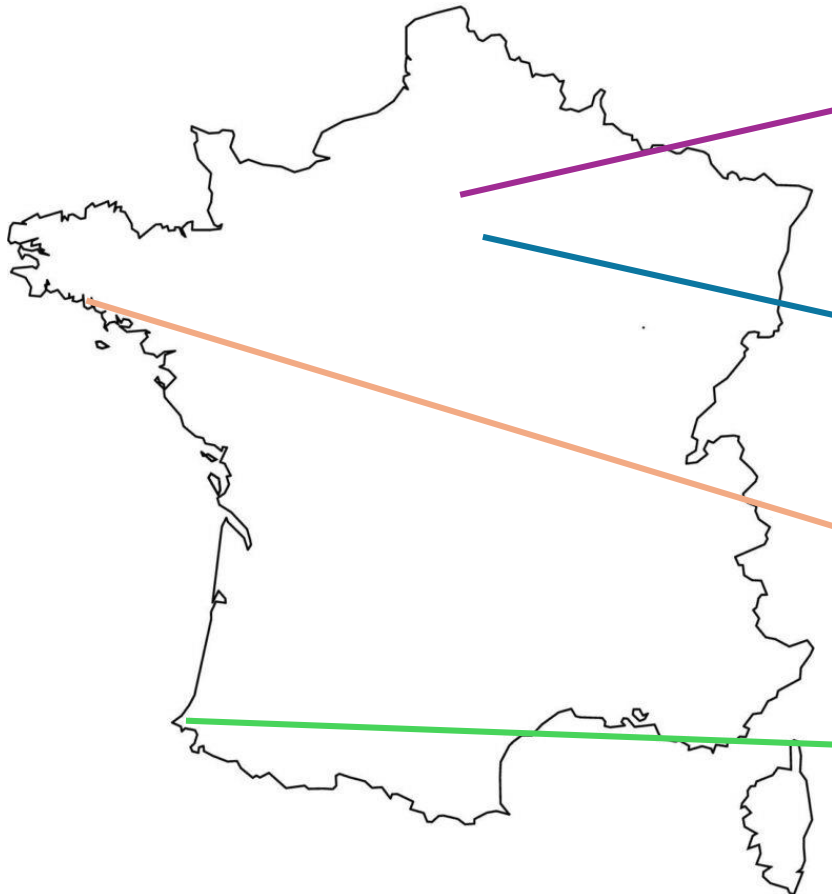


✦ Extraction automatique des quantités de stupéfiants, dans les résumés de procédure enregistrés par les forces de sécurité intérieure, à l'aide de modèles légers

Mathias ROBERT - SSMSI

L'objectif : enrichir les informations sur les procédures



Procédure 123456	12/10/2024	Stupéfiant(s) ? Quantité(s) ?
Trafic de stupéfiants		
PARIS	4 mis en cause	
Procédure 234567	23/12/2024	Stupéfiant(s) ? Quantité(s) ?
Usage de stupéfiants		
MELUN	1 mis en cause	
Procédure 345678	01/06/2024	Stupéfiant(s) ? Quantité(s) ?
Usage de stupéfiants		
LORIENT	2 mis en cause	
Procédure 456789	29/09/2024	Stupéfiant(s) ? Quantité(s) ?
Usage de stupéfiants		
BIARRITZ	1 mis en cause	

- Plusieurs centaines de milliers de procédures.
- Comment enrichir les données avec des types de stupéfiants et des quantités associées ?

Des sources de données non exhaustives

Procédure	PVE	Osiris	Objets	CONCLUSION
Procédure 123456	NA	Cannabis (herbe) 1kg Cocaïne 450g	NA	Cannabis (herbe) 1kg Cocaïne 450g
Procédure 234567	NA	Ecstasy 3,5g	Ecstasy 3,5g	Ecstasy 3,5g
Procédure 345678	Cocaïne 5g	NA	Cocaïne 6g	Cocaïne 5g
Procédure 456789	NA	NA	NA	INCONNU
...

Des sources de données non exhaustives

Procédure	PVE	Osiris	Objets	CONCLUSION
Procédure 123456	NA	Cannabis (herbe) 1kg Cocaïne 450g	NA	Cannabis (herbe) 1kg Cocaïne 450g
Procédure 234567	NA	Ecstasy 3,5g	Ecstasy 3,5g	Ecstasy 3,5g
Procédure 345678	Cocaïne 5g	NA	Cocaïne 6g	Cocaïne 5g
Procédure 456789	NA	NA	NA	INCONNU
...

- Si les sources classiques ne possèdent pas d'information sur le(s) stupéfiant(s) en cause : on fait appel au **résumé de procédure**.

Procédure 456789	29/09/2024	Le mis en cause se trouve le 29/09/2024 à 19h40 sur la commune de Biarritz. Suite à un contrôle il est découvert en sa possession 17 g de produits : un pochon de 12gr de résine de cannabis ainsi que de la cocaïne (5 grammes) , l'individu affirme qu'il s'agit de sa consommation personnelle.
Usage de stupéfiants		
BIARRITZ	1 mis en cause	

Exploiter un résumé de procédure

Le mis en cause se trouve le 29/09/2024 à 19h40 sur la commune de Biarritz. Suite à un contrôle il est découvert en sa possession **17 g** de produits : un pochon de **12gr** de **résine de cannabis** ainsi que de la **cocaïne (5 grammes)**, l'individu affirme qu'il s'agit de sa consommation personnelle.

Type(s) de stupéfiants :

Résine de cannabis
Herbe de cannabis
Cocaïne
Ecstasy
Héroïne
Crack
Amphétamines
Kétamine
Champignons hallucinogènes
LSD
...

- L'extraction du type de stupéfiant peut se faire simplement avec une liste de stupéfiants à détecter.

Exploiter un résumé de procédure

Le mis en cause se trouve le 29/09/2024 à 19h40 sur la commune de Biarritz.
Suite à un contrôle il est découvert en sa possession **17 g** de produits : un pochon de **12gr** de **résine de cannabis** ainsi que de la **cocaïne (5 grammes)**, l'individu affirme qu'il s'agit de sa consommation personnelle.

Type(s) de stupéfiants :

Résine de cannabis
Herbe de cannabis
Cocaïne
Ecstasy
Héroïne
Crack
Amphétamines
Kétamine
Champignons hallucinogènes
LSD
...

- L'extraction du type de stupéfiant peut se faire simplement avec une liste de stupéfiants à détecter.
- Extraire les quantités, cela requiert :
 - De détecter un chiffre
 - De détecter une unité de mesure
 - De les associer correctement au stupéfiant concerné
 - De ne pas prendre en compte les quantités « inutiles »

Plusieurs méthodes légères

1. Méthode déterministe : identifier les quantités par des expressions régulières, et réaliser l'allocation des stupéfiants à leur quantité par un algorithme spécifique.

17 g de produits : un pochon de 12gr de résine de cannabis ainsi que de la cocaïne (5 grammes)

Expression
régulière



Quantités : 17 g,
12gr, 5 grammes
Stupéfiants :
résine de cannabis,
cocaïne

Algorithme spécifique
(Jonker-Volgenant)



Résine de cannabis : 12 g
Cocaïne : 5 g
Ø : 17 g

Plusieurs méthodes légères

1. Méthode déterministe : identifier les quantités par des expressions régulières, et réaliser l'allocation des stupéfiants à leur quantité par un algorithme spécifique.

17 g de produits : un pochon de 12gr de résine de cannabis ainsi que de la cocaïne (5 grammes)

Expression
régulière



Quantités : 17 g,
12gr, 5 grammes
Stupéfiants :
résine de cannabis,
cocaïne

Algorithme spécifique
(Jonker-Volgenant)



Résine de cannabis : 12 g
Cocaïne : 5 g
Ø : 17 g

2. Utiliser un modèle de reconnaissance d'entités nommées (NER).

17 g de produits : un pochon de 12gr de résine de cannabis ainsi que de la cocaïne (5 grammes)

Modèle de reconnaissance
d'entités nommées généraliste



[« Poids de résine de cannabis »] → 12 g
[« Poids de cocaïne »] → 5 g

Plusieurs méthodes légères

1. Méthode déterministe : identifier les quantités par des expressions régulières, et réaliser l'allocation des stupéfiants à leur quantité par un algorithme spécifique.

17 g de produits : un pochon de 12gr de résine de cannabis ainsi que de la cocaïne (5 grammes)

Expression régulière



Quantités : 17 g,
12gr, 5 grammes
Stupéfiants :
résine de cannabis,
cocaïne

Algorithme spécifique
(Jonker-Volgenant)

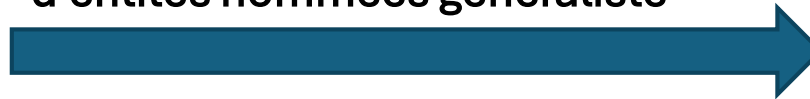


Résine de cannabis : 12 g
Cocaïne : 5 g
Ø : 17 g

2. Utiliser un modèle de reconnaissance d'entités nommées (NER).

17 g de produits : un pochon de 12gr de résine de cannabis ainsi que de la cocaïne (5 grammes)

Modèle de reconnaissance
d'entités nommées généraliste

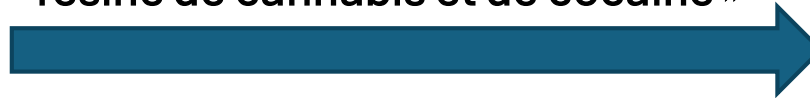


[« Poids de résine de cannabis »] → 12 g
[« Poids de cocaïne »] → 5 g

3. Utiliser un modèle de langage léger (SLM)

17 g de produits : un pochon de 12gr de résine de cannabis ainsi que de la cocaïne (5 grammes)

SLM : « Extrais la quantité de
résine de cannabis et de cocaïne »



1 : 12g de résine de cannabis
2 : 5g de cocaïne



Conclusions sur les méthodes

- Ces méthodes sont prometteuses et fonctionnent pour l'ensemble des cas simples et certains cas complexes.
- Les algorithmes très légers ont du mal à performer.
- En cas de capacités computationnelles limitées, les méthodes déterministes plus simples performant alors mieux.
- Dans ce cas, des approches hybrides peuvent être intéressantes, par exemple :
 1. Simplifier le résumé de procédure avec un **petit modèle de langage**
 2. Utiliser la **méthode déterministe** pour l'extraction sur le résumé de procédure simplifié.