

Projet

Adrien UGON

Dimanche 30 Avril 2023

Ce projet a pour objectif de concevoir un système expert qui va permettre de faire des recommandations de traitements médicamenteux en tenant compte des indications et des effets indésirables. La base de connaissance sera alimentée à partir de la base de données publique des médicaments.

Vous utiliserez également une base de pharmacovigilance qui répertoriera les effets indésirables renseignés par les patients.

Ce projet sera réalisé par binôme.

1 Système à base de connaissance

Vous allez concevoir un système d'aide à la décision de recommandation de traitement médicamenteux. Ce système devra prendre en compte l'indication ainsi que les effets indésirables que le patient a pu avoir avec d'autres traitements pour la même indication. Ainsi, on ne prescrira pas de médicament pour lequel est répertorié un effet indésirable que le patient a déjà eu avec un autre médicament pour la même indication.

La base de règles contiendra, pour chaque spécialité :

- ses indications ;
- ses effets indésirables ;

La base d'observation contiendra :

- la (ou les) indications du patient, pour lesquelles on recherche un traitement ;
- les effets indésirables recensés chez ce patient suite à un traitement médicamenteux.

Le patient sera identifié par son identifiant national de santé (INS).

La spécialité sera identifiée par son code identifiant de spécialité (CIS).

Après avoir formalisé le problème (en explicitant le format des règles de la base de connaissance), vous présenterez la solution technique qui pourra prendre la forme d'une base de données (par exemple au format SQLite).

La base de règles pourra être alimentée en utilisant des techniques de traitement automatique du langage naturel sur les RCP issues de la base de données publique des médicaments ou par la base de pharmacovigilance réalisée par apprentissage par renforcement.

Le système devra donc être capable de proposer une spécialité pharmaceutique en fonction des observations fournies.

2 Apprentissage par renforcement

La pharmacovigilance est une discipline qui s'occupe, entre autres, de surveiller les effets indésirables des spécialités pharmaceutiques ayant une autorisation de mise sur le marché.

Pour cela, nous (citoyens et professionnels de santé) sommes invités à déclarer les effets indésirables survenus suite à la prise d'un médicament en remplissant des formulaires.

<https://www.service-public.fr/particuliers/vosdroits/R48611>

Dans cette partie du projet, vous concevrez une base de pharmacovigilance renseignée par des utilisateurs (patients et/ou professionnels de santé) au travers d'un formulaire simple. Chaque patient renseignera, un à un, les médicaments qu'il a pris dans le cadre d'un traitement, puis il renseignera les effets indésirables survenus suite à ce traitement, éventuellement imputables aux spécialités pharmaceutiques qu'il a pris.

Vous présenterez d'abord la modélisation (États, Actions, Transitions, Récompenses) du problème. Vous discuterez éventuellement les stratégies (exploration, exploitation) dans ce cadre. Puis, vous développerez la solution avec un formulaire python dont le traitement des réponses alimente une base de pharmacovigilance (au format SQLite par exemple).

3 Traitement automatique du langage naturel

Vous exploiterez les RCP (Résumés des Caractéristiques Produit) disponibles sur la base de données publique des médicaments (<https://base-donnees-publique.medicaments.gouv.fr/>) afin d'extraire, pour une spécialité pharmaceutique donnée, ses indications et ses effets indésirables.

Pour cela, vous pourrez utiliser le code identifiant de spécialité (CIS) disponible dans le fichier des spécialités (CIS_bdpm.txt) que vous pourrez télécharger à l'adresse suivante : <https://base-donnees-publique.medicaments.gouv.fr/telechargement.php>

Ce CIS est très utile pour extraire le RCP. En effet, le RCP de la spécialité « DOLIPRANE 1000 mg, comprimé » portant le CIS '60234100' est accessible à l'adresse suivante : <https://base-donnees-publique.medicaments.gouv.fr/affichageDoc.php?specid=60234100&typedoc=R>

Le document est au format HTML. Il est constitué de balises.

Chaque section est annoncée par une balise `<p>` de la classe 'AmmAnnexeTitre2'. Cette balise `<p>` a une balise fille dont l'attribut `name` permet d'identifier la section d'intérêt. Pour identifier la section des indications, on recherchera la balise a portant la valeur 'RcpIndicTherap' pour l'attribut `name`. Pour identifier la section des effets indésirables, on recherchera la balise a portant la valeur 'RcpEffetsIndesirables' pour l'attribut `name`.

Les balises `<p>` suivantes (dont la classe est 'AmmCorpsTexte' ou 'AmmCorpsTexteGras') listent les valeurs de la section. C'est-à-dire, qu'elles fournissent la liste des indications et la liste des effets indésirables.

Votre programme python pourra utiliser les librairies suivantes :

- la librairie `requests` qui permet d'effectuer des requêtes HTTP (et donc de récupérer le contenu HTML d'une page web à partir de son URL)
- la librairie `beautifulsoup` qui permet de travailler sur des fichiers HTML.

