

# Extraction des demandes dans les décisions jurisprudentielles

séminaire LGI2P, IMT Mines Alès – 15 mars 2018

---

Début de thèse: 15 Décembre 2015

**Doctorant:** Gildas Tagny Ngompé

**Direction de thèse:**

- Jacky Montmain (École des mines d'Alès, LGI2P)
- Stéphane Mussard (Université de Nîmes, CHROME)

**Encadrement de proximité:**

- Sébastien Harispe (Ecole des Mines d'Alès, LGI2P)
- Guillaume Zambrano (Université de Nîmes, CHROME)

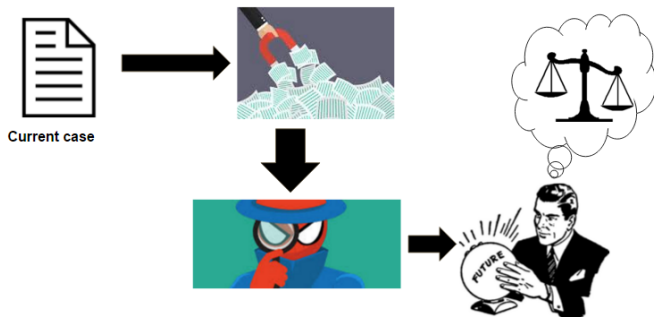


1. Motivations et objectifs
2. Extraction d'information sur les demandes et résultats
3. Une approche basée sur la pondération de termes et le zonage
4. Expérimentations et résultats
5. Conclusion
6. Questions et suggestions

## Motivations et objectifs

---

# Les juristes analysent les décisions



## Pourquoi ?

- comprendre et comparer l'application loi (contentieux, ville, ...)
- estimer le risque judiciaire
- ...

# Motivation : documents non-structurés, langage complexe

ARRÊT N°

R.G : 11/03924

...

COUR D'APPEL DE NÎMES

CHAMBRE CIVILE

1ère Chambre A

ARRÊT DU 20 MARS 2012

APPELANTE :

Madame Michèle A. ...

assistée de la SELARL VAJOU, ...

INTIMES :

Monsieur Martial B ...

assisté de la SCP MARION GUIZARD PATRICIA

SERVAIS, ...

COMPOSITION DE LA COUR LORS DU

DÉLIBÉRÉ :

M. Dominique BRUZY, Président

M. Serge BERTHET, Conseiller

...

FAITS, PROCEDURE, ...

Madame Michèle A. demande :

...

- de condamner Madame JONES-B. à lui payer la somme de 2.500 euros au titre de l'article 700 du Code de Procédure Civile,

PAR CES MOTIFS, LA COUR :

...

Vu l'article 809 du Code de Procédure Civile,

...

Déboute Madame A. de sa demande de provision sur dommages-intérêts.

...

Vu l'article 700 du Code de Procédure Civile,  
Condamne Madame JONES-B. à verser à Madame A. la somme de 2.500 euros.

# Motivation : grand volume de décisions

Plus de 4 millions de décisions prononcées / an

	2010	2011	2012	2013	2014
<b>Justice civile</b>	2 673 131	2 654 179	2 647 813	2 761 554	2 618 374
Justice pénale	1 173 242	1 180 586	1 251 979	1 303 469	1 203 339
Justice administrative	224 787	225 608	228 680	221 882	230 477

Source : <http://www.justice.gouv.fr/budget-et-statistiques-10054/chiffres-cles-de-la-justice-10303/>

TABLE – Nombre de décisions prononcées en France par an

# Motivation : recherches et analyses sémantiques difficiles

## Moteurs de recherche juridique à mots-clés

### Pas d'analyse synthétique des décisions

☐ Recherche simple ☒ Recherche avancée

Mots ou expressions

Ex : gérant **et** pouvoir, bail **s/5** résilt  
[Aide à la recherche](#)

Gestion automatique des :

☒ Singulier / Pluriel ☒ Masculin / Féminin

☐ Verbes conjugués **avoir** cherche **ayons**

Sources

☒ \*Toutes les sources

[Répertoire des sources](#)

ou

☐ Encyclopédies

☐ Codes et Lois

☐ JurisData

☐ Toute la jurisprudence

☐ Revues

☐ Bibliographies

☐ Actualités

☐ Bulletins Officiels

☐ Autorités administratives

☐ Parlement

☐ Europe

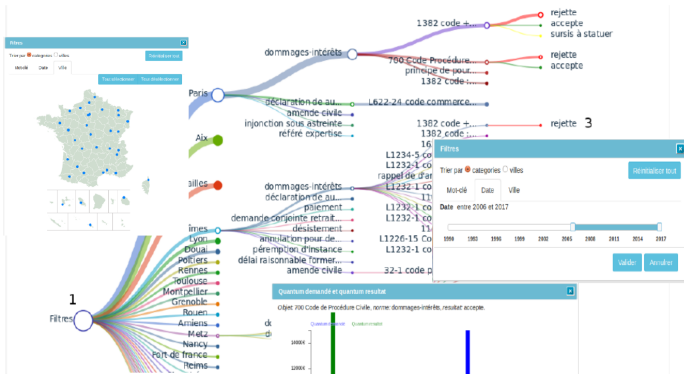
☐ Conventions Collectives

Période

Pas de restriction de date

Source : LexisNexis.com

Stage été 2017 [ PRYSIAZHNIUK Anastasiia ]

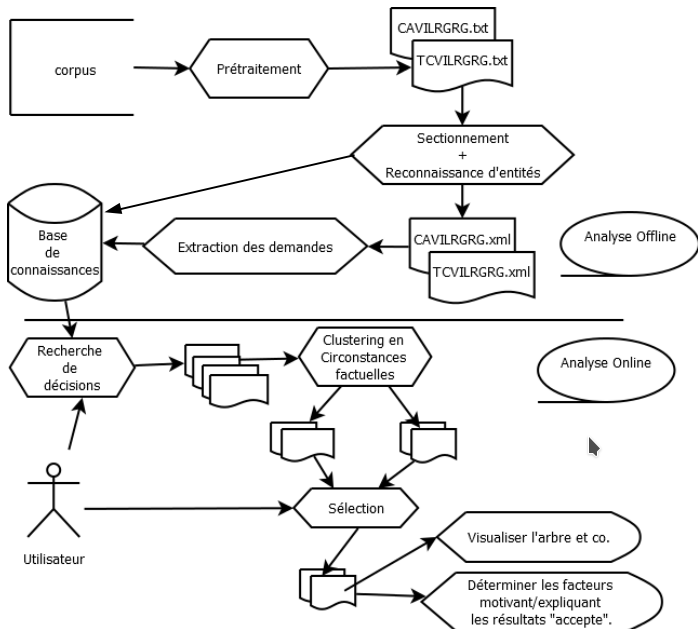


## 2 Tableau de résultats

RG	Ville	Date	Jurisdiction	Quantum demandé	Quantum accordé	Statut	Resultat	Description
15/02360	Metz	Thu, 24 Apr 2003 00:00:00 GMT	CA	-	2	accepte	accepte	<a href="#">Voir</a>
15/00786	Metz	Thu, 21 Jul 2011 00:00:00 GMT	CA	5 000,00 €	5 000,00 €	rejette	rejette	<a href="#">Voir</a>
14/01658	Metz	Sun, 22 Jul 2012	CA	5 000,00 €	5 000,00 €	rejette	rejette	<a href="#">Voir</a>



# Pipeline d'analyse sémantique



# Phase actuelle : extraction d'information



## Utilité

Recherche  
d'information

Prédiction

Exploration

Conseil



Corpus



Chaîne d'Analyse  
Sémantique



Base de Connaissances

## Objectif : Extraction d'Information

**Références :**  
Quand ? Où ? Qui ?

**Demandes :**  
Catégorie ?  
Montant demandé ?

**Résultats :**  
Accepte / rejette ?  
Montant obtenu ?  
Pourquoi ?

**Circonstances factuelles**

## Extraction d'information sur les demandes et résultats

---

# Tâche : extraire les quanta et le sens du résultat

	A	B	C	D	F	H	L	N
1	IDENTIFICATION DE LA DECISION			DESCRIPTION DE LA PRETENTION			DESCRIPTION DU RESULTAT	
2	Type	Ressort	RG	OBJET	NORME	QUANTUM	RESULTAT	QUANTUM RESULTAT (obtenu)
	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
441	CA	Lyon	14/06911	dommages-intérêts	700 Code de Procédure Civile	3,500.00 €	rejette	0.00 €
442	CA	Lyon	14/06911	dommages-intérêts	700 Code de Procédure Civile	2,000.00 €	accepte	1,500.00 €

TABLE – Tableau des informations sur les demandes

## EXEMPLE : INFORMATIONS PERTINENTES à EXTRAIRE

- **Catégorie prédéfinie** : Dommages-intérêts pour procédure abusive
  - **Objet** : Dommages-intérêts
  - **Fondement** : Articles 1382 code civil et 32-1 code de procédure civile
- **Quantum demandé** : 20 000 euros
- **Sens du résultat** : "rejette"
- **Quantum accordé** : 0 euros

# Difficultés (1)

Expressions non structurées, par **référence**, par **agrégation**

## EXEMPLE (SUITE) : EXPRESSION DE DEMANDE

La société A. conclut à la confirmation du jugement entrepris sauf à former appel incident sur la disposition du jugement l'ayant déboutée de sa demande de **dommages-intérêts pour abus de procédure** et elle demande à la cour de condamner l'appelante à lui payer la somme de **20 000 euros** à titre de dommages intérêts ...

## EXEMPLE (SUITE) : EXPRESSION DE RESULTAT

La cour, ...

Confirme **la décision entreprise** en **toutes ses dispositions**,

## Difficultés (2)

- Présence de plusieurs demandes de catégories similaires et/ou différentes dans une même décision
- Toutes les catégories ne sont pas connues d'avance (+500 catégories)
- Difficile d'annoter une base d'évaluation pour toutes les couvrir

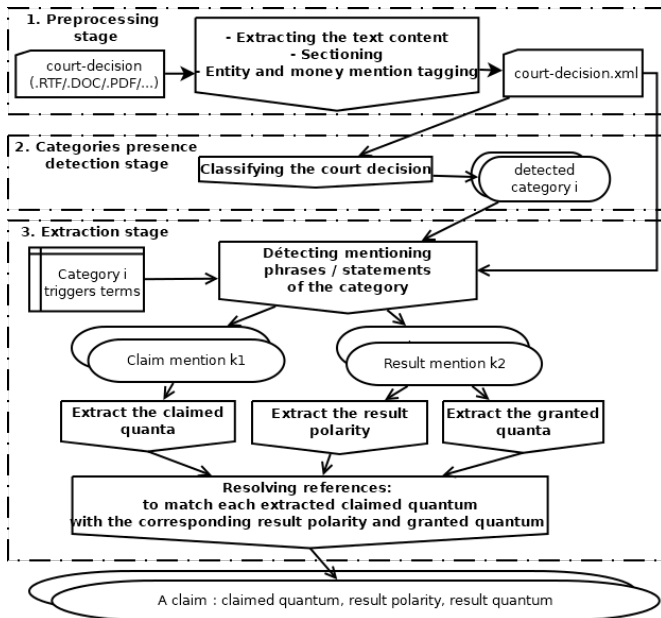
Il faut une approche :

- qui s'adapte à la catégorie à extraire
- qui permette de rajouter de nouvelles catégories

Une approche basée sur la pondération de  
termes et le zonage

---

# Architecture du pipeline d'extraction





# Identification des passages et des informations

Demande dans la section *Litige* (Faits, procédures, et moyens des parties)

Résultat dans la section *Dispositif*

Demande	Résultat (organisé par polarité)		
	accepte	sursis à statuer	rejette
<i>accorder, admettre, admission, allouer, condamnation, condamner, fixer, laisser, prononcer, ramener, surseoir</i>	<i>accorde, accordons, admet, admettons, alloue, allouons, condamne, condamnons, déclare, déclarons, fixe, fixons, laisse, laissons, prononce, prononçons</i>	<i>réserve, réservons, surseoit, sursoyons</i>	<i>déboute, déboutons, rejette, rejettons</i>

TABLE – Mots introduisant les énoncés de demandes et de résultats

- le <demande categorie="acpa">condamner à payer une <trigger categorie="acpa">**amende civile**</trigger> de <argent> **1.500 euros** </argent> pour procédure abusive ...
- le</demande> condamner à payer la somme ..."

FIGURE – Exploiter la proximité entre triggers et sommes d'argent

# Identification des passages et des informations(2)

## ○ Identification des passages :

1. Soit par la seule **présence d'un trigger** : on zone autour des triggers
2. Soit par **pondération des zones à argent** :
  - 2.1 on zone autour des sommes d'argent
  - 2.2 on pondère les zones (par ex. somme des poids des triggers)
  - 2.3 on sélectionne une zone si elle a un poids  $\geq$  POIDS SEUIL

## ○ Identification des informations :

1. quantum : somme d'argent près d'un trigger
2. sens du résultat :
  - soit en fonction du verbe introductif de l'énoncé du résultat
  - soit "*rejette*" si pas d'énoncé du résultat

## ○ Résolution des références :

- matching des énoncés (similarité textuelle)
- matching des quanta (Hypothèse d'apparition dans le même ordre)

# Phase d'entraînement

Catégorie  $c_i$ , Corpus d'entraînement  $D = D_{c_i} \cup D_{\bar{c}_i} = \{D_j\}_{1 \leq j \leq |D|}$

## 1. Détecteur de catégorie :

- vectoriser les décisions de la base d'entraînement :  
 $w(t_k, D_j) = lw(t_k, D_j) \times gw(t_k) \times nf(D_j)$  [?]
- entraîner un algorithme de classification (SVM, Naïf Bayésien, K plus proches voisins ...)

## 2. Extracteur de triplets de quanta et sens du résultat :

- Apprendre les triggers sur la base d'entraînement (passages à quanta vs. passages sans quanta)

2.1 pondération des termes  $t_k$  avec une métrique de RI par ex. :

Métrique non supervisée :

$$idf(t_k) = \log_2\left(\frac{N}{N_{t_k}}\right) \text{ [?]}$$

Métriques supervisées :

$$gss(t_k, c_i) = (N_{t_k, c_i} N_{\bar{t}_k, \bar{c}_i}) - (N_{t_k, \bar{c}_i} N_{\bar{t}_k, c_i}) \text{ [?]}$$

$$ngl(t_k, c_i) = \frac{\sqrt{N}((N_{t_k, c_i} N_{\bar{t}_k, \bar{c}_i}) - (N_{t_k, \bar{c}_i} N_{\bar{t}_k, c_i}))}{\sqrt{N_{t_k} N_{\bar{t}_k} N_{c_i} N_{\bar{c}_i}}} \text{ [?]}$$

2.2 ranking puis sélection

## Expérimentations et résultats

---

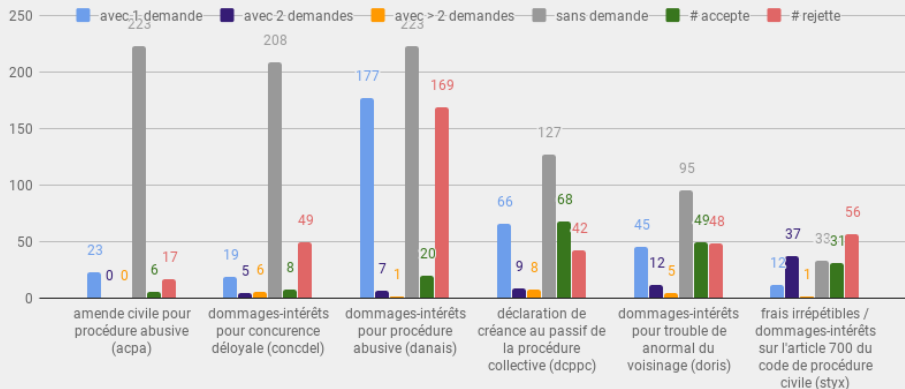


FIGURE – Répartitions des demandes dans les documents annotés pour chaque catégorie.

# Métriques d'évaluation

Catégorie  $c_i$ , tuple d'information  $I \subseteq \{Q_{DMD}, S_{RST}, Q_{RST}\}$

Corpus d'évaluation  $D = D_{c_i} \cup D_{\bar{c}_i} = \{D_j\}_{1 \leq j \leq |D|}$ , où  $D_j$  est un document

$$\text{Nombre de vrais positifs (bons)} : TP_{c_i, I, D} = \sum_{j=1}^{|D|} TP_{c_i, I, D_j}$$

$$\text{Nombre de faux positifs (en trop)} : FP_{c_i, I, D} = \sum_{j=1}^{|D|} FP_{c_i, I, D_j}$$

$$\text{Nombre de faux négatifs (manqués)} : FN_{c_i, I, D} = \sum_{j=1}^{|D|} FN_{c_i, I, D_j}$$

$$Precision_{c_i, I, D} = \frac{TP_{c_i, I, D}}{TP_{c_i, I, D} + FP_{c_i, I, D}}$$

$$Rappel_{c_i, I, D} = \frac{TP_{c_i, I, D}}{TP_{c_i, I, D} + FN_{c_i, I, D}}$$

$$F1_{c_i, I, D} = 2 \times \frac{Precision_{c_i, I, D} \times Rappel_{c_i, I, D}}{Precision_{c_i, I, D} + Rappel_{c_i, I, D}}$$

# Evaluation de la détection des catégories

TABLE – Resultats d’une 5-fold cross-validation sur  $D$  pour la detection categorie à l’aide de Weka [?] (P= Precision, R=Rappel, F1 = F1-mesure)

	Naïf Bayésien			Arbre de décision (J48)			KNN			SVM		
Category	P	R	F1	P	R	F1	P	R	F1	P	R	F1
acpa	1.0	1.0	1.0	0.996	0.955	0.972	1.0	1.0	1.0	0.996	0.955	0.972
concdel	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.995	0.967	0.979
danais	0.988	0.989	0.988	0.996	0.995	0.995	0.995	0.995	0.995	0.993	0.993	0.993
dcppc	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
doris	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
styx	1.0	1.0	1.0	0.984	0.983	0.983	1.0	1.0	1.0	1.0	1.0	1.0

# Quelle pondération des triggers et quel zonage ?

TABLE – Comparaison des métriques de pondération et des stratégies de zonage sur le corpus  $D$  (F1-mesure sur l'extraction du tuple  $(Q_{DMD}, S_{RST}, Q_{RST})$ )

Métrique	acpa		concdel		danais		dcppc		doris		styx	
	tp	zw	tp	zw	tp	zw	tp	zw	tp	zw	tp	zw
CHI2	0.683	0.698	0.061	0.061	0.443	0.411	0.259	0.264	0.187	0.071	0.321	0.366
DBIDF	0.683	0.698	0.076	0.033	0.461	0.416	0.254	0.264	0.084	0	0.331	0.358
DELTADF	0.683	0.698	0.144	0.082	0.443	0.41	0.259	0.264	0.143	0.142	0.334	0.281
DSIDF	0.678	0.698	0.076	0.052	0.399	0.152	0.014	0	0.019	0	0.343	0.33
GSS	0.683	0.698	0.144	0.082	0.443	0.41	0.259	0.264	0.143	0.142	0.334	0.281
IDF	0.067	0	0.033	0	0.04	0	0	0	0	0	0	0
IG	0.011	0.049	0.05	0.034	0.304	0.073	0	0	0.019	0	0.058	0
KLD	0.432	0.398	0.146	0.124	0.459	0.409	0.252	0.254	0.158	0.154	0.243	0.42
MAR	0.683	0.698	0.144	0.091	0.443	0.42	0.259	0.264	0.156	0.146	0.334	0.281
NGL	0.683	0.698	0.061	0.034	0.443	0.411	0.259	0.264	0.122	0.02	0.321	0.347
RF	0.683	0.698	0.202	0.043	0.491	0.367	0.242	0.21	0.101	0.058	0.387	0.351
Max	0.683	<b>0.698</b>	<b>0.202</b>	0.124	<b>0.491</b>	0.42	0.259	<b>0.264</b>	<b>0.187</b>	0.154	0.387	<b>0.42</b>

tp = zonage par la seule présence d'un trigger

zw = zonage par pondération des passages à somme d'argent

La métrique et la stratégie de zonage dépendent de la catégorie



# Exemple de termes sélectionnés

concdel		danais	
NGL	DSIDF	NGL	DSIDF
déloyale	concurrence déloyale	procédure abusive	procédure abusive et injustifiée
perte	déloyale	32-1	fondement de l' article 32-1
actes		abusive	dommages-intérêts pour procédure abusive
50.000	agissements	intérêts pour procédure	titre de dommages-intérêts pour procédure abusive

$$ngl(t_k, c_i) = \frac{\sqrt{N}((N_{t_k, c_i} N_{\overline{t_k}, \overline{c_i}}) - (N_{t_k, \overline{c_i}} N_{\overline{t_k}, c_i}))}{\sqrt{N_{t_k} N_{\overline{t_k}} N_{c_i} N_{\overline{c_i}}}}$$

$$dsidf(t_k, c_i) = \log\left(\frac{(N_{\overline{c_i}} N_{t_k, c_i}) + 0.5}{(N_{c_i} N_{\overline{t_k}, \overline{c_i}}) + 0.5}\right)$$

# Entrainement avec sélection de la meilleure métrique (1)

$c_i$	Tuple d'info ( $I$ )	$P_{c_i, I, D_{c_i}}$	$R_{c_i, I, D_{c_i}}$	$F1_{c_i, I, D_{c_i}}$	Docs. Parfaits	#extraits/#attendus/ $ D_{c_i} $
acpa	( $Q_{DMD}$ )	0.709	0.73	0.705	0.47	5.2/4.6/4.6
	( $S_{RST}$ )	0.691	0.7	0.683	0.48	
	( $Q_{RST}$ )	0.72	0.74	0.716	0.48	
	( $Q_{DMD}, S_{RST}, Q_{RST}$ )	0.651	0.65	0.638	0.43	
concdel	( $Q_{DMD}$ )	0.461	0.393	0.376	0.233	11.6/11.6/6.0
	( $S_{RST}$ )	0.544	0.442	0.427	0.2	
	( $Q_{RST}$ )	0.595	0.482	0.465	0.2	
	( $Q_{DMD}, S_{RST}, Q_{RST}$ )	0.337	0.299	0.28	0.167	
danais	( $Q_{DMD}$ )	0.548	0.516	0.527	0.346	36.6/38.8/37.0
	( $S_{RST}$ )	0.69	0.646	0.661	0.454	
	( $Q_{RST}$ )	0.714	0.666	0.682	0.465	
	( $Q_{DMD}, S_{RST}, Q_{RST}$ )	0.482	0.46	0.466	0.314	
dcppc	( $Q_{DMD}$ )	0.334	0.392	0.358	0.217	26.8/22.2/16.6
	( $S_{RST}$ )	0.665	0.798	0.721	0.544	
	( $Q_{RST}$ )	0.62	0.744	0.672	0.509	
	( $Q_{DMD}, S_{RST}, Q_{RST}$ )	0.22	0.26	0.237	0.181	
doris	( $Q_{DMD}$ )	0.279	0.373	0.314	0.033	26.8/20.0/12.4
	( $S_{RST}$ )	0.391	0.524	0.439	0.146	
	( $Q_{RST}$ )	0.329	0.414	0.361	0.131	
	( $Q_{DMD}, S_{RST}, Q_{RST}$ )	0.177	0.229	0.197	0.017	
styx	( $Q_{DMD}$ )	0.762	0.642	0.695	0.46	15.0/17.8/10.0
	( $S_{RST}$ )	0.701	0.593	0.64	0.34	
	( $Q_{RST}$ )	0.824	0.696	0.752	0.46	
	( $Q_{DMD}, Q_{RST}, S_{RST}$ )	0.44	0.372	0.402	0.28	

TABLE – Zonage par la seule présence d'un trigger (sur le corpus  $D_{c_i}$ )

# Entrainement avec sélection de la meilleure métrique (2)

$c_i$	Tuple d'info ( $I$ )	$P_{c_i, I, D_{c_i}}$	$R_{c_i, I, D_{c_i}}$	$F1_{c_i, I, D_{c_i}}$	Docs. Parfaits	#extraits/#attendus/ $ D_{c_i} $
acpa	( $Q_{DMD}$ )	0.753	0.61	0.672	0.57	3.8/4.6/4.6
	( $S_{RST}$ )	0.92	0.74	0.818	0.7	
	( $Q_{RST}$ )	0.92	0.74	0.818	0.7	
	( $Q_{DMD}, S_{RST}, Q_{RST}$ )	0.753	0.61	0.672	0.57	
concdel	( $Q_{DMD}$ )	0.343	0.128	0.11	0.067	5.6/11.6/6.0
	( $S_{RST}$ )	0.535	0.15	0.17	0.067	
	( $Q_{RST}$ )	0.543	0.17	0.182	0.067	
	( $Q_{DMD}, S_{RST}, Q_{RST}$ )	0.135	0.098	0.079	0.033	
danais	( $Q_{DMD}$ )	0.66	0.296	0.395	0.227	17.8/38.8/37.0
	( $S_{RST}$ )	0.732	0.328	0.438	0.27	
	( $Q_{RST}$ )	0.77	0.348	0.464	0.276	
	( $Q_{DMD}, S_{RST}, Q_{RST}$ )	0.61	0.276	0.367	0.216	
dcppc	( $Q_{DMD}$ )	0.391	0.363	0.372	0.252	21.4/22.2/16.6
	( $S_{RST}$ )	0.732	0.688	0.703	0.532	
	( $Q_{RST}$ )	0.665	0.624	0.638	0.471	
	( $Q_{DMD}, S_{RST}, Q_{RST}$ )	0.275	0.248	0.259	0.204	
doris	( $Q_{DMD}$ )	0.211	0.146	0.171	0.064	9.8/20.0/12.4
	( $S_{RST}$ )	0.418	0.217	0.268	0.114	
	( $Q_{RST}$ )	0.342	0.166	0.211	0.096	
	( $Q_{DMD}, S_{RST}, Q_{RST}$ )	0.095	0.067	0.078	0.017	
styx	( $Q_{DMD}$ )	0.838	0.632	0.718	0.52	13.2/17.8/10.0
	( $S_{RST}$ )	0.772	0.571	0.654	0.36	
	( $Q_{RST}$ )	0.786	0.583	0.666	0.38	
	( $Q_{DMD}, S_{RST}, Q_{RST}$ )	0.573	0.44	0.496	0.32	

TABLE – Zonage par pondération des passages à somme d'argent (sur le corpus  $D_{c_i}$ )

## Conclusion

---

# Conclusion

L'extraction des demandes est une tâche difficile :

- texte non structuré
- extraction dans des documents de plusieurs pages
- complexité du langage : référence, agrégation, implicite

Proposition d'une baseline (pondération des termes et zonage) :

- Détection facile de la présence d'une catégorie
- La métrique et la stratégie de zonage dépendent de la catégorie
- Le sens et le quantum résultat plus faciles à identifier que le quantum demandé
- Difficultés : identification des termes-clés
- Voies d'amélioration : Section des motifs et références aux jugements antérieurs, résolution des références

## Questions et suggestions

---

# References I



Frank, E., Hall, M., and Witten, I. (2016).  
The weka workbench.

*Data mining : Practical machine learning tools and techniques. Burlington : Morgan Kaufmann.*



Galavotti, L., Sebastiani, F., and Simi, M. (2000).  
Experiments on the use of feature selection and negative evidence in automated text categorization.  
In *International Conference on Theory and Practice of Digital Libraries*, pages 59–68. Springer.



Ng, H. T., Goh, W. B., and Low, K. L. (1997).  
Feature selection, perceptron learning, and a usability case study for text categorization.  
In *ACM SIGIR Forum*, volume 31, pages 67–73. ACM.



Salton, G. and Buckley, C. (1988).  
Term-weighting approaches in automatic text retrieval.  
*Information processing & management*, 24(5) :513–523.



Sparck Jones, K. (1972).  
A statistical interpretation of term specificity and its application in retrieval.  
*Journal of documentation*, 28(1) :11–21.