

# Y a-t-il de l'amiante dans les joints de vos fenêtres ? Approches sémantiques de prédiction de présence d'amiante dans les bâtiments

ROCED - MADICS 2021

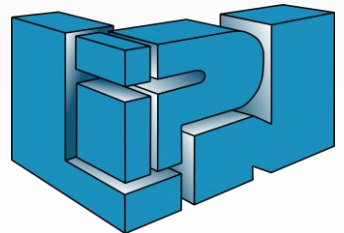
**Thamer Mecharnia**, Nathalie Pernelle, Celine Rouveirol , Fayçal Hamdi, Lydia Chibout Khelifa

**CSTB**  
le futur en construction

**LISN**  
LABORATOIRE INTERDISCIPLINAIRE  
DES SCIENCES DU NUMÉRIQUE

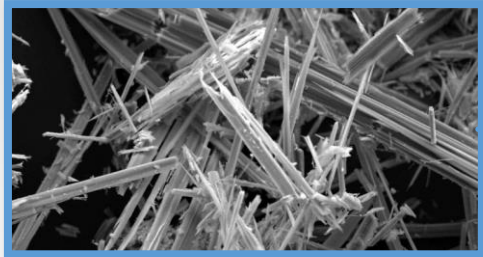
université  
PARIS-SACLAY

Cédric  
le cnam



# Contexte

## Amiante



Construction



- Asbestose ou fibrose pulmonaire
- Le cancer des poumons ou le mésothéliome



1997

# Contexte

## Documents CSTB



Diagnostic  
amiante

Les composants du bâtiment et leurs relations partie-tout: structure, localisation et produit

Données temporelles :  
année de construction

Diagnostiques de présence/absence  
d'amiante dans un produit

# Contexte

## Ressources externes



INRS



ANDEVA

Descriptions des produits commercialisés  
amiantés:

- Nom de produit commercialisé
- Périodes de présence/absence d'amiante

# Problématique

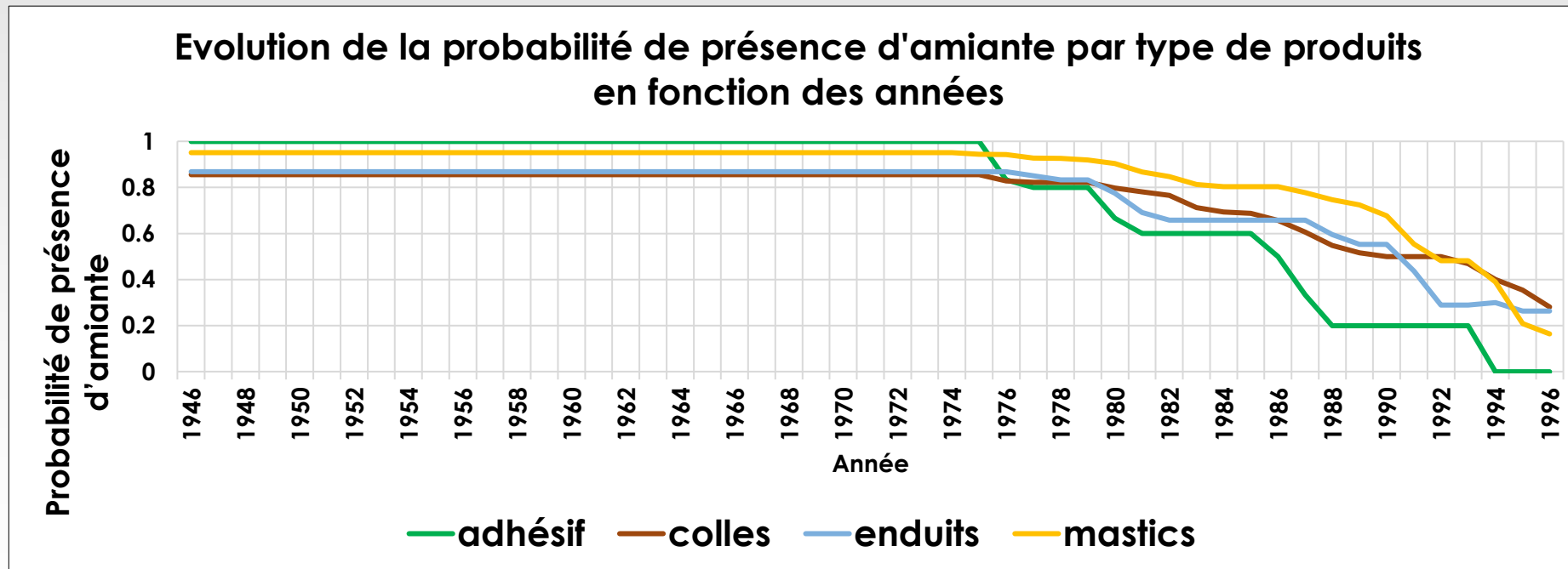
Information incomplète

Structure	Localisations	Produit	Diagnostic
PAROIS VERTICALES INTERIEURES	Murs	<b>Enduit</b>	0

Il existe de nombreux produits commercialisés de caractéristiques différentes de la classe de produits « **Enduit** »

# Problématique

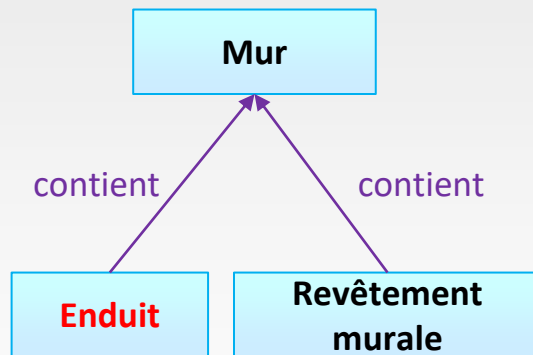
## Influence des données temporelles



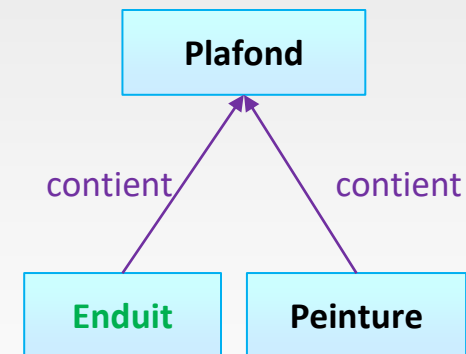
Construit à partir de l'INRS et l'ANDEVA

# Problématique

## Influence du contexte



**Amianté**



**Non amianté**

# Problématique

Ressources externes (INRS et ANDEVA): Informations incertaines et parfois contradictoires

INRS

Produit	Fournisseur	Renseignement divers	Type d'amiante	Type d'utilisation
ARMAZOL	/	Amianté jusqu'en 1982, non renseigné après	AMIANTE	Revêtement de sols

ANDEVA

Classe de produit	Produit	Amianté jusqu'en
Revêtement de sols	ARMAZOL	1990



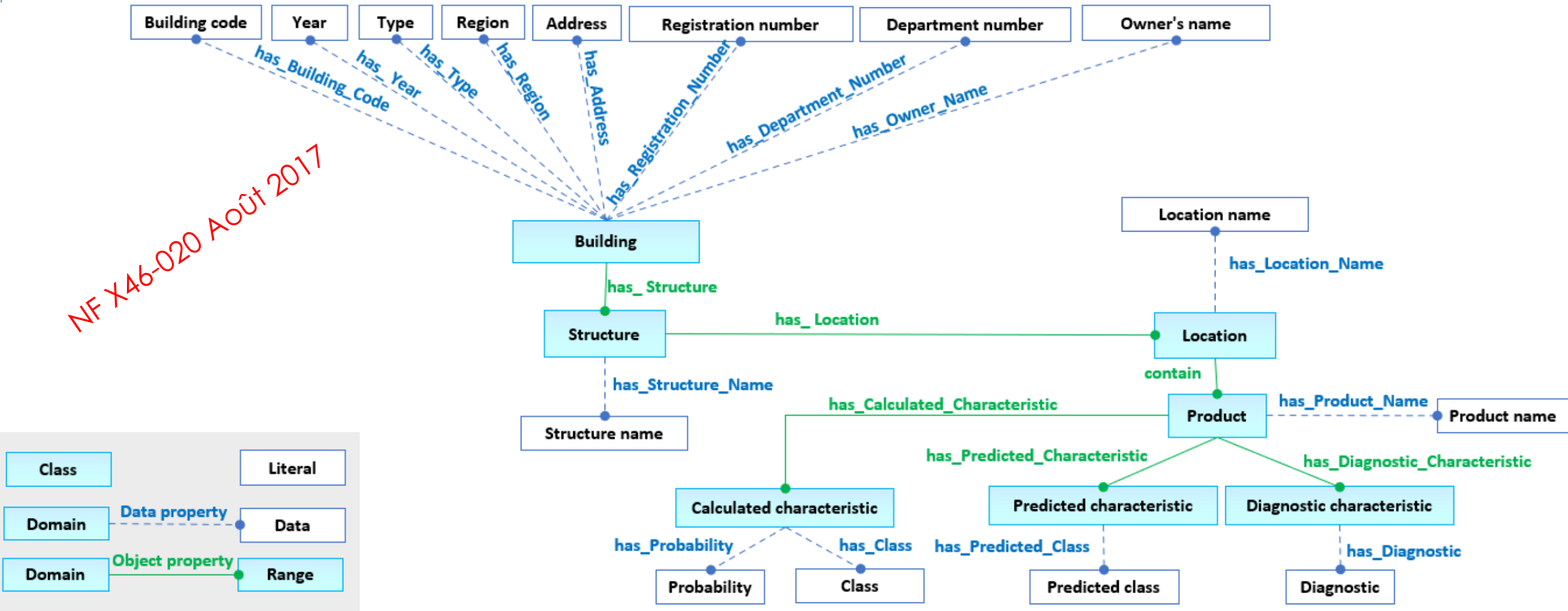
Comment classifier les produits comme amiantés ou non amiantés en exploitant le contexte et les données temporelles, incomplètes, incertaines et contradictoires ?  
Comment générer des règles de classification interprétables par l'expert ?

# Sommaire

- L'ontologie Amiante
- Approche Hybride
- Règles contextuelles
- CRA-Miner
- Expérimentations

# L'ontologie Amiante

NF X46-020 Août 2017



# Approche Hybride [K-Cap2019]

Calcul de la probabilité de présence d'amiante

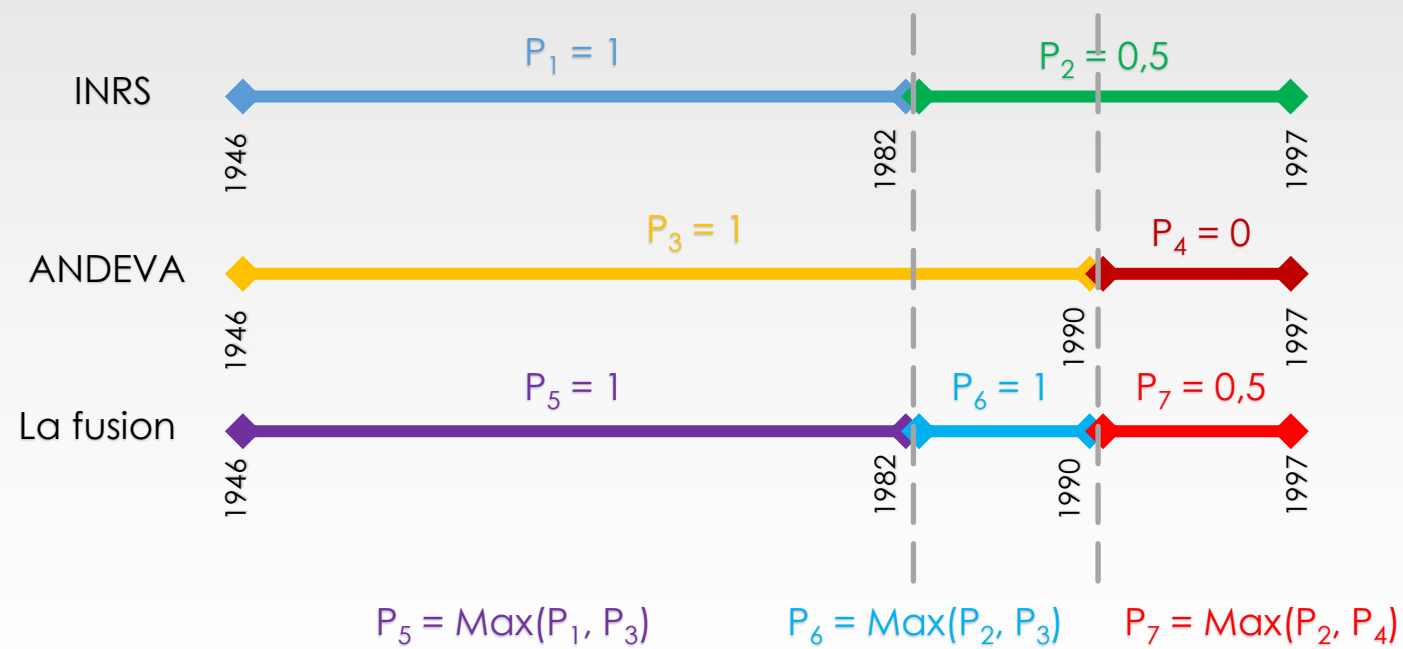
# Fusion des données temporelles

## INRS

Produit	Renseignement divers
ARMAZOL	Amianté jusqu'en 1982, non renseigné après

## ANDEVA

Produit	Amianté jusqu'en
ARMAZOL	1990



# Calcul de probabilité de présence d'amiante : formule

- Nous utilisons les probabilités amiante des produits commercialisés durant l'année  $d$  (caractéristique fusionnée à partir de l'INRS et ANDEVA) pour estimer la probabilité amiante d'un produit utilisé dans un bâtiment construit à l'année  $d$ .
- Pour un produit inconnu  $p_k$  appartenant à la classe  $F$  et utilisé à la date  $d$  :

$$p_a(p_k, d) = \frac{\sum_{p_j \in F(p_k)} p_{ext}(p_j, d)}{|p_j|}$$

$p_{ext}(p_j, d)$  est la probabilité amiante d'une caractéristique fusionnée du produit commercialisé  $p_j$  et qui est de même type  $F$  que  $p_k$ .

# Calcul de probabilité de présence d'amiante : exemple

- Pour calculer la probabilité de présence d'amiante dans « Revêtement mural » en 1982, nous utilisons les informations de l'ANDEVA et l'INRS après la fusion:

Produit	Amianté jusqu'en	Probabilité en 1982
COLOVINYL	1983	1
DECOVER	/	0,5
NOVILON	1981	0
STRAIRLAM	/	0,5

- $p_a(\text{Revêtement Mural}, 1982) = \frac{1+0,5+0+0,5}{4} = 0,5$



# Calcul de probabilité de présence d'amiante : réajustement

- Pour ajuster notre probabilité et approcher au maximum à la réalité, nous avons ajouté une proportion  $\alpha$  qui représente le nombre des produits non mentionnés dans les listes de produits (INRS et ANDEVA):

$$p_a(p_k, d) = \frac{\sum_{p_j \in F(p_k)} p_{ext}(p_j, d)}{|p_j| + (\alpha \times |p_j|)}$$

# Règles contextuelles

Générées pour conclure sur la présence ou l'absence d'amiante dans un produit

# Contexte conceptuel et Règles contextuelles

- Contexte conceptuel  $CO$  : ensemble de classes et de propriétés utilisables dans le corps de la règle.
- Règle Contextuelle : conjonction de prédicats qui conclut sur la présence ou l'absence d'amiante dans un produit :

$\vec{B} \rightarrow h$  avec  $\vec{B} = \{B_1, B_2, \dots, B_n\}$  telle que  $\forall B_i \in \vec{B}, \exists B_j \in CO \cup \{SWRL: CompareTO\}$  s. t.  $B_i \sqsubseteq B_j$

$h$  est le prédicat « has\_diagnostic » qui est instancié par la valeur « positive » ou « négative ».

- Exemple de règle :

$colle(P), contain(L, P), has\_location(S, L), has\_structure(B, S), has\_region(B, "Paris"), has\_diagnostic\_characteristic(P, D) \Rightarrow has\_diagnostic(D, "positive")$

# Mesures de qualité d'une règle contextuelle

- Support: nombre de prédictions correctes générées par la règle.
- Head coverage (hc): ratio entre le support et le nombre de diagnostics de type *val*

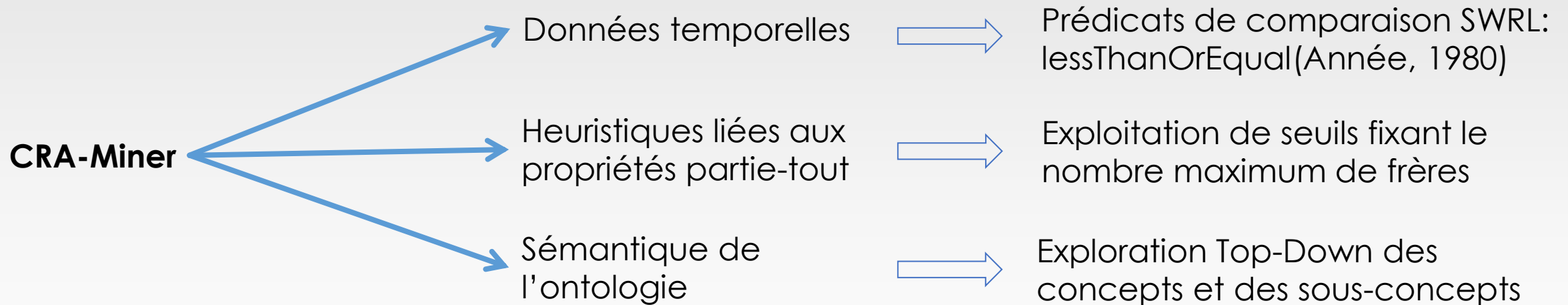
$$hc(\vec{B} \rightarrow has\_diagnostic(D, val)) = \frac{supp(\vec{B} \rightarrow has\_diagnostic(D, val))}{\#(D, val): has\_diagnostic(D, val)}$$

- Confiance (conf) : ratio entre le support de la règle et le nombre de diagnostics différents qui participent à une instantiation du corps de la règle

$$conf(\vec{B} \rightarrow has\_diagnostic(D, val)) = \frac{supp(\vec{B} \rightarrow has\_diagnostic(D, val))}{\#D: \exists x_1, \dots, x_n: \vec{B}}$$

L'objectif est de découvrir toutes les règles qui sont conformes aux contraintes du biais de langage et qui ont  $hc \geq minHC$  et  $conf \geq minConf$ .

# CRA-Miner [IC 2021]



# Etapes de CRA-Miner

CO = { product, location, structure, contain, has\_location, has\_region, has\_year, has\_structure, has\_diagnostic\_characteristic }

1. Spécialisation de T (  $product(P), has\_diagnostic\_characteristic(P, D) \Rightarrow has\_diagnostic(D, Value)$  ) en utilisant des sous-classes de produit.

*enduit*(P),  $has\_diagnostic\_characteristic(P, D) \Rightarrow has\_diagnostic(D, positive)$

2. Ajout d'une contrainte temporelle (année permettant d'obtenir la meilleure confiance pour  $hc > minHC$ ).

*enduit*(P), *contain*(L, P), *has\_location*(S, L), *has\_structure*(B, S), *has\_year*(B, Y), *lessThanOrEqualTo*(Y, "1970"), ...

# Etapes de CRA-Miner

3. Spécialisation par localisation et/ou par structure.

*enduit(P), mur(L), contain(L, P), has\_location(S, L), has\_structure(B, S), has\_year(B, Y), lessThanOrEqual(Y, "1970"), ...*

4. Enrichissement par la région.

5. Spécialisation en ajoutant d'autres composants co-localisés (les frères).

*enduit(P), mur(L), peinture(P2), contain(L, P2), contain(L, P), has\_location(S, L), has\_structure(B, S), has\_year(B, Y), lessThanOrEqual(Y, "1970"), ...*

# Expérimentations

Jeux de données

Analyse des résultats



# Jeux de données



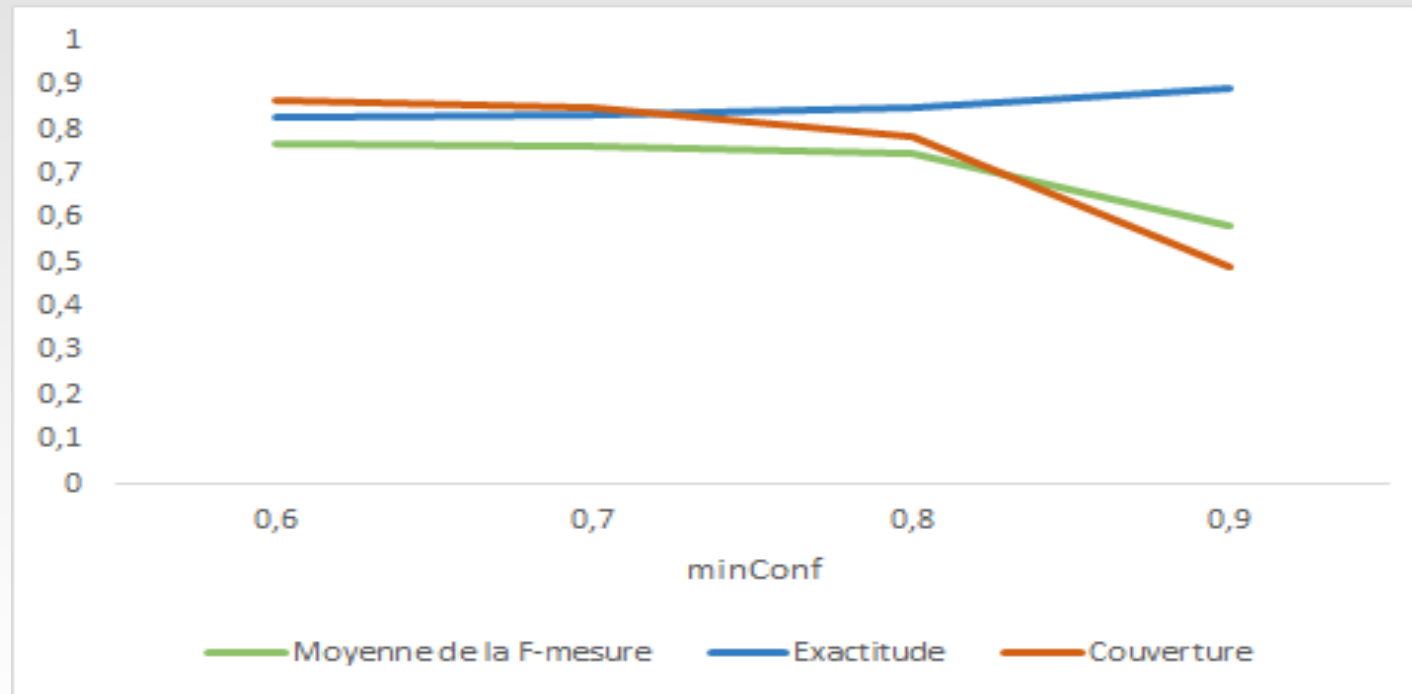
Diagnostic  
amiante

- 51970 triplets.
- 2998 instances de produit, 341 localisations, 214 structures et 94 bâtiments.
- Année de construction  $\in [1948,1997]$ .
- 1525 **produits** contenant de l'amiante et 1473 **produits** sans amiante.

# Paramètre et méthode de test

- CO={product, location, structure, contain, has\_location, has\_region, has\_year, has\_structure, has\_diagnostic\_characteristic }
- maxSibS = 0, maxSibL = 3, maxSibP = 3
- minHC = 0,001
- Validation croisée sur 3 tiers.
- CRA-Miner est parallélisé.

# Validation de minConf = 0,6



Résultats de CRA-Miner selon différents seuils de minConf

# Résultats CRA-Miner

- Approche pessimiste pour les doubles décisions.
- 29 règles comportent au moins un produit frère (au maximum 2).
- 16 règles comportent une localisation frère (au maximum 3).
- 14 règles exploitent une contrainte temporelle.

- Exemple de règle (13 atomes)

enduitsabasedeplatreprojeteslissesoutaloches(?P), Location(?L),  
 Structure(?S), has\_location(?S, ?L), gainesetcoffresverticaux(?L4),  
 has\_location(?S, ?L4),  
 revetementsdemurspoteauxcloisonsgainescoffres(?L3), has\_location(?S,  
 ?L3), cloisonssèchesassembleesprefabriquees(?L2), has\_location(?S, ?L2),  
 contain(?L, ?P), has\_diagnostic\_characteristic(?P, ?D  
 ⇒ has\_Diagnosis(?D, 'positive')

minConf=0.6, minHC=0.001	CRA-Miner
<b># règles</b>	<b>75</b>
<b>Double décision</b>	<b>82</b>
TP	415
TN	303
FP	121
FN	28
UP	66
UN	67
<b>Précision des positives</b>	<b>0.77</b>
<b>Rappel des positives</b>	<b>0.82</b>
<b>F-mesure des positives</b>	<b>0.79</b>
<b>Précision des négatives</b>	<b>0.92</b>
<b>Rappel des négatives</b>	<b>0.62</b>
<b>F-mesure des négatives</b>	<b>0.74</b>
<b>Moyenne de la F-mesure</b>	<b>0.77</b>
<b>Exactitude</b>	<b>0.83</b>
<b>Couverture</b>	<b>87%</b>

# Comparaison avec une Baseline et AMIE3 et l'approche hybride

minConf=0.6, minHC=0.001	Systèmes d'extraction de règles				Systèmes basés sur des ressources externes
	CRA-Miner	AMIE3, l=4	AMIE3, l=6	Baseline	Hybride
# règles	75	45	91	24	/
Double décision	82	50	277	0	0
TP	415	381	473	146	465
TN	303	288	264	257	348
FP	121	146	226	30	16
FN	28	74	32	24	5
UP	66	54	3	338	38
UN	67	58	0	204	127
Précision des positives	<b>77%</b>	72%	68%	83%	97%
Rappel des positives	82%	75%	<b>93%</b>	29%	92%
F-mesure des positives	<b>0.79</b>	0.73	<b>0.79</b>	0.43	0,94
Précision des négatives	<b>92%</b>	80%	89%	91%	99%
Rappel des négatives	<b>62%</b>	59%	54%	52%	71%
F-mesure des négatives	<b>0.74</b>	0.68	0.67	0.66	0,83
Moyenne de la F-mesure	<b>0.77</b>	0.71	0.73	0.55	0,89
Exactitude	<b>0.83</b>	0.75	0.74	0.88	0,97
Couverture	87%	89%	<b>100%</b>	46%	83%

# Conclusion et perspective

- Définition d'une approche hybride qui base sur les ressources externes pour prédire la présence/absence de l'amiante.
- Définition d'une approche CRA-Miner qui utilise les connaissances expert, la sémantique de l'ontologie Asbestos, les informations temporelles et les heuristiques partie-tout pour découvrir des règles pertinentes.
- Les résultats montre une bonne exactitude, avec en particulier une bonne précision pour les négatifs (stratégie pessimiste).
- Nous envisageons de comparer avec d'autres approches de type Divide-and-Conquer qui utilisent un biais de langage comme TILDE [1].
- Combiner les deux approches hybride et CRA-Miner pour améliorer la couverture des données.

[1] Hendrik Blockeel and Luc De Raedt. Top-down in-duction of first-order logical decision trees. *Artificialintelligence*, 101(1-2) :285–297, 1998.

[K-Cap2019] Thamer Mecharnia, Lydia Chibout Khelifa, NathaliePernelle, and Fayçal Hamdi. An approach toward aprediction of the presence of asbestos in buildings ba-sed on incomplete temporal descriptions of marketedproducts. In Mayank Kejriwal, Pedro A. Szekely, andRaphaël Troncy, editors,K-CAP 2019, Marina DelRey, CA, USA, November 19-21, 2019, pages 239–242. ACM, 2019.

Merci