

## **QUATRIÈME PARTIE**

### **L'information juridique et l'intelligence artificielle**



# Apports d'une indexation à facettes pour la représentation et le repérage des décisions de justice

Sabine Mas\*, Michelle Cumyn\*\*, David Lesieur\*,  
Cécile Gaiffe\*, Charles Tremblay-Potvin\*\*

## 1. Introduction

Le droit contemporain a atteint un degré de complexité inédit. L'inflation des sources du droit ne semble pas vouloir s'estomper. À titre d'exemple, le Centre d'accès à l'information juridique (CAIJ), « la plus grande source d'information juridique au Québec » (CAIJ, 2021a), recense en 2020 près de deux millions de documents de jurisprudence québécoise et canadienne (CAIJ, 2021b).

L'explosion de la documentation disponible en ligne et la densification récente de la jurisprudence ont transformé la recherche et l'accès à l'information juridique. La recherche juridique ne peut plus prétendre à l'exhaustivité, mais seulement à la pertinence des résultats obtenus. Paradoxalement, le tri des décisions publiées qu'effectuaient jadis les éditeurs juridiques permettait aux chercheurs de réaliser une recherche considérée comme exhaustive. Aujourd'hui, la diffusion de plus en plus exhaustive des décisions, particulièrement au Québec, oblige les chercheurs à réaliser, tant bien que mal, ce tri. Depuis plusieurs

---

\* École de bibliothéconomie et des sciences de l'information, Université de Montréal.

\*\* Faculté de droit, Université Laval.

décennies déjà, des auteurs appréhendent une crise des sources, trop abondantes et de qualité variable, ce qui mettrait en péril l'accès au droit et son effectivité (Atias, 1985, p. 46 ; Arthurs, 1995).

Les bases de données ont aussi transformé le comportement des chercheurs d'information juridique. La recherche par mots-clés dans le texte intégral (recherche plein-texte) s'est ajoutée à celle qui consistait à repérer des documents à l'aide des termes utilisés pour les classer ou les indexer (recherche à l'aide des termes d'indexation). On observe aussi une tendance chez les utilisateurs de bases de données juridiques à privilégier la recherche à partir des éléments factuels d'un problème (par exemple, « chute trottoir glace »), négligeant la recherche à l'aide de catégories juridiques (par exemple, « responsabilité civile » ou « droit municipal ») (Delgado et Stefancic, 2007, p. 309-310). Or « l'organisation systématique de la documentation dans le domaine s'appuie en partie sur les régimes ou catégories juridiques, dont la connaissance est jugée essentielle à l'efficacité du repérage de sources pertinentes à la résolution d'un problème de droit » (Hudon et Cumyn, 2017, p. 82). Des documents se rapportant aux mêmes questions juridiques, quoique décrivant des faits différents, risquent d'être omis, ce qui dénote une perte de la capacité à qualifier juridiquement les faits et à raisonner par analogie.

L'organisation de la matière juridique à des fins d'enseignement et de recherche reflète la classification des catégories et des règles juridiques dont l'acquisition constitue l'un des objectifs de la formation universitaire. La délimitation des domaines du droit est demeurée sensiblement la même depuis la fin du 19<sup>e</sup> siècle. Elle est souvent critiquée, parce qu'elle tend à marginaliser les catégories juridiques d'apparition récente – pensons aux lois dites « sectorielles » – et qu'elle ne rend pas compte des nombreuses situations qui interpellent plusieurs domaines du droit à la fois. Par exemple, la violence intrafamiliale fait appel au droit pénal (voies de

fait), au droit de la famille (séparation ou divorce ; garde des enfants ; attribution de la résidence familiale), au droit des obligations (bail et attribution du logement), à la protection de la jeunesse et à la responsabilité civile. La formation juridique prépare mal les étudiants à aborder de telles problématiques « transversales ». Ainsi, on reproche aux facultés de droit de développer une pensée en silos. Est-il suffisant de perfectionner et de rendre plus explicites les classifications traditionnelles, ou faut-il envisager une réorganisation plus radicale de la matière juridique à des fins d'exposition et d'enseignement, afin de combattre la pensée en silos ?

Ces éléments contextuels expliquent pourquoi nous avons voulu explorer et développer un modèle original pour la représentation et le repérage de l'information juridique disponible en ligne. Ce modèle permettrait d'accéder à la documentation à partir d'éléments factuels ou de catégories juridiques. Il faciliterait la qualification juridique des faits, c'est-à-dire la détermination de l'ensemble des règles de droit qui sont applicables à la situation à l'étude, par l'association d'une ou plusieurs catégories juridiques aux termes d'indexation de nature factuelle. Il favoriserait le raisonnement par analogie qui consiste à repérer des documents se rapportant aux mêmes questions juridiques, même si les faits sont différents. Enfin, il mettrait en évidence les liens entre certains éléments factuels et une pluralité de catégories juridiques, ce qui permettrait de dépasser les silos.

Le présent texte vise à présenter sommairement la spécificité et les enjeux liés à la recherche d'information juridique, avant de s'attarder sur les fonctions de l'indexation de manière générale et de la documentation juridique en particulier ainsi que sur la nature et l'apport des facettes pour la représentation et le repérage de l'information juridique. Nous terminerons en faisant état du protocole expérimental utilisé dans le cadre d'une étude d'utilisabilité en cours de réalisation visant à vérifier dans quelle mesure un modèle d'indexation à facettes

à la fois basé sur les faits à l'origine d'un litige et sur les catégories juridiques permet de faciliter le raisonnement par analogie et la qualification juridique de ces faits.

## **2. La recherche d'information juridique**

Un juriste est appelé à donner une opinion juridique sur la base du droit applicable à une situation donnée, ce qui rend nécessaire la recherche d'informations. Les décisions de justice sont considérées comme une source privilégiée de compréhension du raisonnement juridique en général (Laprise, 2000, p. 1).

Dans le contexte d'une recherche d'information juridique en ligne, le besoin d'information du juriste doit être traduit par une ou plusieurs requêtes de recherche composées de termes de recherche, soit un ensemble de mots et expressions linguistiques (Guitard, 2018, p. 46) qui sera compris par le système de recherche. Un « terme » est constitué de deux éléments basiques : sa forme linguistique, qui est le mot dans sa réalisation alphabétique, et le contenu, ou concept, sous la forme d'une expression linguistique (Peruginelli, 2011, p. 269). En tant que tels, les termes et la façon dont ils sont utilisés dans la requête reflètent les composants essentiels de la résolution du problème de l'utilisateur et des prises de décision qu'il opère en interaction avec tout système de recherche (Jansen, Spink et Pfaff, 2000, s. p).

Selon Le May et Thiboutot (2014), le travail de qualification juridique d'une situation factuelle est une « condition préalable à toute démarche de recherche » (2014, p. 7 cité dans Lesieur, 2020, p. 20). Or, d'après plusieurs théories sur le comportement informationnel, le processus de recherche réalisé aux fins de la qualification juridique est plutôt un processus itératif, qui est renseigné par la démarche de recherche elle-même. Par exemple, selon le modèle du « *berrypicking* » de Bates (1989), le parcours du chercheur d'information, au lieu de correspondre

à une trajectoire linéaire conduisant à une « destination » prévisible, serpente en fonction de ses trouvailles, au fil de requêtes successives, jusqu'à ce que le besoin d'information soit comblé. Par ailleurs, selon le modèle « *information search process* » de Kuhlthau (2010), le chercheur d'information aborde d'abord son problème avec des idées encore vagues puis dégage du sens et élargit ses connaissances à chaque étape de son processus de recherche, à mesure qu'il recueille de l'information, traversant une succession d'états cognitifs (et affectifs) l'amenant à préciser sa pensée. Autrement dit, la recherche d'information est présentée comme un processus de construction par lequel l'incertitude diminue au fur et à mesure de l'augmentation de la compréhension. Ces théories, lorsque transposées dans le domaine du droit, permettent de penser que la recherche d'information ne sert pas seulement à obtenir des réponses précises : la démarche même de recherche d'information peut aider le juriste à découvrir des enjeux qu'il n'avait pas considérés au début du processus et à construire sa stratégie et son argumentaire (Lesieur, 2020, p. 20).

La recherche d'information est essentielle au travail du juriste, mais relève de processus cognitifs complexes par lesquels le chercheur d'information doit cerner son besoin informationnel et le « traduire » dans une forme lui permettant d'interroger le système d'information, papier ou numérique. C'est ici que le rôle de l'indexation intervient en accompagnant le chercheur d'information dans ces processus.

### **3. Fonctions de l'indexation de la documentation juridique**

L'indexation désigne le processus qui consiste à représenter de manière concise, sous forme de termes d'indexation, le contenu d'un document afin d'en faciliter la recherche ultérieure. Dès leur apparition au 13<sup>e</sup> siècle, les index, outils résultant de l'indexation, comprenaient une liste de termes classés en ordre alphabétique et associés à des localisateurs afin

de retrouver rapidement les passages souhaités (Diepeveen, 2006, p. 74). L'indexation, devenue au fil du temps une activité professionnelle<sup>1</sup>, et la production d'index qui en a découlé sont restées exclusivement manuelles, jusqu'à la première moitié du 20<sup>e</sup> siècle (Maniez et Maniez, 2009).

Par la suite, les techniques d'automatisation offertes par l'ordinateur bénéficient aux index produits à moindre coût (Maniez et Maniez, 2009), rapidement et avec un plus haut niveau de constance que l'indexation manuelle (Shmeer et Sidlo, 2001, p. 87). La technologie numérique a rendu les documents facilement accessibles qui autrement n'étaient pas disponibles et a aussi amélioré la recherche en créant une multiplicité de points d'accès grâce à des techniques de recherche avancées (Cohen, 2007). Elle a rendu possible la recherche d'occurrences de mots ou d'expressions dans les bases de données en texte intégral et remet en question la pertinence de l'indexation.

Tandis que certains pourraient envisager que la technologie remplace l'indexation, d'autres continuent de croire au caractère essentiel des index en rappelant l'enjeu de la croissance exponentielle de la documentation en droit pour en justifier la pertinence (Knapp, 2001, p. 44). De plus, pour qu'une recherche plein-texte soit efficace, les utilisateurs doivent utiliser les mots-clés appropriés. Or, la plupart des mots sont ambigus et ont des synonymes ou des variantes (Shmeer et Sidlo, 2001, p. 86). Par exemple, l'emploi du mot « prescription » dans une recherche plein-texte permet de découvrir des documents où il est question d'une prescription de médicaments, mais aussi des documents qui concernent la prescription au sens juridique, c'est-à-dire l'acquisition ou la perte d'un droit par l'écoulement du temps : ce mot

---

1. L'indexation des documents publiés a fait l'objet de plusieurs normes nationales et internationales en documentation dont les deux principales sont ISO 5963 :1985 et NF Z 47-102 :1993. Selon ces normes, le processus d'indexation peut se décomposer en trois étapes : l'appréhension du contenu du document, l'identification des concepts représentant ce contenu et la sélection des concepts nécessaires à la recherche ultérieure (Guitard, 2018, p. 38).

désigne ainsi deux concepts tout à fait distincts. Par ailleurs, l'emploi du mot « automobile », « véhicule » ou « voiture » dans une recherche est susceptible de mener à des ensembles de documents différents : pourtant, ces trois mots désignent un même concept. Enfin, la recherche exclusivement plein-texte conduit souvent à des documents factuellement semblables tout en manquant certaines références conceptuellement pertinentes (Cohen, 2007, p. 390). L'utilisation judicieuse d'un index permet de réduire ces difficultés en rendant explicites les relations sémantiques entre les termes, notamment à l'aide de renvois (*voir et voir aussi*).

Aujourd'hui, deux méthodes d'indexation sont possibles pour représenter le contenu d'un document textuel, soit l'indexation par extraction et l'indexation par assignation. Dans la première, les termes d'indexation sont extraits du contenu du document, généralement au moyen d'outils informatiques s'appuyant sur des traitements statistiques et linguistiques. La seconde méthode est plus subjective et, au contraire de la première, source de nouvelles informations : les termes d'indexation sont choisis par l'indexeur pour représenter les sujets dont traite le document en s'appuyant sur la langue naturelle (indexation en vocabulaire libre) ou sur un langage documentaire (indexation en vocabulaire contrôlé) (Lesieur, 2020).

Dans le contexte de l'indexation d'un contenu juridique, l'indexeur doit travailler avec le langage du droit qui est une langue de spécialité extrêmement riche, complexe et qui comporte des caractéristiques spécifiques (Gémar, 1980, p. 893). De toutes les langues spécialisées, la langue juridique est peut-être celle où règne la plus grande polysémie<sup>2</sup> (Gémar, 1981, p. 342 et Gémar, 1980, p. 886). Cela ne facilite pas les choses dans le cas d'une recherche d'information juridique, car l'ambiguïté entourant la polysémie des termes est déjà

---

2. Propriété d'un terme qui représente plusieurs concepts distincts et auquel sont par conséquent attribuées plusieurs significations (Hudon, 2013).

identifiée comme un problème dans le domaine de la recherche d'information en général (Maxwell et Schafer, 2010, p. 70).

Le droit est également polyphone<sup>3</sup> et polymorphe (Gémar, 1991, p. 280), car il varie en fonction des domaines du droit et des nombreux types de discours qui le composent. Par exemple, le « vol » qui justifie le congédiement d'un employé n'a pas le sens précis de l'infraction de vol réprimée par le *Code criminel*, dont la portée peut être à certains égards plus large ou plus étroite. De même, la notion d'intention n'a pas la même portée dans tous les domaines du droit. Le langage du droit est donc multiforme. Gémar parle même de « l'incertitude sémantique » qui caractérise le langage du droit (Gémar, 1981, p. 342 et Gémar, 1980, p. 887).

De manière générale et quel que soit le domaine, l'indexeur joue « un rôle de médiateur pour faciliter la communication entre le créateur d'un document et l'utilisateur » (Hudon, 2013, p. 40), avec les responsabilités de représenter adéquatement le contenu et de discerner si le document sera pertinent pour un chercheur d'information utilisant comme clé d'accès l'un ou l'autre des termes assignés au moment de l'indexation. Ainsi, comme la recherche d'information, l'indexation par assignation relève d'un processus cognitif complexe (Milstead, 1994) (Lesieur, 2020, p. 22).

Par conséquent, si l'indexation par extraction est plus facilement automatisée que l'indexation par assignation, on ne peut s'attendre à la même qualité d'indexation, particulièrement dans un contexte juridique. En effet, Pouliquen *et al.* (2003) ont constaté que, dans leur corpus juridique, les termes d'indexation ne figurent habituellement pas dans les documents eux-mêmes, une observation corroborée par Stranieri et Zeleznikow (2005, p. 160), qui expliquent que la classification de textes juridiques implique souvent l'assignation de concepts abstraits dont les

---

3. « Il emprunte, pour exprimer son message, de très nombreuses voies et revêt d'innombrables formes. »

termes n'apparaissent pas forcément dans les textes. Cela renvoie à cette idée de l'indexeur agissant comme médiateur entre les juges, auteurs des décisions, et les personnes à la recherche d'information. Pour remplir son rôle, l'indexeur s'efforce d'assigner les termes d'indexation qui correspondent le plus adéquatement aux concepts à représenter, sans égard à la manière dont ces concepts peuvent être représentés (ou non)<sup>4</sup> dans les textes (Lesieur, 2020, p. 23).

Dès lors que certains mots du langage du droit ne peuvent passer pour des termes uniquement juridiques, étant donné que l'on peut attribuer ces termes-là indifféremment à la langue courante ou à la langue spécialisée (Gémar, 1980, p. 882), se pose également le problème du classement des termes juridiques ambigus :

Dans quelle catégorie, en effet, classer des termes comme acte, accord, droit, équité, crime, infraction, motif, etc. ? Dans celle des termes techniques ou scientifiques ? Ou dans le corpus de la langue courante, car ils ne peuvent passer pour des mots uniquement juridiques. (Gémar, 1979, p. 44)

Une des solutions proposées afin de réduire l'ambiguïté du langage du droit serait d'apporter davantage de contexte aux concepts afin d'en préciser la portée (Peruginelli, 2011, p. 269). C'est d'autant plus pertinent dans un contexte multilingue et bijuridique, comme au Québec, car le langage du droit est intimement lié au système juridique d'un pays. D'ailleurs, la traduction de certains textes juridiques fondamentaux des pays bilingues ou multilingues porte les stigmates de sa complexité (Gémar, 2007, p. 151). Le fait d'ajouter du contexte aux concepts permettrait donc de faciliter la compréhension d'un concept précis, même dans une transposition d'un langage du droit à un autre, d'une culture juridique à une autre.

---

4. Les concepts juridiques sont des principes de raisonnement qui ne sont souvent pas explicites dans le texte (Maxwell et Schafer, 2010, p. 67).

Pour toutes ces raisons, l'indexation basée sur l'emploi de vocabulaires contrôlés bénéficie directement au chercheur d'information. Gross, Taylor et Joudrey (2015, p. 31) énumèrent plusieurs avantages de puiser les termes d'indexation dans des vocabulaires contrôlés. En particulier, l'utilisation d'un vocabulaire contrôlé pour l'indexation permet de regrouper les synonymes et les variations orthographiques et de forme, fournir des références pour les termes obsolètes, distinguer les différents sens des termes polysémiques, fournir des références hiérarchisées, fournir des clés repérables avec les moteurs de recherche, supporter la navigation à facettes, ou guider les calculs de pertinence (Lesieur, 2020, p. 24).

En plus de bénéficier directement au chercheur d'information, l'interposition d'un vocabulaire contrôlé améliore la qualité de l'indexation. Il a en effet été démontré que l'indexation par assignation en vocabulaire contrôlé donne de meilleurs résultats (en termes de rappel et de précision) que l'indexation en vocabulaire libre, qu'elle soit accomplie par les humains ou par les machines (Paroubek, Zweigenbaum, Forest et Grouin, 2012) (Lesieur, 2020, p. 24).

En ce qui a trait aux besoins des juristes, il semble alors assez clair que ces atouts des vocabulaires contrôlés sont susceptibles de mieux répondre à l'exigence d'exhaustivité et de pertinence des résultats dans la recherche d'information. Par ailleurs, les vocabulaires contrôlés au sein desquels les concepts sont hiérarchisés peuvent permettre de naviguer entre des concepts tantôt plus précis, tantôt plus généraux, un atout pour les juristes, puisqu'en droit « il faut constamment passer d'une chose au concept » (Le May et Thiboutot, 2014, p. 18) (Lesieur, 2020, p. 24).

Notre modèle repose sur l'indexation par assignation au moyen d'un vocabulaire contrôlé hiérarchisé. En plus de retenir ces techniques éprouvées, il s'inspire de la théorie des facettes. Celle-ci permet d'envisager l'information sous différentes

perspectives, ce qui pourrait être avantageux pour le juriste à la recherche d'une qualification juridique ou d'analogies (Cumyn, Hudon, Mas et Reiner, 2018).

#### 4. Présentation et apport des facettes

Alors que l'idée de facette remonte à la division logique d'Aristote (Hjørland, 2013), c'est seulement au début du 20<sup>e</sup> siècle qu'apparaissent les premières ébauches<sup>5</sup> de la notion de facette (Côté-Lapointe et Mas, 2017). Il faut cependant attendre la parution des travaux d'un bibliothécaire et mathématicien indien, Shiyali Ramamrita Ranganathan, pour voir apparaître une méthode de classification analytico-synthétique pour décomposer un sujet en facettes.

Le terme « facette », emprunté au vocabulaire courant, est lui-même polysémique et est utilisé dans de multiples domaines tels que les sciences de l'information, l'informatique et la linguistique. Côté-Lapointe et Mas (2017) distinguent trois modalités différentes d'application des facettes : l'analyse par facettes qui désigne le processus analytico-synthétique, les structures à facettes qui désignent les langages et schémas classificatoires (par exemple, vocabulaires contrôlés, plans de classification) résultant de l'analyse par facettes, et les dispositifs à facettes qui désignent les moyens technologiques (par exemple, interfaces de recherche, bases de données, méta-données) qui permettent à un usager d'interagir avec des objets ou des sujets à l'aide de facettes.

Plus spécifiquement, l'analyse par facettes consiste à classifier les termes d'un domaine donné de connaissances dans des classes homogènes, mutuellement exclusives et multidimensionnelles, selon le principe de division logique (classes déduites par une seule caractéristique de division). Elle

---

5. Par exemple, Julius Otto Kaiser (un des premiers à utiliser des catégories de division s'apparentant aux facettes), Paul Otlet (créateur de la classification décimale universelle), ou Henry E. Bliss (auteur du système de classification Bliss).

est basée sur l'identification de catégories ou caractéristiques propres à un domaine donné permettant de rendre compte des dimensions fondamentales de tout objet. Plusieurs catégories fondamentales sont proposées dans la littérature telles que « lieu, temps et forme » (classifications CDU et Bliss), « réalités, processus, pays » (Julius Otto Kaiser), « entité, états, processus » (Jacques Maniez) ou « le système (ou entité), ses types, ses parties, ses matériaux, ses propriétés, ses processus, les opérations effectuées dessus, les agents, les agents du processus, les agents des opérations, le lieu, le temps, les formes de présentation » (Classification Research, Group).

Quant à lui, Ranganathan a défini un nombre limité de cinq catégories fondamentales. Les cinq catégories qu'il privilégiait, et dont il espérait qu'elles auraient une application universelle, étaient la « personnalité » (personne ou objet), la « matière » (substance, propriété ou qualité), l'« énergie » (opération principale, processus ou action), l'« espace » (lieu) et le « temps » (date, durée ou période). Autrement dit, Ranganathan soutenait que tout sujet de la connaissance ou tout sujet d'un document touche à l'un ou plusieurs de ces cinq aspects de la réalité. Ainsi, la notion de facettes désigne la ou les catégories, l'aspect ou « le trait commun à un ensemble d'éléments » (Hudon et El Hadi, 2017, p. 12) qui découlent de l'application des catégories fondamentales à un ensemble donné.

Une structure à facettes ou classification à facettes se présente comme une combinaison de groupes conceptuels complémentaires plutôt que comme une longue liste structurée de sujets. À la différence d'un simple mode hiérarchique et unidimensionnel d'organisation des documents, l'efficacité des classifications à facettes tient à leur capacité à intégrer différentes dimensions d'analyse sur les objets informationnels, à caractériser et à rendre l'accès à l'information plus facile en offrant de multiples voies de navigation vers n'importe quel document. Ainsi, la classification à facettes est considérée comme un langage post-coordonné par opposition à la

classification énumérative et hiérarchique (par exemple, la classification Dewey) qui est un langage documentaire pré-coordonné. Les concepts, représentés par des termes, sont énumérés à l'intérieur de chacune des facettes, et combinés dans une proposition quand la situation le demande (Côté-Lapointe et Mas, 2017).

Les classifications à facettes ont récemment bénéficié d'un regain d'intérêt parce qu'elles sont parfaitement adaptées à l'environnement numérique et qu'elles s'avèrent être un excellent outil d'indexation pour des corpus relativement homogènes (La Barre 2007 ; Denton 2009). Lorsque transposées sur le Web, plusieurs bases de données juridiques sont dotées de dispositifs à facettes, soit des interfaces de recherche à facettes. Soulignons toutefois que plusieurs de ces facettes ou « filtres » ne se rapportent pas au sujet du document, mais s'apparentent plutôt à des métadonnées ou à des éléments de catalogage qui peuvent être utilisés pour préciser la recherche. Ainsi, l'interface de recherche de la base « Recherche juridique » de SOQUIJ présente, dans une colonne située à gauche de l'écran, une liste de termes relevant de différentes facettes. Ces facettes sont la source (par exemple, la législation, la jurisprudence ou la doctrine), la juridiction (par exemple, la Cour d'appel ou le Tribunal des droits de la personne), la date de la décision et les domaines de droit dont relèvent les documents qui contiennent les mots-clés saisis par l'utilisateur. **L'utilisateur** peut alors sélectionner un ou plusieurs des termes associés à ces facettes, afin de préciser sa recherche.

Grâce à la facette correspondant aux domaines du droit, une interface telle que celle de SOQUIJ représente un premier pas dans la direction que nous proposons, car elle permet à l'utilisateur de repérer les catégories juridiques interpellées par les termes de sa recherche et aide ainsi la qualification juridique. Ce genre d'interface demeure toutefois rudimentaire, puisqu'elle repose sur l'utilisation de deux modes d'accès au contenu du document aux contours mal définis : l'une avec des

mots-clés (des termes présents dans l'index ou dans les textes intégraux, selon le cas) et l'autre avec des domaines de droit (chaque document ayant été indexé au regard de ceux-ci).

Dans le cadre de notre projet de recherche, nous avons souhaité déployer des facettes reliées au contenu même des décisions de justice, appelées « facettes-contenu », en laissant de côté les facettes de métadonnées, appelées « facettes-forme ». Le questionnement était alors le suivant : quelles facettes retenir pour l'indexation des décisions de justice ? Quels aspects de leur contenu s'agit-il d'indexer ? Est-ce que les facettes retenues permettent de représenter les faits à l'origine du litige et sur lesquels s'appuie la décision ? Est-ce que les facettes retenues permettent de représenter les catégories juridiques associées aux faits ?

Un élément de réponse à ces questions se trouve dans la célèbre classification proposée par le jurisconsulte Gaius au 2<sup>e</sup> siècle qui distinguait d'abord le droit public et le droit privé, puis divisait le droit privé en trois grandes parties : les « personnes », les « choses » et les « actions » (ces dernières désignaient des remèdes ou des recours au sens très large). Si la classification de Gaius suscite toujours l'admiration aujourd'hui, c'est sans doute pour la raison suivante : lorsqu'on appréhende le droit en général, il est toujours pertinent de s'interroger sur le qui (les personnes), le quoi (les choses) et le comment (les actions).

Ceci a été confirmé lors de l'analyse d'un échantillon de logs<sup>6</sup> d'utilisateurs de SOQUIJ correspondant à toutes les entrées faites dans une journée (en 2016) ou un mois (en 2017) dans la boîte de recherche de sa base de données Recherche

---

6. Le *log* est un terme communément utilisé dans le domaine informatique pour représenter et enregistrer les traces produites lors de toute connexion et d'interaction d'un utilisateur humain ou non avec un système informatique. Ces données de connexion et de transaction peuvent ensuite être utiles, par exemple, à l'étude de l'utilisation d'une application spécifique afin de mieux l'adapter aux objectifs de recherche de l'utilisateur.

juridique. Les logs d'utilisateurs font habituellement référence aux données de connexion et aux requêtes de recherche en tant que telles. Plus précisément, il s'agit d'entrées formulées dans des interfaces de recherche par des usagers en quête d'un retour d'information. Ces recherches peuvent se faire dans des moteurs de recherche Web (par exemple, dans une recherche Google) ou dans des sources propriétaires comme des bases de données privées (Sheble, Wildemuth et Brennan, 2017, p. 185).

L'analyse de logs est basée sur l'approche de la théorisation ancrée (Jansen, 2009a, p. 101). Cette dernière vise à développer des théories fondées empiriquement (Luckerhoff et Guillemette, 2012, p. 6). Ainsi, l'analyse de logs peut relever d'une approche inductive, basée sur les données et caractérisée par une observation empirique (Jansen, 2009b, s. p.). Plus précisément, l'analyse de logs peut être utilisée à la fois dans une démarche inductive, à un niveau exploratoire de la recherche, ou dans une démarche déductive basée sur la vérification d'hypothèses de recherche (Sheble, Wildemuth et Brennan, 2017, p. 186). Malgré les limites relevées dans les écrits<sup>7</sup>, l'identification et l'analyse des logs d'utilisateurs constituent des éléments essentiels pour l'évaluation de la qualité des systèmes et des interfaces de recherche Web ainsi que pour leur amélioration continue. Les logs de recherche, en particulier, sont communément collectés et analysés en sciences de l'information (Sheble, Wildemuth et Brennan, 2017, p. 185).

L'analyse exploratoire<sup>8</sup> des logs fournis par SOQUIJ a consisté à décrire chaque terme ou expression de recherche formulé par les utilisateurs en l'assimilant à une catégorie représentant son contenu. L'analyse a été réalisée par une finissante de la maîtrise en sciences de l'information ayant une

---

7. Les limites que présentent les logs d'utilisateurs sont particulièrement visibles en ce qui concerne l'identification précise de l'utilisateur seulement représenté par une adresse IP (dans le cas où plusieurs utilisateurs utilisent la même adresse IP). De plus, les informations démographiques, par exemple, sont impossibles à obtenir avec cette seule méthode (Jansen et Spink, 2000, p. 15).

8. Environ 21 000 entrées analysées sur les 245 000 entrées fournies par SOQUIJ.

formation juridique de premier cycle et visait à obtenir une première idée des habitudes de recherche des utilisateurs de la base de données de SOQUIJ.

L'analyse a permis d'identifier 19 catégories principales pour un total de 42 catégories et sous-catégories. Les catégories les plus peuplées sont, dans l'ordre décroissant : « Objet » (33,5 %) (comprend des noms communs ou groupes nominaux. Par exemple, discrimination, contrat, chute, drogue, accident de voiture), « Personne morale » (17,3 %) (agent différent d'une personne physique et doté d'une personnalité juridique. Par exemple, Fédération des médecins spécialistes du Québec, Bombardier), « Agent » (15,3 %) (agent de l'action. Par exemple, chien, employeur, ex-conjointe) et « Entité nommée » (11,3 %) (noms de famille, prénoms, ou surnoms. Par exemple, Taylor).

La présence importante des personnes morales et des entités nommées dans les logs s'explique par le fait que l'intitulé d'une décision de justice est constitué du nom des parties à l'action. Ainsi, les utilisateurs qui cherchent une décision dont ils connaissent l'existence et l'intitulé repèrent celle-ci à l'aide du nom des parties, qui peuvent être des personnes physiques ou morales. Même en laissant de côté ces termes, on observe que les concepts qui désignent une personne ou un agent sont fréquents.

La catégorie « Objet », pour sa part, est très large et comprend, dans des proportions à peu près égales, des termes propres au droit (par exemple, appel, action déclaratoire, droit acquis), des termes de la vie courante (par exemple, réunion, pare-brise, drogue) et des termes ambigus, dont la connotation peut être juridique ou non (par exemple, enquête, majorité, discrimination, pension alimentaire). Une analyse plus fine de cette catégorie révèle qu'elle comprend tantôt des choses, tantôt des actions, tantôt des éléments qui dénotent un contexte particulier ou qui caractérisent une personne, une chose ou une action.

En tenant compte de ces observations, nous avons combiné les facettes de Ranganathan avec celles de Gaius, qui sont fondamentales en droit, comme ceci :

<b>Facettes de Ranganathan</b>	<b>Facettes de Gaius</b>	<b>Facettes de notre modèle</b>
Personnalité	Personne	Personne
	Chose	Chose
Énergie	Action	Action
Matière		Contexte
Espace		
Temps		

*Tableau 1. Relations entre les facettes de Ranganathan, celles de Gaius et celles de notre modèle*

La division des personnes et des choses est fondamentale en droit, comme le reflète la classification de Gaius. C'est pourquoi nous avons scindé la facette Personnalité de Ranganathan qui regroupe ces deux catégories. Par ailleurs, les concepts relevant des facettes Matière, Espace et Temps sont moins fréquents dans les logs et ces catégories ne figurent pas dans la classification de Gaius : nous les avons donc regroupés au sein d'une seule facette, Contexte.

L'analyse des logs illustre l'impossibilité de distinguer fermement les concepts juridiques et ceux qui appartiennent à la vie courante. Le droit et la réalité sociale sont fortement imbriqués. On en vient à oublier l'origine juridique de concepts comme le brevet, la personne morale ou la pension alimentaire. De même, le droit s'approprie certains concepts de la vie courante pour leur insuffler un sens juridique précis, comme pour la discrimination et le vol. Nous voulions pourtant représenter la distinction entre « les faits » et « le droit », une distinction fondamentale dans le raisonnement des juristes (Cumyn et Gosselin, 2016, p. 331 et 337-338). Lorsque les

juristes parlent des « faits », des « faits du litige » ou encore de la « situation factuelle » à l'origine du problème juridique à résoudre (*cf.* Cumyn et Gosselin, 2016, p. 334-337), ce sont des « faits » mêlés de droit : une personne a conclu un contrat qui n'a pas été respecté, un employé a été congédié, une société poursuit des activités criminelles. Plusieurs des concepts qui permettent de les décrire reflètent donc cette ambivalence. En revanche, les catégories juridiques qui désignent les règles applicables et qui représentent « le droit » ou, pourrait-on dire, la réponse du droit, sont des concepts avec un sens juridique précis.

Pour en revenir aux facettes, nous avons réalisé, après une nouvelle phase d'expérimentation, que les concepts relevant des facettes de Ranganathan et de Gaius trouvent leur utilité première dans la description des « faits » et non des catégories juridiques mobilisées pour y répondre. C'est pourquoi nous avons réservé ces facettes pour la description des faits à l'origine du litige. Notre projet se démarque ainsi d'autres tentatives de développer une classification à facettes dans le domaine juridique où ce sont principalement des concepts juridiques qui sont organisés en facettes (Broughton, 2010 ; Scott et Smith, 2010 ; *cf.* Reiner, Cumyn, Hudon et Mas, 2019).

Quant aux catégories juridiques mobilisées dans la réponse du droit aux faits à l'origine du litige, nous les avons scindées en deux classes. La première, nommée « Régime juridique », désigne les catégories du droit substantiel, tandis que la deuxième, « Sanction », désigne les recours ou sanctions appliquées par le droit. En effet, nous avons jugé utile, pour les chercheurs d'information juridique, de distinguer ces classes qui reflètent une autre dichotomie importante pour les juristes. Mentionnons que les catégories Régime juridique et Sanction ne sont pas des facettes suivant toutes les acceptions de ce terme (voir ci-dessus), mais plutôt des classes. Elles sont toutefois destinées à figurer dans l'interface de recherche à facettes aux côtés des facettes dites factuelles.

Plus précisément, les six facettes de notre modèle sont définies comme suit :

**Personne** : personne physique, personne morale, organe ou entité auquel le droit reconnaît des droits, accorde des pouvoirs, impose des obligations ou attribue des responsabilités. La personne inclut, par extension, l'organisme, le conseil ou le comité qui ont un pouvoir décisionnel.

**Action** : acte, activité ou décision encadrés par le droit ou qui entraînent des conséquences juridiques. L'action est généralement commise par une personne, volontairement ou non. Par extension, l'action peut inclure l'événement imputable à une personne.

**Chose** : entité tangible ou intangible, concrète ou abstraite qui fait l'objet d'un encadrement ou d'une protection juridique. La chose inclut un bien matériel, une créance, une dette ou un acte juridique. Typiquement, la chose est l'objet d'un litige, l'instrument d'une action ou son résultat. Elle peut désigner l'objet d'un préjudice, comme une partie du corps ou un intérêt protégé.

**Contexte** : élément additionnel dont la présence est essentielle pour le traitement d'un problème juridique. Il s'agit d'un lieu, d'un aspect temporel, d'une circonstance ou d'un événement extérieur, d'une manière d'être de la chose et d'une cause ou conséquence de l'action.

**Régime juridique** : ensemble de règles ou de précédents, correspondant typiquement à une loi ou une partie de loi, un arrêt de principe ou une lignée jurisprudentielle.

**Sanction** : recours, réparation, peine, sanction ou autre solution découlant de l'application des règles de droit (Hudon et Cumyn, 2018).

Afin d'illustrer l'application des facettes, prenons l'exemple d'un jugement qui donne suite à un accident du travail. Le contenu du jugement peut être décrit en précisant le statut des personnes impliquées (médecin, hôpital), l'action à l'origine de l'accident (chute), la chose à l'origine du litige (ne s'applique pas) et certains éléments contextuels tels le lieu et les circonstances de l'accident (stationnement, surface glissante). Enfin, le régime juridique (accident du travail) et la sanction (indemnité de remplacement du revenu) représentent la solution juridique retenue par le tribunal. Cet exemple montre que certains concepts sont susceptibles d'appartenir à plus d'une facette. L'hôpital, qui désigne ici l'employeur, une personne, pourrait également représenter le lieu de l'accident, un élément du contexte. Les facettes jouent un rôle utile en réduisant l'ambiguïté du langage d'indexation.

## 5. Protocole expérimental

Il s'agissait ensuite de tester notre modèle d'indexation à facettes de la documentation juridique. Nous avons construit une banque de données prototype – appelée Gaius – qui comprend un échantillon représentatif de 2500 décisions en droit des obligations, en droit du travail et en droit administratif. Nous avons choisi ces domaines parce qu'on y retrouve du droit civil et de la common law, du droit d'origine législative et d'origine jurisprudentielle, du droit commun et spécialisé. Nous avons établi une procédure d'échantillonnage avec l'aide d'une statisticienne<sup>9</sup> avant de prélever les décisions dans la base de données de SOQUIJ, partenaire du projet.

9. Nous avons adopté une procédure d'échantillonnage à deux niveaux (échantillonnage en grappe). Au premier niveau, nous avons choisi de manière raisonnée trois domaines du droit et identifié les corpus correspondants dans le plan de classification de SOQUIJ. Au deuxième niveau, nous avons choisi de façon aléatoire des décisions au sein des corpus sélectionnés suivant la proportion dans lesquelles ils apparaissent dans la base de données de SOQUIJ (échantillonnage proportionnel). Nous avons toutefois réduit la proportion de documents issus de certains corpus qui comprenaient un grand nombre de décisions redondantes (échantillonnage proportionnel corrigé ou théorique). Ces corpus proviennent d'organismes spécialisés avec lesquels SOQUIJ a conclu des ententes de service pour la gestion de leur documentation,

Nous nous sommes dotés d'une politique d'indexation et toute l'équipe a reçu une formation en indexation dispensée par Michèle Hudon, professeure associée à l'École de bibliothéconomie et des sciences de l'information de l'Université de Montréal. L'indexation a été réalisée dans Gaius grâce à une interface spéciale pour les indexeurs. Celle-ci donne accès au vocabulaire contrôlé, au sein duquel l'indexeur peut sélectionner les termes d'indexation ; il peut aussi en suggérer de nouveaux.

L'analyse des trois domaines du droit couverts par Gaius a permis d'établir le vocabulaire contrôlé relevant des classes Régime juridique et Sanction. Nous avons étudié la structure et les principaux concepts décrivant des lois ou parties de loi de même que des ensembles jurisprudentiels. Nous avons consulté les principaux ouvrages de doctrine afin de choisir des divisions les plus consensuelles possibles. Pour chaque catégorie juridique, nous avons précisé les articles de loi ou les arrêts de principe qui la composent. Nous avons établi une hiérarchie à deux niveaux en privilégiant des catégories ni trop générales ni trop spécifiques (Cumyn et Gosselin, 2016).

Nous avons procédé différemment pour développer le vocabulaire contrôlé appartenant aux quatre facettes factuelles. Ce vocabulaire a été développé au fur et à mesure de l'indexation. Les termes d'indexation proposés par les indexeurs ont été approuvés et intégrés aux vocabulaires sous le contrôle d'un bibliothécaire.

Une fois l'indexation complétée, nous avons conçu l'étude d'utilisabilité qui a pour but de vérifier si l'interface à facettes qui représente notre modèle est plus performante qu'une interface de recherche usuelle. L'interface de recherche qui

---

par exemple la Commission des lésions professionnelles. Voir Cumyn et Tremblay-Potvin (2017) pour plus de précisions. Soixante-seize décisions ont été exclues de Gaius *a posteriori* par les indexeurs. Il s'agissait de jugements en provenance d'une autre province canadienne ou de jugements trop sommaires pour qu'on puisse en comprendre le contenu (par exemple, des jugements de rectification).

représente notre modèle – l'interface A – propose six facettes à l'utilisateur, comme ceci :

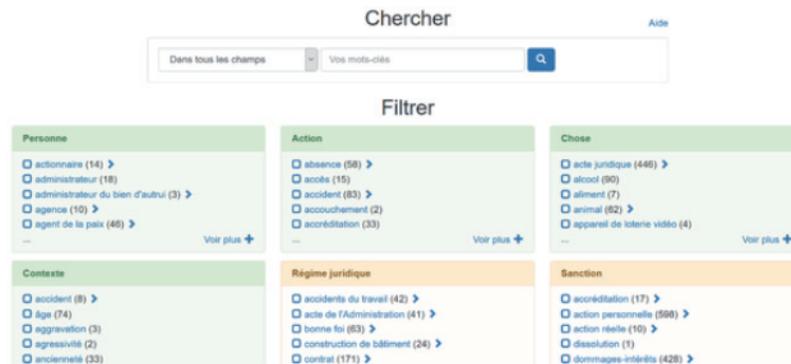


Figure 1. Écran d'accueil de l'interface A dans Gaius

En simplifiant l'interface A, nous avons créé une nouvelle interface de recherche – l'interface D – qui se rapproche de celles qui figurent dans les bases de données juridiques usuelles<sup>10</sup>. Elle comporte uniquement les facettes Régime juridique et Sanction. Dans les bases de données usuelles, on retrouve des facettes qui représentent les domaines de droit. Par ailleurs, plusieurs d'entre elles proposent une facette Quantum qui offre un aperçu des montants accordés par les tribunaux à titre de dommages-intérêts. Notre facette Sanction indique la nature du recours ou de la sanction et non les montants obtenus, mais

10. Nous n'avons pas comparé Gaius avec une base de données « réelle » telle que Recherche juridique de SOQUIJ, comme nous avons l'intention de le faire au début du projet. Nous avons plutôt choisi de créer nous-mêmes les trois interfaces à tester. Ainsi, en manipulant les variables indépendantes (par exemple, nombre ou nature des facettes) nous voulions mesurer avec plus de précision leur effet sur les variables dépendantes, soit la qualité des résultats issus de la recherche d'information (indicateurs de mesure de la performance des interfaces), tout en évitant des facteurs susceptibles de nuire à la validité intrinsèque de nos résultats (par exemple, les différences de styles d'interfaces ou de performance d'algorithmes de recherche). De plus, certaines bases de données juridiques indexent uniquement les décisions ayant un intérêt jurisprudentiel alors que toutes les décisions dans Gaius ont été indexées en appliquant rigoureusement les mêmes règles et vocabulaires. Nous ne voulions pas désavantager une base de données par rapport à une autre sur la base de la portée de l'indexation (représentation des concepts juridiques et des faits) et de son contrôle (utilisation d'un vocabulaire contrôlé).

elle devrait être familière pour les utilisateurs de la fonction Quantum.



Figure 2. Écran d'accueil de l'interface D dans Gaius

Nous avons créé une troisième interface – l'interface E – qui représente une solution mitoyenne entre les deux premières. La facette Situation y regroupe les termes qui relèvent des quatre facettes factuelles. Cette rubrique donne accès à l'indexation des faits sans mettre à profit les facettes inspirées de Ranganathan et Gaius. L'inclusion de l'interface E dans l'étude permettra de vérifier la valeur ajoutée des quatre facettes factuelles présentes dans l'interface A, indépendamment de la valeur ajoutée qu'apporterait l'indexation des faits.



Figure 3. Écran d'accueil de l'interface E dans Gaius

Les interfaces de recherche sont identiques sauf pour le nombre des facettes proposées à l'utilisateur. L'interface constitue la principale variable indépendante de l'étude, dont on cherche à mesurer les effets.

À l'instar des bases de données juridiques usuelles, les trois interfaces proposent quatre fonctions de recherche qui peuvent être combinées. Il est possible de lancer une recherche par mots-clés dans le texte intégral (recherche plein-texte). Il est également possible de lancer une recherche à l'aide des termes d'indexation seulement. Une troisième possibilité consiste à lancer la recherche par mots-clés dans tous ces champs. Enfin, on peut réaliser une recherche à même les facettes, en sélectionnant les termes d'indexation qui y sont présentés. Ainsi, l'utilisateur peut débiter sa recherche par des mots-clés, puis filtrer les résultats à l'aide des facettes. Dans l'exemple suivant, on a recherché les mots-clés « morsure chien » dans tous les champs, puis sélectionné « responsabilité civile » sous la facette Régime juridique.

The screenshot displays the search interface of the Gaius database. It is divided into three main sections: 'Filtrer' (Filter), 'Chercher' (Search), and 'Résultats' (Results).

- Filtrer (Filter):** This section contains three panels of facets:
  - Situation:** Includes 'animal (7)', 'attaque (7)', 'blessure (7)', 'empiètement (1)', and 'voisin (1)'. The 'animal (7)' facet is currently selected.
  - Régime juridique (Legal Regime):** Includes 'responsabilité civile', 'fait d'un animal (7)', and 'partage de responsabilité (1)'. The 'responsabilité civile' facet is selected.
  - Sanction:** Includes 'dommages-intérêts (7)'. This facet is currently unselected.
- Chercher (Search):** Shows the search criteria:
  - Régime juridique:** 'responsabilité civile' is selected.
  - Dans tous les champs (In all fields):** 'morsure chien' is entered in the search box.
  - A search button with a magnifying glass icon is visible.
- Résultats (Results):** Shows the search results:
  - 7 décisions trouvées (7 decisions found):** The total number of results.
  - Case Title:** 'Kumanska-Krawczyk c. Jarry'.
  - Case Details:** '2014-01-06 — Cour du Québec (C.Q.), Petites créances, Montréal — AZ-51049328'.
  - Indexation (Indexing):** A dropdown menu is shown.
  - Situation (Situation):** 'animal' is selected, with sub-facets 'animal de compagnie', 'attaque', and 'blessure' visible.
  - Régime juridique (Legal Regime):** 'responsabilité civile' is selected, with sub-facet 'fait d'un animal' visible.
  - Sanction (Sanction):** 'dommages-intérêts' is selected, with sub-facet 'dommages-intérêts compensatoires' visible.

Figure 4. Recherche réalisée avec l'interface E dans Gaius

Nous avons demandé à des auxiliaires de recherche qui n'étaient pas familiers avec nos vocabulaires d'indexation de nous proposer des recherches dans les domaines de droit couverts par Gaius. À partir de celles-ci, nous avons conçu dix tâches qui représentent trois types de recherche différents et de complexité croissante. Les tâches du premier type consistent à retrouver une décision en particulier, à partir d'un bref résumé

de celle-ci<sup>11</sup>. Les tâches du deuxième type consistent à repérer un ensemble de décisions répondant à des critères de recherche précis<sup>12</sup>. Les tâches du troisième type exigent que l'utilisateur qualifie les faits du problème ou y réponde au moyen d'une analogie<sup>13</sup>.

Notre modèle s'adresse principalement à deux profils d'utilisateurs : les juristes débutants et les chercheurs d'information juridique expérimentés. Par conséquent, nous menons l'étude d'utilisabilité auprès de deux groupes d'au moins 36 participants : des étudiants du baccalauréat en droit et des chercheurs avec plus d'une année d'expérience d'utilisation régulière des banques de données juridiques au Québec. L'étude d'utilisabilité a obtenu l'approbation des comités d'éthique de la recherche de l'Université Laval et de l'Université de Montréal. Durant l'été et l'automne 2020 de même qu'à l'hiver 2021, des appels à participation ont été lancés par courriel et par l'entremise des réseaux sociaux aux étudiants de l'Université Laval, de l'Université de Sherbrooke et de l'Université de Montréal ; ils ont aussi été relayés par des

- 
11. Voici un exemple : « Vous avez suivi une formation récemment. Un jugement mentionné pendant la formation est pertinent pour votre travail. Malheureusement, vous n'avez pas noté la référence. Dans cette affaire, le demandeur habitait un phare situé sur une réserve autochtone. Le conseil de bande a décidé d'autoriser de nouvelles constructions à proximité du phare. Le demandeur a prétendu que le conseil de bande n'avait pas respecté les principes d'équité procédurale à son endroit. Veuillez repérer la décision et la sélectionner. »
  12. Voici un exemple : « Vous préparez un article sur la vente d'entreprise. L'article s'intéresse à la vente des actifs de l'entreprise et non à la vente d'actions. Veuillez trouver toutes les décisions à ce sujet et les sélectionner. »
  13. Voici un exemple (qualification juridique) : « Mon employeur m'a congédiée et mon syndicat refuse de déposer un grief pour contester mon congédiement. Quel est le fondement de mon recours contre le syndicat, en droit québécois ? Veuillez donner la réponse et identifier une décision pertinente. » Voici un second exemple (analogie) : « Vous êtes membre du conseil d'administration d'un club de golf. La direction du club de golf souhaite imposer de nouvelles règles à son personnel, ce qui comprend l'obligation de porter une jupe, pour les femmes, et un pantalon, pour les hommes. Deux membres du personnel ont fait connaître leur opposition à cette nouvelle règle et revendiquent le droit de choisir entre la jupe ou le pantalon. Veuillez trouver une décision qui pourrait éclairer le conseil d'administration concernant l'équilibre à trouver entre les droits de l'employeur et ceux des employés. En quelques mots, expliquez pourquoi cette décision est pertinente. »

réseaux professionnels de bibliothécaires et de chercheurs en droit.

Avant le début du test, les participants lisent et signent un formulaire de consentement. Munis de l'accès Web que nous leur fournissons, ils réalisent les dix tâches de recherche dans la base de données Gaius, ce qui généralement prend moins de deux heures. Les participants complètent les tâches de recherche dans le même ordre, en essayant successivement les trois interfaces de recherche. L'ordre des interfaces varie selon le groupe auquel ont été aléatoirement associés les participants, ce qui permet de contrebalancer l'effet d'apprentissage ou l'effet de fatigue qui pourraient affecter les résultats.

Pour comparer la performance des trois interfaces de recherche, nous avons retenu les indicateurs de performance suivants (variable dépendante) :

- Tâche abandonnée ou complétée (pour toutes les tâches)
- Succès ou échec dans la réalisation de la tâche (pour toutes les tâches)
- Précision, rappel et mesure  $F^{14}$  (pour les tâches du type 2)
- Temps requis pour compléter la tâche, mesuré en secondes (pour toutes les tâches)

Les analyses seront réalisées par une statisticienne.

---

14. La **précision** désigne la proportion des décisions trouvées qui sont pertinentes. Elle mesure la capacité de l'outil de recherche à écarter les décisions non pertinentes. On divise le nombre de décisions pertinentes repérées par le participant par le nombre total des décisions qu'il a repérées. Le **rappel** désigne la proportion des décisions pertinentes présentes dans la banque de données qui sont repérées. Elle mesure la capacité de l'outil de recherche à trouver toutes les décisions pertinentes (exhaustivité de la recherche). On divise le nombre de décisions pertinentes repérées par le participant par le nombre total des décisions pertinentes présentes dans la banque de données. La **mesure F** est la moyenne harmonique de la précision et du rappel. Elle mesure la capacité de l'outil de recherche à trouver les décisions pertinentes tout en écartant celles qui ne le sont pas. On obtient cette mesure comme ceci :  $2 \times (\text{précision} \times \text{rappel}) \div (\text{précision} + \text{rappel})$ .

Les participants complètent un bref questionnaire après avoir essayé chacune des interfaces de recherche. Les réponses obtenues permettent une appréciation qualitative de l'expérience de l'utilisateur, à savoir s'il a trouvé l'interface conviviale et adéquate pour exprimer son besoin d'information.

Des outils d'enregistrement, de gestion, d'enrichissement et d'exportation des transactions des participants (ou logs d'utilisateur) ont été intégrés à Gaius. Ils nous permettent d'automatiser le prélèvement des données nécessaires aux analyses. Ils nous fournissent également des informations sur le comportement de recherche des participants au test, notamment quant à l'utilisation des fonctions de recherche et des facettes. Des analyses plus poussées seront donc possibles pour préciser ou expliquer les résultats de l'étude.

Plusieurs mesures sont prises afin d'assurer la qualité de la recherche et de limiter les biais possibles. La sélection de répondants ayant un profil sociodémographique relativement homogène et leur répartition aléatoire dans plusieurs groupes sont des mesures qui ont pour effet de neutraliser les différences préexistantes entre les individus (par exemple, leur habilité à réaliser les tâches de recherche demandées) et de rendre les groupes équivalents (Fortin et Gagnon, 2016).

Les biais liés à l'effet de fatigue (ou effet de maturation) et à l'effet d'apprentissage (ou accoutumance au test) sont évités grâce à l'alternance des interfaces testées et au nombre significatif visé de répondants (au moins 36) ayant à réaliser les mêmes tâches (Hertzum et Frøkjær, 1996). Ainsi, à mesure qu'un participant au test peu familier avec un modèle d'indexation prend de l'expérience, cette expérience est distribuée également entre les interfaces testées.

Dans notre recherche, en raison de la façon dont le test est réalisé, des mesures inexactes peuvent être collectées qui ne donnent pas une image réelle du comportement de recherche

d'un participant. Ainsi, en raison d'une interruption temporaire du test par le participant sur une base volontaire ou non (par exemple, mauvaise connexion Internet), la mesure du temps requis pour compléter la tâche peut être faussée. Afin de contrôler ce biais, les conditions de réalisation du test à distance ont été clairement énoncées et la réalisation du test a été supervisée par un même assistant de recherche, selon un protocole écrit et décrivant précisément et de manière uniforme la procédure (Fortin et Gagnon, 2016). La formulation des tâches à réaliser et le protocole expérimental ont été prétestés et ont permis à l'assistant de recherche de se familiariser avec la procédure de collecte de données.

Les résultats de l'étude d'utilisabilité en cours permettront de valider (ou non) notre modèle d'indexation en vérifiant s'il atteint les objectifs énoncés, à savoir le repérage efficace de la documentation pertinente, l'aide à la qualification et la découverte d'analogies.

## **6. Conclusion**

Notre projet de recherche est interdisciplinaire (droit et sciences de l'information), et il a principalement pour objet de contribuer à l'avancement des connaissances et à l'amélioration des pratiques dans le domaine de la recherche d'information juridique au moyen de l'indexation en vocabulaire contrôlé de la jurisprudence et de l'usage de techniques de recherche avancées. Plus précisément, nous souhaitons vérifier si l'organisation des termes d'indexation en facettes créant une multiplicité de points d'accès permet d'assister la recherche d'information d'un juriste ou d'un étudiant en droit en lui signalant, à partir des termes de sa recherche, des liens et des recoupements qui émergent de la documentation indexée. En effet, s'il est essentiel pour un juriste de repérer les décisions directement en lien avec une affaire, il est tout aussi pertinent de repérer des décisions reliées à l'affaire, mais pas directement ciblées par la requête. À ce titre, nous souhaitons vérifier au

moyen d'une étude d'utilisabilité en cours de réalisation si notre modèle d'indexation et le système de recherche qui en découle permettent de repérer les décisions directement ciblées par les termes de la requête, mais également les décisions qui seraient plus éloignées des précédents établis. En raison du caractère complet de l'indexation réalisée, qui représente non seulement les catégories juridiques, mais aussi les faits à l'origine d'un litige, nous nous attendons à des résultats statistiquement meilleurs avec notre modèle à facettes. Des différences significatives entre les trois interfaces sont attendues plus précisément dans la réalisation des tâches impliquant le raisonnement par analogie et la qualification juridique.

Notre protocole de recherche est une contribution au développement de méthodologies spécialement adaptées aux études relatives à la recherche d'information juridique. Il existe très peu d'études expérimentales sur la recherche d'information jurisprudentielle et sur les modèles d'indexation à facettes ; notre étude d'utilisabilité permettra, à l'aide d'une méthodologie éprouvée, de vérifier empiriquement l'efficacité d'un tel modèle dans la réalisation de tâches de recherche principalement exploratoires.

Sur le plan sociétal, notre projet de recherche vise à enrichir l'enseignement du droit par un meilleur arrimage entre la structure académique de la discipline et l'organisation de l'information à des fins de recherche. Par ailleurs, nous espérons que l'utilisation des facettes, représentant les faits à l'origine d'un litige, combinées à des classes relatives aux régimes juridiques et aux sanctions permettra d'améliorer les sites de vulgarisation juridique, en rendant le droit plus accessible pour l'ensemble des citoyens. Enfin, notre recherche présente un intérêt certain pour les éditeurs de banques de données juridiques qui pourraient s'appuyer sur les résultats de cette étude pour bâtir des systèmes de navigation à facettes en complément ou en remplacement des systèmes courants d'accès à l'information juridique.

En ce qui concerne les retombées théoriques, une base de données comportant 2500 décisions de justice indexées manuellement en totalité et selon un vocabulaire contrôlé est rare en elle-même, car coûteuse à réaliser. Une telle base de données est cependant précieuse, car elle formalise de façon systématique les relations sémantiques entre les concepts juridiques et les éléments factuels contenus dans un corpus juridique. La base de données Gaius demeure insuffisante pour le fonctionnement efficace des algorithmes d'apprentissage supervisé visant à automatiser ou à soutenir l'indexation en vocabulaire contrôlé de décisions de justice additionnelles. Elle pourrait toutefois être développée davantage dans ce but. Elle pourrait également soutenir le développement d'autres applications d'intelligence artificielle dans le domaine juridique.

### **Remerciements**

Conseil de recherche en sciences humaines du Canada ;  
Michèle Hudon et Günter Reiner, collaborateurs au projet. M<sup>e</sup> Danielle Blondin et M<sup>e</sup> Véronique Abad de chez SOQUIJ, partenaire du projet.

Christine Dufour, professeure à l'EBSI qui nous a aidés avec la procédure d'échantillonnage.

Auxiliaires de recherche de l'équipe Sciences de l'information :  
Camille Demers, Alexandre Fortier, Marjorie Gauchier, Béatrice Lecomte, Claire Nigay, Cécile Pechcontal, Vicky Shank et Marie-Hélène Tanguay-Bérubé.

Auxiliaires de recherche de l'équipe Droit : Christophe Achdjian, Dominique Boulanger, Oksana Burak, Sarah Jolicoeur, Ariane Joncas, Joanie Gosselin-Lepage, Justin Lee, Papa-Adama Ndour, Amélie Roy, Erin Sandberg, Antoine Tousignant, Forrest Wakarchuk et Iryna Zazulya.

**Références citées**

- ARTHURS, Harry W., « A Lot of Knowledge is a Dangerous Thing: Will the Legal Profession Survive the Knowledge Explosion? », (1995) 18 *Dalhousie Law Journal* 295.
- ATIAS, Christian, *Épistémologie juridique*, coll. « Droit fondamental », Paris, Presses universitaires de France, 1985.
- BATES, Marcia J., « The design of browsing and berrypicking techniques for the online search interface », (1989) 13(5) *Online Review* 407.
- BROUGHTON, Vanda, « The use and construction of thesauri for legal documentation », (2010) 10(1) *Legal information management* 35.
- CAIJ (2021a), *La promesse du CAIJ*, en ligne : <<https://www.caij.qc.ca/la-promesse-du-caij>> (consulté le 25 janvier 2021).
- CAIJ (2021b), *Jurisprudence : un corpus québécois et canadien impressionnant*, en ligne : <<https://www.caij.qc.ca/jurisprudence>> (consulté le 25 janvier 2021).
- COHEN, Morris L., « Researching Legal History in the Digital Age Symposium: Legal Information and the Development of American Law: Further Thinking about the Thoughts of Robert C. Berring », (2007) 99(2) *Law Library Journal* 377.
- CÔTÉ-LAPOINTE, Simon et Sabine MAS, « Facettes et archives : modalités et applications » (2017) 248 *La Gazette des archives* 19.
- CUMYN, Michelle et Frédéric GOSSELIN, « Les catégories juridiques et la qualification : une approche cognitive », (2016) 62(2) *Revue de droit de McGill* 329.

CUMYN, Michelle, Michèle HUDON, Sabine MAS et Günter REINER, « Towards a New Approach to Legal Indexing Using Facets », dans Malek MOUHOUB *et al.* (dir.), *Recent Trends and Future Technology in Applied Intelligence*, 31<sup>st</sup> International Conference on Industrial Engineering and Other Applications of Applied Intelligent Systems, Montréal, IEA/AIE, 2018, p. 881.

CUMYN, Michelle et Charles TREMBLAY-POTVIN, *Procédure d'échantillonnage*, 6 avril 2017 (document inédit).

DELGADO, Richard et Jean STEFANCIC, « Why do we ask the same questions? The triple helix dilemma revisited », (2007) 99(2) *Law Library Journal* 307.

DENTON, William, « How to Make a Faceted Classification and Put it On the Web », (2009), en ligne : <<https://www.miskatonic.org/library/facet-web-howto.html>> (consulté le 2 février 2021).

DIEPEVEEN, Caroline, « Continental European indexing: then and now », (2006) 25(2) *The Indexer: The International Journal of Indexing* 74.

FORTIN, Marie-Fabienne et Johanne GAGNON, *Fondements et étapes du processus de recherche : Méthodes quantitatives et qualitatives*, 3<sup>e</sup> éd., Montréal, Chenelière-éducation, 2016.

GÉMAR, Jean-Claude, « La traduction juridique et son enseignement : aspects théoriques et pratiques », (1979) 24(1) *Meta : Journal des traducteurs / Translators' Journal* 35.

GÉMAR, Jean-Claude, « La langue juridique, langue de spécialité au Québec : éléments de méthodologie », (1980) 53(6) *The French Review* 880.

- GÉMAR, Jean-Claude, « Réflexions sur le langage du droit : problèmes de langue et de style », (1981) 26(4) *Meta : Journal des traducteurs / Translators' Journal* 338.
- GÉMAR, Jean-Claude, « Terminologie, langue et discours juridiques. Sens et signification du langage du droit », (1991) 36(1) *Meta : Journal des traducteurs / Translators' Journal* 275.
- GÉMAR, Jean-Claude, « Traduire le droit ou le double langage de Thémis », (2007) 3(49) *Hermès, La Revue* 149.
- GROSS, Tina, Arlene G. TAYLOR et Daniel N. JOUDREY, « Still a Lot to Lose: The Role of Controlled Vocabulary in Keyword Searching », (2015) 53(1) *Cataloging & Classification Quarterly* 1.
- GUIARD, Laure, *Vocabulaire employé pour l'accès thématique aux documents d'archives patrimoniaux : étude linguistique exploratoire de termes de recherche, de description, d'indexation*, Montréal, Thèse (Ph.D.), École de bibliothéconomie et des sciences de l'information, Faculté des arts et des sciences, Université de Montréal, 2018, en ligne : <<https://papyrus.bib.umontreal.ca/xmlui/handle/1866/21122>> (consulté le 25 janvier 2021).
- HERTZUM, Morten et Eric FRØKJÆR, « Browsing and Querying in Online Documentation: A Study of User Interfaces and the Interaction Process », (1996) 3(2) *ACM Transactions on Computer-Human Interaction* 136.
- HJØRLAND, Birger, « Facet analysis: The logical approach to knowledge organization », (2013) 49(2) *Information Processing & Management* 545.
- HUDON, Michèle, *Analyse et représentation documentaires : une introduction à l'indexation, à la classification et*

*à la condensation des documents*, Québec, Presses de l'Université du Québec, 2013.

HUDON, Michèle et Widad Mustafa EL HADI, « Introduction. La classification à facettes revisitée. De la théorie à la pratique », (2017) 13(1) *Les Cahiers du numérique* 9.

HUDON, Michèle et Michelle CUMYN, *Politique et règles d'indexation*, 11 mai 2018 (document inédit).

JANSEN, Bernard J. (2009a), « The Methodology of Search Log Analysis », dans Bernard J. JANSEN, Amanda SPINK et Isak TAKSA (dir.), *Handbook of Research on Web Log Analysis*, Hershey, IGI Publishing, 2009.

JANSEN, Bernard J. (2009b), « What is Web Log Analysis? », Présentation Slideshare dans le cadre du séminaire *The Query Log Analysis: From Research to Best Practice*, Londres, 27 au 28 mai 2009, en ligne : <<https://www.slideshare.net/dj9395/what-is-log-analyis>> (consulté le 25 janvier 2021).

JANSEN, Bernard J., Amanda SPINK et Major Anthony PFAFF, « Linguistic Aspects of Web Queries », dans AMERICAN SOCIETY FOR INFORMATION SCIENCE, *Knowledge innovations, celebrating our heritage, designing our future*, Proceedings of the 63rd ASIS Annual Meeting, vol. 37, Chicago, 13 au 16 novembre 2000, p. 169, en ligne : <<https://faculty.ist.psu.edu/jjansen/academic/pubs/asis2000/asis2000.html>> (consulté le 25 janvier 2021).

KNAPP, Charles, « Indexing Court Cases », dans Peter KENDRICK et Enid L. ZAFRAN (dir.), *Indexing Specialties: Law*, Medford NJ, Information Today, 2001, p. 39.

KUHLTHAU, Carol Collier, « Information search process (ISP) model », dans Marcia J. BATES et Mary Niles MAACK

(dir.), *Encyclopedia of library and information sciences*, Boca Raton, CRC Press, 2010, p. 2586.

LA BARRE, Kathryn, « The Heritage of early FC in Document Reference Retrieval Systems, 1920-1969 », (2007) 23(2) *Library History* 129.

LAPRISE, Gisèle, *Les outils du raisonnement et de la rédaction juridiques*, Montréal, Éditions Thémis, 2000.

LE MAY, Denis et Martin THIBOUTOT, *Documentation juridique : recherche, rédaction et références*, Montréal, Wilson & Lafleur, 2014.

LESIEUR, David, *L'indexation automatique des décisions de justice*, Montréal, Rapport de recherche, École de bibliothéconomie et des sciences de l'information, Faculté des arts et des sciences, Université de Montréal, mai 2020 (document inédit).

LUCKERHOFF, Jason et François GUILLEMETTE, « Méthodologie générale de la théorisation enracinée : un projet épistémologique », dans Jason LUCKERHOFF et François GUILLEMETTE (dir.), *Méthodologie de la théorisation enracinée : Fondements, procédures et usages*, Québec, Presses de l'Université du Québec, 2012, p. 1.

MANIEZ, Jacques et Dominique MANIEZ, *Concevoir l'index d'un livre : histoire, actualité, perspectives*, coll. « Sciences et techniques de l'information », Paris, ADBS Éditions, 2009.

MAXWELL, Tamsin et Burkhard SCHAFFER, « Natural language processing and query expansion in legal information retrieval: Challenges and a response », (2010) 24(1) *International Review of Law, Computers & Technology* 63.

MILSTEAD, Jessica L., « Needs for Research in Indexing », (1994) 45(8) *Journal of the American Society for Information Science* 577.

PAROUBEK, Patrick, Pierre ZWEIGENBAUM, Dominic FOREST et Cyril GROUIN, « Indexation libre et contrôlée », dans les *Actes de la 8<sup>e</sup> édition du défi fouille de textes*, Grenoble, 4 au 8 juin 2012, p. 1.

PERUGINELLI, Ginevra, « Concepts Mapping in the Legal Domain: Some Reflections », (2011) 11(4) *Legal Information Management* 268.

POULIQUEN, Bruno, Ralf STEINBERGER et Camelia IGNAT, « Automatic linking of similar texts across languages », dans Nicolas NICOLOV *et al.* (dir.), *Recent Advances in Natural Language Processing III*, Amsterdam, John Benjamins Publishing Company, 2004, p. 307.

REINER, Günter, Michelle CUMYN, Michèle HUDON et Sabine MAS, « Designing a database to assist legal thinking: a new approach to indexing using facets », dans Erich SCHWEIGHOFER, Franz KUMMER et Ahti SAARENPÄÄ (dir.), *Internet of Things*, Proceedings of the 22nd International Legal Informatics Symposium, IRIS, 2019, p. 345.

SCHMEER, David et Cynthia SIDLO, « Automatic topical indexing at Lexis-Nexis », dans Peter KENDRICK et Enid L. ZAFRAN (dir.), *Indexing Specialties: Law*, Medford NJ, Information Today, 2001, p. 85.

SCOTT, Mark et Nigel SMITH, « Legal Taxonomy from Sweet & Maxwell », (2010) 10(3) *Legal Information Management* 217.

SHEBLE, Laura, Barbara M. WILDEMUTH et Kathy BRENNAN, « Transaction logs », dans Barbara M.

WILDEMUTH (dir.), *Applications of social research methods to questions in information and library science*, 2<sup>e</sup> éd., Santa Barbara, Libraries Unlimited, 2017.

SPITERI, Louise, « A simplified model for facet analysis: Ranganathan 101 », (1998) 23 *Canadian Journal of Information and Library Science* 1.

STRANIERI, Andrew et John ZELEZNIKOW, « Information retrieval and text mining », dans Andrew STRANIERI et John ZELEZNIKOW (dir.), *Knowledge Discovery from Legal Databases*, Dordrecht, Springer, 2005, p. 147.

