

Travail collaboratif entre informaticiens et juristes : l'exemple de la participation du laboratoire IREENAT au projet "ASPHALES, ACI - Sécurité informatique"

Résumé :

Cet article est le premier d'une série de trois faisant état de l'expérience menée par des juristes et des informaticiens sur une lecture conjointe des textes juridiques relatifs à la sécurité informatique en matière de technologies de l'information et de la communication¹. Le premier article porte sur la méthode utilisée, le second sur les constats opérés, le troisième sur les propositions qui peuvent être faites en matière de travail collaboratif portant sur ces matières.

Cette livraison met particulièrement en évidence l'analyse du processus de recherche d'une méthode et d'un outil. La méthode s'applique au choix des textes, à leur suivi, ainsi qu'à leurs commentaires. Elle privilégie particulièrement le choix d'une approche interdisciplinaire et synchrone.

Introduction

Le droit, la médecine et la recherche informatique se rencontrent à un carrefour qui est celui de l'utilisation des nouvelles technologies. Qu'il s'agisse de la mise en place d'une administration électronique, de la circulation des données de santé, de l'imagerie médicale, de l'interopérabilité des fichiers, la sécurité y apparaît comme un enjeu fondamental de la protection des données elles-mêmes, mais encore des droits des personnes, de leurs libertés et du bon fonctionnement du service public. Les questions que se posent les uns et les autres en ce domaine, professionnels de la santé, du droit, de l'informatique sont souvent par la force des choses, l'expression de cultures différentes et se traduisent par des formes d'incompréhension. Il en découle une difficulté de l'adaptation des normes et procédures juridiques à l'évolution de la société et des technologies. Le pari de l'ACI "Asphales" est aussi d'essayer de franchir cette barrière des cultures et finalement d'apprendre à "se parler" puis de tenter de définir les éléments d'un langage compréhensible par l'autre.

L'objectif qui consiste, dans le cadre de l'ACI à tenter de définir un langage commun se situe à plusieurs niveaux. Celui de la formulation des problématiques d'une part ; celui, d'autre part, de la détermination du contenu et de l'adaptation des normes juridiques dans le cadre spécifique de l'étude du lien entre sécurité et informatique. L'une des questions que l'on pourra se poser notamment sera celle de savoir si l'on peut définir pour un concept, utilisé à la fois en droit et en informatique, un dénominateur commun qui soit opératoire et permette au législateur de formuler des normes adaptées. La démarche

¹ Ils sont nés de l'expérience menée par le laboratoire IREENAT de l'Université de Lille 2 dans le cadre du projet de l'ACI "Asphales – Sécurité informatique" initié par Isabelle de Lamberterie, directrice de recherche au CECOJI, unité CNRS. Lien: <http://www.asphales.cnrs.fr/eq-ireenat.html>

a consisté en une lecture et une confrontation des compréhensions des textes juridiques (lois, règlements, directives) relatifs à la sécurité informatique menées par une série de laboratoires².

Parmi les différents acteurs de l'opération, l'IREENAT présentait l'originalité de rassembler dans une même équipe non seulement des juristes et des informaticiens, mais aussi des individualités ayant la double compétence³. Chaque membre prenait part au travail en amenant à la fois, évidemment, ses connaissances approfondies dans son domaine, mais également ses intuitions et interrogations sur la matière en laquelle il est profane. Pour un spécialiste de la cryptographie par exemple, cela a pu se traduire par un sentiment d'abstrusion lors de la première remise des textes juridiques concernant sa matière. Dans la mesure où il n'avait eu l'occasion de les consulter qu'en fonction de besoins très ciblés, celui-ci ne mesurait pas l'ampleur de la réglementation couvrant sa discipline.

Cette démarche commune et synchrone nous a guidés dans la lecture et dans l'effort de compréhension des documents sélectionnés. La détermination des instruments de recherche et de la méthode sont ainsi le résultat d'un processus qui s'est affiné au fil des réunions de travail.

1. La détermination des instruments de recherche

1.1. Les concepts opératoires

1.1.1. Choix des textes

Le corpus des textes à étudier a été compilé par deux doctorants de l'IREENAT. Il s'agit de textes à valeur juridique concernant la sécurité informatique et le droit public.

La constitution de ce corpus s'est heurtée à un écueil principal, celui de l'exhaustivité. Comment, en effet, trouver les textes les plus pertinents sans déborder le sujet ? Afin de contourner cet obstacle, les deux doctorants ont travaillé en trois temps.

Tout d'abord, ils ont constitué un noyau incontournable de la matière. En associant des indices de fond (utilisation du terme "informatique", "sécurité", "réseaux",

² Liste des laboratoires : <http://www.asphales.cnrs.fr/equipes.html>

³Jean-Jacques LAVENUE, Professeur à l'Université Lille 2 (Droit public)

Vincel HOANG NGOC MINH, Professeur à l'Université Lille 2 (Mathématiques, informatique)

Hassan BEZZAZI, Maître de conférence à l'Université Lille 2 (Informatique)

César POVEDA, Maître de Conférence à l'Université Lille 2 (Droit public, informatique)

Jean-Yves ENJALBERT, Docteur, ATER (Mathématiques, cryptologie)

Grégory BEAUVAIS, Doctorant, ATER (Droit public)

Gaylord BAUDEN-HAMEREL, Doctorant, Allocataire de l'ACI (Droit public, informatique)

"automatisation" et "traitement automatique", "fichiers" entre autres) à un critère formel tiré de la hiérarchie des sources, y compris les textes internationaux et communautaires tels qu'ils trouvent leur effectivité dans l'ordre juridique interne. Un certain nombre de ces textes, en particulier, les textes communautaires, ont été écartés dans la mesure où ils faisaient l'objet d'une étude par les autres groupes.

La question de savoir si tel ou tel texte méritait d'être dans ce noyau a été résolue de manière empirique. C'est à partir des connaissances théoriques acquises à l'occasion de recherches antérieures dans le cadre du laboratoire qu'a été opérée une sélection préalable. L'analyse des visas et des dispositifs de ces textes ont permis de mettre en évidence un "réseau" de normes que nous avons soumis à une grille de pertinence. Cette grille tenait compte de la répartition établie au début de l'ACI entre les différentes équipes en fonction des spécialités de chaque laboratoire, en l'espèce, approche de droit public interne (ordre public, télécommunications, sécurité, libertés). Par ailleurs, la matière étant transversale, ont été retenus postérieurement, au fil des séances, des textes qui ne sont pas spécifiquement des textes de droit public, mais qui ont une incidence sur l'objet de la recherche (droit pénal, droit bancaire). La lecture de ce noyau a entraîné la recherche des textes d'application (décrets, arrêtés, circulaires, directives, instructions, notes de service). Il s'agit du deuxième cercle du corpus.

On signalera que, par la suite, l'analyse des textes nous a amenés à découvrir l'existence de textes très divers relevant de procédures de "mise en oeuvre" des mesures pratiques relevant de l'expertise à un moment donné. Ce qui par rapport aux approches classiques, pose le problème de l'origine de la norme. Origine non plus publique, mais relevant éventuellement de l'expertise et du secteur privé. Exemple : lien entre la COFRAC et la DCSSI dans les processus de certification et d'évaluation.

Le corpus a été complété par un dernier champ qui comprend les textes de référence mis en ligne sur le site de la Direction centrale de la sécurité des systèmes d'information.

Il faut noter que ce corpus n'a pas été figé dans le temps et qu'il a évolué au fil des séances de travail.

1.1.2. Appropriation des concepts

Le travail collaboratif entre des chercheurs de disciplines différentes nécessite une explication réciproque des concepts utilisés.

Il est ainsi apparu incontournable pour les juristes d'expliquer la hiérarchie des normes aux informaticiens. En effet, le matériau de base étant constitué de textes juridiques classés selon la hiérarchie des normes, la compréhension par tous de cette articulation juridique était indispensable au bon déroulement de la démarche.

De même, les commentaires des informaticiens portant sur des notions précises, ont remis en cause le discours des juristes. A titre d'exemple, on retiendra les concepts

d'effacement et de fichiers, termes utilisés très souvent dans la législation, sans définition technique précise. L' "effacement" entendu par un juriste ne correspond pas à la réalité informatique. Les données "effacées" peuvent être en effet techniquement récupérées dans certaines conditions, selon l'état logique ou physique du support: ces concepts sont inopérants dans la réflexion du juriste⁴. Un autre problème évoqué fût celui de l'archivage et du chaînage des fichiers⁵.

Progressivement cette démarche heuristique a été intégrée par les membres de l'équipe. Rapidement, les séances de travail ont fait l'objet d'explications réciproques sur les concepts maniés par les uns et les autres. Cet exercice était, non seulement indispensable pour mener à bien les dialogues entre les chercheurs, mais également très enrichissant pour chacune des deux disciplines.

1.2. Les Outils

1.2.1. Environnement de travail - Matériel informatique, Logiciels

Des flux RSS⁶ spécialisés ont été utilisés pour la mise à jour en temps réel des textes et des connaissances. Afin de centraliser les informations de veille (à la fois juridique et technologique) pour toute l'équipe, un script en PHP⁷ d'agrégation des flux RSS (liliana) a été mis en place.

Une des fonctionnalités additionnelle de ce script est l'intégration d'une fonction de recherche via Google, permettant de rapprocher chaque information de ses sources et des liaisons pertinentes renvoyées par le moteur de recherches.

4 CF les 5 niveaux de capacité hostiles définis dans :

http://www.ssi.gouv.fr/fr/documentation/Guide_effaceur_V1.12du040517.pdf

[M1] Accès logique aux données en se servant de privilèges "utilisateur" et des commandes du système d'exploitation courant.

[M2] Accès logique aux données en se servant de privilèges "administrateur" et des commandes du système d'exploitation courant, ou d'un système d'exploitation connu.

[M3] Accès logique aux données en se servant d'un système d'exploitation spécial qui utilise des commandes "bas niveau" spécifiques au disque dur attaqué.

[M4] Accès physique aux signaux analogiques par les têtes, et reconstitution des données.

[M5] Accès physique aux traces laissées sur les plateaux, et reconstitution des données.

5 Un site détruit peut rester accessible via l'archivage de moteurs de recherche comme google qui le gardent en "cache", ou de projets du type Internet Archive <http://www.archive.org/>

6 C'est un moyen simple de recevoir des informations régulièrement mises à jour,

Voir la définition : [http://fr.wikipedia.org/wiki/RSS_\(informatique\)](http://fr.wikipedia.org/wiki/RSS_(informatique))

7 Langage de programmation destiné à créer des applications "en ligne", utilisable avec un navigateur Internet. Voir : <http://fr.php.net>

1.2.2. SPIP

SPIP⁸ est un logiciel libre particulièrement adapté pour l'alimentation d'une base documentaire dans un contexte de travail collaboratif. Plusieurs facteurs ont milité pour le choix de ce logiciel comme outil de travail, implantant à la fois un intranet et un système d'annotations colorées. La réputation de ce logiciel dans le milieu de l'édition électronique (où plusieurs collaborateurs interviennent dans la mise à jour d'un magazine électronique, comme Le Monde Diplomatique⁹, appelé aussi webzine), et son utilisation au sein d'un certain nombre de portails gouvernementaux (dont le site du Cabinet du Premier Ministre¹⁰), ont motivé le choix de cette plateforme. En outre, d'autres raisons ont conforté ce choix :

- SPIP est un logiciel libre pouvant être librement modifié en fonction de nos besoins.
- Le logiciel s'exécute dans un environnement reconnaissant le langage PHP et un service de bases de données MySQL¹¹. Ces deux services sont fournis par le serveur de notre faculté servant à l'hébergement de notre plateforme de travail.
- La connaissance de cet environnement par certains membres de l'équipe.
- La protection des accès bâtie autour de trois niveaux d'utilisateurs : administrateur, administrateur restreint et rédacteur.
- La possibilité d'associer des mots clés aux articles (textes commentés et commentaires) permettant de sélectionner des articles sur la base d'un concept.

Nous avons aussi procédé lors de cette prospection à l'évaluation d'un autre système récent similaire : LODEL¹². LODEL est un logiciel d'édition électronique basé sur les fondamentaux de l'édition papier traditionnelle. Il est donc adapté à la publication d'articles scientifiques en ligne par la puissance de son moteur de conversion des documents conçus sous un traitement de texte vers la publication électronique sans autre intermédiaire technique que la plateforme LODEL (indexation automatique des mots clés et des auteurs). Cependant, LODEL fige le travail, en ce sens qu'il n'est pas possible de lier les informations au sein même du site, ou de produire des modifications directement à partir de l'interface de gestion, sans faire appel à nouveau au traitement de texte. Cette évaluation ne nous a pas permis de déceler des qualités pouvant privilégier, au regard de nos exigences, ce système naissant par rapport à SPIP.

2. La méthode

La méthode de travail que notre équipe a choisie pour commenter les textes est une méthode principalement synchrone dans la mesure où tous les participants parcourent en même temps le même passage en suivant l'animateur.

⁸ Système de Publication pour l'Internet <http://www.spip.net>

⁹ <http://www.monde-diplomatique.fr/>

¹⁰ <http://www.premier-ministre.gouv.fr/>

¹¹ MySQL est un serveur de bases de données relationnelles. Voir: <http://www.mysql.com>

¹² Logiciel d'édition électronique <http://www.lodel.org>

Accessoirement une méthode asynchrone a été utilisée pour assurer, au delà des réunions communes, une actualisation des débats.

2.1. Interdisciplinarité et méthode synchrone

2.1.1. L'interdisciplinarité : droit et informatique

2.1.1.1. A la recherche de l'interdisciplinarité

Le terme "interdisciplinarité" est souvent évoqué pour faire état d'une ouverture méthodologique aux grilles d'analyse, d'interprétation, de systèmes différents. Elle tend à multiplier les regards sur un objet de recherche commun. Ainsi que le souligne Dominique Vinck dans son ouvrage "Pratiques de l'interdisciplinarité"¹³, il doit être distingué entre différents concepts dont on peut donner une présentation hiérarchisée :

- La pluridisciplinarité qui juxtapose des points de vue et des contributions relevant de disciplines distinctes.
- L'interdisciplinarité proprement dite qui suppose un dialogue, un échange ou une confrontation entre plusieurs discipline.
- La transdisciplinarité qui implique la poursuite d'un projet : problématique transversale ou production d'un nouveau savoir autonome qui dépasse les disciplines existantes (Sociologie historique de Norbert Elias).

2.1.1.2. La démarche adoptée

Le point de départ du travail de l'équipe s'est situé dans une optique de pluridisciplinarité en réunissant des juristes et des informaticiens de différentes spécialités (mathématiques, cryptologie, informatique, droit public et droit privé). Le mode d'échange entre les différents participants a évolué rapidement vers une méthode interdisciplinaire à partir notamment de l'interrogation commune sur des concepts homographes¹⁴ des différentes disciplines (eg. signature). Le processus dynamique de la méthode suivie, la mise en évidence de possibles synergies, peuvent s'inscrire dans une trajectoire de transdisciplinarité (définition d'une méthode possible de mise au point de processus normatifs, prise en compte de situations non appréhendées par le législateur, concepts nouveaux, ontologie¹⁵ qui définit des concepts nouveaux).

13 PUG, Grenoble, 2000.

14 <http://dictionnaire.mediadico.com/traduction/dictionnaire.asp/definition/homographe/2006>

homographe (adjectif et nom masculin) : se dit des termes homonymes (même prononciation) qui s'écrivent aussi de la même façon. • Ex.: car (conjonction) et car (véhicule de transport en commun).

Voir aussi: <http://www.med.univ-rennes1.fr/iidris/cache/fr/28/2825>

Libellé homographe (n.m.) Définition Graphème pouvant représenter, suivant les contextes orthographiques, le même phonème. Source BÉRUBÉ, Louise. Terminologie de neuropsychologie et de neurologie du comportement, Montréal, Les Éditions de la Chenelière Inc., 1991, 176 p., p. 65.

15 Ontologie: <http://websemantique.org/Ontologie>: Guarino [guarino95a], [guarino97a] part de sept interprétations possibles pour chercher à clarifier ce qu'est une ontologie. Outre le sens philosophique originel, une ontologie désigne en effet une modélisation conceptuelle, ou une représentation de cette modélisation. Dans les deux cas, on parle d'ontologie formelle pour désigner aussi bien la rigueur de la

2.1.2. Méthodes synchrone et asynchrone

La méthode de travail synchrone¹⁶ s'oppose à la méthode asynchrone¹⁷ où les participants choisissent individuellement les passages à commenter pour les confronter et les synthétiser dans un second temps.

2.1.2.1. Les avantages de la méthode synchrone

L'avantage de notre méthode est la dynamique qu'elle génère, de fait, par la présence collective des participants. Le risque étant cependant que les autonomies individuelles se neutralisent, et qu'une pensée unique s'installe.

Nous avons su être vigilants sur ce point :

- En ayant une attitude constructive envers toutes les propositions de commentaires mêmes celles qui pouvaient sembler hors sujet ou contradictoires,
- En passant suffisamment de temps sur un passage pour voir s'il suscitait, ou non, un commentaire,
- En laissant le soin de la rédaction du commentaire au participant qui en était à l'origine,
- En signant collectivement les commentaires.

La nature pluridisciplinaire du thème de recherche a induit la nécessité de réunions rassemblant simultanément des personnes issues des deux communautés. En effet, au-delà des commentaires mis en place, les échanges transversaux apportent des connaissances dans les divers champs de compétences qui ont pu conduire à une remise en cause de leurs croyances et de quelques idées fausses. On peut évoquer à titre d'exemple la question posée par les informaticiens sur le lien, considéré comme automatique, entre responsabilité et intention, à l'occasion de l'analyse d'un texte qui évoquait la responsabilité de la personne détenant un contenu illicite à son insu à la suite d'un téléchargement.

2.1.2.2. Méthode synchrone et brainstorming

Cette méthode de travail, dont les sessions s'apparentent à bien des égards à des séances de brainstorming, ne convient qu'à un nombre restreint de participants effectifs

modélisation que la structure de sa représentation. Cela nous amène à la notion de convention ontologique, qui est une modélisation conceptuelle du sens du formalisme de représentation. Enfin, on distingue différents types d'ontologies selon le domaine modélisé. À la différence d'un Vocabulaire, une ontologie cherche à représenter le sens des concepts et des relations qui les lient.

16 **Synchrone** : du grec "syn": "avec" (marque l'idée de réunion dans le temps et l'espace) et "chronos" : "temps"

Modalités d'échange d'informations en direct (exemple: téléphone, visioconférence, visiophonie, audiophonie, etc.).

17 **Asynchrone** : *a privatif*, Modalités d'échange d'informations en différé (mél, forum, etc.)

(huit pour notre équipe). L'IREENAT n'étant qu'une des six équipes du projet Asphales, la méthode asynchrone a du être choisie dans la mesure où le nombre de participants au projet avoisinait la cinquantaine de chercheurs géographiquement distants. Quelque soit la méthode choisie par les intervenants au projet, chaque équipe est caractérisée au final par les textes juridiques qu'elle a choisis, les passages retenus pour être commentés et le contenu de ces commentaires.

Ainsi, on peut se trouver dans la situation où un même texte est choisi par plusieurs équipes, et où le même passage d'un texte commenté par plusieurs équipes. Une méthode de synthèse doit être élaborée pour fusionner l'ensemble des commentaires. Les bases mathématiques pour une telle méthode est décrite dans "Pooling commentaries on legal texts from lawyers and experts in a collaborative work"¹⁸.

La banque de connaissances constituée des commentaires doit pouvoir être interrogée d'une façon pertinente. Pour cela les participants sont invités à élaborer en commun et de façon progressive une liste de mots clés représentant des concepts. Un commentaire se verra attribuer un ensemble de mots clés choisis dans cette liste. La recherche dans les commentaires (et les passages associés) utilisera une mesure de similarité (mesure de Jaccard¹⁹ ou mesure du cosinus²⁰) pour comparer les mots clés des commentaires à ceux de la requête dans sa recherche. La mesure de similarité²¹ servira aussi au regroupement des commentaires en fonction de leur ressemblance : un commentaire ressemblera plus à un commentaire de son groupe qu'à un document qui se trouve dans un autre groupe.

2.1.2.3. SPIP, outil au service de la méthode asynchrone

Chaque passage projeté peut interpeller un ou plusieurs participants et faire l'objet d'un commentaire.

Doc. 1. Document de travail sous SPIP.

Les idées échangées à ce propos peuvent aboutir à un commentaire synthétique qui sera lié au passage en mettant en exergue, par surlignage, sa nature "informatique", "juridique" ou "objective" (qui en elle même posait question, sans avoir nécessairement un lien direct avec le droit ou l'informatique).

Le commentaire informatique (surlignage jaune) est émis en principe par des informaticiens, mais il arrive aussi que des juristes aux compétences informatiques confirmées émettent un tel type de commentaire. Le commentaire juridique (surlignage vert) est émis par un juriste. Le commentaire objectif (surlignage violet) est émis par un juriste ou un informaticien et concerne plus particulièrement la formulation des normes qui peuvent être, dans certains cas, ambiguës ou imprécise.

18 <http://portal.acm.org/citation.cfm?doid=1088657>

19 <http://www.limsi.fr/Individu/rosset/similarite2/node6.html#SECTION00062100000000000000>

20 <http://www.limsi.fr/Individu/rosset/similarite2/node6.html#SECTION00062200000000000000>

21 <http://www.limsi.fr/Individu/rosset/similarite2/node6.html>

Doc 2. Commentaire relatif à la partie "défense de ces droits".

Comparée à l'utilisation des notes de bas de page, cette méthode, utilisant le surlignage et les liens, permet de préserver l'authenticité du contenu des textes étudiés. Elle est particulièrement appropriée dans le cas de commentaires longs. Un passage est marqué par une couleur pour faire l'objet d'un premier commentaire qui est le commentaire principal et qui pourrait susciter une chaîne de commentaires qui lui seront rattachés.

Nous avons adapté SPIP en y ajoutant des fonctions permettant d'identifier plus facilement les commentaires, et de les lier automatiquement.

A la racine du site se trouvent six rubriques correspondant aux six équipes d'ASPHALES. Dans chacune de ces rubriques se trouve l'ensemble des textes étudiés par l'équipe correspondante et dans la sous rubrique commentaires, les commentaires relatifs à ces textes. Chaque rubrique est exclusivement gérée par son équipe, même si par la suite, les commentaires des différentes équipes auront vocation à être reliés entre eux sur la base de mots clé, textes de référence communs, chronologie, etc.

Cette approche a permis de mettre en évidence la difficulté pour le législateur, dépendant souvent de l'approche et des référentiels donnés par les experts, de formuler des normes dans un langage clair et accessible, et ce faisant de se conformer à l'exigence d'accessibilité au droit, élevée par le Conseil Constitutionnel au rang de principe à valeur constitutionnelle²². L'éparpillement des textes modificatifs, l'absence de production systématique de textes consolidés, le saupoudrage et l'aspect de "patchwork" des réformes, accentuent encore la difficulté de compréhension des textes étudiés dans leur chronologie.

A cet effet, on observera que le "Guide pour l'élaboration des textes législatifs et réglementaires"²³ présente d'une façon détaillée les règles et les méthodes à suivre lors de la préparation et la rédaction des textes normatifs. Ce guide, à vocation de vadémécum destiné aux parlementaires et au gouvernement dans leur travail de rédaction des normes et des textes d'application, fournit peu d'indications sur l'utilisation des terminologies techniques. Le "modèle" présenté dans le guide²⁴ sur la thématique "Créer un traitement automatisé", par exemple, n'apporte aucun élément permettant au législateur de comprendre le sens technique des termes employés pouvant avoir des conséquences juridiques différentes selon la définition retenue. Ainsi, la notion de suppression de données en informatique peut, dans les faits, correspondre à plusieurs sens (suppression logique, physique partielle ou totale). On peut s'interroger alors sur la possibilité réelle d'utilisation pertinente par le législateur des termes techniques employés dans le texte de la loi. Il peut apparaître nécessaire en ce domaine de réfléchir à la mise au point d'une ontologie de la sécurité informatique afin de disposer d'une terminologie contrôlée.

22 Conseil Constitutionnel – Décision n° 99-421 DC du 16 décembre 1999

23 Consultable à l'adresse http://www.legifrance.gouv.fr/html/Guide_legistique/guide_leg.pdf

24 CF 5.2.9 du "Guide pour l'élaboration des textes législatifs et réglementaires"

2.2. La phase rédactionnelle

L'application de la méthode synchrone au sujet a eu une implication directe sur le contenu même des observations, et sur leur rédaction.

2.2.1. La dominante juridique

Nous avons écarté volontairement les approches didactiques, descriptives, lexicales, historiques, qui n'auraient rien apporté par rapport à la littérature existante en la matière (et qui au mieux n'aurait conduit qu'à la rédaction d'un vademécum)²⁵. La posture choisie a été celle de l'approche analytique et critique mettant en évidence, entre autres, les adéquations ou inadéquations du droit par rapport à la technique, l'absence de mise en oeuvre effective des textes, l'absence de respect de la hiérarchie des normes, les possibilités d'interprétations "grises", ouvertes par la rédaction incertaine des textes.

Privilégiant l'étude des textes juridiques dans leur confrontation à la réalité de la technique informatique, l'apport de la méthode synchrone et interdisciplinaire permet de mettre en évidence des ambiguïtés, des inadaptations, des non-sens, des possibilités d'interprétations "orientées".

2.2.2. Dimension dynamique de la rédaction

Le développement même de la rédaction dans le temps a été l'occasion d'une mise à jour permanente des textes, ainsi que d'une prise en compte des apports de la jurisprudence liée aux matières observées. Le corpus étudié a évolué au fil des séances de travail, et l'un des enjeux de l'expérience était de pouvoir être réactif à tout développement dans le temps du processus normatif. Cette veille juridique nous a amené à constater que, paradoxalement, des services administratifs avaient pu être amenés à prendre des circulaires avant même la publication des décrets. L'équipe a utilisé à cet effet la syndication des flux RSS de divers sites d'information juridiques, et institutionnels, permettant une mise à jour thématique régulière de l'actualité et des publications en la matière, dont nous avons fait état dans la première partie.

La rédaction des commentaires s'est faite en deux temps. Les commentaires injectés lors de la première phase se caractérisaient par une formulation sommaire privilégiant la capture de l'idée exprimée et circonscrite au passage de texte commenté. Ils sont en général très courts et peuvent ne pas montrer de rapport évident avec le thème principal de la critique relative à la sécurité informatique et juridique. C'est lors de la deuxième étape qu'une révision thématique ainsi qu'une synthèse des commentaires a été effectuée. La révision thématique a consisté à mettre en exergue à la fois la coloration juridique, informatique ou mixte du commentaire alors que la synthèse a consisté à regrouper plusieurs commentaires relatifs aux termes d'une même phrase ou d'un même paragraphe. La coloration du commentaire ainsi obtenue est calculée en fonction de la

²⁵ même si naturellement entre les participants au projet, on doit faire état d'une approche pédagogique réciproque rendue nécessaire par la différence de formation initiale

coloration des premiers commentaires. L'étape suivante a été l'annotation des commentaires par des mots clés. L'indexation thématique des commentaires est destinée à faciliter et affiner les recherches, à des fins d'efficacité pour la relecture par un juriste et un informaticien, en fonction de ses besoins.

Les mots clés permettent d'abord d'offrir au visiteur un moyen de naviguer dans le site beaucoup plus souple que la navigation imposée par la structure des rubriques et sous-rubriques. Un mot clé décrira en général une entité qui représente un intérêt particulier dans le commentaire tel qu'un concept (relation public privé) ou un acronyme (CERTA) ou un mot quelconque qui n'a d'autre intérêt que celui de regrouper les commentaires auxquels il aura été associé (jaune, vert, bleu, violet). La navigation induite par les mots clés ignore les frontières imposées par les rubriques et est de ce fait transversale et plus puissante.

Les mots clés sont choisis dans une liste globale et unique de termes prédéfinis. Cette liste peut être enrichie dans le temps par d'autres mots clés en évitant les synonymes et les concepts généraux. Cette liste n'est pas dans l'immédiat structurée sous la forme d'un thésaurus. Nous utilisons pour l'instant des ensembles de mots clés plats (comme un dictionnaire lexical, sans autre structure, telle que celle présente dans un thésaurus). Les mots clés associés manuellement par l'auteur à son commentaire définissent dans un certain sens sa sémantique. Il convient de souligner à ce sujet que dans d'autres contextes de l'analyse textuelle d'un corpus dépourvu de telles annotations, cette sémantique est recherchée à l'aide d'outils linguistiques et statistiques tels que la méthode des segments répétés²⁶. Ces outils servent à l'acquisition et à la structuration automatique des connaissances contenues dans un corpus.

Comme nous allons le voir, les mots clés donnent aussi la possibilité de définir une mesure de distance entre les commentaires. Ils permettront d'extraire de la base des commentaires ceux qui leurs sont connexes, et de remonter éventuellement aux passages de texte qui les ont suscités. Le résultat de cette extraction servira alors à la construction d'une analyse plus riche, synthétisant les commentaires trouvés.

Il existe différentes méthodes d'analyse sémantique (Jaccard, Salem...).

S'agissant d'ensembles de mots clé, il existe dans la littérature un certain nombre de mesures de proximité entre des ensembles finis. C'est la mesure relativement simple de Jaccard qui est souvent utilisée. La proximité entre deux ensembles finis de mots clés A et B est définie par : $\frac{|A \cap B|}{|A \cup B|}$, où $| \cdot |$ dénote la cardinalité d'un ensemble (le nombre d'éléments qu'il contient). Le fait qu'un même mot clé peut être lié à plusieurs commentaires associés à un même passage de texte nous amène à adapter la mesure de Jaccard²⁷ afin de mesurer la proximité entre deux multiensembles ("multiset" en anglais). Un multiensemble est un ensemble où on permet la répétition, c'est à dire où un élément peut apparaître plusieurs fois. Cependant la définition de l'intersection de deux multiensembles est différente de la définition classique. La définition classique de l'intersection de deux multiensembles retient le plus petit nombre d'occurrences pour un élément présent dans chaque multiensemble.

²⁶ SALEM André, *Pratique des segments répétés*, Publications de l'INaLF, collection "St.Cloud", Paris, Klincksieck, 1987

²⁷ <http://www.limsi.fr/Individu/rosset/similarite2/node6.html>

(justification du choix)

Dans la mesure où le nombre d'occurrences d'un élément appartenant aux deux ensembles est additionné au lieu de prendre le plus petit des deux. La définition de l'union correspondrait plutôt à la définition de la somme arithmétique de deux multiensembles.[petrovsky] Ces écarts par rapport à la définition classique qui veille à ce que des propriétés de la théorie des ensembles classique soient préservées est justifié par le souci de mesurer l'importance d'un mot clé de l'intersection ou de l'union de deux ensembles en additionnant ses poids dans ces deux ensembles.

Pour une définition claire de la proximité entre deux multiensembles A et B, nous allons identifier un multiensemble à un ensemble de couples ordonnés où sont retenus les éléments du multiensemble avec leur nombre d'occurrences.

$|A \cap B| = \sum((a_i + b_i)c_i)$ avec $C = A \cap B$ et $|A \cup B| = \sum((a_i + b_i)c_i)$ avec $C = A \cup B$ ²⁸.

L'origine essentielle des mots clés provient de la base juridique constituant notre corpus de travail ; dans le contexte spécifique de cette étude, ce calcul de proximité est peu sensible au bruit généré par les moteurs de recherche en prenant en compte plutôt l'existence des mots que leur contexte et leur sémantique.

Le choix de la méthode a une incidence directe sur le contenu même et la rédaction des observations. Le processus dynamique de recherche de la méthode a permis progressivement aux différents participants (juristes et informaticiens) de procéder à un certain nombre de constats. Ils seront examinés dans notre prochain article.

²⁸ Les a_i (respectivement les b_i , les c_i) sont le nombre d'occurrences des mots de A (respectivement de B, de C).