HAL
open science

CATMInE : la proximité linguistique au service de la justice ?

Tom Barrier

To cite this version:

Tom Barrier. CATMInE : la proximité linguistique au service de la justice ?. Sciences de l'Homme et Société. 2018. dumas-01901870

HAL Id: dumas-01901870

https://dumas.ccsd.cnrs.fr/dumas-01901870v1

Submitted on 23 Oct 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

CATMInE : La proximité linguistique au service de la justice ?

BARRIER
Tom

Sous la direction de Thomas LEBARBÉ

Tuteur entreprise : Emmanuel HUYGHE

UFR LLASIC
Département Informatique intégrée en Langues, Lettres et Langage (I3L)

Mémoire de master 2 mention Sciences du Langage - 20 crédits

Parcours : Industrie de la Langue, orientation professionnelle

Année universitaire 2017-2018

CATMInE : La proximité linguistique au service de la justice ?

BARRIER
Tom

Sous la direction de Thomas LEBARBÉ

Tuteur entreprise : Emmanuel HUYGHE

UFR LLASIC
Département Informatique intégrée en Langues, Lettres et Langage (I3L)

Mémoire de master 2 mention Sciences du Langage - 20 crédits

Parcours : Industrie de la Langue, orientation professionnelle

Année universitaire 2017-2018

Remerciements

Je voulais tout d'abord remercier l'équipe d'IP Trust pour leur accueil et leur bonne humeur tout au long de ce stage.

Un remerciement particulier à Emmanuel HUYGHE pour son soutien sur le projet et sa confiance, et à Constance THEVIER pour son temps et son aide en termes de droits des marques, tu m'as beaucoup éclairci.

Je souhaiterais ensuite remercier Thomas LEBARBÉ, pour ses encouragements et son aide sur le projet, que ce soit pour le projet professionnel ou pour le stage en lui-même. Ainsi qu'à vos enseignements que j'ai pu recevoir depuis la deuxième année de licence en option PHP.

Je remercie également Olivier KRAIF d'avoir accedé à ma demande, en acceptant de faire partie de mon jury.

Enfin, un grand merci à Célia et Mahault de m'avoir supporté ces deux ans, et de continue à le faire. La vie créée parfois de belles amitiés, et la notre en est une très belle.

UNIVERSITÉ Grenoble Alpes

DÉCLARATION

1. Ce travail est le fruit d'un travail personnel et constitue un document original.
2. Je sais que prétendre être l'auteur d'un travail écrit par une autre personne est une pratique sévèrement sanctionnée par la loi.
3. Personne d'autre que moi n'a le droit de faire valoir ce travail, en totalité ou en partie, comme le sien.
4. Les propos repris mot à mot à d'autres auteurs figurent entre guillemets (citations).
5. Les écrits sur lesquels je m'appuie dans ce mémoire sont systématiquement référencés selon un système de renvoi bibliographique clair et précis.

NOM : BARRIER……

PRENOM : Tom……

DATE : 10/09/2018…… SIGNATURE : [redacted]

Sommaire

Remerciements 3
Sommaire 5
Introduction 6
1. LE STAGE 6
2. LE DOMAINE DE DROIT DES MARQUES 7
Partie 1 - Les données 9
CHAPITRE 1. QU'EST-CE QU'UNE DONNEE DE MARQUE ET UNE DONNEE DE JURISPRUDENCE ? 10
1. DONNEES DE MARQUES 10
2. DONNEES DE JURISPRUDENCE 11
3. SELECTION ET TRI DES DONNEES DE MARQUES 12
4. SELECTION ET TRI DES DONNEES DE JURISPRUDENCES 12
CHAPITRE 2. COMMENT AUTOMATISER CE PROCESSUS 14
1. AUTOMATISATION DE LA RECUPERATION 14
2. DIFFICULTES RENCONTREES 15
Partie 2 - Les différentes phases du fonctionnement du programme 17
CHAPITRE 4. LE MODELE LINGUISTIQUE 19
1. LES VECTEURS LINGUISTIQUES 19
2. CALCUL DE LA PROXIMITE LINGUISTIQUE ENTRE DEUX MARQUES 20
3. CALCUL DE LA PROXIMITE LINGUISTIQUE NOTRE CAS D'OPPOSITION FICTIF ET LES CAS D'OPPOSITIONS EXISTANTS 21
4. CALCUL DU POURCENTAGE 22
CHAPITRE 5. LE MODELE INFORMATIQUE 24
1. INTERFACES CLIENT 24
2. INTERFACES ADMINISTRATEUR 27
CHAPITRE 6. L'AFFICHAGE DES RESULTATS 30
1. AFFICHAGE DES CAS D'OPPOSITIONS EXISTANTS LES PLUS PROCHES 31
2. AFFICHAGE DES MARQUES LES PLUS PROCHES 32
3. CALCULS DES COORDONNEES 33
CHAPITRE 7. RESULTATS OBTENUS 39
1. EXEMPLES DE RESULTATS 39
2. UNE CLASSE PEUT TOUT CHANGER 40
Conclusion 42
1. RESUME 42
2. AMELIORATIONS POSSIBLES 42
3. BILAN PERSONNEL 43
Table des illustrations 44
Table des annexes 45
Table des matières 52

Introduction

Ce mémoire est le fruit du travail fourni lors d'un stage effectué au sein de la société IP Trust Innovation[1]. Il s'inscrit dans la continuité des enseignements reçus tout au long de ce master Industrie de la Langue, et montre l'application de ces connaissances théoriques dans le cadre d'un projet professionnel. Nous allons maintenant présenter la société IP Trust Innovation et expliquer le domaine dans lequel se place le projet.

1. Le stage

1.1. La société IP Trust Innovation

IP Trust Innovation est un cabinet de conseil en propriété industrielle qui a pour motivation d'accompagner les acteurs économiques dans la protection et la valorisation de leurs innovations. Cette société est attachée à une approche entrepreneuriale et stratégique de la propriété industrielle et met un point d'honneur sur une organisation et un système d'information privilégiant l'interactivité avec ses clients.

Son siège social se situe à Paris, mais elle s'est aussi implémentée à Grenoble, ainsi qu'à La Rochelle où un cabinet va bientôt ouvrir ses portes. Ils proposent de nombreux services inscrits dans des domaines hétéroclites :

- Protection des inventions et brevets
- Opinions légales
- Conseil en stratégie de PI[2]
- Audit et due diligence
- Evaluations financières de DPI[3]
- Fiscalité
- Contrats
- Protection des marques

Depuis l'été 2018, la société a aussi mis en place le service LABLOCK, qui est un cahier de laboratoire numérique servant à stocker et gérer des documents, pour notamment pouvoir apporter la preuve d'antériorité d'une connaissance technique.

---

[1] https://www.iptrust.fr/
[2] Propriété intellectuelle
[3] Droit de propriété intellectuelle

Dans cette dynamique de proposer toujours plus de services à leurs clients, et affirmer leur place dans le domaine de la propriété industrielle, IP Trust a, en 2003, lancé le projet d'un logiciel d'aide à la décision en droit des marques, se fondant sur une analyse linguistique des marques nominatives.

Ce projet initial avait été réalisé par Thomas LEBARBÉ, qui a ensuite recruté Ni[redacted] C[redacted] et Pi[redacted] R[redacted] pour fournir une deuxième version, avant d'être mis de côté. L'objectif était donc de remettre le logiciel au goût du jour, à partir du cahier des charges et des sources de 2003, et de le refondre afin de le rendre accessible en ligne. C'est donc de ce projet que le nôtre part.

1.2. La mission

Il fallait donc reprendre le projet initial, et le rendre le plus fonctionnel possible. Une des modifications majeures a été la volonté de la société IP Trust Innovation de passer le programme, qui était initialement en langage JAVA[4], en langage Python[5]. Nous restons donc toujours dans le développement d'une application web, mais en changeant de technologie.

Une autre grosse partie de la mission était, en collaboration avec le commanditaire, la récupération automatique des données nécessaires au fonctionnement du programme, car si les données sont erronées, les résultats le seront aussi. Une des contraintes à ce niveau-là était le fait que je n'avais pas de base de données à disposition pour pouvoir les stocker. J'ai donc dû les stocker en local.

Même si la majeure partie du travail sera concentrée sur la partie back-end[6] de la programmation, un travail sera à fournir en front-end[7], afin de pouvoir proposer un affichage des résultats que la partie back-end du programme nous donnera.

Ce stage s'est déroulé du 05 mars au 05 septembre, soit 6 mois, dans les bureaux Isérois d'IP Trust Innovation, situés à Bernin. Au sein de l'entreprise, il était encadré par Emmanuel HUYGHE, ingénieur brevet, et était assisté de Constance THEVIER, juriste junior en alternance.

2. Le domaine de droit des marques

Regardons tout d'abord ce qui peut être protéger sous le terme « marque », car ce ne sont pas que des mots comme on pourrait le penser. Une « marque » peut donc être un mot, un nom, un slogan, des chiffres, des lettres, un dessin, un logo, une combinaison de ces différents éléments (un dessin et un nom dans un graphisme particulier, par exemple), ou encore un signe sonore (à condition qu'il puisse être représenté graphiquement (ex. : une portée musicale).

---

[4] Langage de programmation orienté objet
[5] Langage de programmation objet interprété
[6] « Arrière-plan » du programme, contient généralement le traitement à effectuer, et nous fournit le résultat
[7] Affichage « frontal », la partie que l'utilisateur lambda verra (la partie visible de l'iceberg)

Lorsque l'on souhaite déposer une marque, il faut d'abord vérifier la disponibilité de celle-ci. En citant le site de l'INPI[8], on peut voir que « D'une manière générale, une marque est disponible quand elle ne reproduit ou n'imite pas un signe qui bénéficie d'un droit antérieur pour des produits ou des services, ou des activités qui seraient identiques ou similaires » aux nôtres.

Il faut aussi penser aux classes dans lesquels on souhaiterait déposer notre marque. Les classes sont les différents types de produits et services dans lesquels une marque peut être posé, et pour lesquels elle souhaite une protection. Il existe un total de 45 classes. Ils sont accompagnés des libellés, qui permettent d'indiquer plus précisément ce que nous protégeons. Là où les classes sont des catégories de produits et services, les libellés sont eux renseigner par le dépositaire, et sont donc personnalisables, afin de préciser au maximum le service ou produit que notre marque couvre.

Après cela, lorsque la marque est déposée, le propriétaire d'une marque antérieure dispose de 2 mois pour faire opposition et ainsi s'opposer à l'enregistrement de notre marque, car il estime que celle-ci porte atteinte à ses droits. Après ce délai, il dispose de 5 ans pour attaquer notre marque en contrefaçon, se tournant alors vers une procédure judiciaire beaucoup plus lourde et beaucoup plus coûteuse.

Lorsque les personnes souhaitant déposer une marque vérifient la disponibilité de celle-ci, ils peuvent faire appel à des sociétés comme celle d'IP Trust Innovation, dont les juristes conseillent ces personnes sur le risque de contrefaçon et si la marque qu'ils souhaitent déposer posera un problème à ce niveau-là.

C'est là que notre solution rentre en jeu, avec pour objectif de faciliter les recherches concernant d'éventuelles marques similaires à la marque voulue, ainsi que d'afficher les cas de jurisprudences existants pertinents pour le cas fictif étudié. Le juriste pourra donc avoir accès à ces informations plus rapidement, qui pourront l'aiguiller dans sa décision.

Dans un second temps, notre solution servira aussi en interne à la société, si les associés de la société souhaite se renseigner sur une marque précise, et accéder à ces informations pour eux-mêmes.

---

[8] L'Institut National de la Propriété Industrielle | https://www.inpi fr/fr

Partie 1

-

Les données

Chapitre 1. Qu'est-ce qu'une donnée de marque et une donnée de jurisprudence ?

Une des nécessités pour notre solution idéale est d'avoir assez de données pour comparer et placer notre cas fictif de possible opposition. Pour cela, la société IPTrust Innovation a pris contact avec l'INPI (l'Institut National de la Propriété Industrielle), qui a différentes responsabilités et missions dans le domaine de la propriété industrielle en France. Parmi leurs missions, ils sont chargés de recevoir et de délivrer les titres de propriété industrielle (brevets, marques, dessins et modèles), mais aussi de centraliser dans des bases de données, les informations relatives aux marques et à la jurisprudence. Et ce sont ces informations qui vont nous être utiles, notamment les cas d'oppositions pour la jurisprudence.

Nous avons donc obtenu l'accès aux données via le FTP[9] de l'INPI, pour d'un côté les marques pour pouvoir comparer les marques existantes et la marque que le client souhaite déposer, et de l'autre les cas d'oppositions existants afin de les comparer à notre cas fictif et d'en sortir les plus pertinents pour le calcul du pourcentage de risque.

1. Données de marques

La base de données des marques contient toutes les marques qui ont été déposées en France depuis la création de l'INPI. Pour les plus anciennes, elles sont regroupées par année (2014 pour les premières, car le FTP a été mis en place en 2014) dans un dossier « Archives », dans des sous-fichiers zippés. Pour les marques de l'année en cours, un nouveau fichier zippé est ajouté chaque vendredi, contenant les marques déposées dans la semaine passée. Une fois qu'une année est passée, tous les fichiers de celle-ci sont regroupés dans un dossier dont le nom est l'année en question, afin de les archiver.

Dans ces fichiers zippés, on retrouve des fichiers XML, contenant les marques déposées, mises à jour et supprimées de la semaine. Chaque entrée[10] est composée de ces informations (dans l'ordre d'apparition) :

- Le code du pays pour lequel on enregistre la marque (sera toujours FR car nous ne disposons que des marques déposées en France)
- Le numéro d'enregistrement de la marque
- La date d'enregistrement de la marque
- Le code de la langue de dépôt
- La date d'expiration de la marque

[9] Serveur utilisant la technologie File Transfer Protocol, qui est destiné au partage de fichiers sur un réseau
[10] Voir Annexe 1, p.46

- Le statut de la marque (demande publiée, marque renouvelée, marque expirée)
- Le type de marque[11]
- Les détails de la description de la marque
- Le nom de la marque (sauf si figurative)
- Les détails de l'image à protéger (s'il y en a une)
- Les différentes classes et libellés associés à la marque
- Description du dépositaire (nom, adresse, pays, type de société)
- Description du représentant du dépositaire (si société)
- Description et date des différentes étapes et modifications de la marque (publication, marque renouvelée, changement de propriétaire…)

2. Données de jurisprudence

Comme dit plus haut, les données de jurisprudence qui nous intéressent sont celles concernant les cas d'oppositions. Là où les données de marques sont actualisées toutes les semaines, celles d'oppositions sont actualisées tous les mois. De la même manière, les cas d'oppositions sont stockés sous forme de fichiers zippés, dont les noms sont pour chacun de ces fichiers, le mois et l'année des cas.

Dans ces fichiers zippés, on retrouve des fichiers XML qui contiennent l'ensemble des cas d'oppositions ayant eu lieu dans le mois en question, ainsi que l'ensemble des fichiers PDF détaillant chaque cas listé dans ces fichiers XML.

Comme pour les marques, les fichiers des cas d'oppositions sont séparés entre les nouvelles entrées, et les entrées mises à jour. Chaque entrée[12] est composée de ces informations (dans l'ordre d'apparition) :

- La référence juridique du cas (qui sera le nom du PDF associé)
- Le ou les articles qui ont permis de prendre la décision
- La date où la décision a été prise
- Le type de conflit (nous n'aurons ici que les OPPOSITIONS)

---

[11] Voir Introduction, p.8
[12] Voir Annexe 2, p.50

- Les différents cas en lien avec celui étudié, s'il y en a (avec la date, la référence et le cour associé)

- Les noms des deux marques en opposition

- La nationalité, ou type, des deux marques en opposition

- La référence de l'INPI du cas

- Les numéros d'enregistrement des deux marques en opposition

- Les noms des organismes ou personnes dépositaires des deux marques en opposition

- Le chemin relatif de l'emplacement du fichier PDF du cas

C'est donc parmi ces informations que nous allons devoir sélectionner et trier les données qui nous semblent nécessaires à notre solution, et réussir à faire en sorte que ce processus n'est pas à être touché, et puisse ainsi fonctionner de manière autonome.

3. Sélection et tri des données de marques

Concernant les données de marques, étant donné qu'elles sont séparées dans de nombreux fichiers XML[13], et que beaucoup d'entre elles ont expirées, il ne suffit pas de parcourir les fichiers et de prendre toutes les marques comme ça : j'ai donc dû les trier et garder celles qui sont encore actives, et de faire attention de ne pas prendre deux fois la même.

Pour cela, j'ai eu besoin du nom de la marque (évidemment) afin de pouvoir la comparer graphiquement et phonétiquement à la marque que l'on entrera. Mais il me faudra aussi, afin de pouvoir ensuite vérifier que cette marque n'est pas déjà présente dans nos données, son numéro d'enregistrement, ainsi que la date de sa dernière mise à jour : pour deux entrées qui ont le même numéro d'enregistrement, on gardera donc celle qui a la date de dernière mise à jour la plus récente. J'aurais aussi besoin de sa date de dépôt, ainsi que de ses classes, sa date d'expiration, et enfin sa transcription phonétique (qui ne se trouve pas dans les fichiers XML, mais que l'on transcrira lors de l'automatisation de la récupération des données).

4. Sélection et tri des données de jurisprudences

Concernant les données de jurisprudence, J'aurais besoin du nom des deux marques en opposition, ainsi que leurs numéros d'enregistrement. On peut remarquer que les classes des deux marques en opposition ne sont pas renseignées dans les fichiers XML associés, or ce sont des informations qui peuvent être intéressantes et qu'il serait bien d'afficher pour l'utilisateur. Il faudra donc aller les chercher dans les données des marques, en utilisant les

[13] Langage de balisage, souvent utilisé pour stocker des données

numéros d'enregistrement des deux marques, afin de trouver les classes correspondantes à celles-ci.

La deuxième chose que l'on remarque c'est que, pour chaque cas, la décision qui a été prise n'est pas renseignée dans le fichier correspondant. Or, c'est une information qu'il est capital d'avoir pour notre traitement, afin de pouvoir établir le pourcentage de risque et afficher à l'utilisateur si les cas proches sont des cas d'oppositions s'étant soldés par un rejet de l'opposition, ou un rejet de la demande d'enregistrement de la marque. Pour récupérer la décision d'un cas, je parcours donc le fichier PDF associé, dont on récupère le nom dans le fichier XML approprié.

A noter qu'il existe trois décisions possibles : opposition justifiée, opposition rejetée, et opposition partiellement rejetée. Dans le premier cas, la demande d'enregistrement est rejetée car l'opposition a amené assez de preuves pour être justifiée. Dans le deuxième cas, la demande d'enregistrement de la marque est acceptée, car ne constituait pas un risque de contrefaçon suffisant pour la marque faisant opposition. Et dans le dernier cas, la demande d'enregistrement est acceptée, mais seulement pour les classes listées dans le fichier PDF de la décision : on a donc un rejet partiel de l'opposition, ou une demande d'enregistrement partiellement acceptée.

Pour savoir dans quelle catégorie se classe notre cas, je convertis le fichier PDF en chaîne de caractères afin de parcourir celle-ci avec des expressions régulières (une pour chacun des trois cas possibles), et ainsi noter la décision prise dans le fichier contenant le cas associé.

Chapitre 2. Comment automatiser ce processus

1. Automatisation de la récupération

Comme il y a énormément de données, surtout du côté des marques qui compte plus d'un million d'entrées, et qui est mis à jour chaque semaine, il est nécessaire d'automatiser cette récupération, car cela serait beaucoup trop long et laborieux de parcourir les fichiers XML et récupérer les données à la main. Pour cela, j'ai écrit et mis en place des scripts autonomes qui seront lancés à intervalle régulier (un pour les marques, et un pour les cas d'oppositions existants, qui eux ne s'élèvent « qu'à » un peu plus de quatre millions d'entrées).

Mais, il faut séparer la création de la « base » des données, qui nous servira de données de départ, et la mise à jour de cette « base » (une fois par semaine pour les marques, et une fois par mois pour les cas d'oppositions) notamment pour les marques, qui peuvent périmées. On considère qu'un cas d'opposition ne périme pas car il sera toujours intéressant de le garder. Il peut cependant être mis à jour, si la décision change par exemple. Le script de création de « base » ne sera donc lancé qu'une seule fois.

En effet, pour la première étape qui est de se connecter au serveur FTP de l'INPI et de récupérer les fichiers, pour créer la « base », il faudra donc prendre tous les fichiers zippés présents sur le serveur ftp au moment de la création de notre fichier, ainsi que ceux des archives pour les marques les plus anciennes. Mais lors de la mise à jour hebdomadaire, il ne faudra prendre que le ou les fichiers zippés ajoutés dans la semaine courante, afin d'actualisé le fichier de « base ».

Ensuite, j'extrait les fichiers XML présents dans les fichiers zippés que l'on a téléchargés, et je supprime ces derniers, car ils ne nous seront d'aucune utilité et ils prennent de la place, or seuls les fichiers XML nous servent. Je continue en effectuant le traitement de récupération des données sur chacun de ces fichiers XML, dans lesquels on récupère les données souhaitées dont on a parlé plus haut. Pour la création du fichier de « base » on prend donc tous les fichiers présents et on construit le fichier de sortie petit à petit. Alors que lorsque qu'on le met à jour, on parcourt le fichier « base » créer par le premier script, pour enlever les marques périmées depuis, mettre à jour celles qui sont présentes dans la base et le nouveau fichier de la semaine, et intégrées les nouvelles.

Heureusement il est facile de parcourir et récupérer les informations que l'on désire dans un fichier XML, grâce à la structuration de ce type de fichier, qui utilise des balises, pour stocker de manière hiérarchisée les données.

Attention cependant, car comme expliqué plus haut, il faut non seulement ne pas prendre en compte les marques périmées, mais aussi faire en sorte de ne pas avoir plusieurs fois la même marque dans notre fichier de données de sortie, ce qui rajoute du temps de traitement.

Par même marque, j'entends deux entrées ayant le même numéro d'enregistrement. A noter qu'il peut y avoir plusieurs fois le même nom de marque, mais avec des numéros d'enregistrement différents : il s'agit bien de la même marque, du même dépositaire, mais enregistrée dans des classes différentes. Ceci est dû au fait qu'une marque, une fois déposée, ne peux pas être modifiée. Si une entreprise veut déposer sa marque dans des

classes qui n'étaient pas indiquées dans celle qu'elle a déjà déposée, elle doit donc refaire une demande d'enregistrement. On peut donc retrouver plusieurs fois le même nom de marque, mais associés avec des classes et des numéros d'enregistrement différents. C'est pour cette raison que le tri des éventuelles marques redondantes s'effectue par rapport au numéro d'enregistrement et non par rapport au nom de la marque.

On ne prendra pas aussi en compte la marque si elle est figurative, car nous ne pourrons pas la comparer linguistiquement. Si l'on se réfère aux types de marques que l'on a énoncé dans l'introduction, on ne prendra donc pas en compte les dessins, les logos, ou les signes sonores. En revanche, s'il s'agit d'une combinaison d'éléments, tant qu'il y a une donnée écrite, on la prendra en compte.

Suite à ce traitement, on effectue la transcription phonétique, qu'on rajoute en fin de ligne. Cette transcription phonétique nous est fournie par un script JAVA qui se lance en prenant en paramètre le nom de la marque dont on souhaite la transcription, et nous sors son équivalent en phonétique. Ce script était présent dans l'ancienne version dont on a parlé et a été récupéré tel quel.

Voici deux extraits de nos fichiers de données en sortie

```
ADDVALENTIAM    3620646 2008-12-31  09 35 36 41 42  NONE    2018-12-31  advalâtyam
GALLINETTE  3620647 2008-12-31  28 35 38 41 NONE    2018-12-31  gallinètt
MYRABILIA   3620648 2008-12-31  09 35 38 42 2014-05-07  2018-12-31  mirabilya
E-NAUTIA    3620649 2009-01-01  35 38 41    2010-02-09  2019-01-01  nosya
IOBACKUP    3620650 2009-01-01  35 38 39 42 NONE    2019-01-01  yobakup
SECTEURX    3620651 2009-01-01  35 38 43    NONE    2019-01-01  sèkteriks
```

Figure 1 - Extrait de 6 lignes du fichier contenant nos données de marques

```
ADIDAM      16171704    DIDAM   16 35 41    4387713 2018-06-04  O20174837   2
ADISTAR         14221601    ADISTAR 09 38 41 42 4337826 2017-10-24  O20171942   2
ADL     800730  ADL. MARIE ADELINE  16 21 24 25 27  4409788 2018-07-06  O20180717   2
ADLP DIGITAL    35 38 41 42 3938374 ADLP    35 38 42    4267399 2017-01-16  O20163157   1
ADOPTE UN MEC   03 09 14 16 18 20 25 28 35 38 41 45 4051410 ADOPTE UN CONCIERGE 35 36   4394303 2018-06-08  O20175114   0
```

Figure 2 - Extrait de 5 lignes du fichier contenant nos données de Jurisprudence

Si dans la Figure 1 il y a parfois des « NONE », c'est parce qu'à cet emplacement, on met la date de la dernière mise à jour de la marque, sans prendre en compte le dépôt qui n'est pas vraiment considérer comme une mise à jour mais comme la création de la marque. J'ai donc ajouté dans le traitement une condition qui dit que s'il n'y a pas de date de mise à jour à prendre, alors on met NONE, de sorte à avoir la même structure dans tous les cas.

2. *Difficultés rencontrées*

Tout d'abord, le temps de traitement est à prendre en compte : en effet, le fait de devoir télécharger, extraire et parcourir tous les fichiers, et certains de ces fichiers XML pouvant atteindre les 500 mégaoctets, rend ce temps de traitement long, même si toujours plus rapide que de le faire à la main. De plus, la mise à jour des données nécessitant de

reparcourir le fichier de donnée existant à chaque nouvelle entrée, on obtient un temps de traitement qui peux des fois atteindre une demi-journée. D'où l'utilité de le lancer pendant la nuit.

Concernant les libellés, il est pour l'instant difficile de les prendre en compte, car cela implique de prendre beaucoup plus de données (car les libellés peuvent parfois être très longs). Mais aussi parce que, les libellés étant des phrases personnalisées, où le dépositaire décrit avec précision le service ou produit qu'il fournit, il n'y a pas de phrases prédéfinies, et donc pas de pattern facilement démarqué. On serait donc tenté de prendre des mots clés dans ces libellés, afin de les comparer, mais j'ai considéré que ce n'était pas une bonne solution, car cela ramènerait à rendre ces données approximatives, alors que tout le but de ces libellés, c'est d'être précis. Je pense qu'en perdant de l'information, le résultat ne serait plus aussi fiable, car on se contenterait de comparer des mots-clés, rassemblant ainsi plus à des catégories. Or les classes remplissent déjà ce rôle de catégorisation.

Pour rester dans les classes, un problème s'est posé lors de la création du fichier contenant les données de Jurisprudence. En effet, comme expliqué plus haut, les classes des deux marques en oppositions n'étant pas renseignées dans les fichiers XML de l'INPI, nous sommes allés chercher classes associées à ces marques dans nos données de marque que l'on a récolté au préalable. Malheureusement, comme nous ne disposons que des marques déposées en France, certaines marques dans ces cas d'oppositions ne s'y trouvent pas, rendant ainsi impossible de les renseigner à l'utilisateur. Une possibilité aurait pu être de garder les marques périmées dans un autre fichier, mais cela rajouterait du temps de traitement, et ferait beaucoup de données en plus. Comme ces données ne servent pas dans nos calculs, et ne servent que pour un affichage, elles peuvent être mises de côté.

Un autre point qu'il faut prendre en compte, même s'il est indépendant de notre volonté, c'est la manière utilisée par l'INPI pour stocker les données. En effet, même s'ils ne l'ont pour l'instant jamais changé, si l'INPI décide de changer sa façon de fournir les différentes données dont on a besoin, il faudra alors reprogrammer entièrement cette automatisation automatique des données.

Il faut vraiment bien comprendre à quel point il est important que les données utilisées dans notre solution soient fiables, car c'est ce qui donne du crédit à nos résultats : si les données que l'on utilise pour calculer la proximité linguistique ainsi que le pourcentage sont erronées, alors notre résultat le sera aussi. C'est pour ça que, malgré le temps que cela peut mettre, les données se doivent d'être à jour et venir d'une source reconnue (ici l'INPI).

Partie 2

-

Les différentes phases du fonctionnement du programme

Dans cette partie, je vais m'attarder sur le modèle linguistique que j'ai utilisé dans notre solution, avec lequel nous calculons donc la proximité linguistique, ainsi que sur le modèle informatique, pour voir ce que nous j'ai choisi d'afficher, les différents modules ainsi que leurs relations. Nous verrons ensuite comment j'ai choisi d'afficher les résultats et nous regarderons enfin ces derniers.

Il faut tout d'abord savoir que dans notre solution, il y a deux types de recherche : une que nous appellerons « simple » et l'autre que nous appellerons « globale ».

- Pour la version simple, c'est à l'utilisateur de rentrer, non seulement la marque qu'il souhaite déposer, mais aussi celle avec laquelle il souhaite la comparer (il peut aussi rentrer les différentes classes de ces deux marques, mais celles-ci ne seront pas prises en compte dans les calculs de proximité).
  Ici nous n'utiliserons donc pas les données des marques car nous n'utiliserons que les deux rentrées par l'utilisateur. Pour ce type de recherche, nous voulons obtenir les dix cas de jurisprudence existants les plus linguistiquement proches de notre cas fictif.

- Pour la version globale, l'utilisateur ne rentre que la marque qu'il souhaite déposer, ainsi que les classes dans lesquelles il veut la déposer. A noter qu'ici, contrairement à la version simple, les classes ont un rôle important. En effet, lors du traitement, si l'utilisateur a rentré des classes pour la marque qu'il souhaite déposer, les marques que l'on analyse dans nos données n'ayant pas au moins une classe en commun avec celles-ci ne sont pas prises en compte. A l'inverse, si l'utilisateur n'a pas rentré de classes, on prend tous les marques des données sans distinctions.
  Pour ce type de recherche, nous voulons obtenir les dix marques existantes les plus proches de notre marque à déposer, ainsi que les dix cas d'oppositions existants les plus proches pour chacune de ces dix marques les plus proches. Nous obtiendrons ainsi 100 cas en tout, et pourront établir un pourcentage en fonction de leurs décisions associées, mais nous reviendrons sur cela plus tard.

L'objectif de notre solution, et donc le but des calculs de proximité linguistiques, est de nous renseigner, dans un premier temps, si la marque que l'on souhaite déposer est plus ou moins linguistiquement proche de l'autre marque rentrée, dans le cas de la version simple, ou des différentes marques existantes en France dans le cas de la version globale.

Et dans un second temps, de nous renseigner si chaque cas d'opposition fictif que l'on créé (marque à déposer/marque existante), est plus ou moins linguistiquement proches des différents cas d'opposition déjà existants. Nous allons maintenant détailler comment sont calculées ces proximités.

Chapitre 4. Le modèle linguistique

1. Les vecteurs linguistiques

Pour évaluer la proximité linguistique entre deux marques, il nous faut d'abord calculer des vecteurs entre les deux entrées (la marque que l'on souhaite déposer et la marque avec laquelle on la compare) afin de mettre sous forme numérique les graphèmes et les phonèmes que l'on compare. Ces vecteurs sont les suivants :

Caractères de Marque 1 dans Marque 2 :

- Indique la quantité de caractères de la Marque 1 que l'on retrouve également dans la Marque 2. Pour le calculer, nous parcourons simplement la Marque 1 et nous récupérons le nombre de caractères identiques à la Marque 2. On divisera ensuite cette valeur par la longueur de la Marque 1, pour avoir un nombre décimal entre 0 et 1.

Caractères de Marque 2 dans Marque1 :

Indique la quantité de caractères de la Marque 2 que l'on retrouve également dans la Marque 2. Ce vecteur se calcule de la même façon que celui juste au-dessus, sauf que l'on parcoure cette fois la Marque 2. On divisera encore une fois la valeur obtenue par la longueur de la Marque 2, pour avoir un nombre décimal entre 0 et 1.

Plus grande sous chaine commune de Marque 1 dans Marque 2 :

Indique la longueur de la plus longue sous chaîne commune entre la Marque 1 et la Marque 2, relativement à la Marque 1. Pour cela, on trouve cette sous-chaine commune et on calcule sa position dans les deux marques : si elle se trouve au début, au milieu, à la fin, ou bien s'il s'agit de la marque entière. On calcule également un coefficient en fonction de la différence de positionnement de cette sous-chaine dans les deux marques. On divise ensuite cette valeur par la longueur de la Marque 1 multiplié par le coefficient calculé.

Plus grande sous chaine commune de Marque 2 dans Marque 1 :

Indique la longueur de la plus longue sous chaîne commune entre la Marque 1 et la Marque 2, relativement à la Marque 2 cette fois. On divise ensuite cette valeur par la longueur de la Marque 2 multiplié par le coefficient calculé, comme pour celui juste au-dessus.

Ces quatre vecteurs sont calculés dans un premier temps avec les graphèmes, la façon dont le mot est écrit, et dans un second temps avec les phonèmes, la façon dont il se prononce. On obtient donc 8 valeurs, que l'on stocke avec la décision qui a été prise dans ce cas d'opposition : « 0 » si l'opposition a été rejetée, « 1 » si l'opposition a été justifiée, et « 2 » si l'opposition a été partiellement rejetée.

Ces chiffres sont arbitraires et nous permettront juste de récupérer la décision prise pour chaque cas lors de l'affichage des résultats. On ajoute également à la fin de la ligne les deux marques en conflit pour savoir de quel cas on parle.

```
1.0 1.0 0.3333333333333333  0.30000000000000004 1.0 1.0 1.5 1.5 1   ACROPOLIS   ACCROPOLIS
0.7142857142857143  1.0 0.6428571428571428  0.8999999999999999  0.75    1.0 0.5625  0.75    2   ACTERIM ACTIM
```

Figure 3 - Extrait du fichier JurisVecto.data contenant les cas d'opposition existants vectorisés

Il est nécessaire de calculer le vecteur des caractères du Mot 1 dans le Mot 2, et le vecteur des caractères du Mot 2 dans le Mot 1 séparément car il ne suffit pas de simplement prendre les caractères communs aux deux mots. En effet, si on prend les exemples (tirés des données des cas d'oppositions) de la marque « Acterim » et la marque « Actim », on peut voir que tous les caractères de « Actim » sont également présents dans « Acterim », mais que les caractères de « Acterim » ne se retrouvent pas tous dans « Actim ». Nous aurons donc bien deux vecteurs qui auront des valeurs différentes.

A noter que pour nos données des cas d'oppositions, ces vecteurs sont calculés en avance et stockés dans un fichier JurisVecto.data[14], car ces données n'étant pas si importantes en termes de taille, on peut se permettre d'effectuer ces calculs lors de la mise à jour des données et ainsi réduire, même si très légèrement, le temps de traitement général. Pour les données des marques, on ne peut malheureusement pas faire la même chose. Déjà parce qu'il y a beaucoup plus de données, mais surtout parce que ces vecteurs se calculant entre deux marques, on ne peut pas savoir à l'avance quelle marque l'utilisateur va rentrer, et ainsi effectuer les calculs entre celle-ci et les marques déjà existantes. Nous sommes donc obligés de le faire lors du traitement, ce qui est en partie responsable du problème du temps de traitement dont on a parlé plus haut.

2. *Calcul de la proximité linguistique entre deux marques*

Une fois ces valeurs obtenues, nous pouvons tout d'abord calculer si les deux marques étudiées sont linguistiquement proches ou non. Pour cela on ajoute tout simplement ces valeurs les unes aux autres, et on obtient ainsi un nombre qui peux aller jusqu'à 10. En sachant que lorsque ce résultat obtenu est 10 les deux marques sont identiques, on sait donc que plus cette valeur tendra cers 10, plus les deux marques seront proches.

Si on prend un exemple avec ces données, toujours prises dans les données enregistrées pour la solution :

```
0.7142857142857143  1.0 0.6428571428571428  0.8999999999999999  0.75    1.0 0.5625  0.75    2   ACTERIM ACTIM
1.0 0.6086956521739131  0.6428571428571428  0.19565217391304346 1.0 0.6842105263157895  0.5625  0.23684210526315788 2   ACTERIM ACTIM, L INTERIM MUTUEL
```

Figure 4 – Autre extrait du fichier JurisVecto.data contenant les cas d'opposition existants vectorisés

Et que nous effectuons les calculs : pour le cas « Acterim | Actim » on additionne donc les vecteurs, ce qui nous donne 6,32 si l'on arrondi au centième près. Et pour le cas « Acterim | Actim, l'interim mutuel », nous obtenons 4,93 lorsque l'on arrondi au centième près. On peut donc voir que si « Acterim » était notre marque à déposer, la marque existante « Actim » en serait plus linguistiquement proche que « Actim, l'interim mutuel ». En effet, même s'il y a « Actim » dans les deux, le fait que dans le deuxième cas on ait « l'interim mutuel » à la suite, le rend moins proche.

[14] Voir Figure 3, p.21

Comme on peut le voir dans l'image extraites des données, les vecteurs peuvent être des nombres décimaux très longs, il est donc nécessaire de les prendre en entier et d'être précis dans le résultat de cette addition, car des fois cela peut se jouer à un millième près pour savoir si une marque est plus proche qu'une autre.

Si on prend le cas de la recherche globale, lorsque nous parcourons les marques existantes, nous effectuons donc ces mêmes calculs entre notre marque à déposer et chacune des marques existantes, et nous stockons les résultats dans un tableau, que nous trions ensuite par ordre décroissant. On peut ainsi prendre les dix premières entrées de ce tableau trié, pour obtenir les dix marques les plus proches de notre marque à déposer (et ainsi créer dix cas fictifs d'oppositions).

3. Calcul de la proximité linguistique notre cas d'opposition fictif et les cas d'oppositions existants

Ensuite, pour obtenir les dix cas de jurisprudence existants les plus proches, de chacun de ces cas fictifs créés, nous nous servons de ces vecteurs pour calculer une « distance », et ainsi évaluer la proximité linguistique, entre un cas fictif et un cas existant.

Pour calculer cette distance, il nous faut les vecteurs d'un cas fictif dont on veut les cas existants les plus proches, et parcourir le fichier de données contenant les vecteurs des cas existants, calculés en avance. On obtient ainsi deux jeux de 8 vecteurs, que l'on peut comparer l'un à l'autre. Pour cela on soustrait chaque vecteur avec son correspondant : on soustrait le !!!, pour voir la différence de chaque vecteur.

Chaque différence est ensuite mise à la puissance 2, avant d'être additionné l'une à l'autre (on utilise donc le théorème de Pythagore $AB^2 = AC^2 + BC^2$). On obtient ainsi un nombre dont on prend la racine carrée et nous donne ainsi la distance que l'on recherchait (car dans la formule précédente, AB^2 est la distance que l'on souhaite trouver au carré, il faut donc prendre la racine carrée de cette valeur).

Étant donné que lorsque le résultat obtenu est 0, cela signifie que les deux cas sont identiques, plus ce résultat tendra vers 0, plus le cas existant sera proche de notre cas fictif. J'entends par identiques qu'ils auront la même logique linguistique, par exemple : les cas « Mama Shelter | Mama Home » et « Attitude Alpes | Attitude Pyrénées » seront potentiellement proches à 100% car dans les deux cas le premier mot est le même, mais le deuxième change. Ces résultats obtenus sont ainsi stockés dans un tableau, que l'on trie par ordre croissant, nous permettant ainsi, de la même façon que pour les marques, de récupérer les dix premières entrées du tableau, qui seront les dix cas existants les plus proches pour ce cas fictif.

Si on prend un exemple avec ces données :

```
0.7142857142857143  1.0 0.6428571428571428  0.8999999999999999  0.75    1.0 0.5625  0.75    2   ACTERIM ACTIM

1.0 0.38461538461538464 1.5 0.576923076923077   1.0 0.5 1.2000000000000002  0.6000000000000001  2   HUSKY   HUSKY-VINTAGE

1.0 0.75    0.5 0.3125  1.0 0.875   0.5 0.375   1   STORY   MY STORY
```

Figure 5 - Données utilisées pour exemple de calcul de proximité entre deux cas d'oppositions

Imaginons que le cas « Acterim | Actim » soit notre cas fictif, on compare donc « Acterim | Actim » à « Husky | Husky-vintage », puis « Acterim | Actim » à « Story | My story ».

Pour la première comparaison, en effectuant les calculs expliqués plus haut, on obtient une distance de 1,43 arrondi au centième près. Et pour la deuxième comparaison on obtient 0,86 arrondi au centième près. Le cas « Acterim | Actim » est donc plus proche du cas « Story | My story » que du cas « Husky | Husky-vintage ».

La proximité linguistique entre les cas d'oppositions n'a donc rien avoir avec si les marques 1 et les marques 2 des cas d'oppositions et du cas fictif se ressemblent entre elles ou non, mais plutôt de calculer si la différence entre les deux marques du cas fictif et la différence entre les deux marques du cas existant se ressemblent ou pas.

4. Calcul du pourcentage

Maintenant que nous avons obtenus, pour une recherche simple les dix cas d'oppositions les plus proches de notre cas fictif, et pour une recherche globale les dix marques les plus proches, et pour chacune de ces marques les dix cas d'oppositions existants les plus proches, Il nous faut calculer le pourcentage. Pour cela on utilise la décision qui a été prise pour chaque cas d'oppositions existants que l'on a trouvé.

En effet, pour une recherche simple, on aura dix cas d'oppositions en tout, dont on récupère la décision au moment du traitement, et qui aura une valeur en fonction de la décision prise, que l'on additionnera avec les autres : si la décision du cas a été la justification de l'opposition, c'est qu'il y a bien eu un risque, on ajoutera 1 à la somme ; si l'opposition a été rejetée on n'ajoutera rien car il n'y a pas eu de risques ; et quand la demande d'enregistrement a été partiellement acceptée, on ajoutera 0,5 car il y a quand même eu un certain risque mais la demande a été acceptée sous conditions. On obtient donc un total qui peut valoir au maximum 10, auquel on multiplie 10 pour avoir un résultat sur 100.

Par exemple, si l'on a :

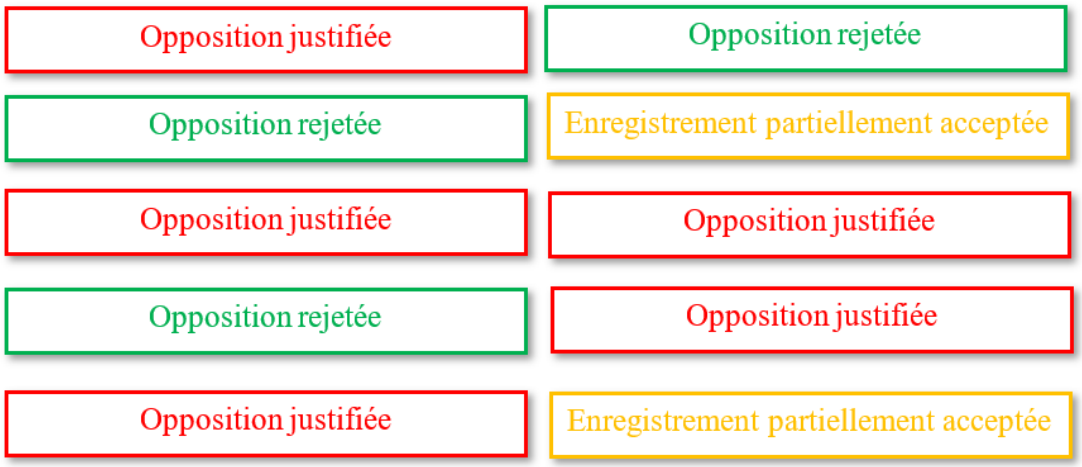

Figure 6 - Exemple de calcul du pourcentage, dans le cas d'une recherche simple

Alors le pourcentage de risque de confusion avec une marque existante sera de 60%.

Pour une recherche globale, étant donné que nous aurons cent cas en tout, dix pour chacune des dix marques les plus proches, nous n'aurons donc pas besoin de multiplié le résultat par 10. Sinon, nous effectuons le même calcul que pour la recherche simple, en faisant la somme des valeurs en fonction des décisions. A noter que, si l'une des marques les plus proches obtient au moins 7/10 à la somme des valeurs en fonctions des décisions, on estimera qu'il existe forcément un risque, et le pourcentage de risque de confusion sera donc de minimum 70%, et ce peu importe le reste des cas existants.

Chapitre 5. Le modèle informatique

Pour encadrer le projet, qui a été réalisé en langage de programmation Python, nous avons utilisé le Framework souvent associé à ce langage, qui s'appelle Django. Un Framework est un cadre de développement qui sert à créer les fondations ainsi que les grandes lignes de tout ou d'une partie d'un logiciel. Il permet de faciliter la conception et la mise en place d'une applications web, en fournissant le moyen de naviguer entre des pages, et une interface administrateur. Cette interface va nous être utile pour notre solution, car c'est dans celle-ci que l'on va retrouver les résultats enregistrés lors des recherches des utilisateurs.

Django nous permet aussi de créer des modèles, qui sont les sources d'informations à propos de nos données, il contient les champs et les types de données que l'on souhaite stocker. Un modèle créé ainsi un patron que l'on peut ensuite appeler dans le traitement, pour créer un objet de ce type, en remplissant les champs qui leurs sont associés. Cela nous sera particulièrement utile pour séparer les résultats des recherches simples et les résultats des recherches globales.

Mais regardons d'abord comment le programme est structuré, notamment la distinction entre interface client et interface administrateur. En effet, il faut bien distinguer le côté client, qui est celui sur lequel arriverons les utilisateurs lambdas, qui souhaitent obtenir des informations concernant le risque de dépôt de la marque qu'il aimerait voir naître, et le côté administrateur, qui est celui où seront stockées les données des clients, et auquel on ne pourra bien sûr accéder seulement avec un identifiant et un mot de passe. Un organigramme des différentes pages et leurs liaisons se trouve en Annexe 3 (p.51).

1. Interfaces client

On retrouvera dans ces interfaces les deux recherches possibles, la simple et la globale, avec un lien en haut à gauche de chacune de ces deux pages pour passer de l'une à l'autre, ainsi qu'un lien pour se connecter à l'interface administrateur en haut à droite.

1.1. Recherche simple :

L'interface de recherche simple côté client se présente de cette manière (A noter que c'est la page index du programme, c'est donc la première page sur laquelle on atterrit lorsque qu'on ne rajoute rien dans l'URL)

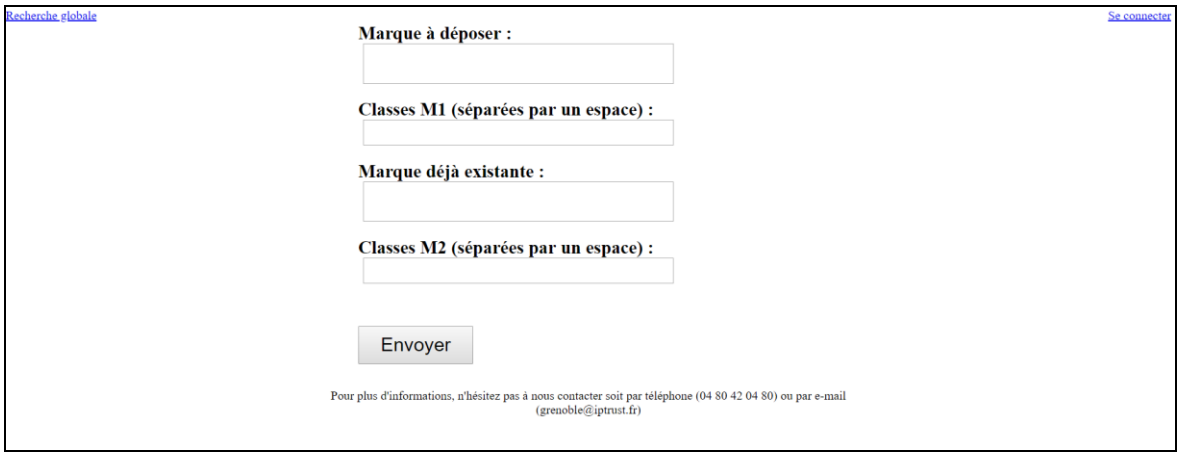

Figure 7 - Index recherche simple côté client

Lorsque le client remplit les zones de textes, et clique sur envoyer, le traitement se lance. Comme dit plus haut, la recherche simple n'ayant pas besoin de parcourir le fichier des données de marques, le traitement est quasi-instantanée et nous emmène donc sur la page affichant les résultats. Le client peut ne pas rentrer de classes, mais doit obligatoirement rentrer deux marques.

Figure 8 - Résultats recherche simple côté client

On obtient donc le pourcentage de risque de confusion, qui sera inscrit en vert s'il est inférieur à 50%, en orange s'il est égal à 50%, et enfin en rouge s'il est supérieur à 50%.

On remarque qu'ici, il n'y a pas beaucoup d'informations qui sont affichées, car en effet le but un peu plus marketing de notre solution, est que le client voulant en savoir plus sur le pourcentage de risque de confusion qui lui est fourni, rentre ensuite en contact avec la société IPTrust afin de voir en détails avec le juriste (qui lui aura toutes les informations nécessaires côté administrateur) s'il y a bien risque ou non, pourquoi, et comment essayer d'y remédier.

1.2. Recherche globale :

L'interface de recherche globale côté client se présente de cette manière

Figure 9 - Index recherche globale côté client

Ici, le client a donc juste besoin de remplir les champs appropriés avec la marque qu'il souhaite déposer, et les classes associées à celle-ci (comme pour la recherche simple, l'utilisateur peut ne pas rentrer de classes). Encore une fois contrairement aux recherches simples, les recherches globales nécessitent le parcours des données de marques, et sont donc beaucoup plus longues au niveau du traitement.

Ainsi, pour faire patienter le client lors du traitement, nous avons mis en place une page de chargement, avec un lien menant à LABLOCK[15].

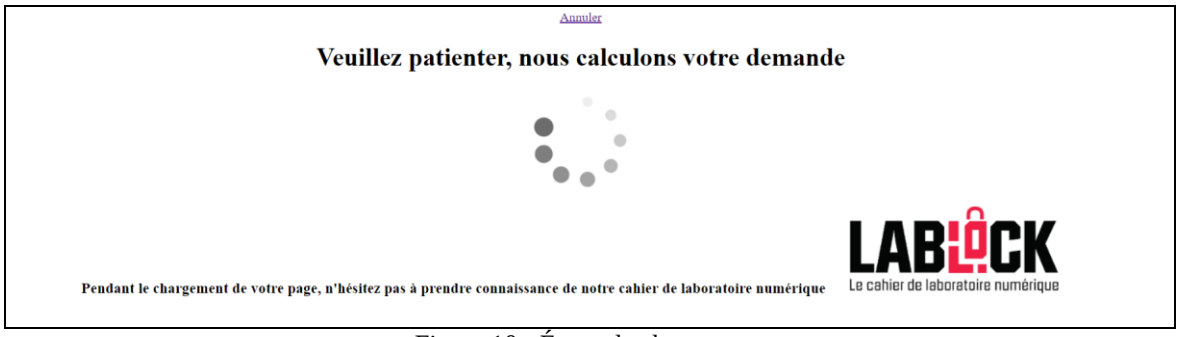

Figure 10 - Écran de chargement

[15] Voir Introduction, p.7

Une fois la recherche terminée, le client arrive sur cette page

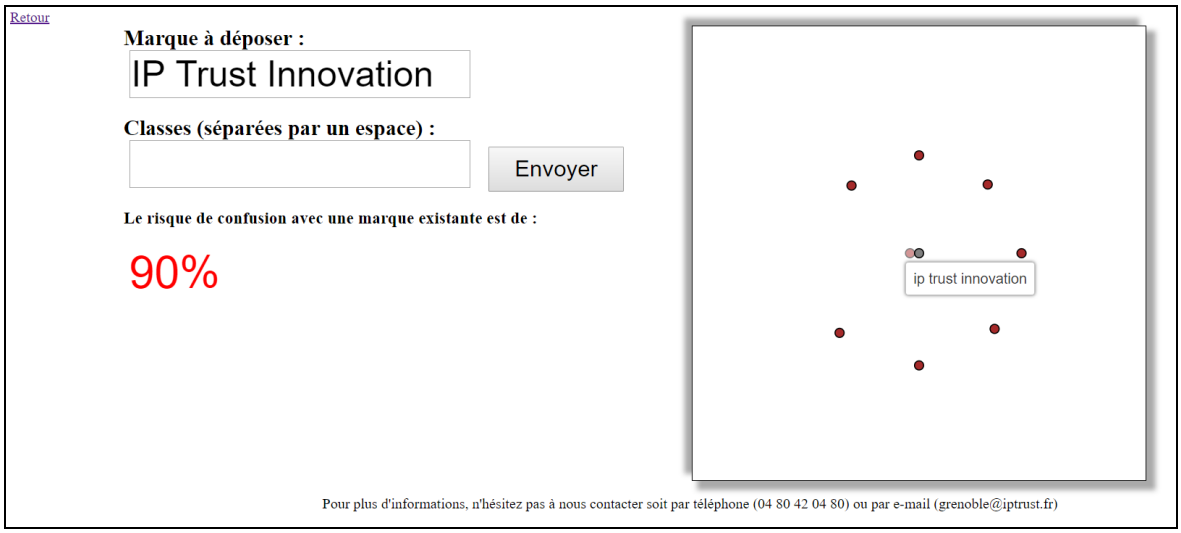

Figure 11- Résultats recherche globale côté client

Ici, en plus du pourcentage de risque de confusion, on obtient aussi un plan avec des points qui représente les marques les plus proches calculées lors du traitement par rapport au point central qui lui est notre marque à déposer, afin de les renseigner au client. Nous reviendrons dans le chapitre 6 (p.30) sur la manière dont on les affiche, et la raison pour laquelle il y en a huit et pas dix.

Voilà pour les interfaces clients, sur lesquels les résultats des recherches sont donc enregistrés en fonction du modèle utilisé et donc du type d'objet créé (recherche simple ou globale). Pour qu'ensuite le juriste ait toutes les informations nécessaires, on passe à l'interface administrateur.

2. *Interfaces administrateur*

La page sur laquelle nous arrivons après nous être connecté comporte plusieurs catégories différentes, avec une mise en page fournie par le Framework Django.

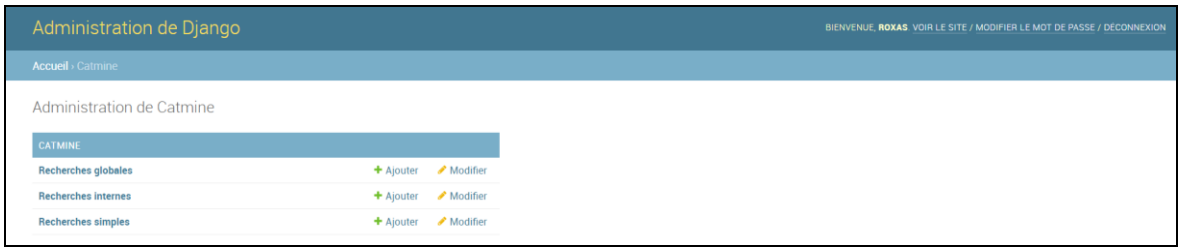

Figure 12 - Interface administrateur après identification

2.1. Recherche simple :

Cette catégorie regroupera toutes les recherches simples faites pas les utilisateurs côté client. Chaque entrée est identifiée par :

- Les deux marques rentrées par l'utilisateur
- La date à laquelle la recherche a été effectuée
- Le pourcentage résultant du traitement

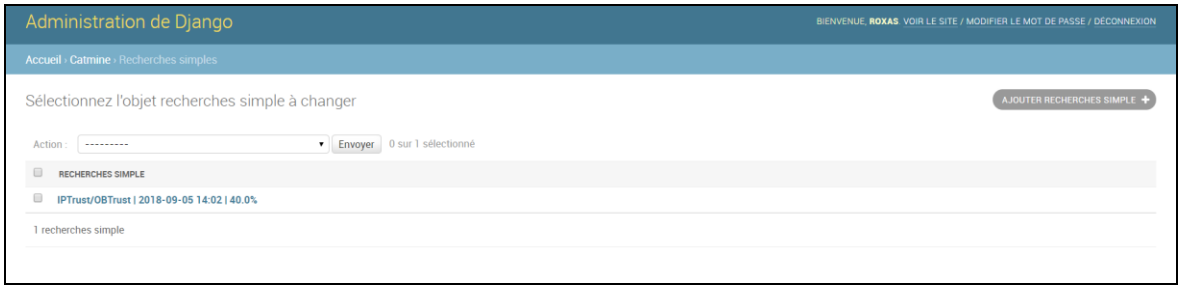

Figure 13 - Interface administrateur recherches simples

Le Framework nous permet par ailleurs de supprimer les données sélectionnées si on le souhaite.

2.2. Recherche globale :

Cette catégorie regroupera toutes les recherches globales faites pas les utilisateurs côté client. Chaque entrée est identifiée par :

- La marque rentrée par l'utilisateur
- La date à laquelle la recherche a été effectuée
- Le pourcentage résultant du traitement

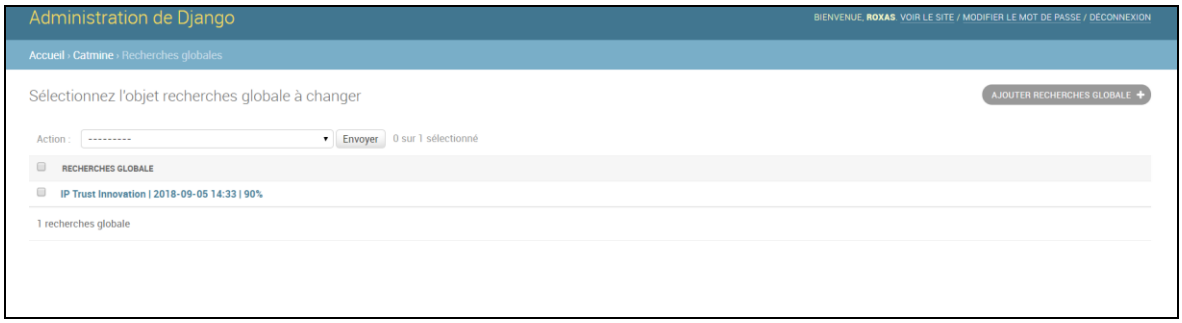

Figure 14 - Interface administrateur recherches globales

Le Framework nous permet par ailleurs de supprimer les données sélectionnées si on le souhaite.

2.3. Recherche internes :

En cliquant sur cette catégorie, l'utilisateur enregistré est emmené sur une page identique à l'interface client des recherches simples, à quelques détails près. Tout d'abord, même si le lien en haut à gauche sert toujours à passer aux recherches globales (mais dans les recherches internes dans ce cas), le lien en haut à droite lui, est cette fois un lien pour retourner sur l'interface administrateur, car cette page de recherche interne n'étant accessible que si l'on est enregistrés, il n'est pas nécessaire de se reconnecter.

Le deuxième détail qui a quand même une grande importance c'est que sur ces interfaces de recherche internes, les résultats des recherches ne sont pas enregistrés. En effet, elles ont été mises en place pour permettre aux administrateurs chez IPTrust, d'effectuer les recherches qu'ils souhaitent, sans « polluer » les données enregistrées par les utilisateurs clients. Sans ça, il y aurait beaucoup plus de données qui n'ont pas à voir avec les clients. A part cela, ce sont les mêmes interfaces que pour les utilisateurs clients.

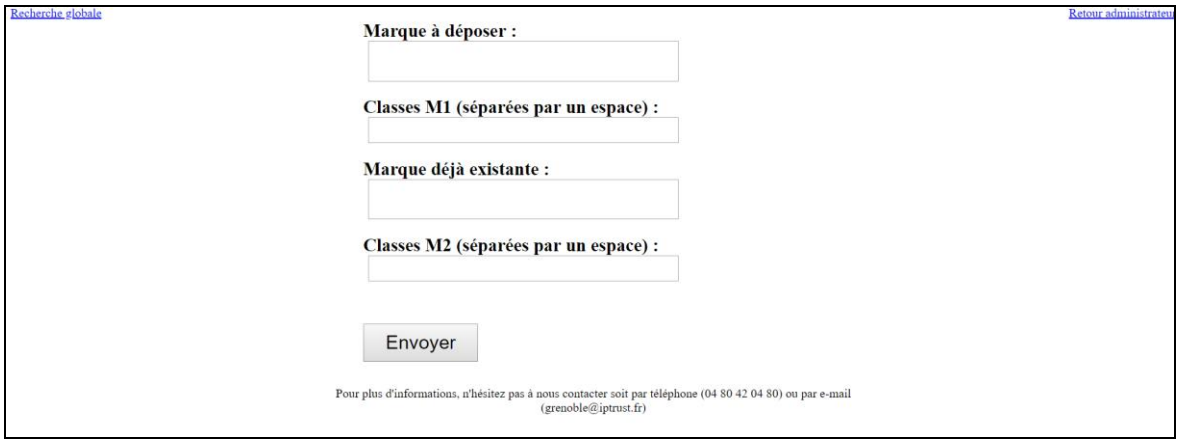

Figure 15 - Index recherche simple en interne

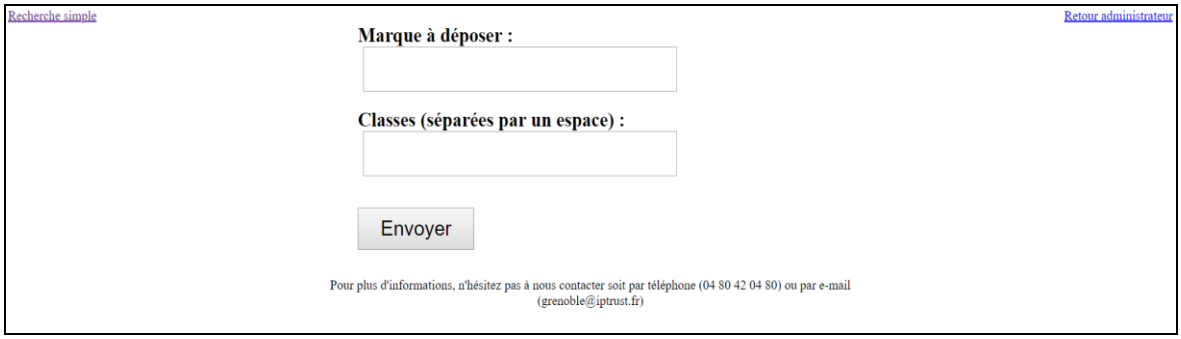

Figure 16 - Index recherche globale en interne

Nous allons maintenant voir comment sont affichés ces résultats, surtout pour les recherches globales, pour lesquelles il faut afficher non seulement les marques les plus proches mais aussi les cas d'opposition existants pour chacune de ces marques les plus proches. Pour les recherches simples, il faudra juste afficher les cas d'oppositions existants les plus proches, car il n'y a pas de recherche de marques les plus proches.

Chapitre 6. L'affichage des résultats

Comme on a pu en partie voir dans la figure 8 pour l'affichage du résultat de la recherche simple côté client, les résultats complets comprennent non seulement l'affichage du pourcentage mais aussi un plan avec des points.

A noter que dans le chapitre qui suit, nous partirons du principe que ce qui est affiché dans les résultats enregistrées côté administrateur, sera aussi affiché dans les résultats des recherches internes, les deux étant du côté de l'administrateur et nécessitant donc d'afficher toutes les informations pour celui-ci. Ces plans sont réalisés grâce au langage SVG[16], qui permet de créer dessins vectoriels, avec des formes que l'on choisit et que l'on place dans un espace donné. Ici, le carré ayant une dimension de 20*20, le centre se trouve au point x=10 et y=10.

Pour créer ce plan, on ajoute donc une balise <svg> dans le fichier HTML[17] où l'on souhaite l'afficher, puis à l'intérieur de cet élément svg, on ajoute des balises <circle> qui seront des points qui correspondront donc aux marques les plus proches, ou au cas les plus proches. L'objectif est donc de donner les bonnes coordonnées à chacun de ces cercles, pour qu'ils se placent à une distance représentative de leur proximité linguistique avec le centre, par rapport à cet axe :

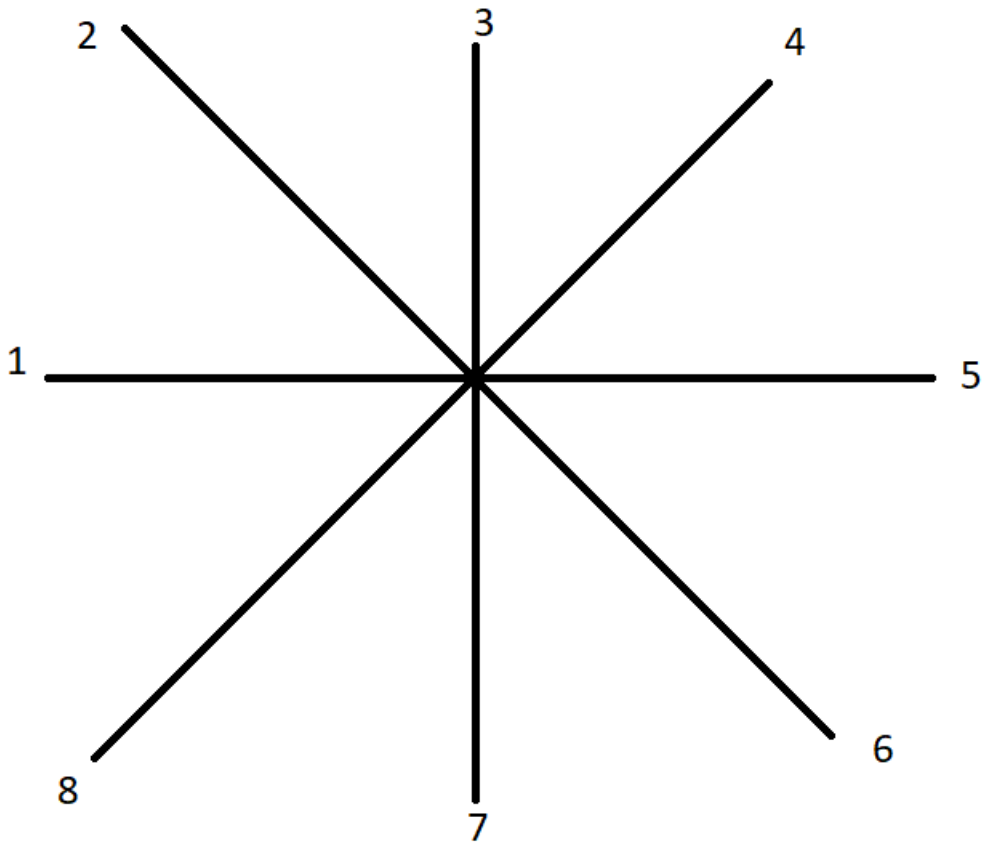

Figure 17 - Axe de positionnement des points, le centre étant notre cas fictif

[16] Scalable Vector Graphics
[17] Langage de programmation permettant de mettre en forme une page web

On aura donc le point au milieu qui sera soit notre marque à déposer dans le cas de l'affichage des marques les plus proches, soit notre cas fictif dans l'affichage des cas d'oppositions existants les plus proches. Les autres points resteront ainsi sur les axes, et seront plus ou moins proches du centre en fonction de leur proximité linguistique (du plus au moins proche en partant de l'axe horizontal gauche, jusqu'à l'axe diagonal en bas à gauche, dans le sens des aiguilles d'une montre).

Il y a un point par axe, c'est pour cela que pour le calcul nous prenons dix marques, ou cas existants, les plus proches, mais que nous n'en affichons que huit. Pour cela il va donc falloir calculer les coordonnées pour que les différents points restent sur leur axe, et que l'on ne perde pas la distance relative à la proximité linguistique. Parce que, le plan étant un carré, il faut faire en sorte que lorsque tous les points sont à la même distance du centre, mais doivent rester sur leurs axes, cela ne fasse pas un carré (auquel cas les points présents dans les diagonales, seront plus éloignés du centre que les autres). Il faut donc trouver le coefficient qui nous permette de mettre les points en cercle.

Regardons d'abord les deux types de plan différents :

1. *Affichage des cas d'oppositions existants les plus proches*

Le plan pour les cas d'oppositions existants les plus proches de notre cas fictif, qui sera donc affiché dans les résultats des recherches globales côté client, dans les résultats des recherches simples côté client, dans l'interface des résultats des recherches globales enregistrées, côté administrateur, et dans l'interface des résultats des recherches simples enregistrées, côté administrateur également.

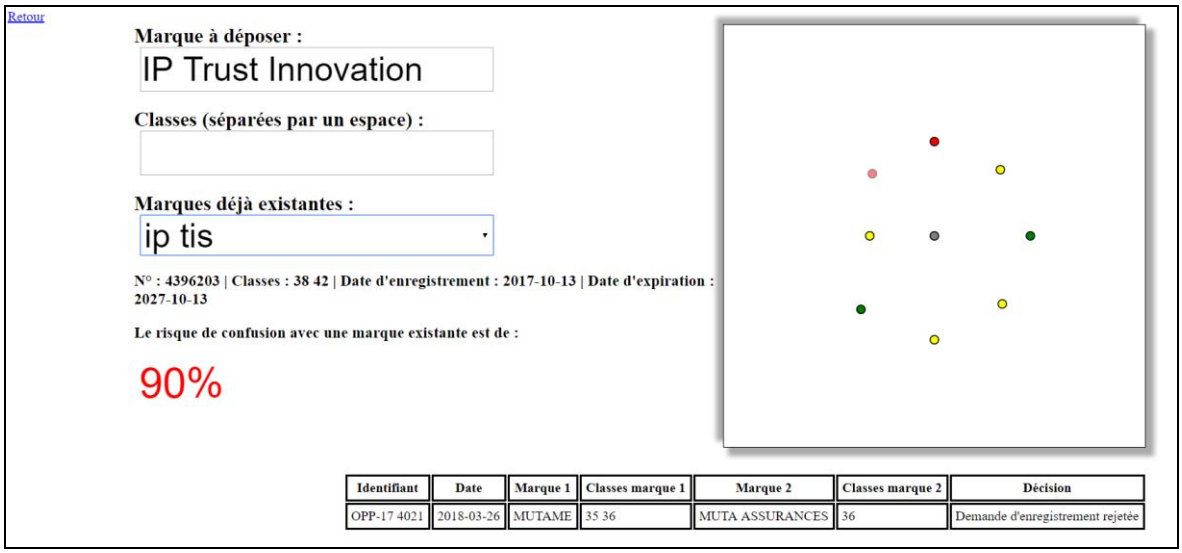

Figure 18 - Résultats recherches globales enregistrées côté administrateur

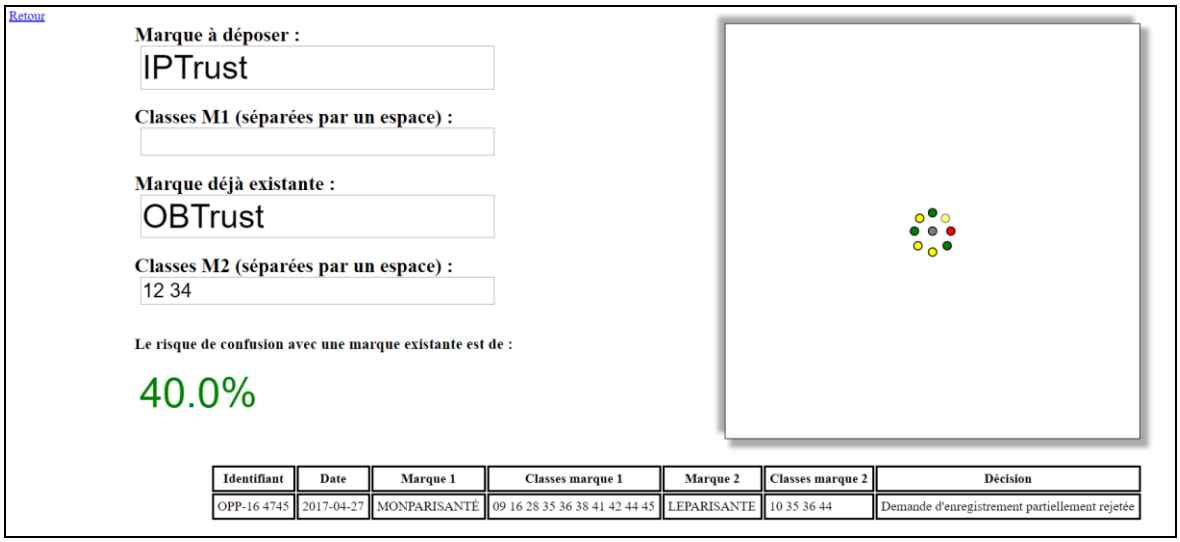

Figure 19 - Résultats recherches simples enregistrées côté administrateur

Lorsque l'on passe la souris sur un des cas, le tableau se trouvant en dessous est actualisé avec les informations relatives à ce cas. Et lorsque l'on clique dessus, un nouvel onglet s'ouvre sur le fichier PDF associé au cas existant, qui se trouve dans nos données. L'administrateur peut ainsi le consulter et le télécharger s'il le souhaite. Les couleurs des points quant à elles, permettent de facilement identifier la décision du cas en question (rouge si l'opposition a été justifiée, vert si l'opposition a été rejetée, et jaune si la demande d'enregistrement a été partiellement rejetée).
Dans le cas des recherches globales, la liste déroulante permet de circuler entre les différentes marques les plus proches et permet d'actualiser la map svg avec les cas existants proches associés à chaque marque proche, en fonction de la sélection. On affiche également les informations liées à la marque sélectionnée.

2. *Affichage des marques les plus proches*

Le plan pour les marques les plus proches de notre marque à déposer, qui sera donc affiché dans les résultats des recherches globales côté client, ainsi que dans l'interface des résultats des recherches globales enregistrées, accessibles côté administrateur (Et donc dans les résultats des recherches globales internes). Cet affichage n'apparaitra donc que dans les recherches globales, car encore une fois il n'y a pas de calcul de marques les plus proches dans les recherches simples.

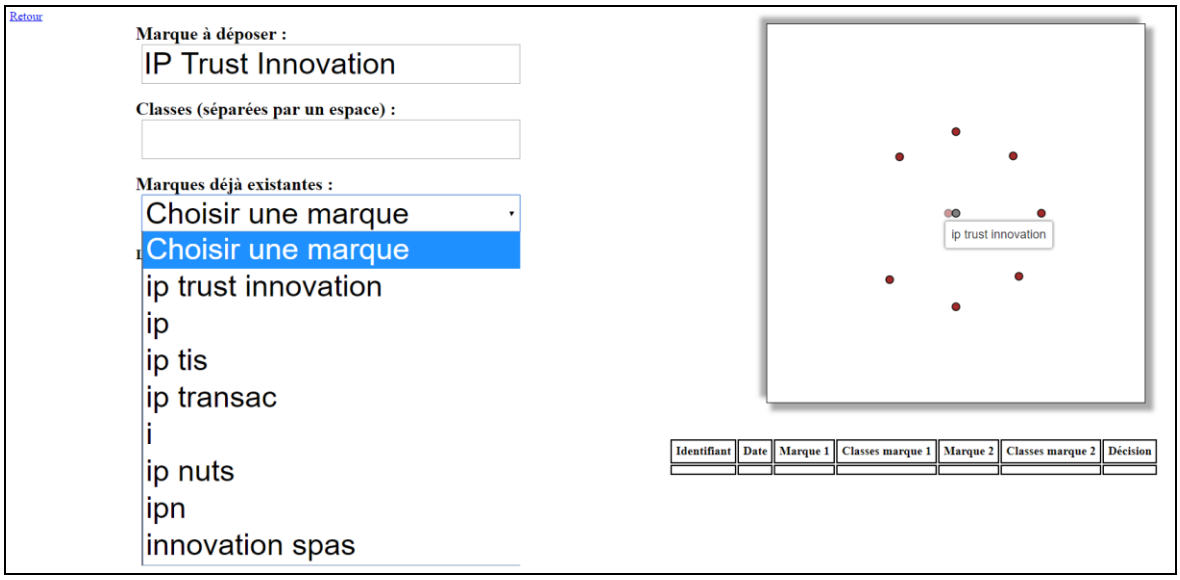

Figure 20 - Affichage des marques les plus proches

Lorsque l'on passe la souris sur une des marques du plan, le nom de cette marque s'affiche à côté du curseur. Si l'option par défaut « Choisir une marque » de la liste déroulante est sélectionnée, cela affiche le plan des marques les plus proches.

3. Calculs des coordonnées

3.1. Coordonnées pour les cas d'oppositions existants les plus proches

Maintenant regardons comment l'on calcule ces coordonnées. Nous prendrons comme exemple la recherche globale de la marque « IP Trust Innovation » avec la marque proche sélectionnée « Ip Tis », comme dans la figure 18 (p.32).

Commençons par le calcul et l'affichage des cas existants les plus proches de ce cas fictif. Tout d'abord, il nous faut les distances des cas existants les plus proches, calculés entre le cas fictif « IP Trust Innovation | Ip Tis » et les cas d'oppositions existants. On obtient donc les valeurs suivantes :

- 0,15351381314928292
- 0,21053545301808807
- 0,22368847483025203
- 0,22403331037623103
- 0,2280436280008136
- 0,23052781801894423
- 0,24639966622549236
- 0,24915719803414837
- 0,25130671953840517
- 0,2565975427819415

On ne garde donc que les huit premières comme expliqué plus haut, et on les multiplie par 20 pour pouvoir plus facilement les comparer au centre du plan qui est 10. On obtient donc :

- 3,070276262985658
- 4,21070906036176
- 4,47376949660504
- 4,48066620752462
- 4,560872560016272
- 4,610556360378884
- 4,927993324509846
- 4,983143960682966

On applique ensuite ces distances à ce tableau, qui rassemble les calculs pour positionner le point en fonction de l'axe sur lequel il doit se trouver :

POSITION	Valeur de X	Valeur de Y
1er cas le plus proche	10-distance	10
2ème cas le plus proche	10-(distance*0,7)	10-(distance*0,7)
3ème cas le plus proche	10	10-distance
4ème cas le plus proche	10+(distance*0.7)	10-(distance*0,7)
5ème cas le plus proche	10+distance	10
6ème cas le plus proche	10+(distance*0.7)	10+(distance*0.7)
7ème cas le plus proche	10	10+distance
8ème cas le plus proche	10-(distance*0,7)	10+(distance*0.7)

Comme on peut le voir, on alterne entre soustraire et additionner au point du milieu 10, afin de graviter autour de ce point central, tout en gardant la distance calculée. La multiplication par 0,7 est le coefficient dont on a parlé plus tôt, qui nous permet de garder les proportions adéquates.

En utilisant nos distances, on obtient donc ces coordonnées :

POSITION	Valeur de X	Valeur de Y
1er cas le plus proche	6,929723737014342	10
2ème cas le plus proche	7,052503657746767	7,052503657746767
3ème cas le plus proche	10	5,52623050339496
4ème cas le plus proche	13,136466345267234	6,863533654732766
5ème cas le plus proche	14,560872560016271	10
6ème cas le plus proche	13,22738945226522	13,22738945226522
7ème cas le plus proche	10	14,92799332450985
8ème cas le plus proche	6,511799227521923	13,488200772478077

Qui nous donne ce plan ci, que l'on retrouve bien dans la figure 18 (sans les couleurs des décisions ici) :

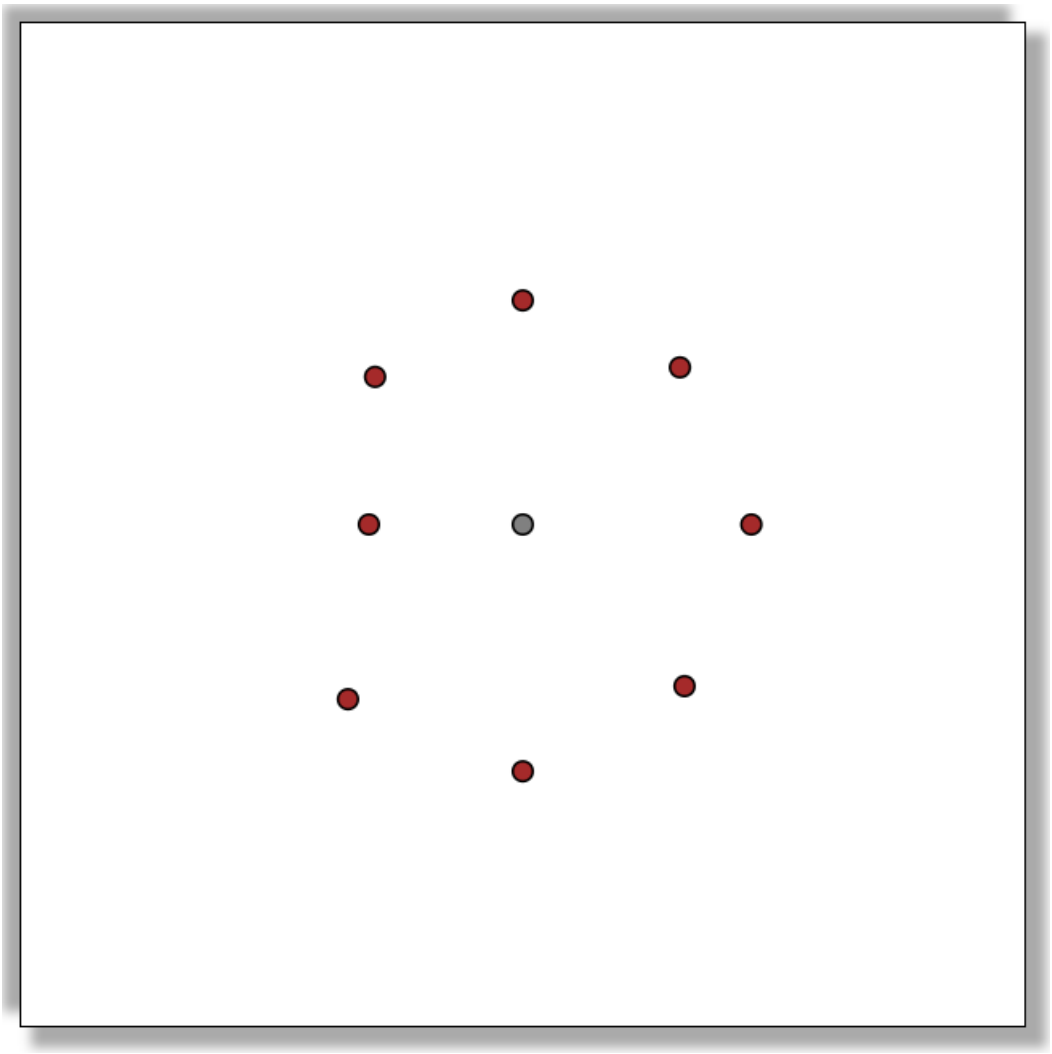

Figure 21 - Calcul des cas d'oppositions existants proches du cas fictif « IP Trust Innovation | Ip Tis » avec le coefficient

Pour appuyer sur le fait que le coefficient est nécessaire pour ne pas perdre de données concernant la proximité pour les points en diagonales, essayons de faire les mêmes calculs, sans ce coefficient. On obtient les coordonnées suivantes :

POSITION	Valeur de X	Valeur de Y
1er cas le plus proche	6,929723737014342	10
2ème cas le plus proche	5,78929093963824	5,78929093963824
3ème cas le plus proche	10	5,52623050339496
4ème cas le plus proche	14,48066620752462	5,51933379247538
5ème cas le plus proche	14,560872560016271	10
6ème cas le plus proche	14,61055636037888	14,61055636037888
7ème cas le plus proche	10	14,92799332450985
8ème cas le plus proche	5,016856039317034	14,98314396068297

Qui cette fois nous donne ce plan ci :

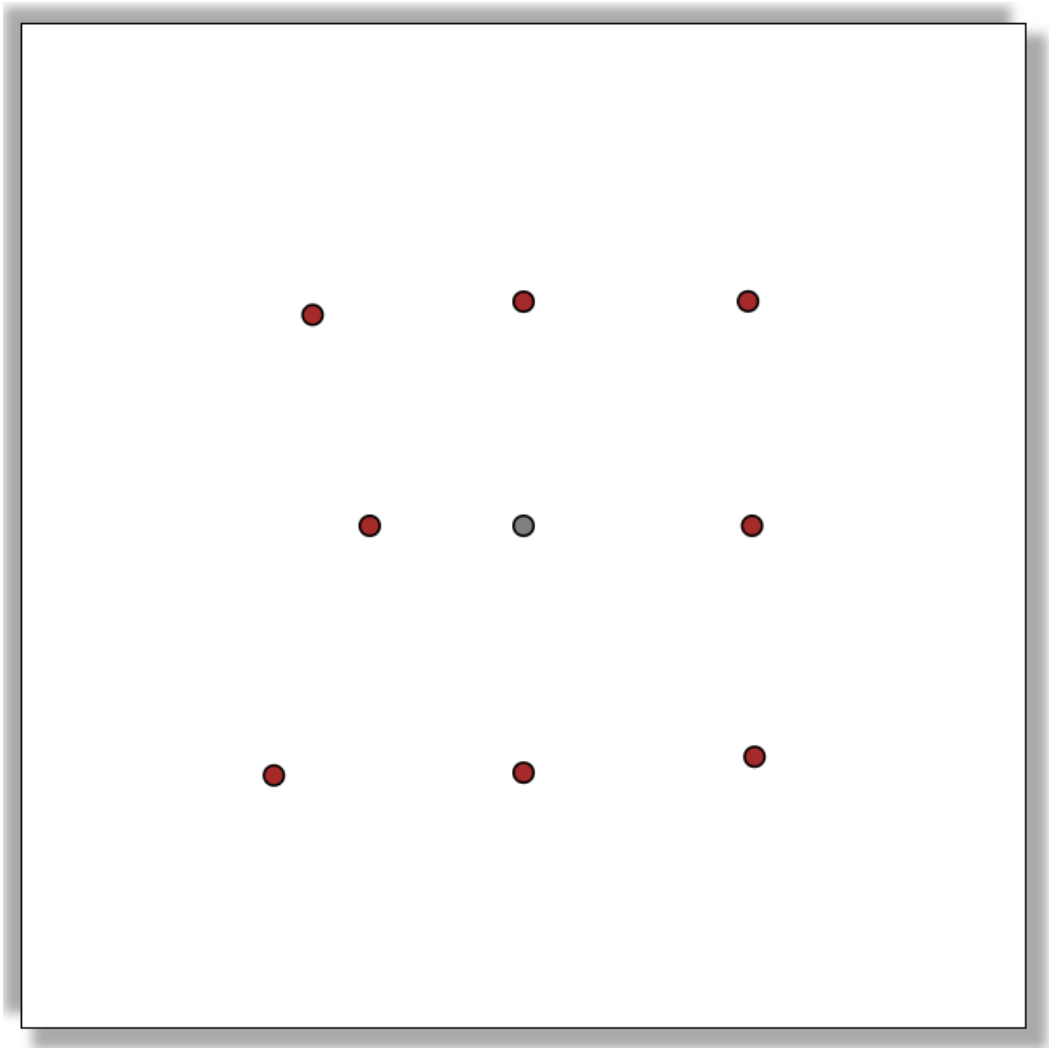

Figure 22 - Calcul des cas d'oppositions existants proches du cas fictif « IP Trust Innovation | Ip Tis » sans le coefficient

On voit bien que les positions des points en diagonales sont erronées, car ne représente pas la distance par rapport au point central.

3.2. Coordonnées pour les marques les plus proches

Pour l'affichage des marques les plus proches calculées, on utilisera la même technique, ainsi que les mêmes calculs, à un détail prêt. En effet, si la distance calculée pour les cas existants tend vers 0 pour les cas les plus proches, la distance calculée pour les marques elle, tend vers 10 pour les marques les plus proches.

Donc, là où l'on a multiplié la distance par 20 pour les cas les plus proches, on soustraira notre distance pour les marques à 10, afin de savoir la différence par rapport au point du milieu 10, et avant de faire les mêmes calculs. Si l'on garde le même exemple, on

prend donc les distance des marques les plus proches et le nom de ces marques, et l'on obtient :

- 10 | ip trust innovation
- 5,7368421052631575 | ip
- 5,684210526315789 | ip tis
- 5,668421052631579 | ip transac
- 5,473684210526316 | i
- 5,233082706766917 | ip nuts
- 5,052631578947368 | ipn
- 4,982456140350878 | innovation spas
- 4,947368421052632 | ipo
- 4,909090909090909 | innovations

Comme tout à l'heure on ne garde que les huit premières, et on les soustrait donc à 10 pour obtenir :

- 0 | ip trust innovation
- 4,2631578947368425 | ip
- 4,315789473684211 | ip tis
- 4,331578947368421 | ip transac
- 4,526315789473684 | i
- 4,766917293233083 | ip nutes
- 4,947368421052632 | ipn
- 5,017543859649122 | innovation spas

Puis on utilise le même tableau de calcul :

POSITION	NOM DE LA MARQUE	Valeur de X	Valeur de Y
1ere marque la plus proche	ip trust innovation	10-distance = 10	10
2ème marque la plus proche	ip	10-(distance*0,7) = 7,015789473684211	10-(distance*0,7) = 7,015789473684211
3ème marque la plus proche	ip tis	10	10-distance = 5,684210526315789
4ème marque la plus proche	ip transac	10+(distance*0.7) = 13,0321052631579	10-(distance*0,7) = 6,967894736842105
5ème marque la plus proche	i	10+distance = 14,52631578947368	10
6ème marque la plus proche	ip nuts	10+(distance*0.7) = 13,33684210526316	10+(distance*0.7) = 13,33684210526316
7ème marque la plus proche	ipn	10	10+distance = 14,94736842105263
8ème marque la plus proche	innovation spas	10-(distance*0,7) = 6,487719298245615	10+(distance*0.7) = 13,51228070175439

Ce qui nous donne le plan suivant :

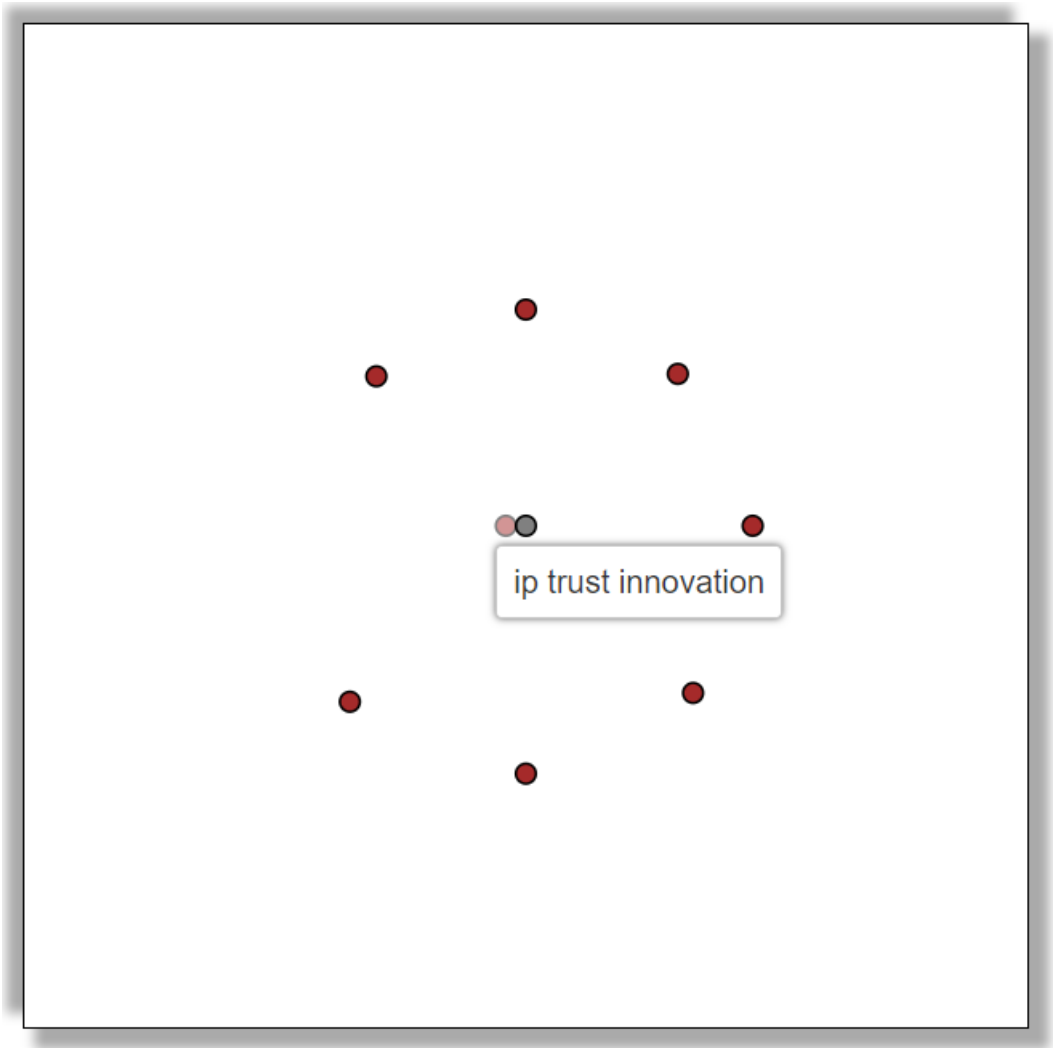

Figure 23 – Affichage des marques les plus proches après calcul des coordonnées

On remarque que la marque la proche est exactement la même que notre marque à déposer, à savoir « ip trust innovation ». Un problème se pose alors : si une marque est identique à notre marque à déposer, alors ses coordonnées seront X = 10 et Y = 10, comme le point central. Les deux points seraient donc superposés, et compliqueraient la compréhension du plan. Pour remédier à cela, lors du calcul de ces coordonnées, on précise que si la valeur de X ou la valeur de Y sont égales à 10 (sauf pour celles où les valeurs sont forcément 10, pour être sur le bon axe, et qui ne passe pas par un calcul), on passe cette valeur à 9,6.

C'est pour cela que sur la figure juste au-dessus, on peut voir que le premier point (celui sur l'axe horizontal gauche) est contre le point central, pour ainsi éviter la superposition de ces deux points, et de pouvoir quand même garder la compréhension que cette marque est extrêmement proche de notre marque à déposer.

Chapitre 7. Résultats obtenus

1. Exemples de résultats

Maintenant que nous avons toutes les clés en main, essayons de voir des exemples de résultats. Imaginons par exemple que nous voulions tenter de déposer la marque « BARRIER » dans la classe 20, dont l'intitulé est « Meubles ; glaces (miroirs) ; cadres (encadrements) ».

Voilà ce qu'on obtient :

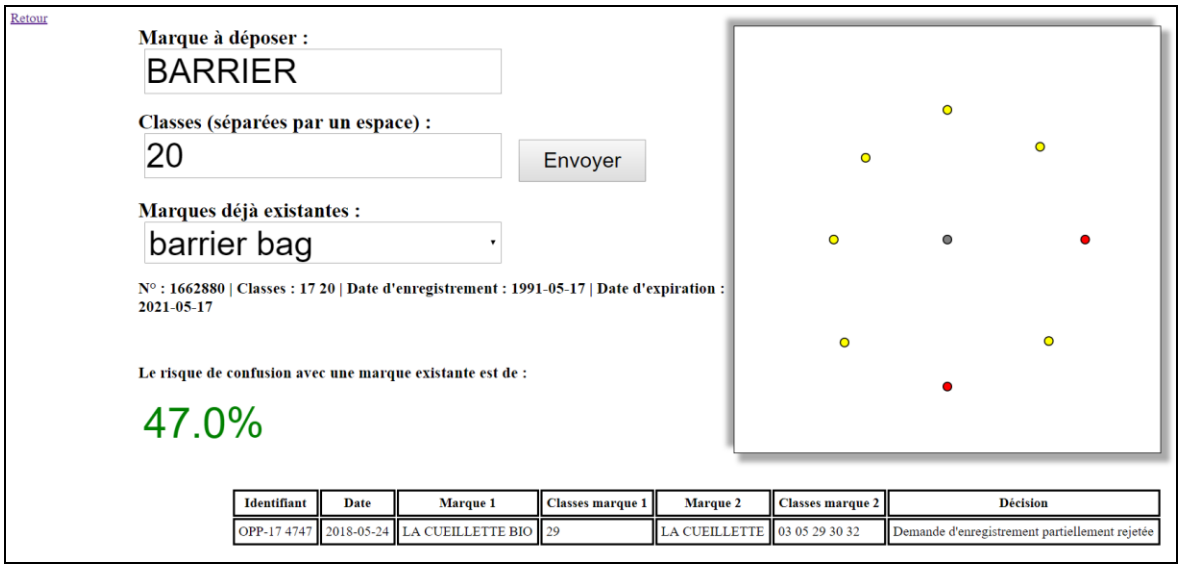

Figure 24 - Recherche globale de la marque à déposer « BARRIER »

On peut voir que la marque la plus proche est « barrier bag » enregistrée dans les classes 17 et 20, Mais que les cas d'oppositions existants les plus proches nous donne un total de 47% de risque, ce qui ne correspond pas à un grand risque de confusion. Il y a donc une bonne chance que notre marque soit déposée sans problèmes, surtout si l'on précise bien le libellé associé à la classe.

Essayons maintenant de déposer la marque « Tom BARRIER » dans cette même classe 20. Voici le résultat que nous obtenons :

Figure 25 - Recherche globale de la marque à déposer « Tom BARRIER »

On peut donc voir que cette fois le pourcentage de risque de confusion est de 70%, ce qui rend le dépôt de cette marque risqué, alors que nous n'avons fait que rajouté « Tom » (car en effet, beaucoup de marques contiennent le prénom Tom, et représentent donc un risque).

Mais il faut savoir que notre solution n'est pas un programme décisionnel, mais une aide à la prise de décision. En ce sens, il ne remplacera jamais l'intervention et l'avis du juriste, mais sera là pour l'aider à accéder plus facilement aux données qui pourrait poser un problème au dépôt de la marque. Là où avant, le juriste devait rechercher minutieusement les marques proches et étudier le cas qui lui était donné en partant de rien, notre solution lui propose un accès plus rapide aux données qui pourrait l'intéresser et l'aiguiller dans son jugement, et ainsi renseigner de manière plus rapide le client.

2. *Une classe peut tout changer*

Essayons maintenant de démontrer le fait de prendre les classes en compte est important, et que cela peut vraiment changer le pourcentage de risque de confusion, d'où l'intérêt de bien préciser les classes, pour avoir un résultat plus précis. Reprenons l'exemple de la marque à déposer « IP Trust innovation ». On se souvient que le résultat était de 90% car la marque la plus proche était exactement la même, et nous n'avions pas précisé de classes.

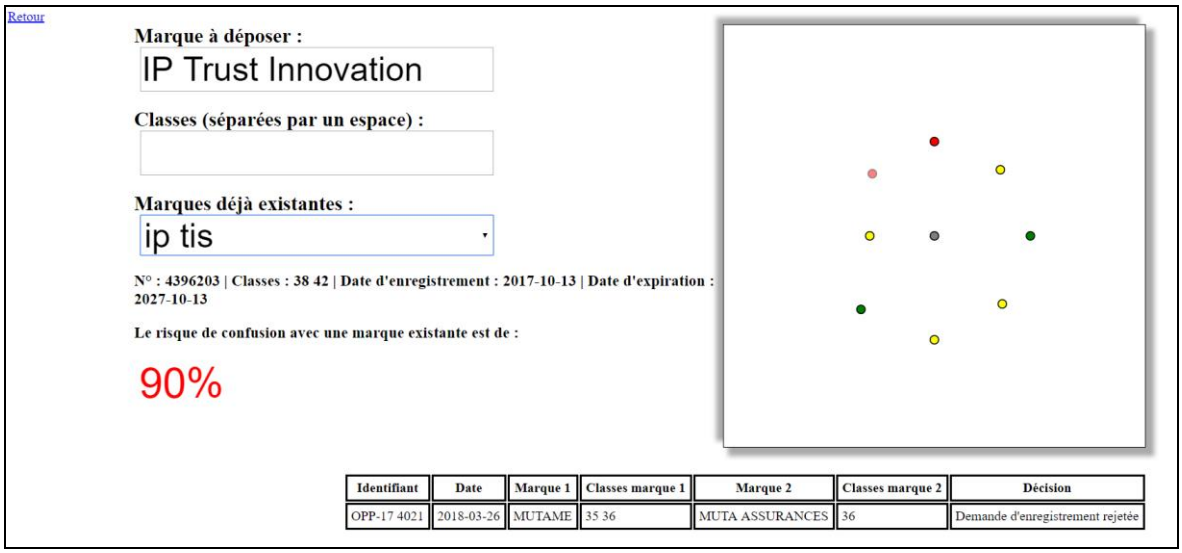

Figure 26 - Résultat recherche globale IP Trust Innovation sans classe

Maintenant réessayons mais en précisant les classes. Imaginons que l'on souhaite déposer cette marque dans la classe 42, qui s'intitule « Evaluations techniques concernant la conception (travaux d'ingénieurs) ; recherches scientifiques ; recherches techniques ». Voici ce qu'on obtient comme résultat :

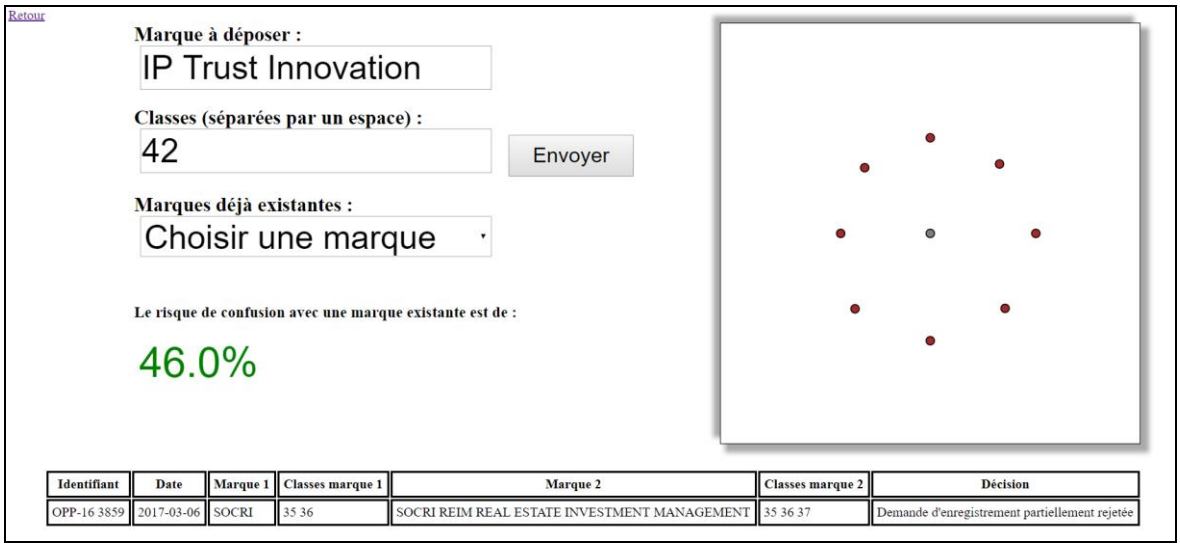

Figure 27 - Résultat recherche globale IP Trust Innovation avec classe

On remarque que le pourcentage de risque à 46%, ce qui est 44% de risque en moins que lorsque l'on n'avait pas précisé la classe associée à notre marque à déposer. Ce qui confirme que le résultat sera toujours plus précis si l'utilisateur rentre aussi des classes associées à leurs marques.

Conclusion

1. Résumé

L'objectif de ce travail était donc de fournir un logiciel d'aide à la décision en droit des marques, afin de faciliter, ne serait-ce qu'un peu le travail des juristes.

Dans la première partie j'ai montré les données brutes que j'ai dû exploiter, et comment j'ai fait pour automatiser la récupération des données nécessaires au bon fonctionnement du logiciel. Nous avons vu les étapes de cette récupération, et les problèmes qu'elle pouvait poser.

Dans la deuxième partie, nous avons pu voir comment était organisé le logiciel, et rentrer dans les détails du modèle linguistique, qui nous renseignait sur la façon dont la proximité était évaluée. Mais aussi les détails du modèle informatique, afin de rentrer dans un aspect plus technique de notre logiciel, et bien comprendre la différenciation entre le côté client et le côté administrateur. Ensuite, nous avons vu comment les résultats étaient affichés dans le plan SVG, ainsi que comment les coordonnées étaient calculées. Enfin, nous nous sommes penchés sur des types de résultats que notre solution pouvait produire.

2. Améliorations possibles

La première amélioration qui me vient à l'esprit, afin d'essayer de répondre au problème du temps de traitement, c'est de stocker les données que nous récupérons (des marques et des cas d'oppositions) afin de ne pas avoir à parcourir le fichier contenant les plus d'1 millions de marques de la même façon. Au-delà de ça, un stockage mieux organisé ne peut être que bénéfique, étant donné la quantité de données.

Une autre amélioration possible serait de laisser la possibilité à l'utilisateur de transcrire phonétiquement les marques avec un dictionnaire, et des règles, basées sur une autre langue. En effet, les transcriptions ne sont pour l'instant que faites par le biais d'un dictionnaire de règles basé sur la langue française. Le résultat pourrait être plus précis si on souhaite déposer une marque mais qu'elle ne se prononce pas à la française, mais à l'anglaise par exemple.

Enfin, une dernière amélioration possible qui pourrait améliorer les résultats donnés, serait de trouver un moyen de prendre en compte les libellés, dans les données, mais aussi en entrée dans les recherches. En effet, ceux-ci étant vraiment présents pour spécifier la marque, il serait intéressant de pouvoir les utiliser, pour avoir d'un côté un résultat un peu large (sans les libellés) et un résultat vraiment très précis sur ce qu'il existe comme marques avec presque le même libellé.

3. Bilan personnel

D'une manière générale, ce stage a été une bonne expérience pour moi. L'accueil de l'équipe aux bureaux était vraiment agréable et c'était plaisant de travailler là-bas. La plus grosse difficulté a été de travailler seul sur un projet assez conséquent je pense. J'ai trouvé cela assez difficile de réussir à tout gérer seul, mais cela a été très formateur, et m'a appris à mieux m'organiser et à travailler en autonomie. De plus, le fait de travailler sur un projet concret, le voir terminé et en être satisfait est assez agréable. Enfin, il est toujours intéressant d'avoir une expérience dans un organisme, afin de se rendre compte de comment peut se passer la vie en entreprise.

Table des illustrations

Figure 1 - Extrait de 6 lignes du fichier contenant nos données de marques 15
Figure 2 - Extrait de 5 lignes du fichier contenant nos données de Jurisprudence..... 15
Figure 3 - Extrait du fichier JurisVecto.data contenant les cas d'opposition existants vectorisés.... 20
Figure 4 – Autre extrait du fichier JurisVecto.data contenant les cas d'opposition existants vectorisés 20
Figure 5 - Données utilisées pour exemple de calcul de proximité entre deux cas d'oppositions.... 22
Figure 6 - Exemple de calcul du pourcentage, dans le cas d'une recherche simple 23
Figure 7 - Index recherche simple côté client 25
Figure 8 - Résultats recherche simple côté client..... 25
Figure 9 - Index recherche globale côté client 26
Figure 10 - Écran de chargement 26
Figure 11- Résultats recherche globale côté client..... 27
Figure 12 - Interface administrateur après identification 27
Figure 13 - Interface administrateur recherches simples..... 28
Figure 14 - Interface administrateur recherches globales 28
Figure 15 - Index recherche simple en interne..... 29
Figure 16 - Index recherche globale en interne..... 29
Figure 17 - Axe de positionnement des points, le centre étant notre cas fictif..... 30
Figure 18 - Résultats recherches globales enregistrées côté administrateur 31
Figure 19 - Résultats recherches simples enregistrées côté administrateur..... 32
Figure 20 - Affichage des marques les plus proches..... 33
Figure 21 - Calcul des cas d'oppositions existants proches du cas fictif « IP Trust Innovation | Ip Tis » avec le coefficient..... 35
Figure 22 - Calcul des cas d'oppositions existants proches du cas fictif « IP Trust Innovation | Ip Tis » sans le coefficient 36
Figure 23 – Affichage des marques les plus proches après calcul des coordonnées..... 38
Figure 24 - Recherche globale de la marque à déposer « BARRIER » 39
Figure 25 - Recherche globale de la marque à déposer « Tom BARRIER »..... 40
Figure 26 - Résultat recherche globale IP Trust Innovation sans classe 41
Figure 27 - Résultat recherche globale IP Trust Innovation avec classe..... 41

Table des annexes

Annexe 1 Entrée de marque déposée extrait d'un des fichiers XML 46
Annexe 2 Entrée de cas d'opposition extrait d'un des fichiers XML 50
Annexe 3 Organigramme des différentes pages du logiciel 51

Annexe 1
Entrée de marque déposée extrait d'un des fichiers XML


```
<TradeMark operationCode="Insert">
   <RegistrationOfficeCode>FR</RegistrationOfficeCode>
   <ApplicationNumber>4475903</ApplicationNumber>
   <ApplicationDate>2018-08-09</ApplicationDate>
   <FilingPlace>92 INPI - Dépôt électronique</FilingPlace>
   <ApplicationLanguageCode>fr</ApplicationLanguageCode>
   <ExpiryDate>2028-08-09</ExpiryDate>
   <MarkCurrentStatusCode>Demande publiée</MarkCurrentStatusCode>
   <MarkFeature>Combined</MarkFeature>
   <MarkDescriptionDetails>
      <MarkDescription>Traduction de la marque : Golf de Roissy Couleur(s)
de la marque : N° pantone vert 60/0/100/0 bleu 76/58/0/0</MarkDescription>
   </MarkDescriptionDetails>
   <WordMarkSpecification>
      <MarkVerbalElementText>Golf de Roissy</MarkVerbalElementText>
   </WordMarkSpecification>
   <MarkImageDetails>
      <MarkImage>
         <MarkImageFilename>FMARK0000000004475903</MarkImageFilename>
         <MarkImageFileFormat>JPEG</MarkImageFileFormat>
         <MarkImageColourIndicator>true</MarkImageColourIndicator>
      </MarkImage>
   </MarkImageDetails>
   <GoodsServicesDetails>
      <GoodsServices>
         <ClassificationKindCode>Nice</ClassificationKindCode>
         <ClassDescriptionDetails>
            <ClassDescription>
               <ClassNumber>16</ClassNumber>
               <GoodsServicesDescription        languageCode="fr">affiches ;
albums ;    cartes  ;    livres ;    journaux ;    prospectus  ;    brochures  ;
calendriers ;  instruments d'écriture ;</GoodsServicesDescription>
            </ClassDescription>
            <ClassDescription>
               <ClassNumber>18</ClassNumber>
               <GoodsServicesDescription     languageCode="fr">malles     et
valises ;      parapluies   et   parasols ;      portefeuilles ;      porte-
monnaie ;</GoodsServicesDescription>
            </ClassDescription>
            <ClassDescription>
               <ClassNumber>25</ClassNumber>
               <GoodsServicesDescription       languageCode="fr">Vêtements ;
gants (habillement) ;</GoodsServicesDescription>
```

```
            </ClassDescription>
            <ClassDescription>
               <ClassNumber>28</ClassNumber>
               <GoodsServicesDescription languageCode="fr">balles et ballons
de  jeux ;  rembourrages  de  protection  (parties  d'habillement  de
sport) ;</GoodsServicesDescription>
            </ClassDescription>
            <ClassDescription>
               <ClassNumber>35</ClassNumber>
               <GoodsServicesDescription  languageCode="fr">Publicité ;
gestion  des  affaires  commerciales ;  diffusion  de  matériel  publicitaire
(tracts, prospectus, imprimés, échantillons) ;  présentation de produits sur
tout  moyen  de  communication  pour  la  vente  au  détail ;  location  de  temps
publicitaire  sur  tout  moyen  de  communication  ;  publication  de  textes
publicitaires ;  location d'espaces publicitaires ;  diffusion d'annonces
publicitaires ;  relations publiques ;</GoodsServicesDescription>
            </ClassDescription>
            <ClassDescription>
               <ClassNumber>41</ClassNumber>
               <GoodsServicesDescription  languageCode="fr">formation ;
divertissement ;  activités  sportives  et  culturelles ;  informations  en
matière  de  divertissement ;  informations  en  matière  d'éducation ;  mise  à
disposition  d'installations  de  loisirs ;  organisation  et  conduite  de
colloques ;  organisation  et  conduite  de  conférences ;  organisation  et
conduite  de  congrès ;  organisation  d'expositions  à  buts  culturels  ou
éducatifs ;</GoodsServicesDescription>
            </ClassDescription>
            <ClassDescription>
               <ClassNumber>43</ClassNumber>
               <GoodsServicesDescription  languageCode="fr">Services  de
restauration  (alimentation) ;  services  de  bars ;  services  de
traiteurs ;</GoodsServicesDescription>
            </ClassDescription>
            <ClassDescription>
               <ClassNumber>45</ClassNumber>
               <GoodsServicesDescription  languageCode="fr">location  de
vêtements.</GoodsServicesDescription>
            </ClassDescription>
         </ClassDescriptionDetails>
      </GoodsServices>
   </GoodsServicesDetails>
   <ApplicantDetails>
      <Applicant>
         <ApplicantIdentifier/>
         <ApplicantSequenceNumber>1</ApplicantSequenceNumber>
         <ApplicantLegalEntity>Collectivité
territoriale</ApplicantLegalEntity>
         <ApplicantAddressBook>
            <FormattedNameAddress>
```

```
                <Name>
                   <FormattedName>
                      <LastName>Mairie de Roissy-en-France</LastName>
                      <IndividualIdentifier>219505278</IndividualIdentifier>
                      <OrganizationName>Mairie          de          Roissy-en-
France</OrganizationName>
                   </FormattedName>
                </Name>
                <Address>
                   <AddressCountryCode>FR</AddressCountryCode>
                   <FormattedAddress>
                      <AddressStreet>40        avenue        Charles        de
Gaulle</AddressStreet>
                      <AddressCity>Roissy-en-France</AddressCity>
                      <AddressPostcode>95700</AddressPostcode>
                   </FormattedAddress>
                </Address>
             </FormattedNameAddress>
          </ApplicantAddressBook>
       </Applicant>
    </ApplicantDetails>
    <RepresentativeDetails>
       <Representative>
          <RepresentativeIdentifier/>
          <RepresentativeSequenceNumber>1</RepresentativeSequenceNumber>
          <RepresentativeAddressBook>
             <FormattedNameAddress>
                <Name>
                   <FormattedName>
                      <LastName>Mairie  de  Roissy-en-France,  Monsieur  André
Toulouse</LastName>
                      <OrganizationName>Mairie de Roissy-en-France, Monsieur
André Toulouse</OrganizationName>
                   </FormattedName>
                </Name>
                <Address>
                   <AddressCountryCode>FR</AddressCountryCode>
                   <FormattedAddress>
                      <AddressStreet>40        Avenue        Charles        de
Gaulle</AddressStreet>
                      <AddressCity>Roissy-en-France</AddressCity>
                      <AddressPostcode>95700</AddressPostcode>
                   </FormattedAddress>
                </Address>
             </FormattedNameAddress>
          </RepresentativeAddressBook>
       </Representative>
    </RepresentativeDetails>
    <MarkRecordDetails>
```

```xml
      <MarkRecord>
         <BasicRecord>
            <BasicRecordKind>Publication</BasicRecordKind>
            <RecordPublicationDetails>
               <RecordPublication>
                  <PublicationIdentifier>2018-36</PublicationIdentifier>
                  <PublicationDate>2018-09-07</PublicationDate>
               </RecordPublication>
            </RecordPublicationDetails>
         </BasicRecord>
      </MarkRecord>
   </MarkRecordDetails>
</TradeMark>
```

Annexe 2
Entrée de cas d'opposition extrait d'un des fichiers XML


```xml
<JURISPRUDENCE REF="O20162965" ETAT="ACTIF">
        <ABS>
            <P>Décision après projet - R 712-16, 3° alinéa 2</P>
        </ABS>
        <DAT>2016-12-23</DAT>
        <DCLAIR>23 décembre 2016</DCLAIR>
        <DOM>OPPOSITION</DOM>
        <MARQ>
            <P>SAILOR JERRY ; BLACK SAILOR</P>
        </MARQ>
        <NAT>Marque communautaire ; Marque française</NAT>
        <NRG>2016-2965</NRG>
        <OBJ>
            <P>6600571 ; 4263848</P>
        </OBJ>
        <PARTAN>William Grant &amp; Sons Irish Brands Limited / SOCIETE AGRICOLE DE BOLOGNE</PARTAN>
        <LOTS><LOT LABEL="O20162965"><DOCUMENT>/Documents/Oppositions1/O20162965/O20162965.pdf</DOCUMENT></LOT></LOTS>
    </JURISPRUDENCE>
```

Annexe 3
Organigramme des différentes pages du logiciel

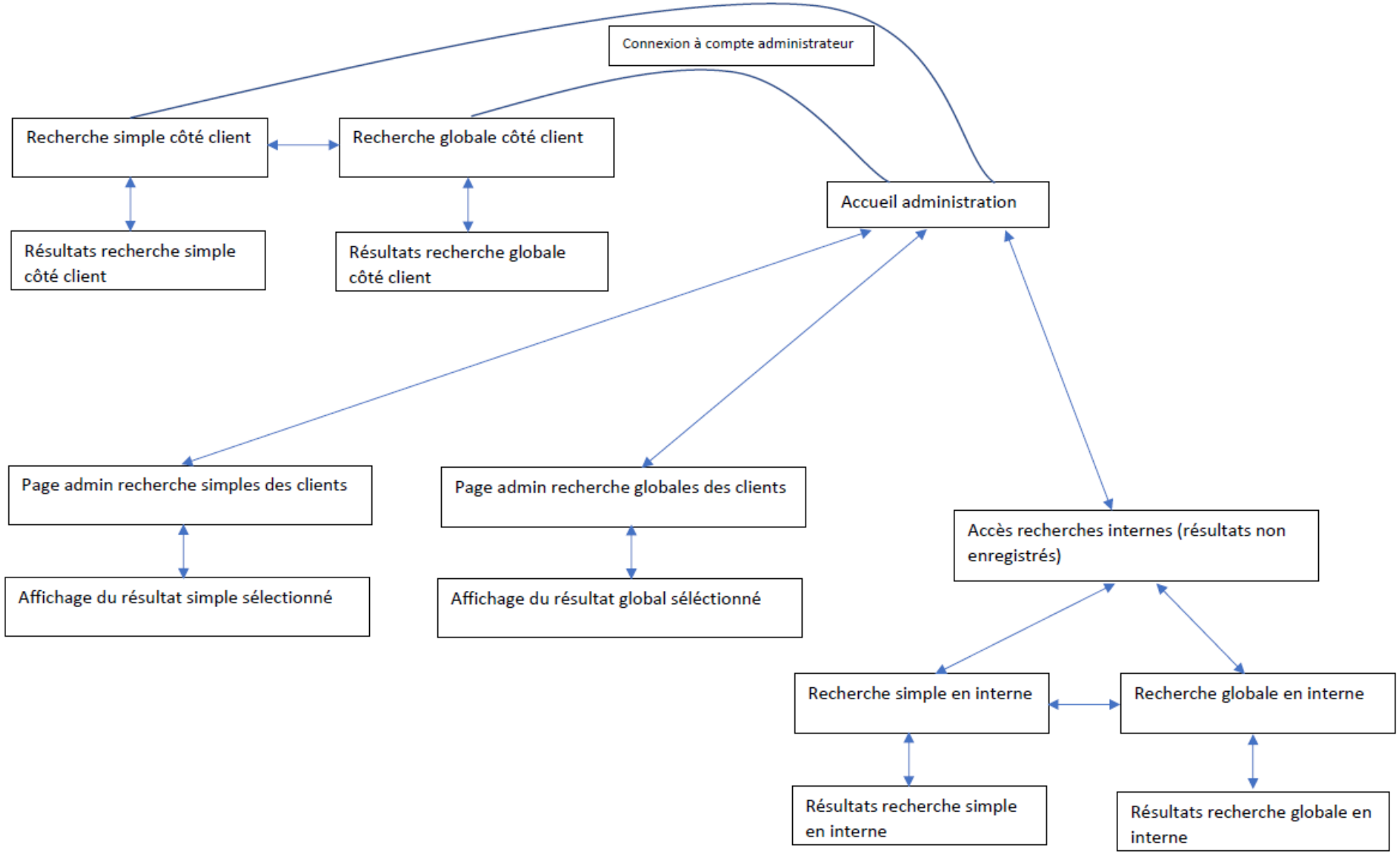

Table des matières

Remerciements 3
Sommaire 5
Introduction 6
1. Le stage 6
1.1. La société IP Trust Innovation 6
1.2. La mission 7
2. Le domaine de droit des marques 7
PARTIE 1 - LES DONNEES 9
CHAPITRE 1. QU'EST-CE QU'UNE DONNEE DE MARQUE ET UNE DONNEE DE JURISPRUDENCE ? 10
1. Données de marques 10
2. Données de jurisprudence 11
3. Sélection et tri des données de marques 12
4. Sélection et tri des données de jurisprudences 12
CHAPITRE 2. COMMENT AUTOMATISER CE PROCESSUS 14
1. Automatisation de la récupération 14
2. Difficultés rencontrées 15
PARTIE 2 - LES DIFFERENTES PHASES DU FONCTIONNEMENT DU PROGRAMME 17
CHAPITRE 4. LE MODELE LINGUISTIQUE 19
1. Les vecteurs linguistiques 19
2. Calcul de la proximité linguistique entre deux marques 20
3. Calcul de la proximité linguistique notre cas d'opposition fictif et les cas d'oppositions existants 21
4. Calcul du pourcentage 22
CHAPITRE 5. LE MODELE INFORMATIQUE 24
1. Interfaces client 24
1.1. Recherche simple : 24
1.2. Recherche globale : 26
2. Interfaces administrateur 27
2.1. Recherche simple : 28
2.2. Recherche globale : 28
2.3. Recherche internes : 29
CHAPITRE 6. L'AFFICHAGE DES RESULTATS 30
1. Affichage des cas d'oppositions existants les plus proches 31
2. Affichage des marques les plus proches 32
3. Calculs des coordonnées 33
3.1. Coordonnées pour les cas d'oppositions existants les plus proches 33
3.2. Coordonnées pour les marques les plus proches 36
CHAPITRE 7. RESULTATS OBTENUS 39
1. Exemples de résultats 39
2. Une classe peut tout changer 40
Conclusion 42
1. Résumé 42
2. Améliorations possibles 42
3. Bilan personnel 43
Table des illustrations 44
Table des annexes 45
Table des matières 52

MOTS-CLÉS : proximité linguistique, propriété industrielle, droits des marques, marques, jurisprudence

RÉSUMÉ

Ce mémoire rapporte le travail effectué dans le cadre d'un stage de six mois, effectué au sein de la société en conseil de propriété industrielle IP Trust Innovation. Le projet derrière ce travail était de fournir un logiciel d'aide à la décision en droit des marques, basé sur une analyse linguistique de ces marques, afin d'aider les juristes dans leurs prises de décisions en droit des marques, afin de pouvoir renseigner au mieux les clients. Nous présenterons comment les données brutes ont été exploitées, et s'est mise en place l'automatisation de la récupération des données. Puis nous verrons comment est organisé le logiciel, et rentrer dans les détails des modèles linguistique et informatique, afin de comprendre comment est organisé le logiciel, et comment les résultats qu'il produit sont affichés.

KEYWORDS : linguistic proximity, industrial property, trademark rights, trademarks, jurisprudence

ABSTRACT

This paper reports on the work done in a six-month internship at the IP Trust Innovation company. The project behind this work was to provide a trademark-rights decision support software, based on a linguistic analysis of the trademarks, to help jurists into their decision-making, to be able to inform at best the customers. We will present how the raw data was exploited, and set up the automation of data recovery. Then we will see how the software is organized, and go into the details of the linguistic and computer models, in order to understand how is organized the software, and how the results it produces are displayed.