

# Vers une recherche sémantique et à base de graphe dans les systèmes d'accès à l'information juridique

Nada Mimouni, Adeline Nazarenko et Sylvie Salotti

LIPN, CNRS (UMR 7030), Université Paris 13  
Sorbonne Paris Cité, F-93430 Villetaneuse

Nada.Mimouni, Adeline.Nazarenko, Sylvie.Salotti@lipn.univ-paris13.fr

## Résumé :

Ce papier montre que les systèmes d'accès à l'information juridique pourraient être étendus pour permettre une recherche sémantique mais aussi à base de graphes. Le défi est de retrouver des documents non seulement en fonction de leurs descripteurs de contenu, mais aussi sur la base des relations intertextuelles qu'ils entretiennent. Le papier présente une formalisation logique pour décrire les collections documentaires et un langage de requêtes simple basé sur les graphes qui est défini pour répondre aux besoins du projet Légilocal.

**Mots-clés :** Recherche d'information sémantique, analyse des besoins, langage de requêtes.

## 1 Introduction

La modélisation de l'intertextualité forme un enjeu dans le domaine juridique où les nouveaux documents (décisions administratives, les jugements, les provisions) reposent sur ceux déjà existants dont ils modifient, annulent, confirment ou appliquent dans un nouveau contexte. L'intertextualité a été identifiée comme une source majeure de complexité de la documentation juridique (Bourcier, 2011). Elle a été analysée au niveau global du réseau documentaire mais pas au niveau local qui est plus pertinent dans le cas de production de textes législatifs (*legal drafting*) ou l'analyse des cas.

La majorité des systèmes d'accès à l'information juridique permettent aux praticiens de retrouver des documents, mais pas sur des critères intertextuels. Ce phénomène d'intertextualité est apparu comme une limitation importante dans le contexte du projet Légilocal<sup>1</sup> où les agents administratifs sont amenés de façon quotidienne à analyser et à produire des documents juridiques (Amardeilh *et al.*, 2013). Lors de la rédaction d'une ordonnance, les secrétaires de mairies doivent généralement identifier la législation à laquelle il faut se référer, les anciens actes publiés sur le même sujet et en particulier ceux qui ont fait l'objet d'un recours.

---

1. Ce travail a été partiellement financé par le projet LEGILOCAL (FUI-9, 2010-2013) et par le Labex EFL (ANR-10-LABX-0083).

Ce papier présente une formalisation logique qui a été définie pour décrire et interroger les collections documentaires sur des bases intertextuelles. Il ouvre la voie vers de nouvelles fonctionnalités de recherche prenant en compte les relations intertextuelles. L'opérationnalisation a été faite par deux approches : une approche de classification conceptuelle utilisant l'analyse formelle et relationnelle de concept (Mimouni *et al.*, 2015a) et une approche utilisant les technologies du web sémantique (Mimouni *et al.*, 2015b).

La suite du papier est organisée comme suit. La section 2 définit l'objectif de cette proposition vis à vis des travaux précédents dans la recherche d'information et l'analyse des liens. Les sections 3 et 4 présentent la collection Légilocal comme un réseau sémantique de documents et le type du langage de requêtes qui peut être utilisé pour l'interroger. La section 5 montre comment ce langage de requêtes peut être utilisé pour exprimer un échantillon de requêtes exprimées par des praticiens juridiques et collectées dans le cadre du projet Légilocal. La section 6 discute les limitations du langage et les extensions possibles.

## **2 L'intertextualité dans la recherche d'information juridique**

L'« intertextualité », dans un usage restreint, est traditionnellement définie comme « une relation de co-présence entre deux ou plusieurs textes, à savoir, le plus souvent, par la présence effective d'un texte dans un autre » (Genette, 1982, p.8). Les citations - forme explicite de l'intertextualité - sont importantes à prendre en compte lorsque les textes cibles contribuent à l'interprétation des textes sources, comme il est souvent le cas pour les textes juridiques (Bhatia, 1998).

Plusieurs travaux ont reconnu l'importance de l'intertextualité des sources juridiques qui représente un facteur majeur de complexité de la documentation (Bourcier, 2011). L'analyse de réseaux a été considérée comme un moyen puissant pour modéliser les collections juridiques (Fowler *et al.*, 2007; Romain *et al.*, 2011; Winkels & de Ruyter, 2011). Cependant, comme dans l'analyse de citations et de réseaux sociaux (Rubin, 2010), l'accent a été mis sur le niveau du réseau afin d'identifier les sous-collections les plus fortement connectées ou les sources de loi les plus influentes. Moins d'attention a été accordée à l'analyse détaillée de l'intertextualité et la sémantique des liens intertextuels.

En recherche d'information, les liens intertextuels ou les citations sont généralement modélisés comme des métadonnées associées aux documents qui sont présentés aux utilisateurs. Ces derniers peuvent naviguer d'un document aux sources qu'il cite, puis à partir de ces sources vers les documents auxquels ils se réfèrent, et ainsi de suite. Ceci représente la façon hypertextuelle commune de manipulation de l'intertextualité, où l'on se perd rapidement dans l'hyperespace (Conklin, 1987).

L'intertextualité a été aussi utilisée pour le tri des documents, comme dans Brin & Page (1998), ce qui est moins pertinent dans le domaine juridique, où l'exhaustivité de la

recherche est plus importante que le classement des résultats.

Les recherches récentes dans l'analyse socio-sémantique exploitent à la fois la topologie des réseaux et la sémantique de leurs noeuds et liens (par ex. Cointet & Roth (2009)). Cela ouvre la voie à de nouvelles fonctionnalités de recherche (par ex. le *Graph Search* de Facebook). En se basant sur ces premières expériences, nous défendons l'idée que les systèmes d'accès à l'information juridique peuvent aller plus loin et exploiter sémantiquement l'intertextualité, c.à.d. comme un critère de recherche.

Dans ce qui suit, nous montrons l'avantage que les praticiens du droit peuvent tirer d'un langage de requêtes relationnelles lors de la rédaction des actes administratifs.

### 3 La collection Légilocal comme un réseau sémantique de documents

Nous considérons une collection juridique comme étant l'ensemble des documents reliés par des liens intertextuels de types différents. La collection de documents Légilocal consiste en un nombre croissant de documents produits ou cités par le réseau des secrétaires de mairies dans Légilocal qui collaborent pour l'élaboration des actes administratifs.

#### 3.1 Description de la collection

Nous considérons tout fragment de document qui peut être mis à jour ou retourné indépendamment de son document comme une unité documentaire. Dans Légilocal, tout document ou article appartenant à un document juridique est une unité documentaire et chaque unité a un identifiant unique. Dans ce qui suit,  $d_i$  se réfère à l'unité documentaire  $i$ .

La collection est composée de différents types de documents : actes administratifs utilisés comme exemples positifs ou négatifs dans le processus de rédaction, divers textes législatifs issus de juridictions supérieures et utilisées comme référence, ainsi que des documents éditoriaux (exemples et directives). Tout document  $d_i$  possède un type unique  $j$  :  $Type(d_i, t_j)$  (dans ce qui suit,  $t_j$  se réfère au type  $j$ ).

Les documents sont annotés. Nous nous focalisons ici sur les annotations sémantiques, qui sont les mots-clés du texte ou les tags des utilisateurs associés aux documents à des fins de recherche. Tout document est associé à un nombre quelconque d'attributs. Dans ce qui suit,  $Att(d_i, s_k)$  indique que le descripteur  $s_k$  est attaché au document  $d_i$ .

Les documents (ou unités documentaires) sont reliés les uns aux autres par différents types de liens dont la sémantique dépend des types des documents reliés et de la valeur du lien lui-même. Ces liens sont orientés :  $Rel(d_i, r_l, d_{i'})$  indique que  $d_i$  est la source d'un lien  $r_l$  dont la cible est  $d_{i'}$ <sup>2</sup>.

---

2. Nous simplifions cette représentation de plusieurs manières : i) les types et les descripteurs sémantiques sont en fait organisés en hiérarchies, ii) des relations ternaires existent aussi, par exemple lorsqu'un

| Types                                       | Descriptifs  |
|---|--|
| Décision                                    | décision   |
| ArrêtéMun                                   | arrêté municipal   |
| ArrêtCcass                                  | Arrêt de Cour de cassation   |
| ArrêtCappel                                 | Arrêt de Cour d'appel  |
| ArticleCode                                 | Article de code  |
| Relations                                   | Descriptifs  |
| application                                 | un texte législatif en <i>applique</i> un autre<br>ou une décision <i>applique</i> une autre décision ou un texte législatif |
| décision                                    | un jugement fait une <i>décision</i> sur un jugement précédent   |
| annulation                                  | la décision ci-dessus est une <i>annulation</i>  |
| confirmation                                | la décision ci-dessus est une <i>confirmation</i>  |
| composition                                 | un document <i>se compose d'</i> articles  |
| Descripteurs                                | Equivalents terminologiques  |
| cheminR                                     | « chemins rural »  |
| véhiculeAMoteur                             | « véhicule à moteur »  |
| Identifiants                                | Référents  |
| Code <sub>Env</sub>                         | Code de l'environnement  |
| Code <sub>CV</sub> _Article <sub>1382</sub> | Article 1382 du Code civil   |
| ArrêtCcass <sub>A</sub>                     | Arrêté A de la Cour de cassation   |
| ArrêtCappel <sub>X</sub>                    | Arrêté X de la Cour d'appel  |

TABLE 1 – Vocabulaire utilisé pour la formation de la collection Légilocal et des requêtes associées

Le tableau 1 présente une sélection du vocabulaire (types de documents et identifiants, descripteurs sémantiques et relations) utilisé pour la description de la collection Légilocal.

### 3.2 Formalisation de la collection

À partir de cette analyse, une collection documentaire  $C$  peut être modélisée comme un graphe orienté, étiqueté et attribué  $C = \mathcal{G}(D, R, A)$  où

- les noeuds sont des unités documentaires de  $D$  ;
- les unités documentaires sont décrites par des attributs, types de  $T$  et descripteurs sémantiques de  $S$  ( $A = T \cup S, T \cap S = \emptyset$ ) ;
- les arcs sont des relations binaires typées et orientées, avec des types appartenant à  $R$ .

---

document  $d_1$  indique qu'un document  $d_2$  est modifié et remplacé par un document  $d_3$  et iii) nous ne considérons pas la distinction œuvre vs. expression (Rubin, 2010; Sartor *et al.*, 2011).

$graph_{coll} \rightarrow predicate_c [ \text{'\wedge'} predicate_c ]^*$   
 $predicate_c \rightarrow \text{'Type' } ( \text{'('id_{doc}', id_{type} )'} \mid \text{'Att' } ( \text{'('id_{doc}', id_{sem} )'} \mid \text{'Rel' } ( \text{'('id_{doc}', id_{rel}', id_{doc} )'} )$   
 $id_{doc} \rightarrow \text{'d}_1' \mid \text{'d}_2' \mid \dots$   
 $id_{type} \rightarrow \text{'t}_1' \mid \text{'t}_2' \mid \dots$   
 $id_{sem} \rightarrow \text{'s}_1' \mid \text{'s}_2' \mid \dots$   
 $id_{rel} \rightarrow \text{'r}_1' \mid \text{'r}_2' \mid \dots$   
 where  $(\forall i, j, k, l) (d_i \in D, s_j \in S, t_k \in T \text{ and } r_l \in R)$ .

FIGURE 1 – Langage modélisant les collections documentaires. Les éléments du vocabulaire terminal sont notés entre guillemets simples (ex.  $\text{'('}$ ), les non-terminaux sont en italiques (ex. *prédicat*) et les métasymboles utilisés sont la flèche de réécriture ( $\rightarrow$ ), les crochets pour former les groupes ( $[ ]$ ), la barre d'alternative ( $\mid$ ) et l'étoile de Kleene pour marquer la répétition de l'élément ou du groupe précédent pour un nombre quelconque d'occurrences ( $*$ ).

Notons qu'il n'existe aucune contrainte sur le nombre de noeuds, attributs et liens dans le graphe ni sur la combinaison des attributs et liens pour une unité documentaire donnée.

Une telle collection est décrite par une formule du langage présenté dans la figure 1. La figure 2 montre le graphe d'un exemple de collection associé avec sa formule.

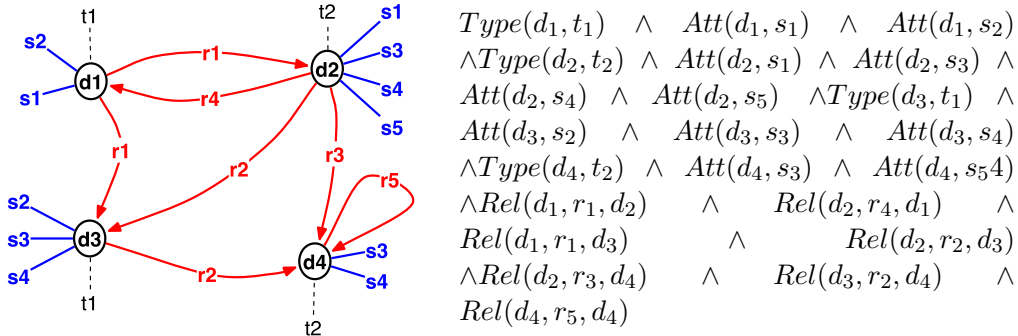


FIGURE 2 – Exemple de graphe modélisant une collection documentaire comportant 4 unités documentaires. Les attributs et relations partagés par plusieurs documents sont représentés en double. Les unités documentaires sont représentées par des cercles. Les relations sont notées comme des flèches. Les attributs sont reliés aux documents par des traits pleins (descripteurs sémantiques) ou pointillés (types de documents).

$\text{'graph}_{query} \rightarrow [\text{'focus' ':'}]? \text{graph}_q [\text{'with' constraint} [\text{'\wedge' constraint} ]^*]?$   
 $\text{focus} \rightarrow \text{'(' variable} [\text{'\wedge' variable} ]^* \text{'')}$   
 $\text{graph}_q \rightarrow \text{predicate}_q [\text{'\wedge' predicate}_q]^*$   
 $\text{predicate}_q \rightarrow \text{'Att' ' (' document ';' attribute ')'} \mid \text{'Type' ' (' document ';' type ')'} \mid \text{'Rel' ' (' document ';' relation ';' document ')')}$   
 $\text{document} \rightarrow \text{id}_{doc} \mid \text{var}_{doc}$   
 $\text{attribute} \rightarrow \text{id}_{sem} \mid \text{var}_{sem}$   
 $\text{type} \rightarrow \text{id}_{type} \mid \text{var}_{type}$   
 $\text{relation} \rightarrow \text{id}_{rel} \mid \text{var}_{rel}$   
 $\text{id}_{doc} \rightarrow \text{'d}_1' \mid \text{'d}_2' \mid \text{'d}_3' \mid \dots$   
 $\text{id}_{sem} \rightarrow \text{'s}_1' \mid \text{'s}_2' \mid \text{'s}_3' \mid \dots$   
 $\text{id}_{type} \rightarrow \text{'t}_1' \mid \text{'t}_2' \mid \text{'t}_3' \mid \dots$   
 $\text{id}_{rel} \rightarrow \text{'r}_1' \mid \text{'r}_2' \mid \text{'r}_3' \mid \dots$   
 $\text{constraint} \rightarrow \text{variable} \neq \text{variable}$   
 $\text{variable} \rightarrow \text{var}_{doc} \mid \text{var}_{sem} \mid \text{var}_{type} \mid \text{var}_{rel}$   
 where  $\text{var}_{doc} \in \mathbf{D}$ ,  $\text{var}_{sem} \in \mathbf{S}$ ,  $\text{var}_{type} \in \mathbf{T}$ ,  $\text{var}_{rel} \in \mathbf{R}$   
 and  $(\forall i, j, k, l) (\text{d}_i \in \mathbf{D} \wedge \text{s}_j \in \mathbf{S} \wedge \text{t}_k \in \mathbf{T} \wedge \text{r}_l \in \mathbf{R})$

FIGURE 3 – Langage de requêtes. La convention de notation est la même que pour la figure 1. Le méta-symbole ? indique que l'élément ou le groupe précédent se produit au plus une fois.

#### 4 Requetes relationnelles et structurées

Modéliser les collections documentaires comme des graphes nous amène à considérer l'interrogation avec des graphes (Khan *et al.*, 2012) comme une approche de recherche d'information où les requêtes se formalisent elles même sous forme de graphes. Un graphe de requête est similaire à celui de la collection mais qui peut contenir :

- des variables à la place des identifiants des documents, attributs, types et relations,
- des contraintes d'inégalité sur ces variables,
- une cible pour restreindre la réponse à un sous ensemble des variables de la requête.

Un graphe requête est donc décrit par une formule du langage donné par la grammaire de la figure 3.

L'appariement entre les requêtes et les documents revient à instancier le graphe de requête sur le graphe de la collection. Une réponse est :

- un ensemble contenant tous les sous-graphes du graphe de la collection qui instancient le graphe requête s'il ne possède pas une cible explicite,
- un ensemble contenant tous les tuples d'identifiants qui instancient la cible de la requête si elle possède un,

- un ensemble vide si le graphe requête ne peut être instancié.

Prenons quelques exemples de requêtes (avec et sans cible et contraintes) ainsi que les réponses produites par leur appariement sur l'échantillon de la collection de la figure 2 :

1.  $Att(x, s_1) \wedge Rel(x, r_1, d_2)$   
*Trouver tous les sous-graphes composés d'un document décrit par  $s_1$  et ayant pour cible  $d_2$  par la relation  $r_1$ .*  
 Résultat :  $\{Att(d_1, s_1) \wedge Rel(d_1, r_1, d_2)\}$  (1 graphe)
2.  $Rel(x, y, x)$   
*Trouver tous les documents en relation avec eux-même.*  
 Résultat :  $\{Rel(d_4, r_5, d_4)\}$  (1 graphe)
3.  $(y) : Rel(x, y, x)$   
*Trouver tous les types de relation reliant un document à lui-même.*  
 Résultat :  $\{r_5\}$  (1 relation)
4.  $(x, y) : Rel(x, y, x)$   
*Trouver tous les couples composés d'un document lié à lui-même et du type de la relation.*  
 Résultat :  $\{(d_4, r_5)\}$  (1 couple composé d'un document et d'une relation)
5.  $(x, y) : Att(x, z) \wedge Rel(x, r_1, y) \wedge Att(y, z)$   
*Trouver tous les couples de documents décrits par un même descripteur sémantique et tels que le second est la cible du premier par la relation  $r_1$ .*  
 Résultat :  $\{(d_1, d_2), (d_1, d_3)\}$  (2 couples)
6.  $Att(x, s_2) \wedge Att(x, y)$  avec  $y \neq s_2 \wedge y \in S$   
*Trouver tous les sous-graphes composés d'un document décrit par  $s_2$  et un autre descripteur sémantique différent.*  
 Résultat :  $\{Att(d_1, s_1) \wedge Att(d_1, s_2), Att(d_3, s_2) \wedge Att(d_3, s_3), Att(d_3, s_2) \wedge Att(d_3, s_4)\}$   
 (3 graphes)
7.  $Type(x, y) \wedge Type(x, z)$  avec  $y \neq z$   
*Trouver les documents de deux types différents.*  
 Résultat :  $\emptyset$

## 5 L'intertextualité dans les requêtes des praticiens

Le langage de requête ci-dessus offre une manière homogène et naturelle pour exprimer un large éventail de requêtes des praticiens, qui combinent souvent le contenu et les critères intertextuels. Les exemples de requêtes suivantes ont été toutes recueillies dans le cadre de l'analyse des besoins du projet Légilocal. Nous montrons comment elles peuvent

être formalisées utilisant le langage de requêtes ci-dessus, en laissant de côté les questions relatives à la formulation des requêtes et leur réponses. Cette formalisation suppose que les documents sont correctement analysés, leurs types sont identifiés, ils sont annotés avec des descripteurs sémantiques et les liens intertextuels sont eux-mêmes identifiés et sémantiquement typés.

1. "Quelles sont les décisions de jurisprudence qui citent l'article 1382 du code civil ?"  
 $(x) : Type(x, Decision) \wedge Rel(x, application, Code_{CV\_Article1382})$   
 Le terme générique "cite" est interprété comme une relation d'application à cause des types des documents reliés, une décision de jurisprudence et un texte législatif.
2. "Quels sont les articles du code de l'environnement qui parlent de véhicules à moteurs ?"  
 $Type(x, ArticleCode) \wedge Rel(x, composition, Code_{Env}) \wedge Att(x, vehiculeAMoteur)$
3. "Je voudrais la décision qui fait l'objet de l'arrêt A de la Cour de cassation."  
 $(x) : Type(x, Decision) \wedge Rel(ArretCass_A, decision, x)$
4. "Je cherche les décisions qui ont été annulées par la Cour de cassation."  
 $(x) : Type(x, Decision) \wedge Rel(y, annulation, x) \wedge Type(y, ArretCass)$   
 $Type(x, Decision) \wedge Rel(y, decision, x) \wedge Type(y, ArretCass)$   
 Ces deux formules de requêtes ne diffèrent que par leurs cibles : dans la première, la requête devrait avoir comme réponse une ou plusieurs décisions tandis que la deuxième aurait comme réponse une liste de graphes quiinstancient le graphe requête (voir figure 4).

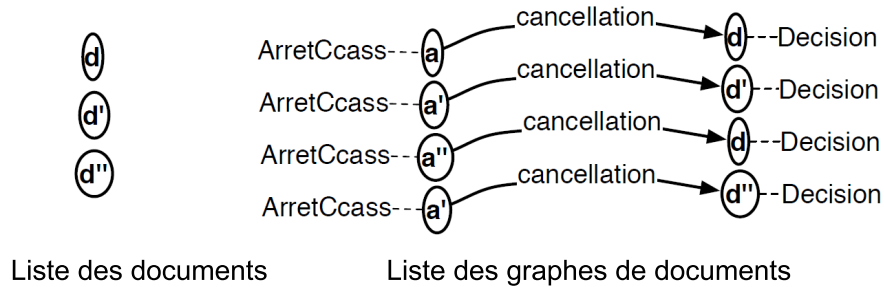


FIGURE 4 – Exemples de réponses associées à la requête 4 : une liste de documents (1ère formule) ou une liste de graphes de documents (2ème formule).

5. "Je voudrais savoir si cet arrêt de la cour d'appel a fait lui-même objet d'un recours."  
 $Rel(x, decision, ArretCappel) \wedge Type(x, ArretCass)^3$
6. "Je cherche des arrêtés municipaux concernant les chemins ruraux qui ont fait l'objet d'un recours et ont été annulés par une décision de jurisprudence."  
 $(x) : Type(x, ArreteMun) \wedge Att(x, cheminR) \wedge Rel(y, annulation, x)$

---

3. Le type du jugement final ( $Att(x, ArretCass)$ ) dépend de la procédure d'appel suivie.



7. Quels sont les articles de code qui ont été confirmés et qui sont cités par les arrêtés municipaux parlant de chemins ruraux ?"

$$(x) : Type(x, ArticleCode) \wedge Type(y, ArreteMun) \wedge Att(y, cheminR) \wedge Rel(y, application, x) \wedge Rel(z, confirmation, y)$$

8. "Je souhaite savoir si les textes visés par des arrêtés municipaux parlant de chemins ruraux sont aussi cités par ceux concernant les véhicules à moteurs."

$$Type(x, ArreteMun) \wedge Att(x, cheminR) \wedge Rel(x, application, y) \wedge Rel(z, application, y) \wedge Type(z, ArreteMun) \wedge Att(z, vehiculeAMoteur)$$

9. "Quels sont les arrêtés municipaux qui ont fait l'objet de deux recours ?"

$$(x) : Type(x, ArreteMun) \wedge Rel(y, decision, x) \wedge Rel(z, decision, x) \text{ with } y \neq z$$

## 6 Discussion

Pour des fins de démonstration, nous gardons le langage de requête aussi simple que possible. Nous nous sommes concentrées sur l'intertextualité, mais le langage de requête ci-dessus doit naturellement être étendu pour interroger les documents avec leurs métadonnées (par ex. la date de publication ou l'auteur), les documents sources (les œuvres) doivent être différenciés de leurs versions (expressions), tel que proposé par Sartor *et al.* (2011).

Le langage de requêtes ci-dessus permet de traiter l'intertextualité mais présente des limites.

*Quantification* Les requêtes en langage naturel impliquent des hypothèses de (non-)unicité qui ne sont pas exprimables dans le langage proposé. Par exemple, les variantes de requêtes suivantes sont considérées comme équivalentes dans notre langage de requête, où les variables sont quantifiées de manière existentielle : « Quels sont les jugements qui confirment une/des/plusieurs décision(s) ... ». Même si la quantification universelle permettrait d'exprimer des requêtes telles que « Y a-t-il un article de code cité par tous les arrêtés portant sur les chemins ruraux ? », nous avons choisi de ne pas l'inclure dans un premier temps, car elle est difficile à maîtriser pour les utilisateurs et qu'elle n'apparaissait pas dans les requêtes recueillies.

*Négation et disjonction* Pour préserver la simplicité du langage pour les utilisateurs, nous avons choisi de ne pas inclure la négation ou la disjonction des opérateurs dans la spécification du langage de requête, ce qui est une limitation en ce qui concerne les besoins des praticiens. Par exemple, la requête « Quelles sont les décisions antérieures à la décision D ? » doit être formulée de la manière suivante :

$$(x) : Type(x, Decision) \wedge (Rel(decision_D, decision, x) \vee (Rel(decision_D, decision, y) \wedge Rel(y, decision, x)))$$

pour prendre en compte différentes longueurs de chaînes de décision. Aussi, sans opérateur de négation, une requête comme « Quels sont les articles qui ne sont pas annulés ? » ne peut être formalisée que comme « Quels sont les articles qui ont été confirmés ? », qui est plus restrictive.

*Cible de requête* Il est souvent difficile d'identifier si une requête en langage naturel est ciblée ou non. Même si on est habitué à avoir des listes de documents, nous nous attendons à ce que les utilisateurs spécialisés apprécient un large éventail de types de réponses. Les graphes réponses donnent plus de contexte et peuvent être affinés grâce à une interface interactive. La différence ne réside pas dans la mise en correspondance du graphe de la requête et de la collection, mais dans la présentation des résultats.

*Opérateur de comptage* Jusqu'à présent, nous n'avons recueilli aucune requête nécessitant un opérateur de comptage, mais ce point doit être étudié davantage.

*Topologie de graphe* Nous n'avons mis aucune contrainte sur la taille des graphes de requêtes ni sur la présence de cycles. Même si les exemples ci-dessus de graphes de requêtes sont simples, nous nous attendons à ce que les utilisateurs spécialisés entrent progressivement des requêtes plus complexes.

## 7 Conclusion

L'analyse des besoins dans le projet Légilocal a révélé les limitations des systèmes existants d'accès à l'information juridique, qui permettent aux utilisateurs de retrouver des documents en fonction de leurs métadonnées et les descripteurs de contenu mais pas en fonction des relations intertextuelles qui sont néanmoins critiques dans l'analyse juridique.

En outre, l'état de recherche et des technologies permet aujourd'hui de développer des approches socio-sémantiques de recherche (Cointet & Roth, 2009) qui permettent de rechercher dans les grands graphes attribués exploitant à la fois les attributs des noeuds et la structure du réseau. Nous soutenons l'idée qu'une approche similaire peut être adoptée pour la recherche d'information juridique.

Les collections juridiques peuvent être modélisées comme des réseaux sémantiques de documents, où les noeuds (les documents) sont associés à des attributs sémantiques et sont reliés les uns aux autres par divers types de liens sémantiques.

Ce papier montre l'avantage qui peut être tiré d'un simple langage de graphe de requêtes et ouvre la voie vers une nouvelle forme de recherche sémantico-relationnelle dans les sources juridiques.

Beaucoup de travail reste à faire. Nous considérons que les premiers outils de démonstration doivent être très simples et seront enrichis progressivement par de nouvelles fonctionnalités lorsque les utilisateurs prennent l’habitude des premiers tests. Toutefois, exploiter un langage de graphe de requêtes, aussi simple qu’il soit, nécessite une interface utilisateur adéquate. Parmi les différents types de modes d’interrogation, ceux basés sur un formulaire ou sur la technique "*fill-in-the-blank*" sont probablement les plus commodes, mais cela reste à évaluer. L’approche proposée doit être testée sur une collection de taille croissante au sein du projet Légilocal. L’interrogation est actuellement mise en œuvre avec SPARQL et des outils sont développés en parallèle pour annoter les documents de la collection et construire le réseau sémantique des documents Légilocal.

## Références

- AMARDEILH F., BOURCIER D., CHERFI H., DUBAIL C., GARNIER A., GUILLEMIN-LANNE S., MIMOUNI N., NAZARENKO A., PAUL ÈVE., SALOTTI S., SEIZOU M., SZULMAN S. & ZARGAYOUNA H. (2013). The légilocal project : the local law simply shared. In K. D. ASHLEY, Ed., *Legal Knowledge and Information Systems - JURIX 2013 : The Twenty-Sixth Annual Conference*, volume 259, p. 11–14, University of Bologna, Italy : IOS Press.
- BHATIA V. K. (1998). Intertextuality in legal discourse. *JALT Journal Online*, (1).
- BOURCIER D. (2011). Sciences juridiques et complexité. un nouveau modèle d’analyse. *Droit et Cultures*, **61**(1), 37–53.
- BRIN S. & PAGE L. (1998). The anatomy of a large-scale hypertextual web search engine. In *Proceedings of the seventh international conference on World Wide Web (WWW7)*, p. 107–117, Amsterdam, The Netherlands : Elsevier Science Publishers B. V.
- COINTET J. & ROTH C. (2009). Socio-semantic dynamics in a blog network. In *Computational Science and Engineering, 2009. CSE '09. International Conference on*, volume 4, p. 114–121.
- CONKLIN J. (1987). Hypertext : An introduction and survey. *IEEE Computer*, **20**(9), 17–41.
- FOWLER J. H., JOHNSON T. R., SPRIGGS J. F., JEON S. & WAHLBECK P. J. (2007). Network analysis and the law : Measuring the legal importance of precedents at the u.s. supreme court. *Political Analysis*, **15**, 324–346.
- GENETTE G. (1982). *Palimpsestes*. Poétique. Le Seuil.
- KHAN A., WU Y. & YAN X. (2012). Emerging graph queries in linked data. In *Data Engineering (ICDE), 2012 IEEE 28th International Conference on*, p. 1218–1221, USA.
- MIMOUNI N., NAZARENKO A. & SALOTTI S. (2015a). A conceptual approach for relational IR : application to legal collections. In J. BAIXERIES, C. SACAREA & M. OJEDA-ACIEGO, Eds., *Formal Concept Analysis - 13th International Conference, ICFCA 2015, Nerja, Spain, June 23-26, 2015, Proceedings*, volume 9113 of *Lecture Notes in Computer Science*, p. 303–318 : Springer.
- MIMOUNI N., NAZARENKO A. & SALOTTI S. (2015b). Une ontologie documentaire pour l’accès aux contenus juridiques. In *26es Journées francophones d’Ingénierie des Connaissances (IC), Vers le traitement de la masse de données disponibles sur le web, 29 Juin - 3 Juillet, 2015*,

Rennes - France.

- ROMAIN, MAZZEGA P. & BOURCIER D. (2011). A network approach to the french system of legal codes- part i : Analysis of a dense network. *Journal of Artificial Intelligence and Law*, **19**, 333–355.
- RUBIN R. (2010). *Foundations of Library and Information Science*. Neal-Schuman Publishers.
- SARTOR G., PALMIRANI M., FRANCESCONI E. & BIASIOTTI M. A. (2011). *Law, Governance and Technology : Legislative Xml for the Semantic Web : Principles, Models, Standards for Document Management*. Law, Governance and Technology Series, 4. Springer London, Limited.
- WINKELS R. & DE RUYTER J. (2011). Survival of the fittest : Network analysis of dutch supreme court cases. In *AICOL*, p. 106–115.