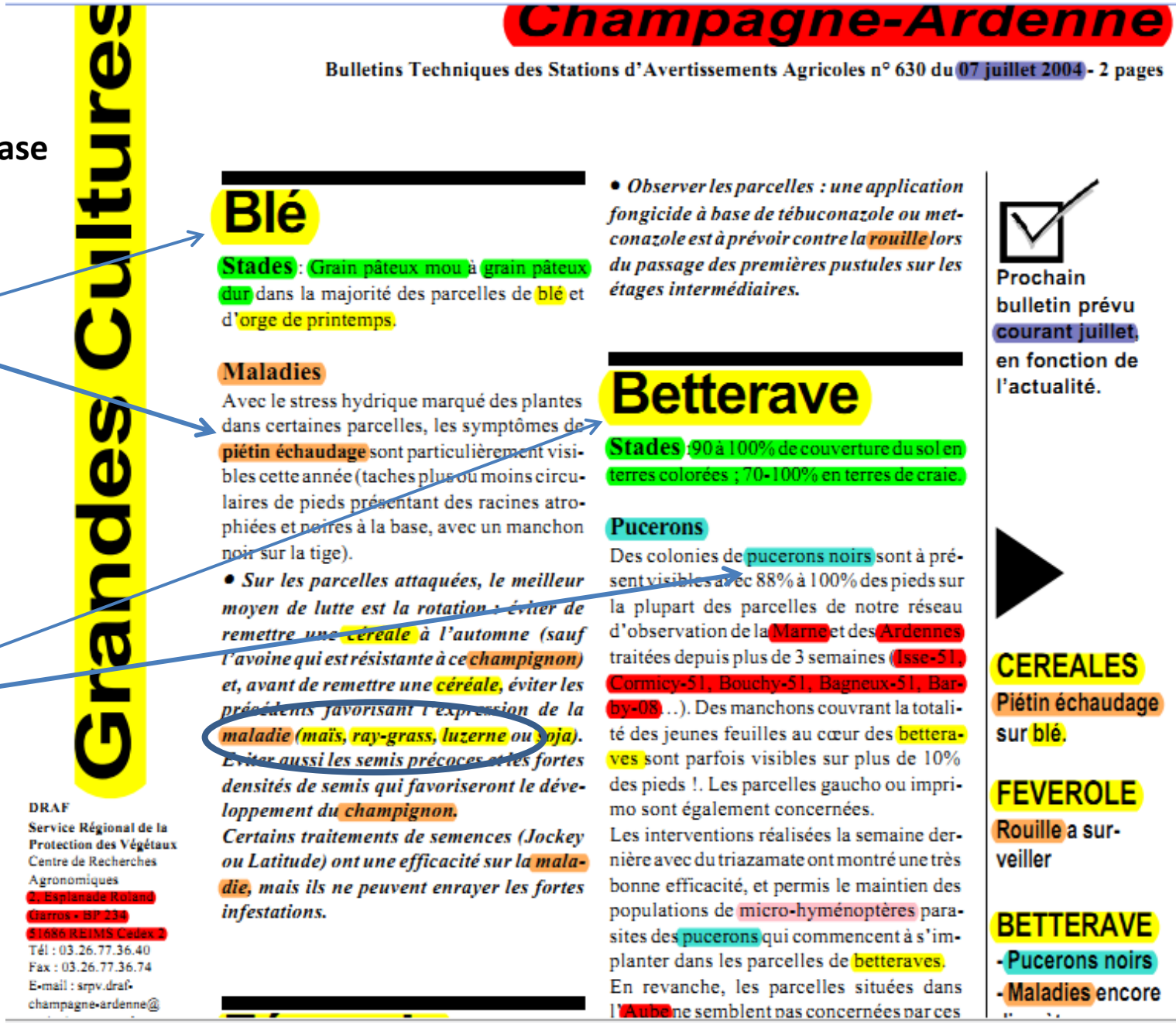
Information Extraction: relation extraction

Relation Crop-Disease

Instance 1
wheat/ 'take-all-disease

Relation Crop-Pest

Instance
Beetroot/black aphid



Champagne-Ardenne

Bulletins Techniques des Stations d'Avertissements Agricoles n° 630 du 07 juillet 2004 - 2 pages

Grandes Cultures

Blé

Stades : Grain pâteux mou à grain pâteux dur dans la majorité des parcelles de blé et d'orge de printemps.

Maladies

Avec le stress hydrique marqué des plantes dans certaines parcelles, les symptômes de piétin échaudage sont particulièrement visibles cette année (taches plus ou moins circulaires de pieds présentant des racines atrophiées et noires à la base, avec un manchon noir sur la tige).

• *Sur les parcelles attaquées, le meilleur moyen de lutte est la rotation : éviter de remettre une céréale à l'automne (sauf l'avoine qui est résistante à ce champignon) et, avant de remettre une céréale, éviter les précédents favorisant l'expression de la maladie (maïs, ray-grass, luzerne ou soja). Eviter aussi les semis précoces et les fortes densités de semis qui favoriseront le développement du champignon.*

Certains traitements de semences (Jockey ou Latitude) ont une efficacité sur la maladie, mais ils ne peuvent enrayer les fortes infestations.

• *Observer les parcelles : une application fongicide à base de tébuconazole ou metconazole est à prévoir contre la rouille lors du passage des premières pustules sur les étages intermédiaires.*

Betterave

Stades : 90 à 100% de couverture du sol en terres colorées ; 70-100% en terres de craie.

Pucerons

Des colonies de pucerons noirs sont à présent visibles avec 88% à 100% des pieds sur la plupart des parcelles de notre réseau d'observation de la Marne et des Ardennes traitées depuis plus de 3 semaines (Isse-51, Cormicy-51, Bouchy-51, Bagneux-51, Barby-08…). Des manchons couvrant la totalité des jeunes feuilles au cœur des betteraves sont parfois visibles sur plus de 10% des pieds !. Les parcelles gaucho ou imprimo sont également concernées.

Les interventions réalisées la semaine dernière avec du triazamate ont montré une très bonne efficacité, et permis le maintien des populations de micro-hyménoptères parasites des pucerons qui commencent à s'implanter dans les parcelles de betteraves.

En revanche, les parcelles situées dans l'Aube ne semblent pas concernées par ces

Prochain bulletin prévu courant juillet, en fonction de l'actualité.

CEREALES

Piétin échaudage sur blé.

FEVEROLE

Rouille a surveiller

BETTERAVE

- Pucerons noirs
- Maladies encore

DRAF
Service Régional de la Protection des Végétaux
Centre de Recherches Agronomiques
2, Esplanade Roland Garros - BP 234
51686 REIMS Cedex 2
Tél : 03.26.77.36.40
Fax : 03.26.77.36.74
E-mail : srpv.draf-champagne-ardenne@

Information Extraction: NER our approach

- Dictionary domain-based
  - Crops, diseases, pests, auxiliaries, region, towns , chemicals

7 concepts

blé:N:blé:BLE:blés:Triticum:blé dur:blé tendre:
blé dur:L:BLE DUR:T. durum:Triticum durum:bles durs:blés durs:blé dur:
blé noir:L:BLE NOIR:f. esculentum:fagopyrum esculentum:sarrasin:bles noirs:blés noirs:blé noir:sarrasins:
blé tendre:L:BLE TENDRE:T. aestivum:Triticum aestivum:blé froment:blés froments:ble froments:blé tendre:blés tendres:bles tendres:

wheat *(species)*
durum wheat *(variety)*

buckwheat *(variety)*

soft wheat *(variety)*

	entities						
	auxiliaries	crops	pests	diseases	chemicals	regions	towns
#entries	28	114	373	275	4968	26	33161
#leafs	28	103	334	241	4968	26	33161
#concepts	0	18	53	40	0	0	0
#lexems	107	727	2673	1846	4968	869	89603

Information Extraction: NER our approach

- Evaluation

37 annotated files (with entities and relations)
Annotation Cost: 1000 files ~5 months work by 1 person

CoNLL or BIO/BILOU format

Champagne-Ardenne		REG
.	O	
BSV	O	
du	O	
09/06/2011	O	
--	O	
semaine	O	
23	O	
A	O	
RETENIR	O	
CETTE	O	
SEMAINE	O	
.	O	
TOURNESOL	PLA	
:	O	
Pucerons	BIO	
:	O	
fin	O	
du	O	
risque	O	
.	O	
Absence	O	
de	O	
maladies	O	
.	O	
MAïS	PLA	
:	O	
Pyrale	BIO	
:	O	

Our format

5_-_BSV_CHAMPAGNE-ARDENNE_COLZA_2010_03_25_cle09c576:r:$:CHAMPAGNE-ARDENNE:
5_-_BSV_CHAMPAGNE-ARDENNE_COLZA_2010_03_25_cle09c576:p:$:colza:
5_-_BSV_CHAMPAGNE-ARDENNE_COLZA_2010_03_25_cle09c576:b:$:charançon de la tige du colza:charançon de la tige du chou:Méligèthes:mouche du chou:
5_-_BSV_CHAMPAGNE-ARDENNE_COLZA_2010_03_25_cle09c576:d:$:25.03.2010:
5_-_BSV_CHAMPAGNE-ARDENNE_COLZA_2010_03_25_cle09c576:v:$:St Dizier:Reims:Charleville:Langres:TROYES:
5_-_BSV_CHAMPAGNE-ARDENNE_COLZA_2010_03_25_cle09c576:n:$:peu nuisible:à risque:
5_-_BSV_CHAMPAGNE-ARDENNE_COLZA_2010_03_25_cle09c576:s:$:c2:c1:d1:f1:
5_-_BSV_CHAMPAGNE-ARDENNE_COLZA_2010_03_25_cle09c576:p:b:$$:colza:charançon de la tige du colza:1
5_-_BSV_CHAMPAGNE-ARDENNE_COLZA_2010_03_25_cle09c576:p:b:$$:colza:mouche du chou:1
5_-_BSV_CHAMPAGNE-ARDENNE_COLZA_2010_03_25_cle09c576:p:b:$$:colza:Méligèthes:1
5_-_BSV_CHAMPAGNE-ARDENNE_COLZA_2010_03_25_cle09c576:p:s:$$:colza:c2:1
5_-_BSV_CHAMPAGNE-ARDENNE_COLZA_2010_03_25_cle09c576:p:s:$$:colza:d1:1
5_-_BSV_CHAMPAGNE-ARDENNE_COLZA_2010_03_25_cle09c576:p:s:$$:colza:c1:1
5_-_BSV_CHAMPAGNE-ARDENNE_COLZA_2010_03_25_cle09c576:p:b:n:$$:colza:charançon de la tige du chou:peu nuisible:1
5_-_BSV_CHAMPAGNE-ARDENNE_COLZA_2010_03_25_cle09c576:p:b:n:$$:colza:mouche du chou:à risque:1

Information Extraction: NER our approach

- Evaluation

$$P = \frac{\#correct\_answers}{\#produced\_answers}$$

$$R = \frac{\#correct\_answers}{\#possible\_correct\_answers}$$

$$F - score = \frac{(\beta^2 + 1)P.R.}{(\beta^2 R + P)}$$

SNER - Average						
Entity	**P**	**R**	**F1**	**TP**	**FP**	**FN**
BIO	92,66%	71,41%	**80,52%**	184,32	14,79	76,18
MAL	95,46%	77,38%	**85,38%**	98,46	4,68	30,39
PLA	93,99%	82,68%	**87,94%**	277,14	16,86	56,11
REG	93,20%	73,73%	**81,92%**	40,18	3,36	16,82
Totals	93,68%	76,85%	**84,41%**	600,11	39,68	179,50

X.ent						
Entity	**P**	**R**	**F1**	**X_Y**	**X**	**Y**
BIO	96,46%	95,52%	**95,98%**	299,00	313,00	310,00
MAL	96,97%	95,53%	**96,24%**	192,00	201,00	198,00
PLA	88,80%	98,67%	**93,47%**	222,00	225,00	250,00
REG	100%	100%	**100%**	37,00	37,00	37,00
Totals	94,33%	96,67%	**95,48%**	750,00	766,00	795,00

Information Extraction: NER our approach

- Rule-based

  other kinds of extraction, not specifically named entities

  5 concepts

  – Developmental stages, Risk assessment, Climat, Number of issue , Date

  – For towns , mixed approach rule-based and dictionary-based.

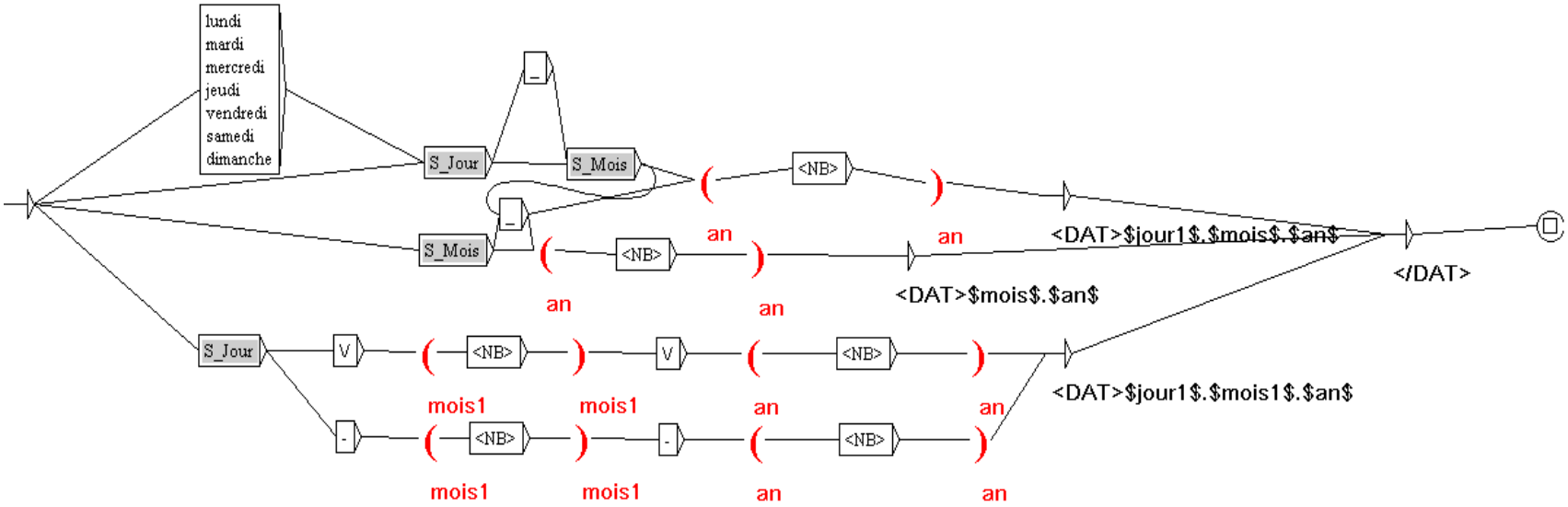
- No use of dictionaries but local grammar
  – Handcrafted-rules with markers : such as « week » for *Number of issue*, of « xx {January | February…} xxxx » for a *date*
  – We use UNITEX tool to make rule . It is a Finite-State Automata Graphical Tool
  – Easy to make language models.

Information Extraction: NER our approach

- Unitex rules

http://www-igm.univ-mlv.fr/~unitex/

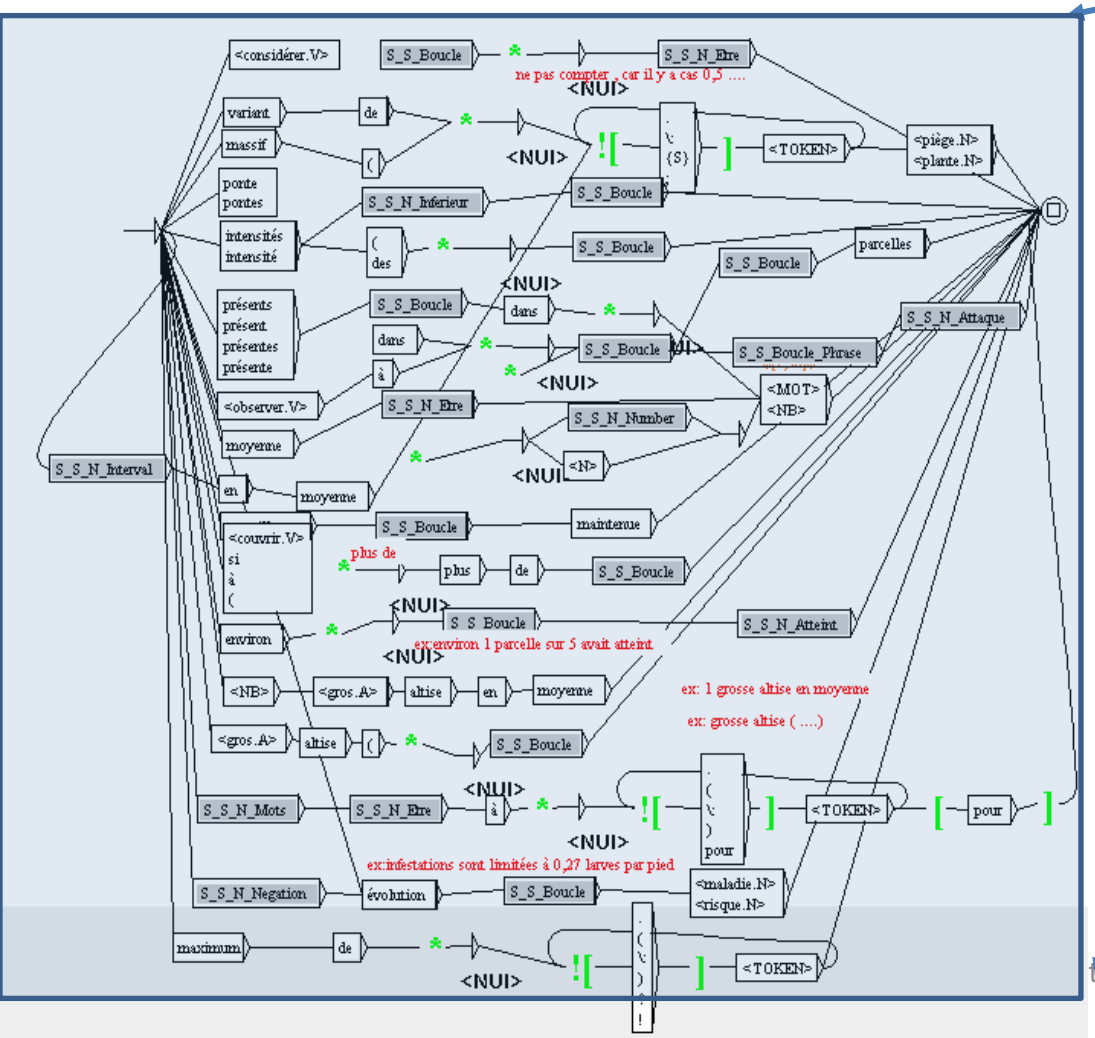
  - Dates detection

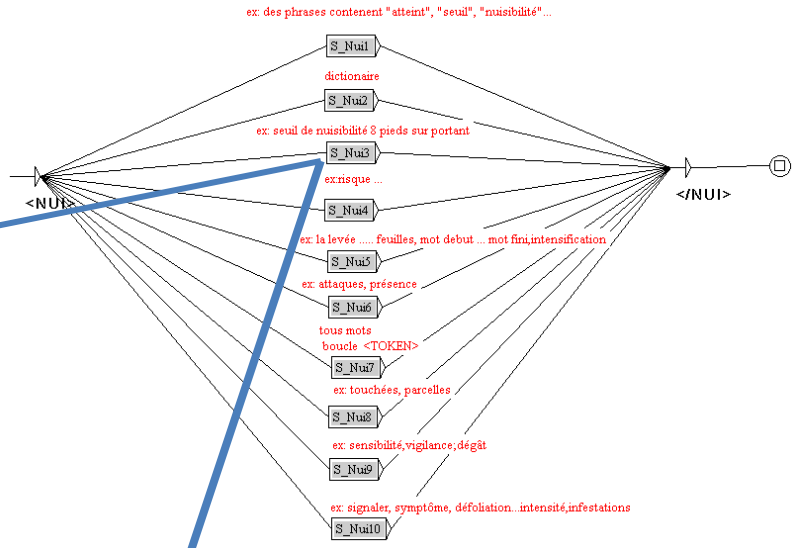


« 15 janvier 1992 » ou « 10-2012 »

Information Extraction: NER our approach

- Unitex rules
  - Risk Assessment detections





infestations sont limitées à 0,27 larves par pied
environ 1 parcelle sur 5 avait atteint
1 grosse altise en moyenne

infestations are limited to 0.27 larvae per foot
about 1 parcel in 5 had reached
1 large flea beetle in average

Information Extraction: relation extraction

These are the main approaches for relation extraction.

Qualitative approaches:

- Exact analysis from lexical dictionary ('Exact Dictionary-Based Chunking')
- Hand-crafted pattern definition techniques

And optimisation approaches:

- Symbolic Learning Models ('inductive-logic programming')
- Statistical Learning Models ('Bayesian network analysis')
- Unsupervised Learning Models ('co-occurrence analysis')

Information Extraction: relation extraction

Our approach is a combination between :

- *Hand-crafted pattern definition techniques*

In the sense that we consider some **rules about the document design**.

&

- *Unsupervised Learning Models*

In the sense that we considere **detection of cooccurrence** without POS tagging attachment.

Information Extraction: relation extraction

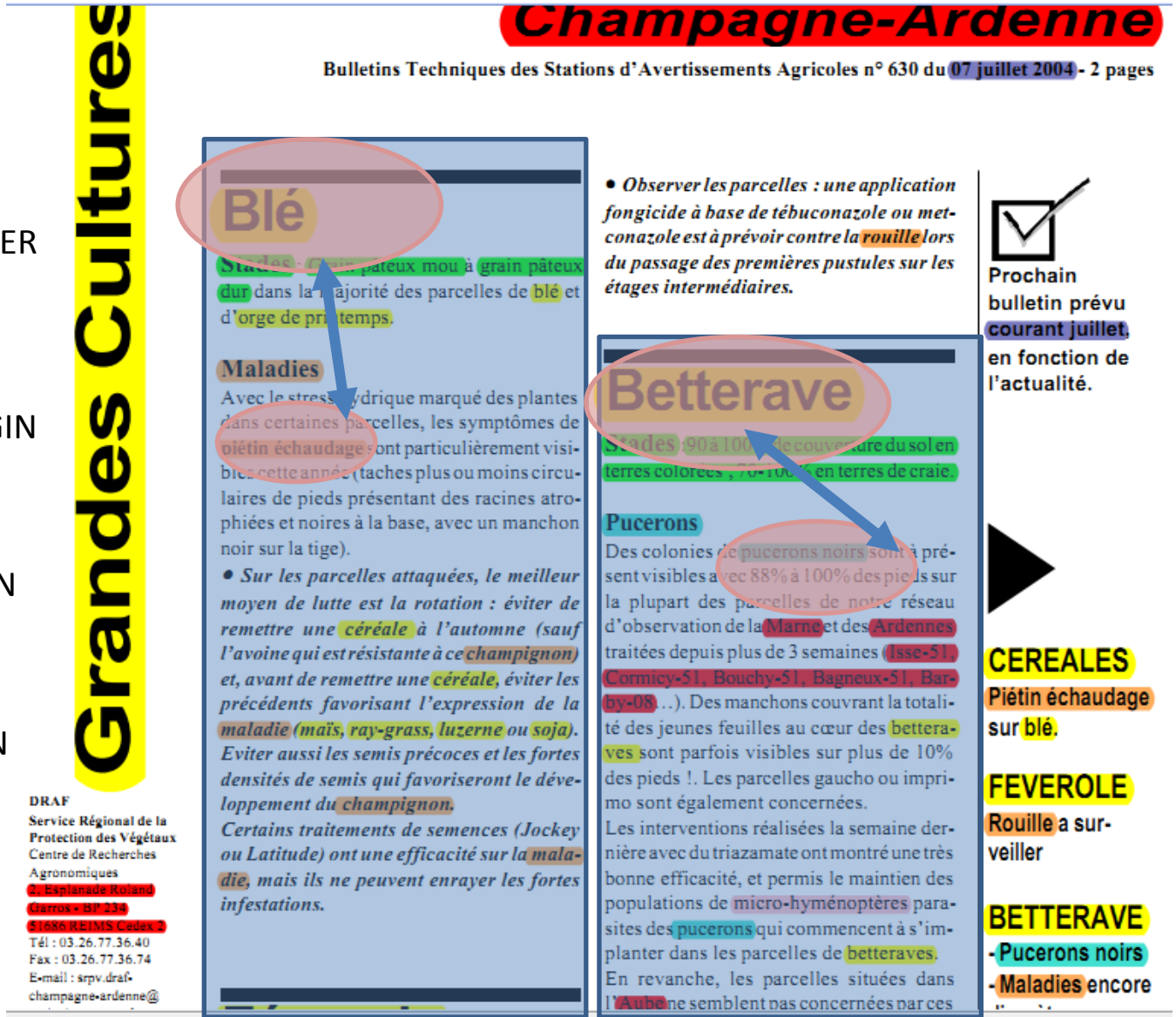
Rule 1:
Named Entity(crop) @ HEADER

Rule 2:
Named Entity(region) @ BEGIN

Rule 3:
Named Entity(issue) @ BEGIN

Rule 4:
Named Entity(date) @ BEGIN

Rule 5:
Reject Named Entity(*)
IN AVOID_PART



Champagne-Ardenne

Bulletins Techniques des Stations d'Avertissements Agricoles n° 630 du 07 juillet 2004 - 2 pages

Grandes Cultures

Blé

Stades : Grain pâteux mou à grain pâteux dur dans la majorité des parcelles de blé et d'orge de printemps.

Maladies

Avec le stress hydrique marqué des plantes dans certaines parcelles, les symptômes de piétin échaudage sont particulièrement visibles cette année (taches plus ou moins circulaires de pieds présentant des racines atrophiées et noires à la base, avec un manchon noir sur la tige).

• *Sur les parcelles attaquées, le meilleur moyen de lutte est la rotation : éviter de remettre une céréale à l'automne (sauf l'avoine qui est résistante à ce champignon) et, avant de remettre une céréale, éviter les précédents favorisant l'expression de la maladie (maïs, ray-grass, luzerne ou soja). Eviter aussi les semis précoces et les fortes densités de semis qui favoriseront le développement du champignon.*

Certains traitements de semences (Jockey ou Latitude) ont une efficacité sur la maladie, mais ils ne peuvent enrayer les fortes infestations.

• *Observer les parcelles : une application fongicide à base de tébuconazole ou metconazole est à prévoir contre la rouille lors du passage des premières pustules sur les étages intermédiaires.*

Betterave

Stades : 90 à 100 % de couverture du sol en terres colorées ; 70-100 % en terres de craie.

Pucerons

Des colonies de pucerons noirs sont à présent visibles avec 88% à 100% des pieds sur la plupart des parcelles de notre réseau d'observation de la Marne et des Ardennes traitées depuis plus de 3 semaines (Isse-51, Cormicy-51, Bouchy-51, Bagneux-51, Barby-08…). Des manchons couvrant la totalité des jeunes feuilles au cœur des betteraves sont parfois visibles sur plus de 10% des pieds !. Les parcelles gaucho ou imprimo sont également concernées.

Les interventions réalisées la semaine dernière avec du triazamate ont montré une très bonne efficacité, et permis le maintien des populations de micro-hyménoptères parasites des pucerons qui commencent à s'implanter dans les parcelles de betteraves.

En revanche, les parcelles situées dans l'Aube ne semblent pas concernées par ces

Prochain bulletin prévu courant juillet, en fonction de l'actualité.

CEREALES

Piétin échaudage sur blé.

FEVEROLE

Rouille a surveiller

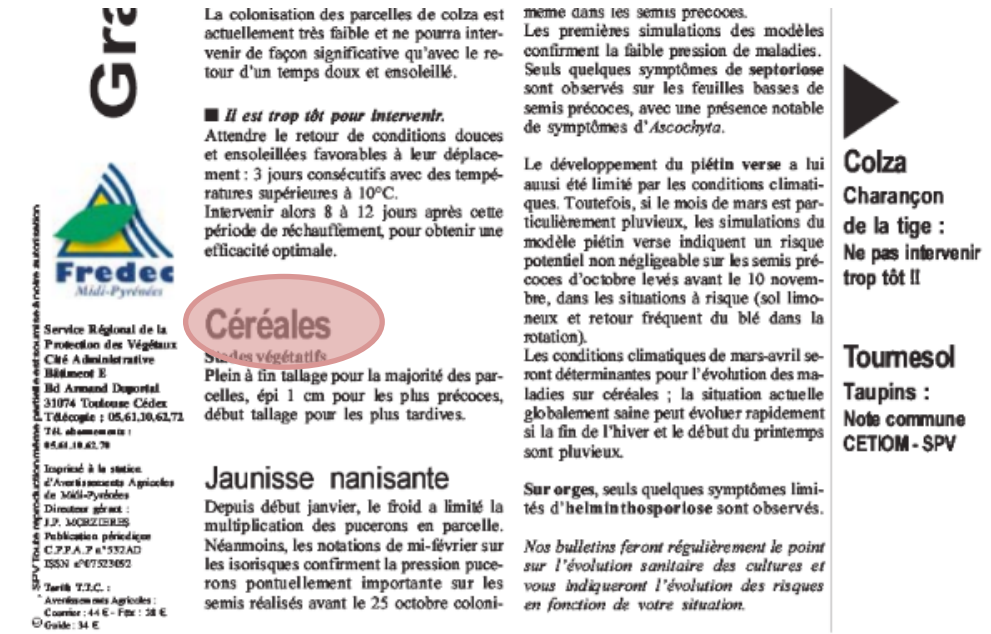
BETTERAVE

- Pucerons noirs
- Maladies encore

DRAF
Service Régional de la Protection des Végétaux
Centre de Recherches Agronomiques
2, Esplanade Roland Garros - BP 234
51686 REIMS Cedex 2
Tél : 03.26.77.36.40
Fax : 03.26.77.36.74
E-mail : srpv.draf-champagne-ardenne@

Information Extraction: relation extraction

Rule 5:
Reject Named Entity(*)
IN AVOID_PART



Gra

La colonisation des parcelles de colza est actuellement très faible et ne pourra intervenir de façon significative qu'avec le retour d'un temps doux et ensoleillé.

■ ***Il est trop tôt pour intervenir.***
Attendre le retour de conditions douces et ensoleillées favorables à leur déplacement : 3 jours consécutifs avec des températures supérieures à 10°C.
Intervenir alors 8 à 12 jours après cette période de réchauffement, pour obtenir une efficacité optimale.

Fredec
Midi-Pyrénées

Service Régional de la Protection des Végétaux
Cité Administrative
Bâtiment E
Bd Armand Duportal
31074 Toulouse Cédex
Tél.écogie : 05.61.10.62.72
Tél. abonnements : 05.61.10.62.70

Imprimé à la station d'Avertissements Agricoles de Midi-Pyrénées
Directeur gérant : J.P. MORZIERES
Publication périodique
C.P.P.A.P n°532AD
ISSN n°07523652

Tarifs T.T.C. :
Avertissements Agricoles :
Courrier : 44 € - Fax : 38 €
Guide : 34 €

Céréales

Stades végétatifs

Plein à fin tallage pour la majorité des parcelles, épi 1 cm pour les plus précoces, début tallage pour les plus tardives.

Jaunisse nanisante

Depuis début janvier, le froid a limité la multiplication des pucerons en parcelle. Néanmoins, les notations de mi-février sur les isorisques confirment la pression pucerons ponctuellement importante sur les semis réalisés avant le 25 octobre colonisation même dans les semis précoces.
Les premières simulations des modèles confirment la faible pression de maladies. Seuls quelques symptômes de **septoriose** sont observés sur les feuilles basses de semis précoces, avec une présence notable de symptômes d'*Ascochyta*.

Le développement du **piétin verse** a lui aussi été limité par les conditions climatiques. Toutefois, si le mois de mars est particulièrement pluvieux, les simulations du modèle piétin verse indiquent un risque potentiel non négligeable sur les semis précoces d'octobre levés avant le 10 novembre, dans les situations à risque (sol limoneux et retour fréquent du blé dans la rotation).
Les conditions climatiques de mars-avril seront déterminantes pour l'évolution des maladies sur céréales ; la situation actuelle globalement saine peut évoluer rapidement si la fin de l'hiver et le début du printemps sont pluvieux.

Sur orges, seuls quelques symptômes limités d'**helminthosporiose** sont observés.

Nos bulletins feront régulièrement le point sur l'évolution sanitaire des cultures et vous indiqueront l'évolution des risques en fonction de votre situation.

Colza
Charançon de la tige :
Ne pas intervenir trop tôt !!

Tournesol
Taupins :
Note commune CETIOM - SPV

Reasoning control against

wireworms

Instance: crop/wireworms is false

Raisonner la lutte contre les larves de taupins en tournesol
Note commune CETIOM - SPV : Février 2005

L'objectif de cette note commune CETIOM-SPV est de présenter le raisonnement de la lutte contre les larves de taupins sur tournesol, tel qu'il peut être proposé pour la campagne 2005, en l'état actuel de nos connaissances.

De nouvelles données, recueillies dans le cadre d'un réseau de surveillance du risque taupins (dispositif présenté à la fin de cette note) mis en place dès cette année, sont susceptibles de faire évoluer ces conseils de lutte.

1/ Evaluation actuelle du risque taupin en tournesol

Dans le contexte actuel de production du tournesol, **le risque d'attaque est globalement faible pour cette culture**. Ce risque est limité pour deux raisons : le tournesol est peu attractif pour les larves de taupins et les situations favorables sont peu fréquentes dans les systèmes de culture actuels incluant le tournesol.

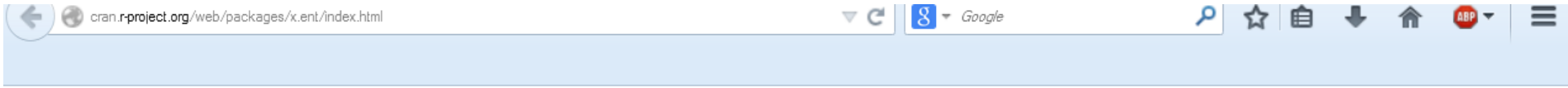
- **Une assez faible sensibilité du tournesol :**

Le tournesol fait partie des grandes cultures susceptibles de subir des attaques de larves de taupins. Cependant les références disponibles montrent, d'une part, que cette culture est faiblement attractive pour les larves et que, d'autre part, la période de sensibilité aux attaques est relativement brève (de la germination de la graine au stade cotylédons). La sensibilité globale du tournesol aux larves de taupins est donc assez faible, dans tous les cas **bien inférieure au maïs**.

- **Des systèmes de cultures incluant le tournesol peu favorables :**

Le tournesol est majoritairement cultivé dans des rotations à base de cultures annuelles. Dans ces rotations, le

Implementation: x.ent



`x.ent: eXtraction of ENTity`

x.ent is a system for extracting information (entities and relations between them) in text datasets. It also emphasizes results exploration with graphical displays. It is a rule-based system and works with hand-made dictionaries and local grammars defined by users. x.ent has been written in perl and use javascript to define user preferences through a browser. Local grammars are defined and compiled with the tool Unitex, developed by University Paris Est, and supporting multiple languages. See ?xconfig for an introduction.

Version:	1.0.6
Depends:	R (≥ 3.0.0), opencpu, rJava
Imports:	stringr, xtable, jsonlite, venneuler, ggplot2
Published:	2014-11-03
Author:	Nicolas Turenne, Tien T. Phan
Maintainer:	Tien T. Phan <phantien84 at gmail.com>
License:	GPL-3
NeedsCompilation:	no
SystemRequirements:	Perl (>= 5.0), Unitex (>= 3.0 http://www-igm.univ-mlv.fr/~unitex/)
Materials:	README NEWS
CRAN checks:	x.ent results

`Downloads:`

Reference manual:	x.ent.pdf
Package source:	x.ent_1.0.6.tar.gz
Windows binaries:	r-devel: x.ent_1.0.6.zip, r-release: x.ent_1.0.6.zip, r-oldrel: x.ent_1.0.6.zip
OS X Snow Leopard binaries:	r-release: x.ent_1.0.6.tgz, r-oldrel: x.ent_1.0.6.tgz
OS X Mavericks binaries:	r-release: x.ent_1.0.6.tgz

x.ent xconfig()

configuration settling (in a browser)
is set in 5 parts : paths, dictionaries, Unitex graphs, Relations and special files

Directory corpus:* E:\DropBox\Dropbox\projet_VESPA\data\test_x.ent\data ex: C:\data\corpus

File result:* E:\DropBox\Dropbox\projet_VESPA\data\test_x.ent\output.txt ex: C:\data\out\output.txt

File evaluation: E:\DropBox\Dropbox\projet_VESPA\data\test_x.ent\eval\eval_utf8.txt I don't use this feature

(*) *This field is mandatory! The file or folder have a full path.*

Dico

Tag result	File dico	Node	Column key	Comumn value	Get value	Action
	Parcourir... Aucun fichier sélectionné.	false				Add
p	dico-p_v3.txt	true	1	3..*	all	Delete
b	dico-b_v3.txt	true	1	3..*	all	Delete
m	dico-m_v3.txt	true	1	3..*	all	Delete
a	dico-a_v2.txt	true	1	3..*	all	Delete
r	dico-r_v2.txt	false	5	4,6..*	1	Delete
v	dico-d_v1.txt	false	6	3,5	all	Delete
f	dico-f_v3.txt	false	1	2..*	all	Delete
s	dico-s_v2.txt	false	1	2..*	all	Delete

Unitex *If you don't have Unitex, you can download here* I don't use Unitex

Tool unitex:* E:\Unitex3.1beta\App\UnitexToolLogger.exe

Main graph:* E:\DropBox\Dropbox\projet_VESPA\data\test_x.ent\unitex\VESPA.grf

Directory local unitex:* E:\unitex\Resu\French

Dico unitex		Action
Path:		Add
	E:\DropBox\Dropbox\projet_VESPA\DicoUnitex\dela-fr-public.bin	Delete
	E:\DropBox\Dropbox\projet_VESPA\DicoUnitex\Delaf_Communes_France_FR_utf8.bin	Delete
	E:\DropBox\Dropbox\projet_VESPA\DicoUnitex\Delaf_Toponyme_Departement_France_FR_utf8.bin	Delete
	E:\DropBox\Dropbox\projet_VESPA\DicoUnitex\Delaf_Toponyme_Region_France_FR_utf8.bin	Delete

Mapping between tag of result with tag of unitex

Tag result	Tag Unitex	Get value	Action
			Add
n	NUI	all	Delete
s	STA	all	Delete
z	NEG	all	Delete
c	CLI	all	Delete
d	DAT	1	Delete
v	VIL	all	Delete

Relation I don't use this feature

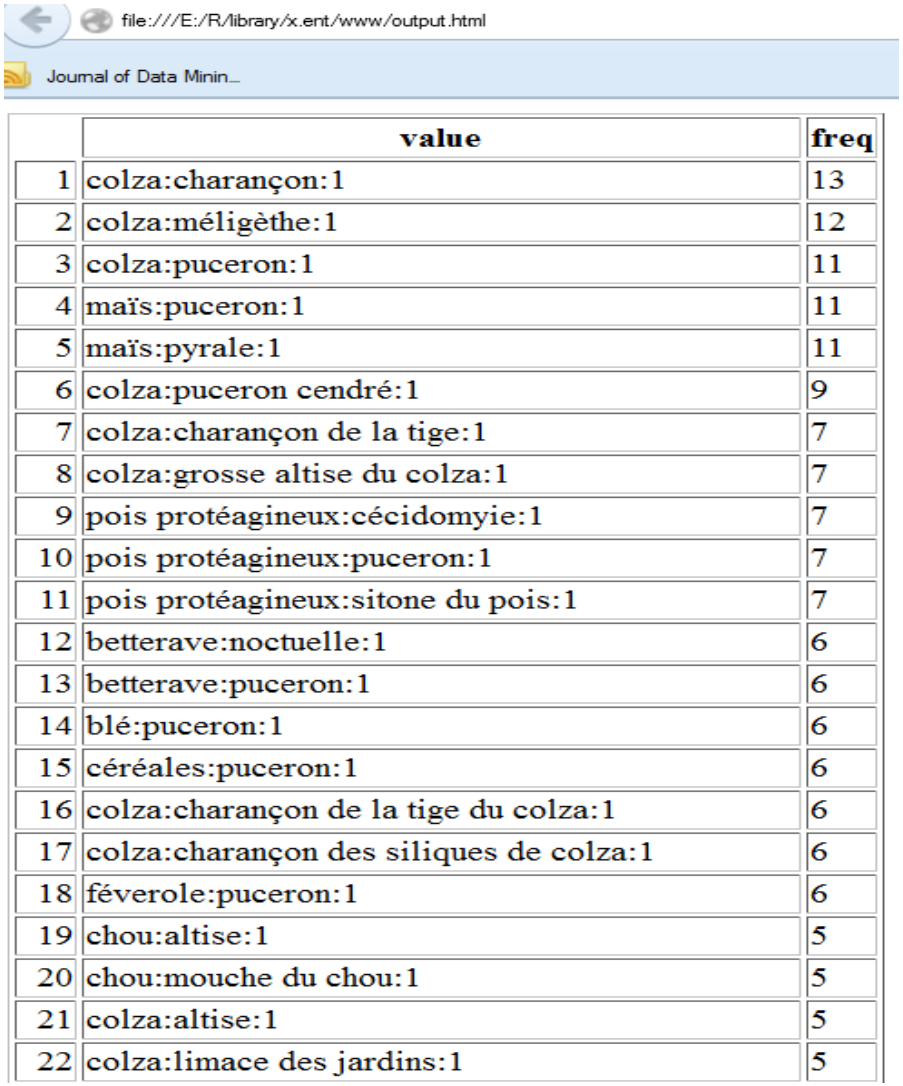
Root:* p

Negation:* z

Relation between entities		Action
Relation:		Add
	p:b	Delete
	p:s	Delete
	p:m	Delete
	p:c	Delete
	p:a	Delete
	p:b:n	Delete
	p:m:n	Delete

x.ent xshow()

```
> xshow("p:b",sort="f" )
```

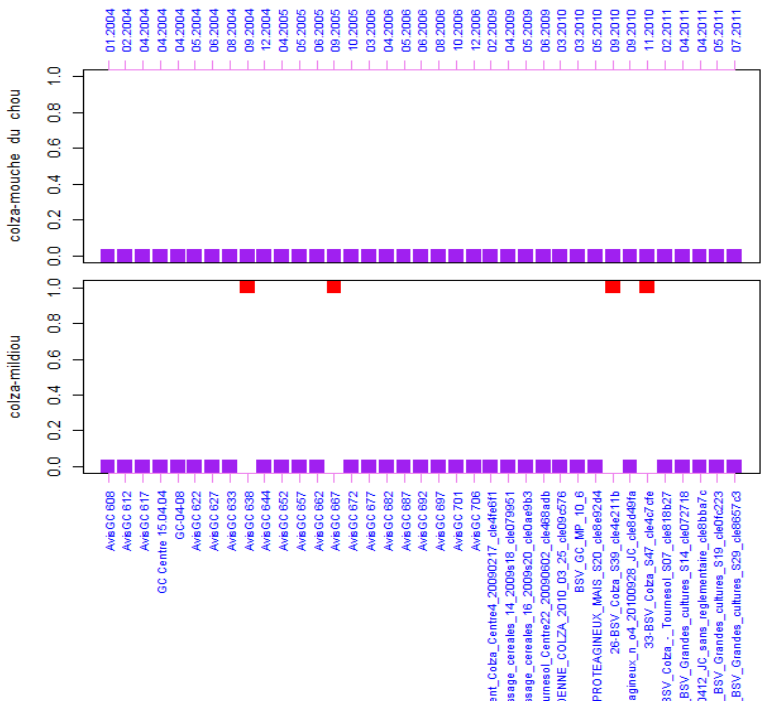


	value	freq
1	colza:charançon:1	13
2	colza:méligèthe:1	12
3	colza:puceron:1	11
4	maïs:puceron:1	11
5	maïs:pyrale:1	11
6	colza:puceron cendré:1	9
7	colza:charançon de la tige:1	7
8	colza:grosse altise du colza:1	7
9	pois protéagineux:cécidomyie:1	7
10	pois protéagineux:puceron:1	7
11	pois protéagineux:sitone du pois:1	7
12	betterave:noctuelle:1	6
13	betterave:puceron:1	6
14	blé:puceron:1	6
15	céréales:puceron:1	6
16	colza:charançon de la tige du colza:1	6
17	colza:charançon des siliques de colza:1	6
18	féverole:puceron:1	6
19	chou:altise:1	5
20	chou:mouche du chou:1	5
21	colza:altise:1	5
22	colza:limace des jardins:1	5

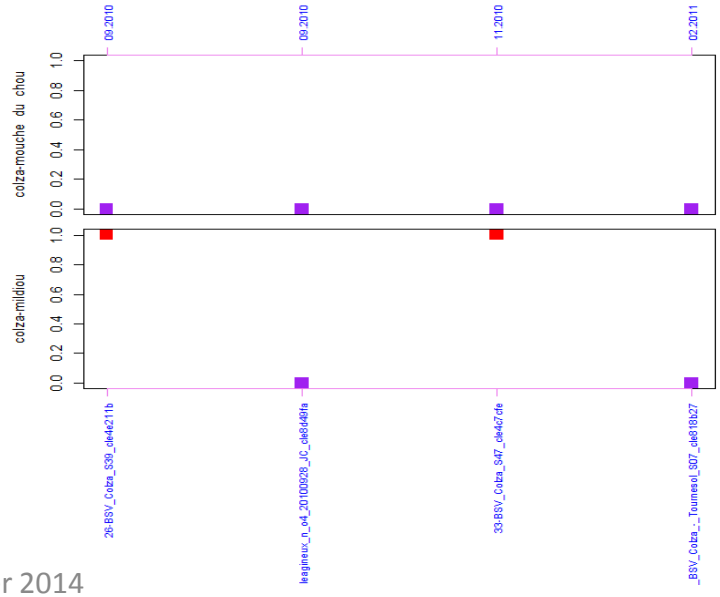
x.ent xplot()

Parallel coordinate display

xplot(e1="colza",e2=c("mouche du chou", "mildiou"))

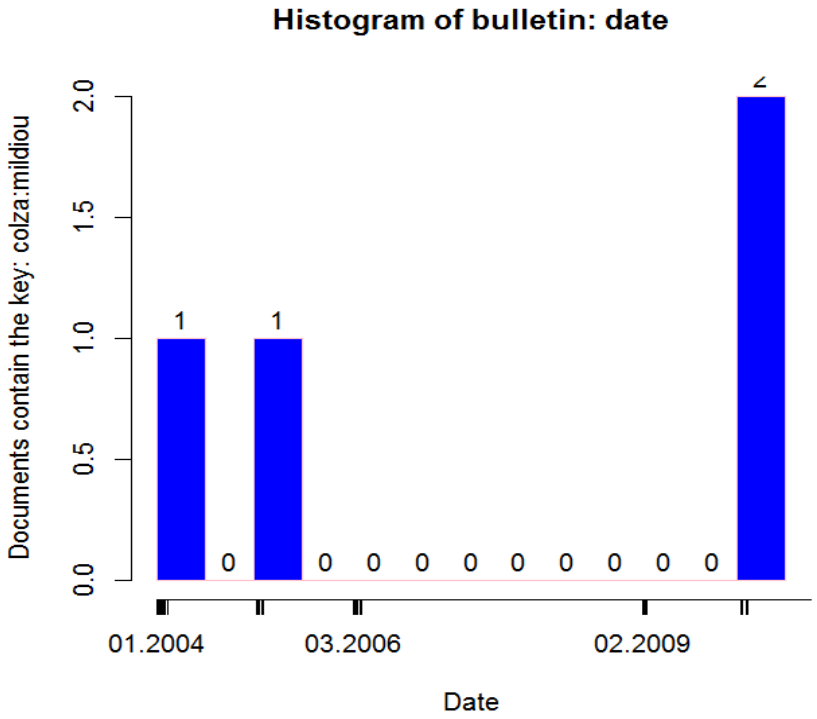


xplot(e1="colza",e2=c("mouche du chou", "mildiou"),t=c("09.2010","02.2011"))



Chronological histogram

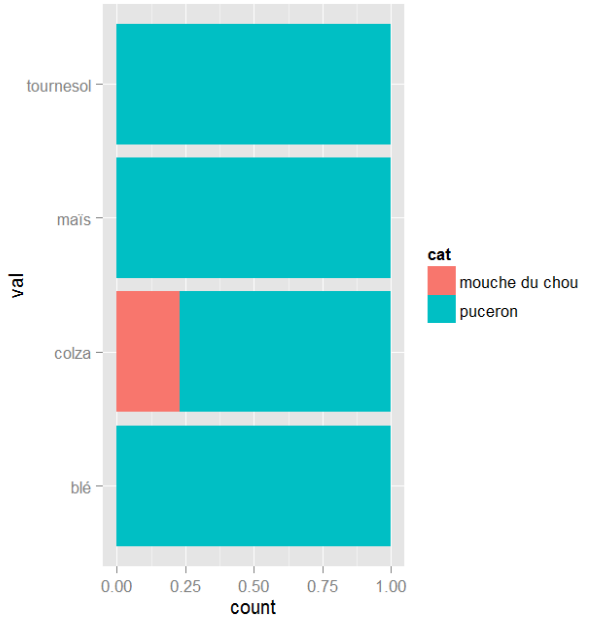
xhist("colza:mildiou")



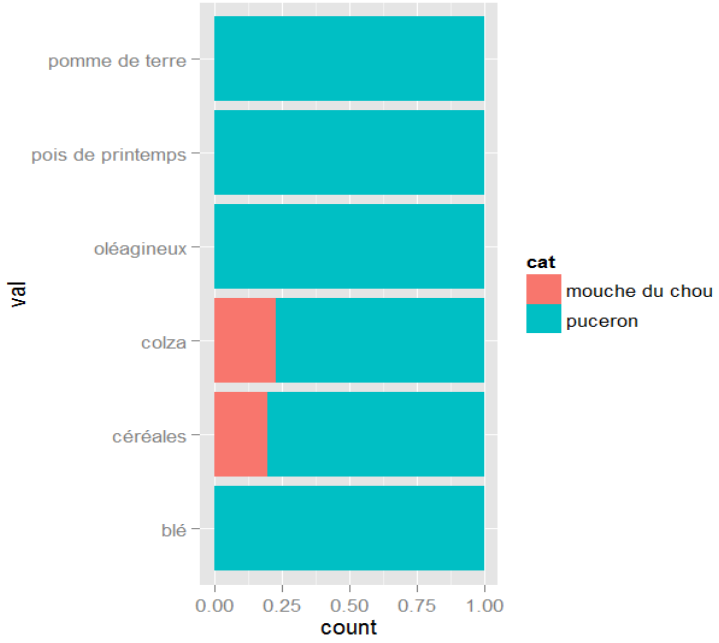
x.ent xprop()

stacked bar graph

xprop(c("blé", "maïs", "tournesol", "colza") , c("mouche du chou", "puceron"))

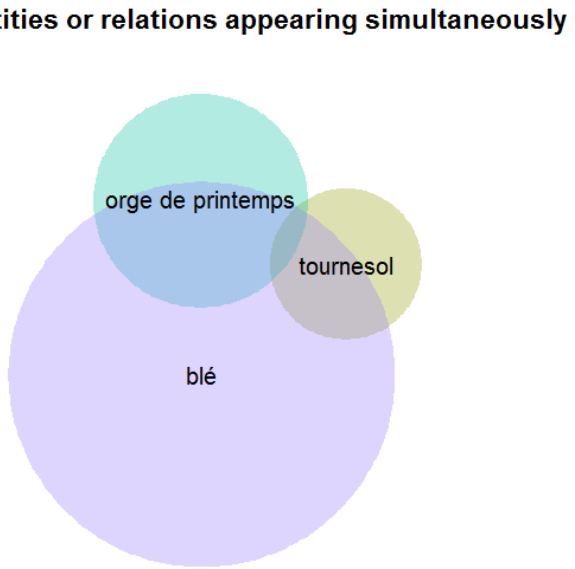


xprop(as.vector(xdata_value("p")$value[xdata_value("p")$freq>2]) , c("mouche du chou", "puceron"))



Venn diagram

xvenn(v=c("blé","orge de printemps","tournesol"),e=c("b","m"))



xtest()

Pairwise relations comparison

	relation	KOLMOGOROV	WILCOXON	STUDENT	GrowthCurves
700	blé:méligèthe/blé:thrips	1.00	0.13	0.13	0.02
543	blé:cicadelle/blé:pyrale	1.00	0.00	0.00	0.02
613	blé:criocère/blé:thrips	1.00	0.00	0.00	0.02
689	blé:méligèthe/blé:puceron des épis de céréales	0.91	0.00	0.00	0.02

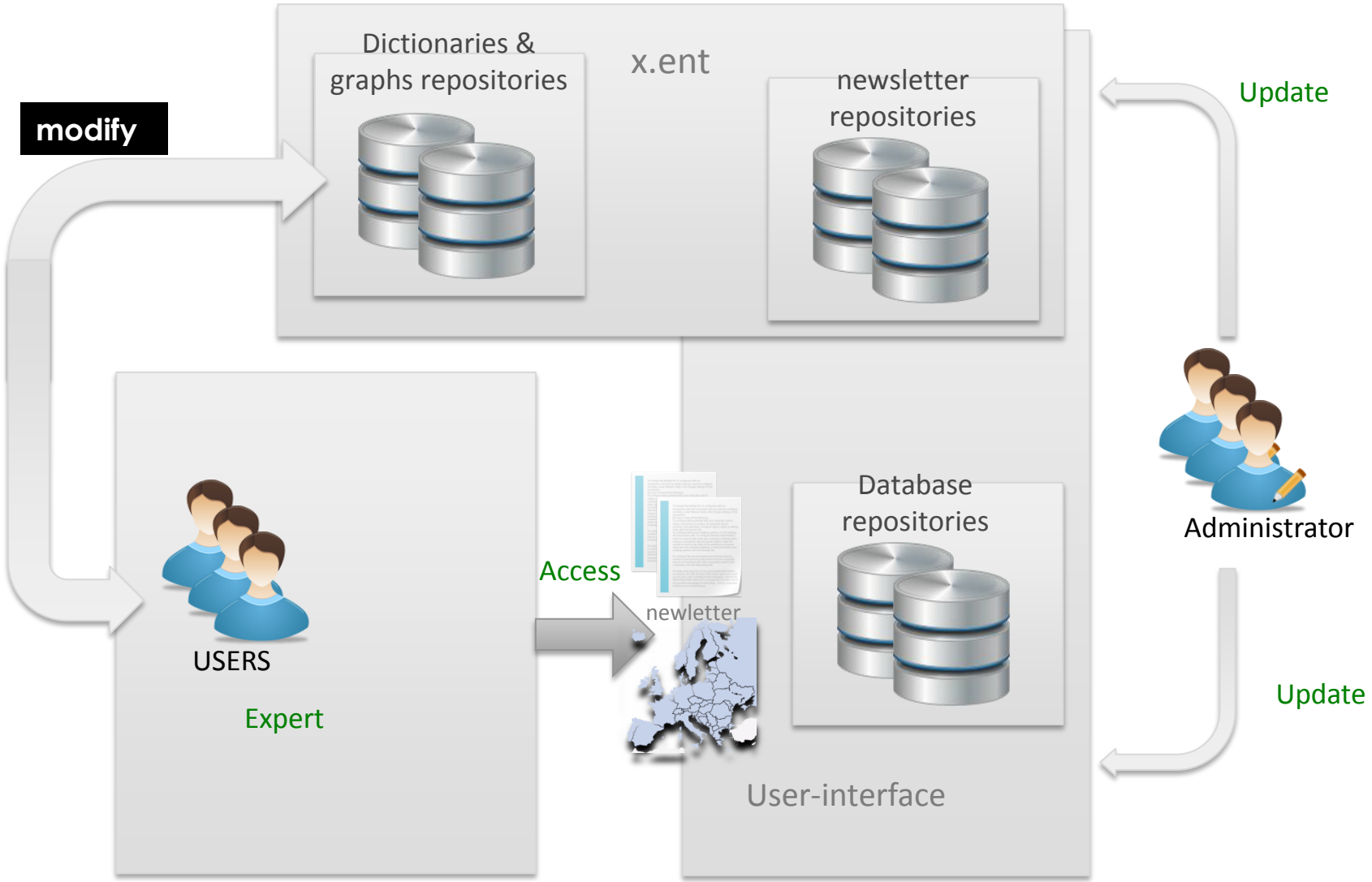
blé:adventice/blé:limace des jardins
blé:adventice/blé:puceron des céréales et du rosier
blé:campagnol des champs/blé:corbeau freux
blé:campagnol des champs/blé:pyrale
blé:campagnol des champs/blé:zabre des céréales
blé:cécidomyie jaune du blé/blé:charançon
blé:cécidomyie jaune du blé/blé:charançon de la tige
blé:cécidomyie jaune du blé/blé:mouche grise des céréales
blé:cécidomyie jaune du blé/blé:noctuelle
blé:cécidomyie jaune du blé/blé:oscinie de l'avoine

Global saturation of all tests

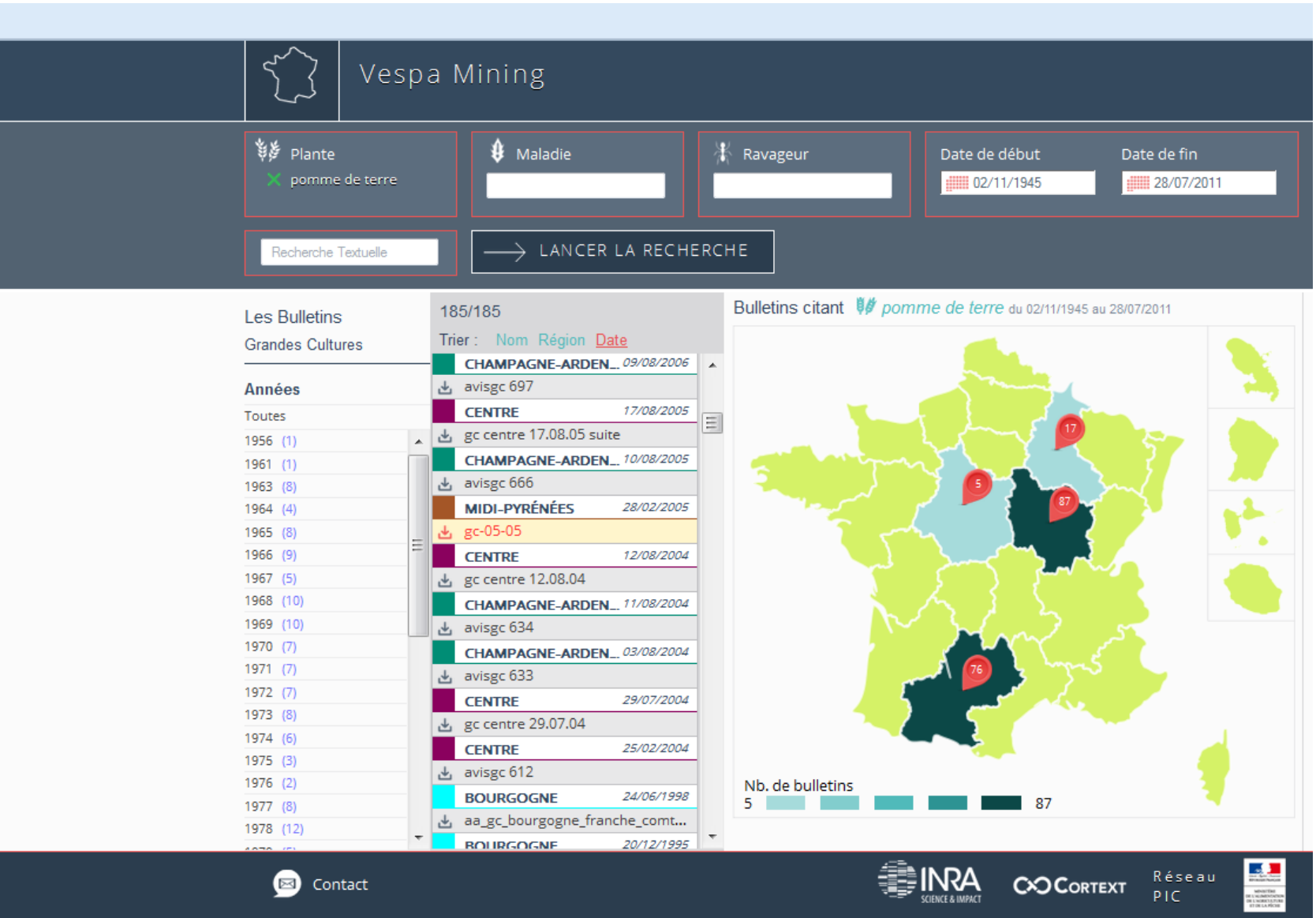
Relation Extraction: evaluation

	relations		
	crops-diseases	crops-pests	*total*
P	48.9	48.5	***48.6***
R	61.8	68.1	***65.4***
F-score	54.6	56.6	***55.8***

Vespa platform



VESPA platform



VESPA platform

http://213.229.108.100/Vespa/Site/Vespa.html

DEMO

TEST 1

Crop: wheat
Disease: rust
Pest: ./.
On map : risk assessment
region burgondy

Culture: blé
Maladie: rouille
Ravageur: ./.
sur carte : nuisibilité
région bourgogne

TEST 2

Crop: rapeseed
Disease: ./.
Pest: cabbage maggot
On map: Date Sort
region burgondy

Culture : colza
Maladie: ./.
ravage: mouche du chou
sur carte : tri par date
région bourgogne

TEST 3

Crop: ./.
Disease: mildew
Pest: ./.
On map: region burgondy
crop potato

Culture : ./.
Maladie : mildew
Ravageur: ./.
sur carte : région bourgogne
culture pomme de terre

Conclusion

- Concrete real-world issue about damage on crops
- Implementation of an original tool to extract relations (x.ent)
  - F-55%
  - crops/diseases & crops/pests
- Integration of the tool in a user-friendly platform with geolocalization and feedback to original documents

Perspectives

- X.ENT
  - Add a cooccurrence analysis for unformatted documents
  - Evaluation of relation with know set of relations (extern ontology)
  - Refine extraction of risk factor with a scale.
  - Add probabilistic approach for unknown relationship detection, with unknown named entities

- Vespa interface
  - Add a multiuser collaborative interface to modify ontology
  - Perhaps add other languages and documents (polish?)
  - Fusion with a meteorological ontology and database

Staff

Vespa platform

Nicolas Turenne
Research fellow

Tien Phan
Computer scientist

Alexandre Louchart
Computer scientist

Chloe Duloquin
designer

OCR dataset

Vincent Cellier
Research engineer

Mathieu Andro
PhD student & engineer

Danke für Ihre Aufmerksamkeit

Fragen ?