



IA et Data Mining au service de l'accès thématique aux ressources des bibliothèques



TECH'advantage-Progilone

Depuis
2004
Certification
ISO9001

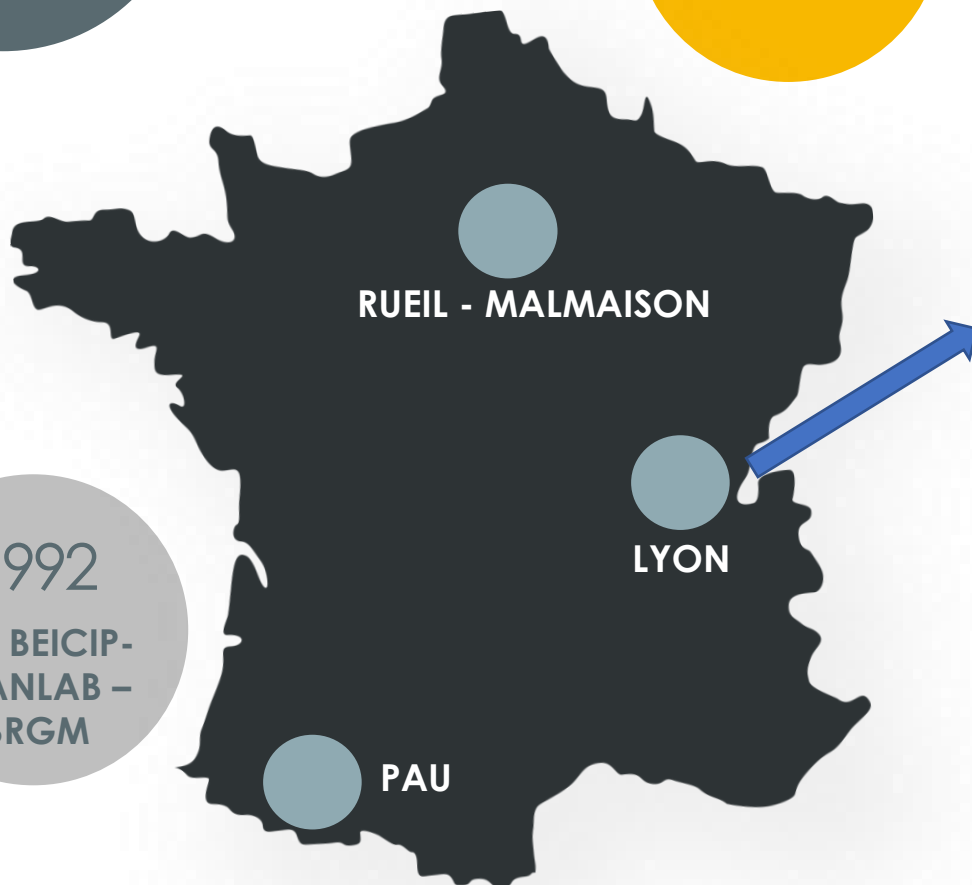
2020
Rachat
PROGILONE

CA
12 M€

2001
Filiale à
100% IFPEN

1996
SA

1992
GIE BEICIP-
FRANLAB –
BRGM



RUEIL - MALMAISON

LYON

PAU

Pôle Édition



Système
d'information
documentaire



Plate-forme de gestion
des chantiers de
numérisation



Solution
collaborative pour la
conception d'une
documentation
utilisateur contextuelle



Moteurs de recherche
GED
Portails

Fouille de données
Classification
Web sémantique





Le Contexte

Les notices bibliographiques s'avèrent souvent riches d'informations saisies par des catalogueurs expérimentés dans des zones de saisies libres comme des résumés ou des notes.

Pour autant, ces données sont sous-employées dans l'apport que leur contenu peut apporter en termes de recherche sur les catalogues des bibliothèques de lecture publique, dont la finalité est l'accès à l'information et la culture, pour tous et sur l'ensemble du territoire.

En effet, ces informations sont restituées lors de recherche, mais elles n'interviennent pas ou mal dans les critères ayant permis une recherche.

Le Projet

Le projet a pour objectif la fouille de ces données pour en extraire tous les points d'intérêts à même de permettre une classification et un rapprochement vers un ou des thésaurus de référence.

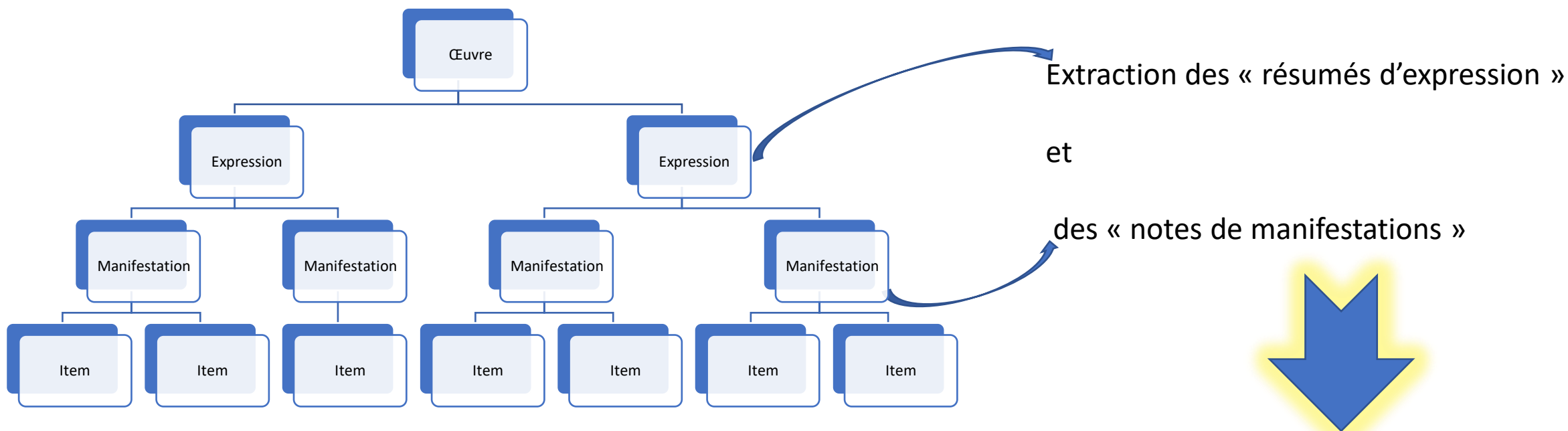
Analyser de manière automatisée les zones de résumés (ou d'extraits, de tables des matières, de biographie de l'auteur etc.) pour les caractériser, les classer et les rattacher à un thésaurus utilisé par la Médiathèque Départementale du Puy de Dôme et/ou à une indexation matière basée sur les vedettes RAMEAU (Répertoire d'autorité-matière encyclopédique et alphabétique unifié).

L'objectif est de mettre en place prioritairement un classement thématique des informations et, dans certains cas, d'enrichir également l'indexation sujet par des points d'accès.



DONNEES A TRAITER

Le catalogue bibliographique de la
Médiathèque Départementale du Puy de Dôme
est au modèle **IFLA LRM**

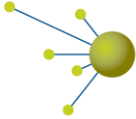


144 427 résumés

Les notes de manifestations s'avérant très « techniques » n'ont finalement pas été incluses

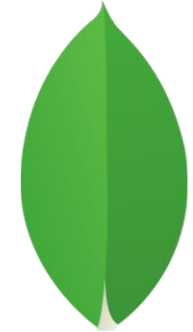


ENVIRONNEMENT TECHNIQUE

ETL  **talend**

BASE DE
DONNEES

mongoDB®



FRAMEWORKS



Natural Language Analysis
with Python NLTK

LANGAGE
PROGRAMMATION



CamemBERT

SIGB IFLA LRM



syrtis
by progilone



NLP-Pré traitement

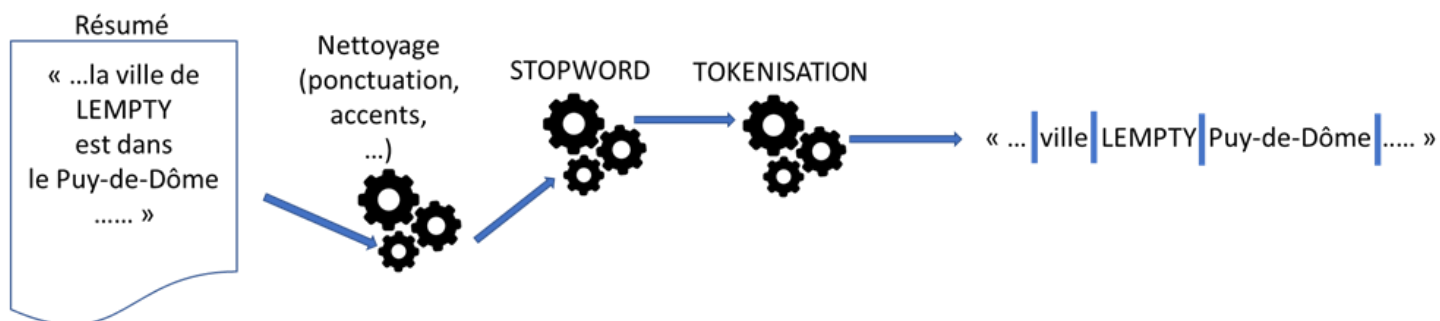
NLP (Natural Language Processing ou Traitement Numérique du Langage)

Partie « linguistique » du NLP qui consiste à prétraiter et transformer les informations en entrée en un jeu de données exploitable :

☐ Nettoyage (majuscules/minuscules, accents, ponctuation,..)

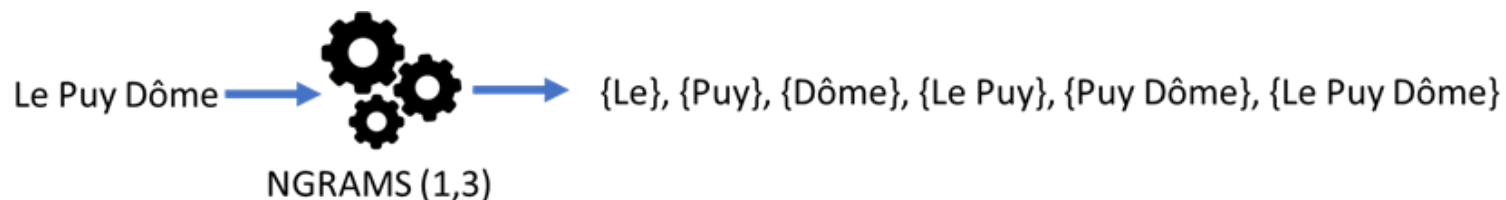
☐ Stop words

☐ Tokénization



☐ Stemming(racinisation) ou Lemmatization

☐ N-Gramms





L'APPRENTISSAGE AUTOMATIQUE

Partie « Data Science » du NLP qui consiste au traitement de « l'apprentissage automatique » et qui porte sur l'application de modèles de Machine Learning ou Deep Learning

Deux types d'apprentissage :

- ☐ Apprentissage supervisé : automatisation des processus de décision en généralisant des données fournies à partir d'exemples connus
- ☐ Apprentissage non supervisé : seules les données en entrées sont déterminées et aucune sortie connue n'est fournie à l'algorithme

Dans le cadre de notre projet, nous avons utilisé un apprentissage supervisé.



L'APPRENTISSAGE SUPERVISE

Deux types majeurs de problèmes d'apprentissage supervisé :

- ☐ La Classification : prédire une *étiquette de classe* qui est un choix possible parmi une liste prédéfinie de possibilité
La Classification peut être binaire (uniquement 2 classes) ou multiclass
- ☐ La Régression : prédire une valeur continue (nombre)

Dans le cadre de notre projet, nous avons utilisé une Classification multiclasse.



L'APPRENTISSAGE SUPERVISE

Les différents types d'algorithmes testés :

❑ La régression linéaire

Il s'agit d'un des modèles les plus simples.

L'apprentissage consiste à calculer les coefficients qui minimisent les erreurs de données sur un jeu d'apprentissage.

La régression linéaire est utilisée pour l'estimation de certaines tendances en économétrie et dans le marketing notamment lorsqu'il est probable qu'il existe une relation linéaire entre variable explicative et cible.

❑ Les k plus proches voisins

Cette méthode est un algorithme de classification supervisé et non paramétrique.

L'idée consiste à supposer qu'une observation est similaire à celle de ses voisins.

Ce type d'algorithme est par exemple utilisé par Netflix pour prédire les scores qu'un utilisateur attribuera à un film en fonction des scores qu'il a déjà attribué à des films similaires.

❑ La classification naïve bayésienne

Le classifieur de type naïf bayésien est un modèle d'apprentissage automatique de type supervisé permettant la classification en se basant sur le théorème de Bayes avec une forte indépendance des hypothèses (c'est-à-dire pas de corrélation entre les caractéristiques d'un ensemble de données).

Le classifieur de type naïf bayésien peut être appliqué dans différents scénarios, un des cas d'utilisation classiques pour ce modèle d'apprentissage est la classification des documents. Il s'agit de déterminer si un document correspond à certaines catégories ou pas (analyse de spam, de sentiments, ..)



L'APPRENTISSAGE SUPERVISE

Les différents types d'algorithmes testés :

☐ La régression logistique

L'algorithme de régression logistique mesure la relation entre la variable dépendante catégorielle et une ou plusieurs variables indépendantes en donnant une estimation à la probabilité d'occurrence d'un événement à travers l'usage de sa fonction logistique.

La régression logistique peut être mise en œuvre pour des campagnes de ciblage marketing. Le but ainsi est d'attirer de nouveaux clients. La construction d'un modèle de régression logistique permettra d'expliquer et de prédire les clients susceptibles d'acheter le produit ou le service offert par l'entreprise (Booking.com).

☐ Les arbres de décision

Les arbres de décision sont des modèles de Machine Learning supervisés et non paramétriques extrêmement flexibles. Ils sont utilisés aussi bien pour la classification que pour la régression.

Les problèmes de classification sont traités dans toute leur généralité, par contre il existe un fort risque de sur-apprentissage.

☐ Les forêts aléatoires

Une forêt aléatoire est un méta-estimateur qui ajuste un certain nombre de classificateurs d'arbres de décision sur divers sous-échantillons de l'ensemble de données et utilise la moyenne pour améliorer la précision prédictive et contrôler le surajustement.

C'est un algorithme qui s'avère à la fois très précis et très robuste, et c'est aujourd'hui un des meilleurs algorithmes disponible pour ce qui est de la précision des prédictions, cependant la complexité de son algorithme rend son implémentation délicate.



L'APPRENTISSAGE SUPERVISE

Les différents types d'algorithmes testés :

❑ Les réseaux de neurones

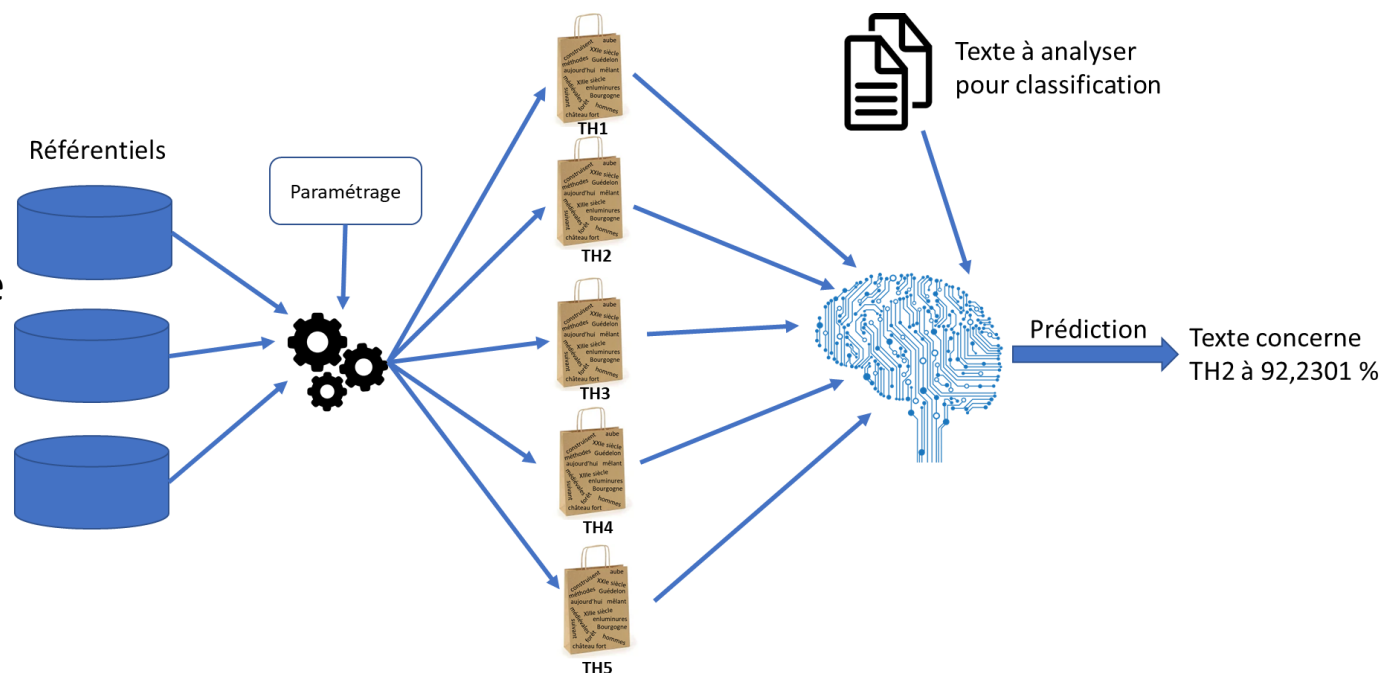
Les réseaux de neurones tirent leur nom des premières tentatives dans les années 1960 de modéliser mathématiquement les traitements de l'information dans le cerveau.

Ces algorithmes ont récemment trouvés une nouvelle jeunesse sous l'appellation d'apprentissage profond ou Deep Learning.

Bien qu'ils soient très prometteurs, ces algorithmes sont cependant réservés à des cas d'utilisation très spécifiques.

❑ Les sacs de mots

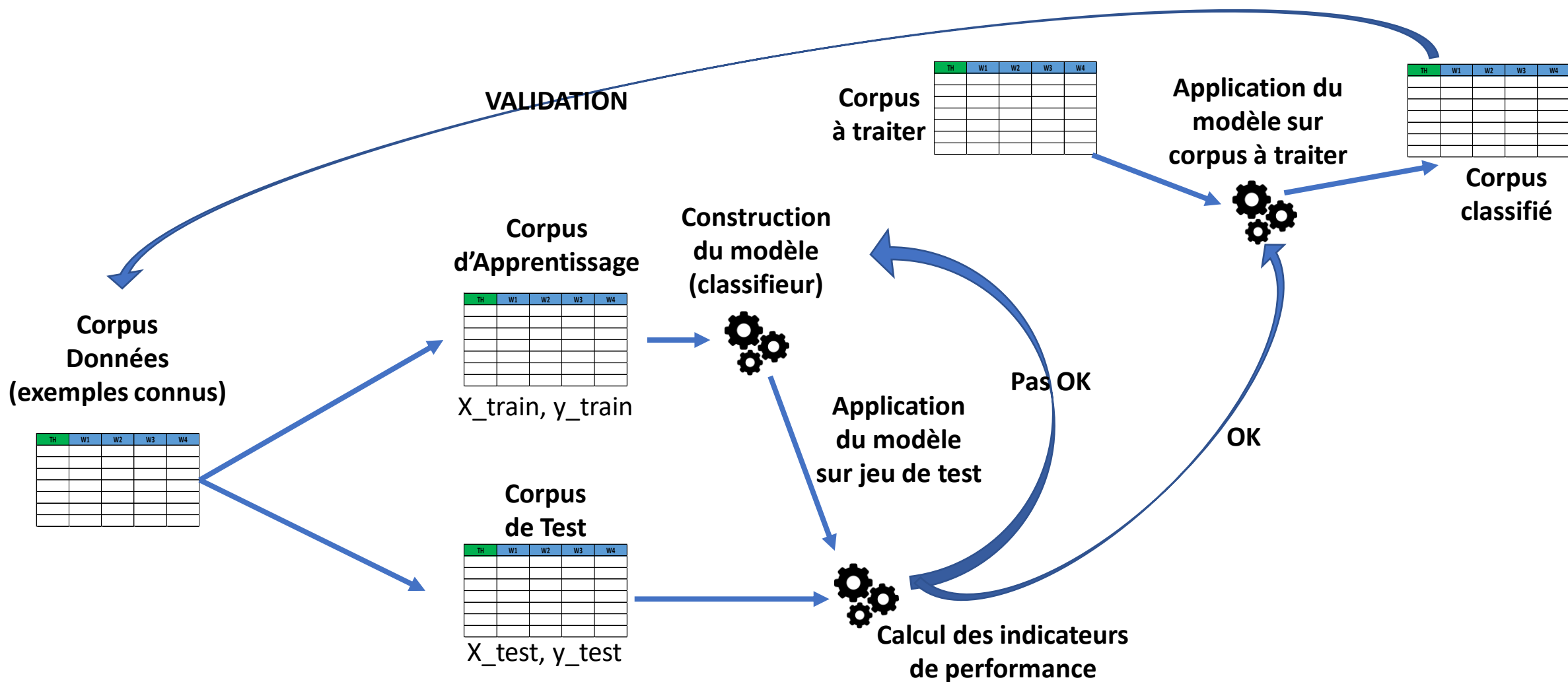
L'une des méthodes les plus simples, mais aussi les plus efficaces et les plus courantes, servant à transformer les textes afin de leur donner un format reconnaissable pour l'apprentissage automatique consiste à utiliser une représentation dite par « sacs de mots » (bag of words en anglais).





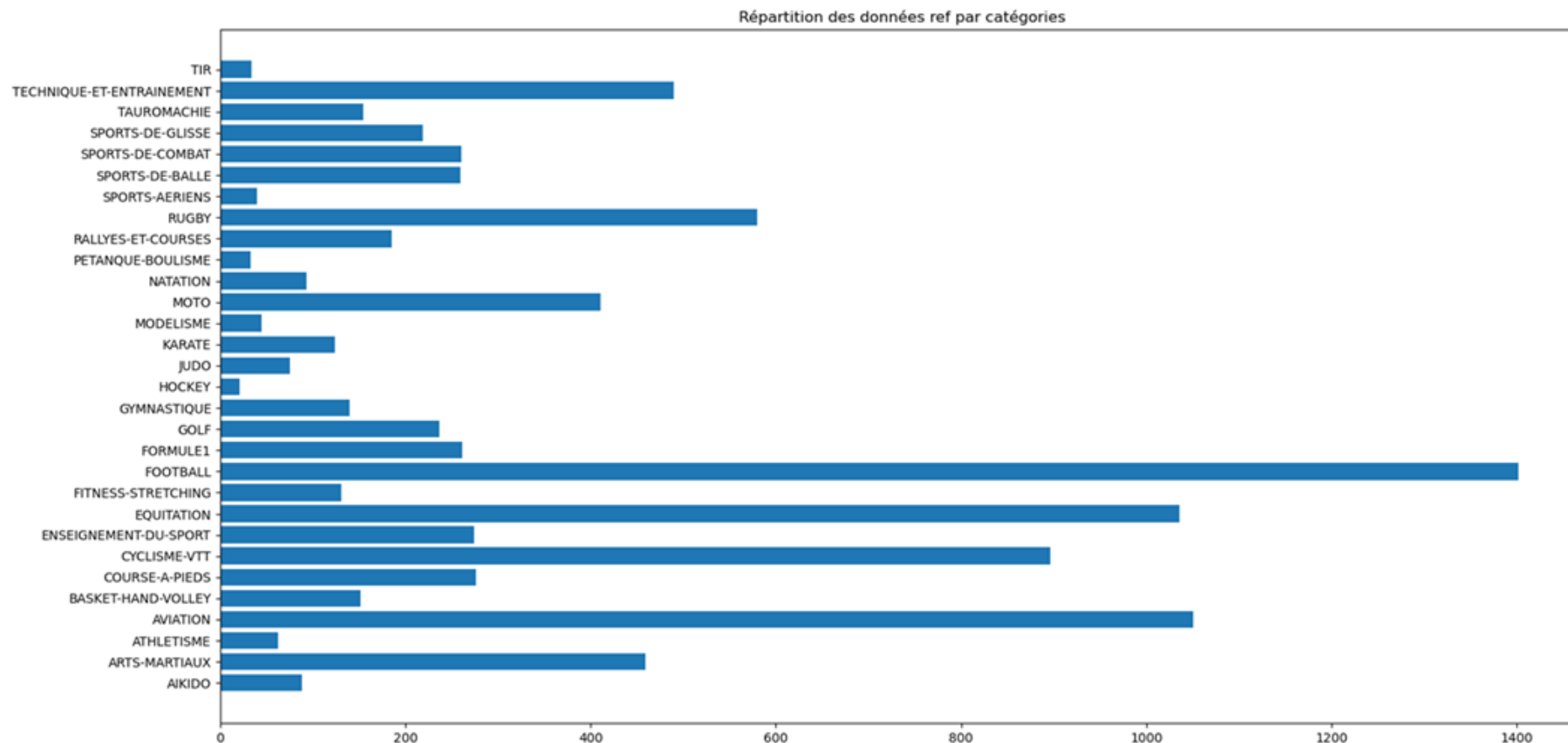
L'APPRENTISSAGE

Construire le bon modèle d'apprentissage





ANALYSE DU CORPUS D'APPRENTISSAGE



Eviter le sur-apprentissage :

- ☐ faire une nouvelle catégorisation avec les catégories les moins représentées
- ☐ trouver de nouvelles sources de données pour faire un entraînement plus précis
- ☐ modifier la source de données pour équilibrer les catégories
- ☐ augmenter le poids dans la fonction de loss des catégories avec moins d'exemple



LES METRICS D'ANALYSE

Global performance:

Accuracy: 0.887

Aera Under Curve: 0.987

Detail:

	precision	recall	f1-score	support
AIKIDO	0.82	0.88	0.85	26
ATHLETISME	0.50	0.50	0.50	18
BASKET-HAND-VOLLEY	0.93	0.93	0.93	45
COURSE-A-PIEDS	0.76	0.82	0.79	83
CYCLISME-VTT	0.99	0.82	0.90	143
EQUITATION	0.98	0.94	0.96	148
FOOTBALL	0.96	0.91	0.94	139
FORMULE1	0.77	0.99	0.87	78
GOLF	0.90	0.99	0.94	70
GYMNASTIQUE	0.67	0.80	0.73	41
HOCKEY	0.55	1.00	0.71	6
JUDO	0.84	0.73	0.78	22
KARATE	0.95	0.95	0.95	37
MODELISME	0.58	0.85	0.69	13
MOTO	0.95	0.85	0.90	123
NATATION	0.66	0.89	0.76	28
PETANQUE-BOULISME	0.83	1.00	0.91	10
RUGBY	0.96	0.97	0.96	145
SPORTS-DE-GLISSE	0.94	0.71	0.81	65
TAUROMACHIE	0.91	0.93	0.92	46
TIR	0.82	0.90	0.86	10
accuracy			0.89	1296
macro avg	0.82	0.87	0.84	1296
weighted avg	0.90	0.89	0.89	1296

- ❑ **Accuracy:** c'est la plus simple d'entre elle. L'accuracy est un métrique qui permet de mesurer le pourcentage de réussite du classifieur sur le jeu de données d'évaluation. Sur N prédictions, quel pourcentage de prédictions ont été correctes.
- ❑ **Area Under Curve:** c'est l'aire sous la courbe RoC. Cette courbe représente en 'y' le nombre de prédiction juste (j'ai prédit la classe 'C' et c'est juste) et en 'x' le nombre de prédiction fausse (j'ai prédit la classe 'C' alors que c'est autre chose). Dans le cas d'une multi-classe comme le notre, cette mesure n'est pas forcément la bonne mesure car on fait une moyenne globalisée sur l'ensemble des classes.
- ❑ **precision:** La précision est le pourcentage d'éléments correctement associés à un certain label par rapport à tous les éléments associés à ce label. Elle peut être vue comme la capacité du modèle à ne pas classer dans une catégorie donnée un élément qui ne doit pas y être.
- ❑ **recall:** c'est le pendant de la précision. Le rappel est le pourcentage de classifications correctes suivant un label donné. Soit la capacité du modèle à classer correctement les éléments d'une catégorie donnée. C'est le pourcentage de bonne prédiction par rapport au nombre d'élément de la classe 'C'. Une précision < 0.5 signifie que l'on se trompe plus d'une fois sur 2.
- ❑ **f1-score:** Le F1-score est une métrique qui prend en compte la précision et le rappel sur une catégorie donnée. Elle peut être vue comme une combinaison de ces deux métriques. Elle vaut la moyenne harmonique du rappel et de la précision. Le F1-score est compris entre 0 et 1. Le meilleur score est donc 1 et le pire 0.
- ❑ **support:** le nombre l'élément de la classe.



LA MATRICE DE CONFUSION

La matrice de confusion est une matrice qui mesure la qualité d'un système de classification. Chaque ligne correspond à une classe réelle, chaque colonne correspond à une classe estimée. La cellule ligne L, colonne C contient le nombre d'éléments de la classe réelle L qui ont été estimés comme appartenant à la classe C.

L'intérêt principal de la matrice de confusion est qu'elle montre rapidement si un système de classification parvient à classer correctement.

		Confusion matrix																				
True	AIKIDO -	15	0	0	1	0	2	1	0	1	1	0	1	2	0	0	0	0	1	0	0	1
	ATHLETISME -	0	4	0	3	1	2	3	0	0	0	0	0	0	0	0	0	0	5	0	0	0
	BASKET-HAND-VOLLEY -	0	0	25	0	1	2	6	0	0	0	0	0	0	0	0	0	0	11	0	0	0
	COURSE-A-PIEDS -	0	0	0	38	17	6	4	0	0	8	0	0	0	0	2	1	0	6	1	0	0
	CYCLISME-VTT -	0	0	0	2	120	6	2	1	1	0	0	0	0	0	4	0	0	6	1	0	0
	EQUITATION -	0	0	0	0	11	130	1	0	0	0	0	0	0	0	0	0	0	5	1	0	0
	FOOTBALL -	0	0	0	1	6	1	117	0	1	0	0	0	0	0	1	0	0	10	2	0	0
	FORMULE1 -	0	0	0	0	10	0	4	56	0	0	0	0	0	0	6	0	0	2	0	0	0
	GOLF -	0	0	0	0	2	4	2	2	42	0	0	0	0	0	1	0	0	16	1	0	0
	GYMNASTIQUE -	0	0	0	6	1	7	1	0	0	24	0	0	0	0	0	0	0	2	0	0	0
	HOCKEY -	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	3	0	0
	JUDO -	0	0	0	0	3	1	7	0	0	1	0	6	3	0	0	0	0	0	1	0	0
	KARATE -	1	0	0	2	1	3	2	0	0	0	0	0	28	0	0	0	0	0	0	0	0
	MODELISME -	0	0	0	0	1	1	1	1	0	0	0	0	0	6	3	0	0	0	0	0	0
	MOTO -	0	0	0	0	7	1	0	2	1	0	0	0	0	0	108	0	0	4	0	0	0
	NATATION -	0	0	0	2	6	4	1	0	0	0	0	0	0	0	0	8	0	6	1	0	0
	PETANQUE-BOULISME -	0	0	0	0	3	0	6	0	0	0	0	0	0	0	0	0	0	1	0	0	0
	RUGBY -	0	0	2	0	6	0	4	0	0	0	0	1	0	0	2	0	0	130	0	0	0
	SPORTS-DE-GLISSE -	0	0	0	0	10	3	4	0	1	0	0	0	0	0	1	0	0	5	41	0	0
	TAUROMACHIE -	0	0	0	0	2	4	5	0	0	0	0	0	0	0	1	0	0	5	0	29	0
	TIR -	0	0	0	0	0	4	4	0	0	0	0	0	0	0	0	0	0	0	0	0	2
		AIKIDO	ATHLETISME	BASKET-HAND-VOLLEY	COURSE-A-PIEDS	CYCLISME-VTT	EQUITATION	FOOTBALL	FORMULE1	GOLF	GYMNASTIQUE	HOCKEY	JUDO	KARATE	MODELISME	MOTO	NATATION	PETANQUE-BOULISME	RUGBY	SPORTS-DE-GLISSE	TAUROMACHIE	TIR

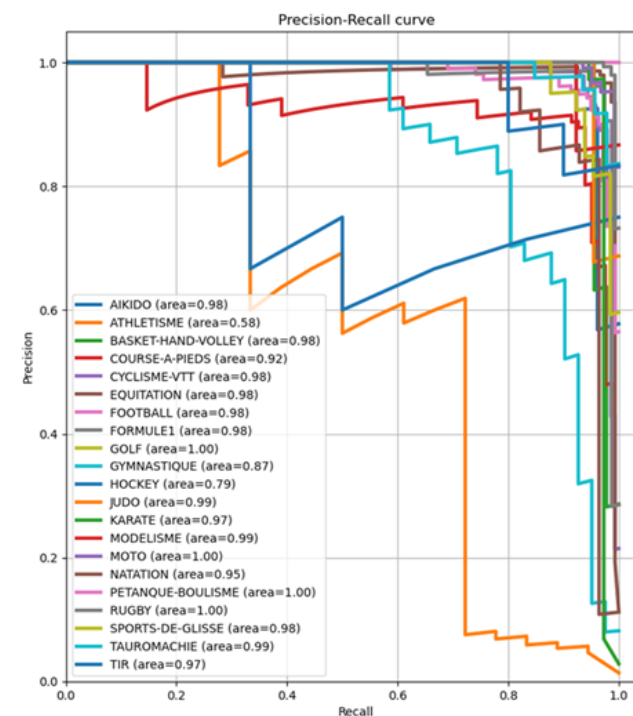
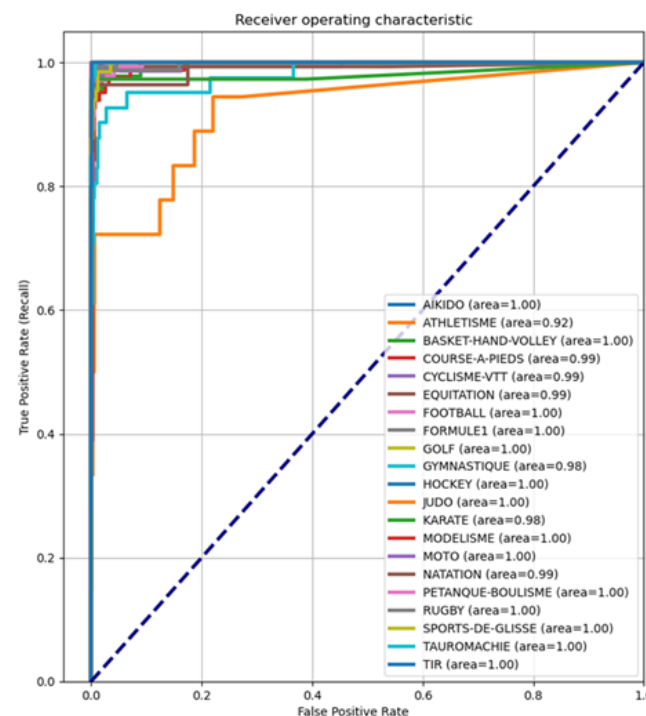


LES COURBES DE PRECISION

Receiver operating characteristic (Roc_curve) :

Une courbe ROC est un graphique représentant les performances d'un modèle de classification pour tous les seuils de classification. Cette courbe trace le taux de vrais positifs en fonction du taux de faux positifs. Les courbes ROC sont généralement utilisées dans la classification pour étudier la sortie d'un classificateur.

Precision-recall curve area : Precision-Recall est une mesure utile du succès de la prédiction lorsque les classes sont très déséquilibrées. Dans la recherche d'informations, la précision est une mesure de la pertinence des résultats, tandis que le rappel est une mesure du nombre de résultats vraiment pertinents qui sont renvoyés.





IMPORTANCE DES PARAMETRES

☐ La perte (loss) : permet de minimiser le surapprentissage

fonction qui évalue l'écart entre les prédictions réalisées par le modèle et les valeurs réelles des observations utilisées pendant l'apprentissage

☐ La régularisation (pénalité) de données

La régularisation est un concept par lequel les algorithmes d'apprentissage automatique peuvent être empêchés de surappliquer un ensemble de données. La régularisation y parvient en introduisant un terme qui attribue une pénalité plus élevée aux courbes complexes.

☐ Constante de multiplication

Constante qui multiplie le terme de régularisation. Plus la valeur est élevée, plus la régularisation est forte.

☐ Critère d'arrêt

Arrêter l'entraînement après une certaine valeur

☐ Nombre d'itérations

Nombre d'itérations sans amélioration à attendre avant d'arrêter l'ajustement

☐ Mélange des données

☐ Suppressions

- ☐ Termes qui apparaissent dans moins de x ou % documents,

- ☐ Termes qui apparaissent dans plus de x ou % documents.



IMPORTANCE DES PARAMETRES

De petites modifications.....avec de grands effets

☐ Test avec une constante de multiplication = 0,01

```
Global performance:  
Accuracy: 0.927  
Aera Under Curve: 0.986
```

☐ Test avec une constante de multiplication = 0,1

```
Global performance:  
Accuracy: 0.569  
Aera Under Curve: 0.942
```



IMPORTANCE DES PARAMETRES ET DES ALGORITHMES

A partir du même corpus d'apprentissage :

Global performance:
Accuracy: 0.669
Area Under Curve: 0.706
Detail:



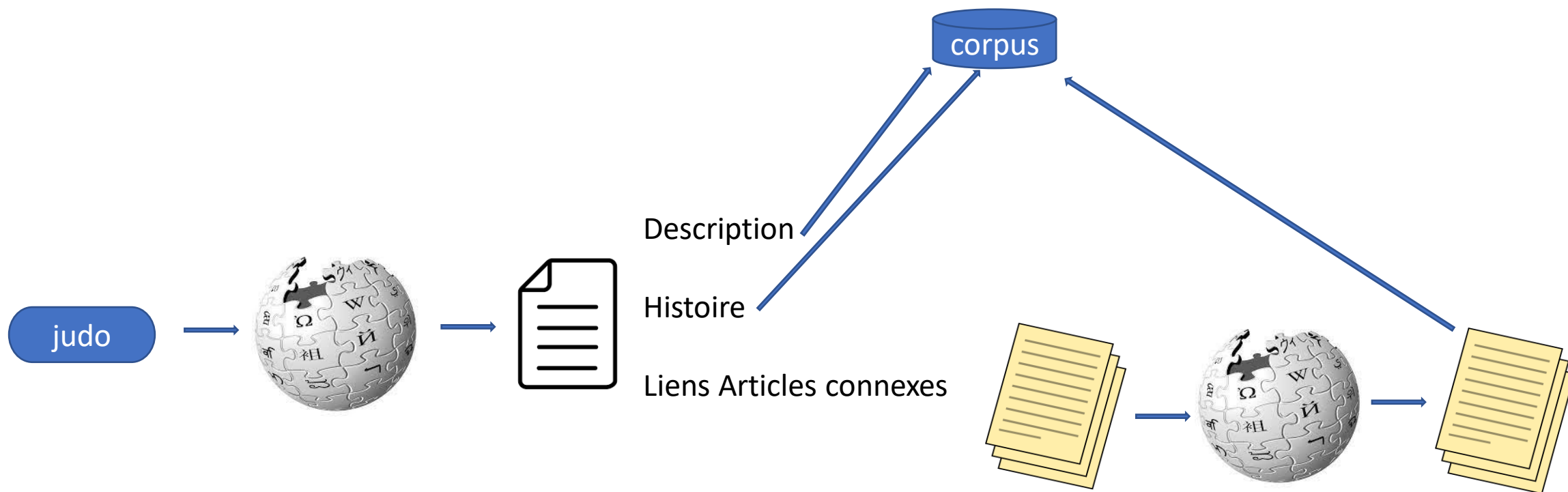
Global performance:
Accuracy: 0.958
Area Under Curve: 0.993
Detail:



ENRICHISSEMENT

Possibilité d'enrichir les corpus à partir :

- D'autres référentiels
- Par complément d'informations (ajouter titre, note, ...)
- De wikipédia





LES RESULTATS

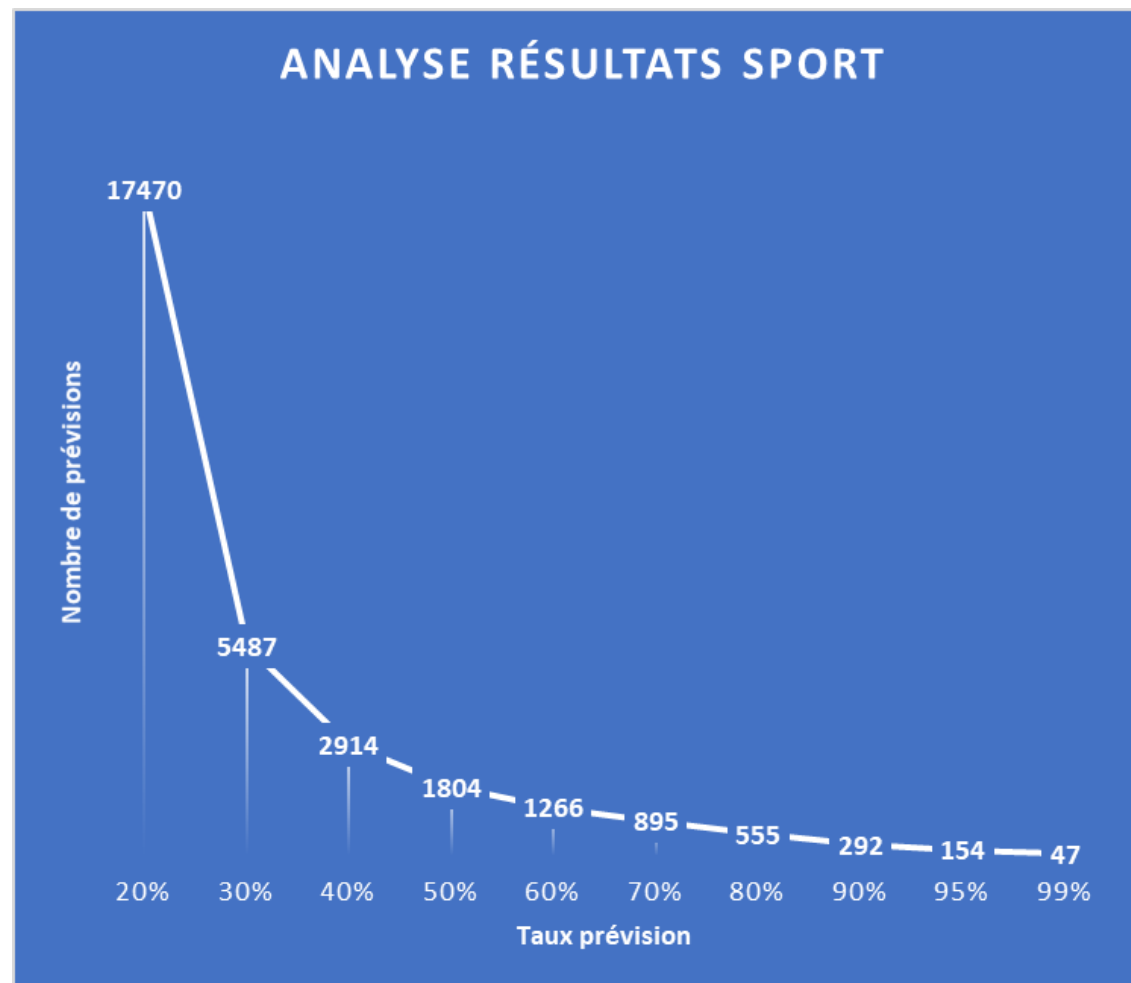
- ☐ Ne pas s'attendre à des résultats exhaustifs ni sans erreurs, tout comme peut l'être l'analyse humaine
- ☐ En fonction du besoin, du volume de données, bien qualifier son modèle
- ☐ Importance des référentiels dans une approche supervisée
- ☐ Déterminer le taux d'erreur que l'on souhaite (je ne prends que les résultats dont le taux de succès est $> 0,xx \%$)
- ☐ Des résultats impossibles avec une approche traditionnelle basée sur une seule recherche de mots.

Paradoxe : la création d'un modèle demande plus de temps humain de réparation/validation que de temps machine de traitement



LES RESULTATS

Déterminer si l'on souhaite privilégier la quantité (mots clés par exemple) ou la qualité (rattachement à un thésaurus par exemple)





LES MODELES CREES

Modèle Sport
AIKIDO
ATHLETISME
BASKET-HAND-VOLLEY
COURSE-A-PIEDS
CYCLISME-VTT
EQUITATION
FOOTBALL
FORMULE1
GOLF
GYMNASTIQUE
HOCKEY
JUDO
KARATE
MODELISME
MOTO
NATATION
PETANQUE-BOULISME
RUGBY
TIR

Modèle Sport1
ENSEIGNEMENT-DU-SPORT
FITNESS-STRETCHING
SPORTS-AERIENS
SPORTS-DE-BALLE
SPORTS-DE-COMBAT
SPORTS-DE-GLISSE
RALLYES-ET-COURSES

Modèle Droit
DROIT-AFFAIRES
DROIT-CIVIL
DROIT-INTERNATIONAL
DROIT-PENAL
DROIT-PRIVE
DROIT-PUBLIC
DROIT-TRAVAIL

Modèle Histoire
ANTIQUITE
MOYEN-AGE
PREHISTOIRE
RENAISSANCE
XIX-SIECLE
XVI-SIECLE
XVII-SIECLE
XVIII-SIECLE
XX-SIECLE

Modèle Religion
BOUDDHISME
CHRISTIANISME
CONFUCIANISME
HINDOUISME
ISLAM
JUDAISME
ORTHODOXIE
PROTESTANTISME
TAOISME

Modèle Guides
GUIDES-ALSACE
GUIDES-AQUITAINE
GUIDES-AUVERGNE
GUIDES-BOURGOGNE
GUIDES-BRETAGNE
GUIDES-CENTRE
GUIDES-CHAMPAGNE-ARDENNE
GUIDES-CORSE
GUIDES-FRANCHE-COMTE
GUIDES-ILE-FRANCE
GUIDES-LANGUEDOC-ROUSSILLON
GUIDES-LIMOUSIN
GUIDES-LORRAINE
GUIDES-MIDI-PYRENEES
GUIDES-NORD_PÄS_CALAIS
GUIDES-NORMANDIE
GUIDES-PACA
GUIDES-PAYS_LOIRE
GUIDES-PICARDIE
GUIDES-POITOU-CHARENTES
GUIDES-RHONE-ALPES

Modèle Psychopathologie
ALCOOLISME
ANOREXIE
AUTISME
DEPRESSION
DEUIL
ETATS-LIMITES
MALTRAITANCE-ENFANT
PHOBIES
SCHIZOPHRENIE
TOXICOMANIE
TROBLES_BIPOLAIRES
TROUBLES-COMPORTEMENT

Modèle Informatique
ELECTRONIQUE
IA
INTERNET
LANGAGES
MATERIEL
PROGICIEL
RESEAU-TELECOM
SYSTEME-EXPLOITATION

Modèle Santé
ALLERGIES
ALZHEIMER-PARKINSON
ANXIETE-PHOBIES
ASTHME
CANCER
CARDIO-VASCULAIRE
CHOLESTEROL
DENTS
DEPENDANCE
DEPRESSION
DIABETE
MAL-DOS
MALADIE-PEAU
MENOPAUSE
MIGRAINE
MST-SIDA
OBESITE
ORL
RHUMATISMES
TROUBLES-SOMMEIL

Modèle Sciences de la Terre
CARTOGRAPHIE
CLIMATOLOGIE
GEOLOGIE
GEOMORPHOLOGIE
GÉOPHYSIQUE-CHIMIE
GEOTECHNIQUE
HYDRO-GEOLOGIE
METEOROLOGIE
OCEANOGRAPHIE
SEDIMENTOLOGIE
SISMOLOGIE
TECTONIQUE
VOLCANOLOGIE

Modèle Sciences de la vie
AGRICULTURE
AQUACULTURE
BIOLOGIE-ANIMALE
BIOTECHNOLOGIE
BOTANIQUE
ENTOMOLOGIE
ETHOLOGIE
MYCOLOGIE
PHYSIOLOGIE-ANIMALE
SOLS-IRRIGATION
SYLVICULTURE
ZOOTECNIE

Modèle Médecine
ANESTHESIE
CARDIOLOGIE
CHIRURGIE
DERMATOLOGIE
ENDOCRINOLOGIE
GASTRO-ENTEROLOGIE
GERIATRIE
NEUROLOGIE
OBSTETRIQUE
ONCO-HEMATOLOGIE
OPHTALMOLOGIE
ORL
ORTHOPEDIE
PEDIATRIE
PNEUMOLOGIE
PSYCHIATRIE
RHUMATOLOGIE
UROLOGIE



LES PERSPECTIVES

- ☐ Automatisation des classification par traitement journalier sur les nouvelles entrées du catalogue
- ☐ Traitements par enchainement de modèles (histoire+religion → guerres de religion)
- ☐ Gestion en temps réel des classifications à partir d'un résumé :
→ label ayant le meilleur taux de réussite par modèle
- ☐ Potentiellement modèle utilisable sur n'importe quel corpus (n'est pas lié à la bibliothèque pour lequel le projet a été effectué)
A adapter toutefois aux corpus spécialisés (scientifique par exemple)
- ☐ Création de nouveaux modèles avec les mêmes principes

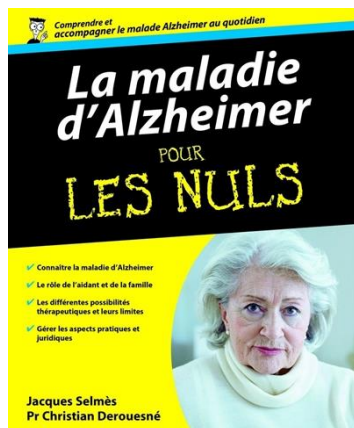


LES PERSPECTIVES

- ❑ Permettre une recherche contextuelle (moteur de recherche web sémantique)
- ❑ Mise en place de points d'accès spécialisés sur le portail lecteur



DÉTERMINER LE SEUIL D'EFFICACITÉ...

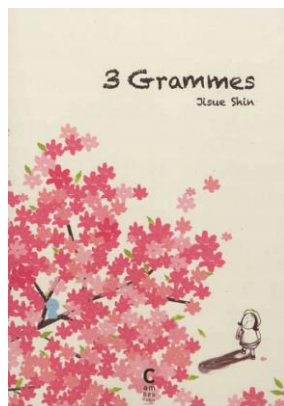


100% - Alzheimer-parkinson

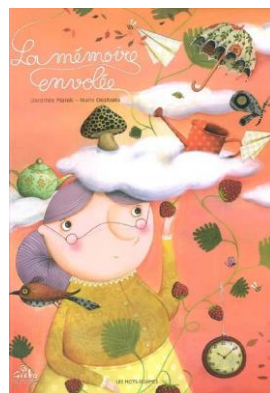
Modèle Santé		
ALLERGIES		DIABETE
ALZHEIMER-PARKINSON		MAL-DOS
ANXIETE-PHOBIES		MALADIE-PEAU
ASTHME		MENOPAUSE
CANCER		MIGRAINE
CARDIO-VASCULAIRE		MST-SIDA
CHOLESTEROL		OBESITE
DENTS		ORL
DEPENDANCE		RHUMATISMES
DEPRESSION		TROUBLES-SOMMEIL



90%
Troubles-sommeil



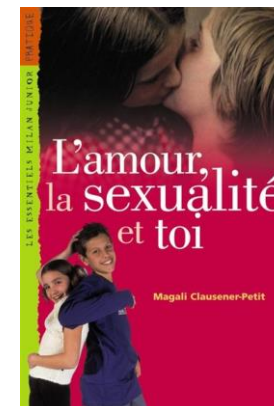
80%
cancer



70%
Alzheimer-parkinson



60%
Cardio-vasculaire



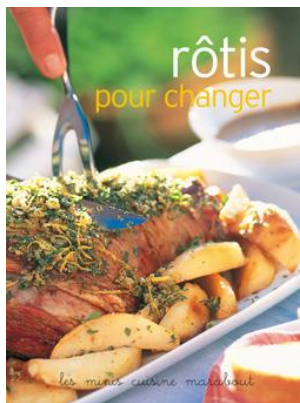
50%
MST-sida



40%
Migraine



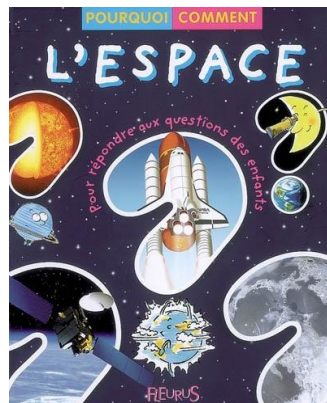
... UNIQUE POUR CHAQUE MODÈLE



51%



60%



70%

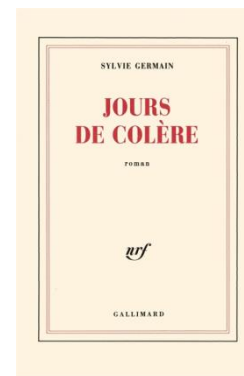
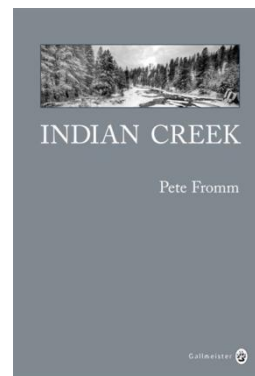
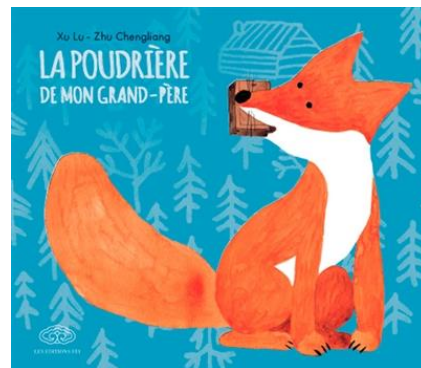
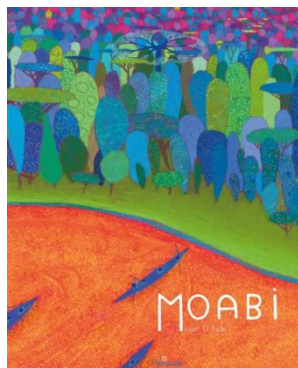


80%



93%

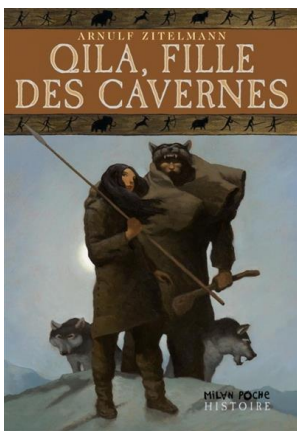
Agriculture ?



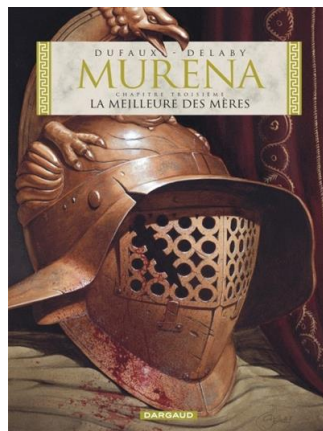
Sylviculture +/- 50%



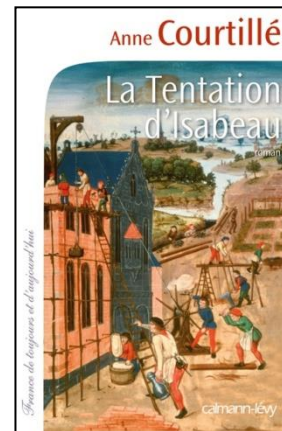
TRAVAILLER LES CATÉGORIES



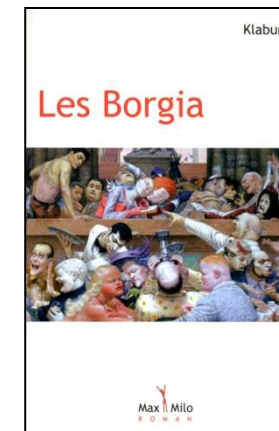
Préhistoire: 50%



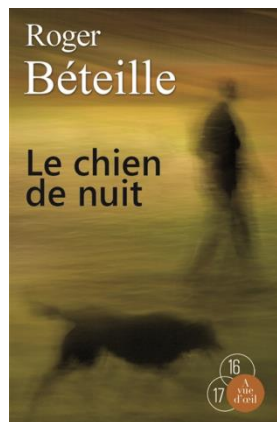
Antiquité : 50%



Moyen-âge: 50%



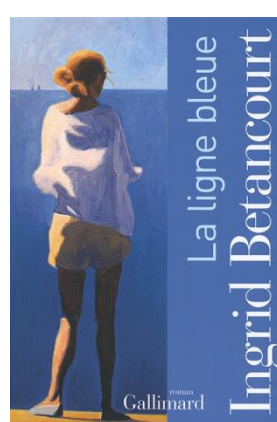
Renaissance : 50%



1° GM



2° GM



1970...



1980...



2010...

**XX°
siècle
50%**

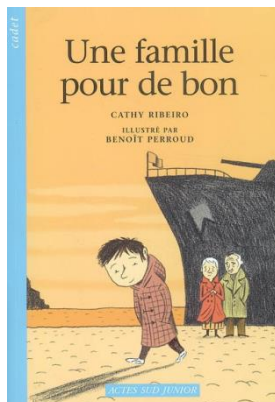


REFLEXIONS SUR LE VOCABULAIRE

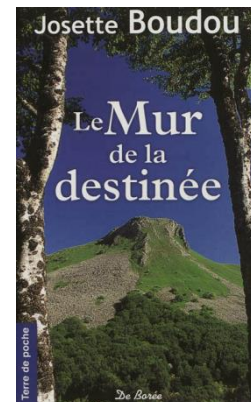
Revoir les terminologies



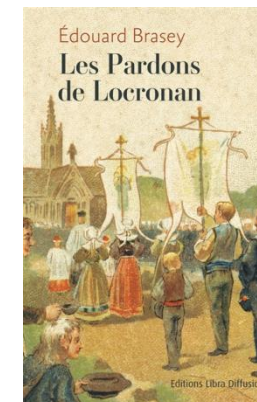
Droit pénal
=> Crimes et délits



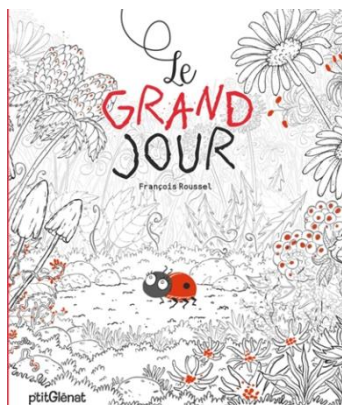
Droit civil
=> Famille



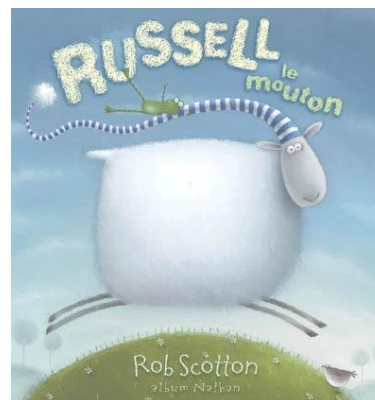
Guide Auvergne
=> Auvergne



Guide Bretagne
=> Bretagne



Anxiété phobie
=> Peur (ark:/12148/cb11953276d)



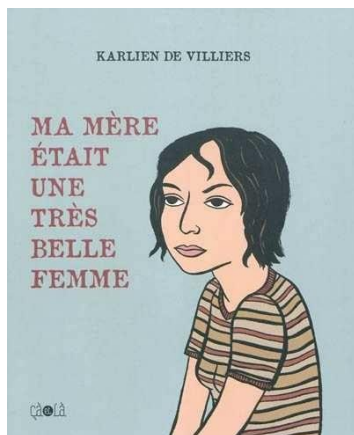
Troubles-sommeil
=> Sommeil ([/ark:/12148/cb11933879p](https://ark:/12148/cb11933879p))



aquaculture
=> Milieu aquatique
([/ark:/12148/cb177650761](https://ark:/12148/cb177650761))



TRAVAILLER LE CORPUS À TRAITER



Agriculture : 52%
Pas de résumé



Sylviculture : 59%
« 20 pièces en bois »

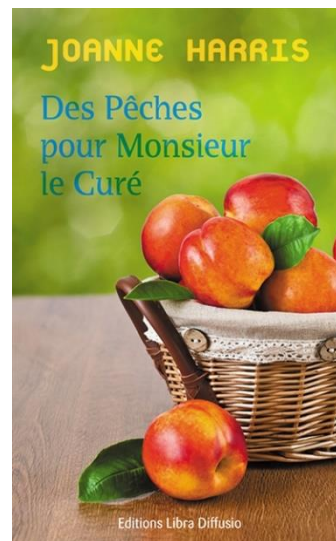
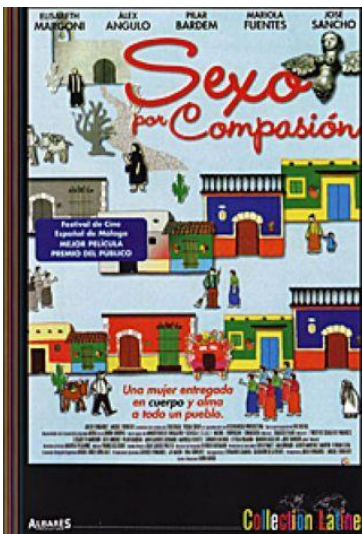


Moyen-âge : 60 %
+ genre/forme
Fantastique

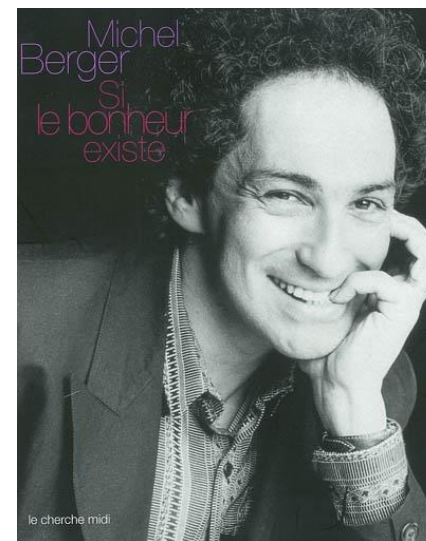


DES LIMITES DE L'AUTOMATISATION 1/2

Homonymie



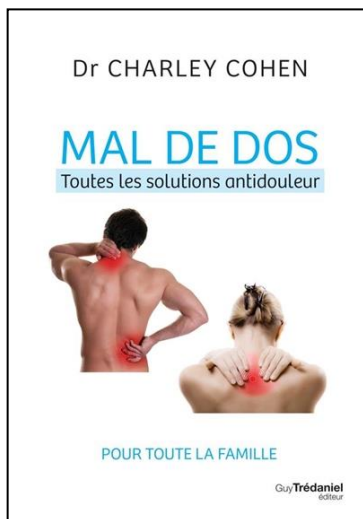
Aquaculture



Agriculture



DES LIMITES DE L'AUTOMATISATION 2/2



progilone
innovative technologies

100%
Mal de dos

{ BnF

Sujet(s) : Dorsalgie -- Prévention ::

Dorsalgie -- Thérapeutique par l'exercice ::

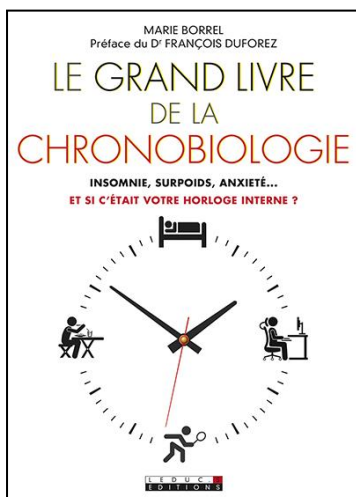
Indice(s) Dewey : 617.564 052 (23e éd.) ::

Electre
DATA SERVICES

617.56 Dos

Dorsalgie -- Prévention

Dorsalgie -- Thérapeutique
dorsalgie, douleur, santé, prévention



50%
Troubles-
sommeil

Sujet(s) : Chronobiologie ::

Rythmes biologiques ::

Indice(s) Dewey : 571.77 (23e éd.) ::

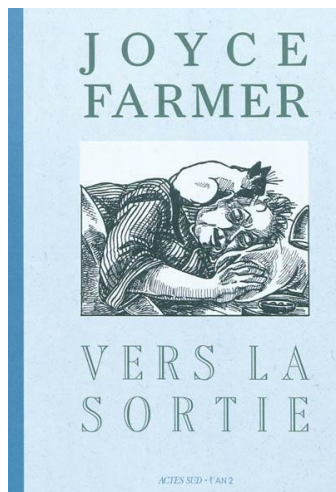
613.04 Promotion de la santé selon le sexe et l'âge

Chronobiologie

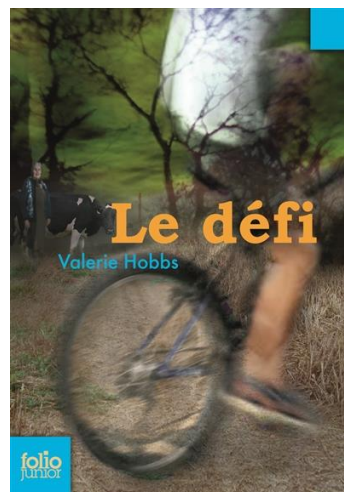
Rythmes biologiques
diététique, chronobiologie



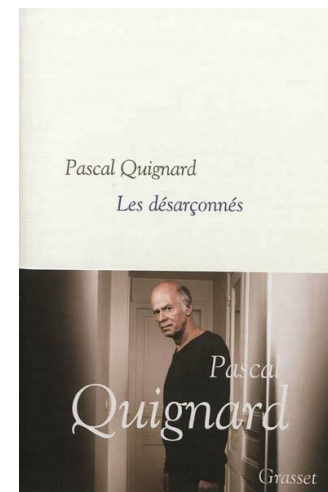
DES PERSPECTIVES DE RECHERCHE POUR LES USAGERS



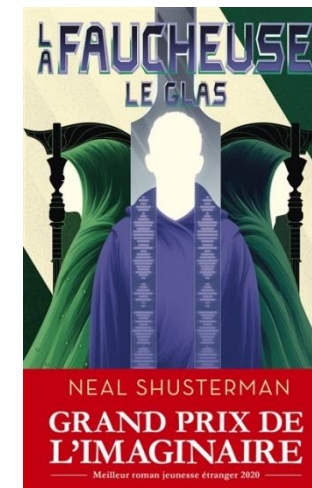
BD + Alzheimer



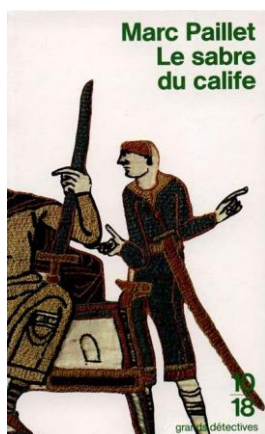
Roman + Cancer



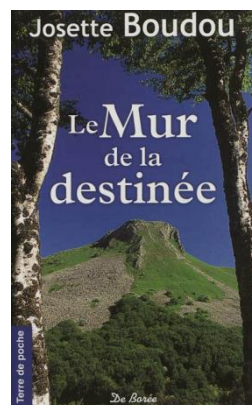
**Roman/essai +
Dépression**



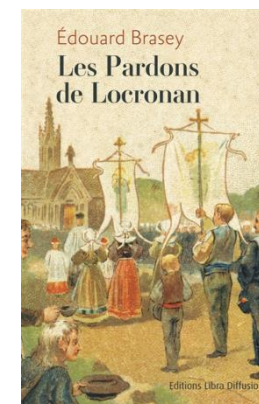
RSF + IA



**Roman policier +
Moyen-âge**



**Auvergne + XX°s +
famille**



**Roman + Bretagne + XX°s
= Romans régionaux, Terroir**

Merci pour votre écoute

Ronald GANIER

ronald.ganier@tech-advantage.com



www.tech-advantage.com



Julien COLIN

julien.colin@puy-de-dome.fr

