

# NLP expression des besoins

## 1 OBJECTIF

Notre objectif est de pouvoir repérer au milieu d'une phrase les mots qui ont du sens par rapport à notre glossaire et de leur faire correspondre des numéros dans la base de données.

Ces numéros sont récupérés par l'API.

A partir de ces numéros, l'interrogatoire est fait par MedVir en posant des questions auxquelles on répond en cochant des cases. Au terme de cet interrogatoire des hypothèses sont posées avec un niveau de gravité.

L'interface NLP est donc seulement un outil pour reconnaître des mots dans une phrase et de les signaler à l'API en amont de l'interrogatoire.

## 2 INTERFACE ACTUELLE

L'accès se fait par l'url [healthcenter.mis-medvir.fr](http://healthcenter.mis-medvir.fr)

- Identifiant : .....
- Mot de passe ....

On s'intéresse pour l'instant uniquement à l'interface MedVir prediag urgences qui gère 400 diagnostics d'urgence.

La page d'accueil propose l'âge, le sexe et la durée des signes (choisir 1 jour)

The screenshot shows the MedVir prediag urgences interface. At the top, there is a navigation bar with the MedVir logo and the text "EDITED BY MEDICAL INTELLIGENCE SERVICE". Below the navigation bar, there are three main sections: "Age (en années)", "Durée des signes (en jours)", and "Sexe". The "Age" section shows a slider with the value 36. The "Durée des signes" section shows a slider with the value 1. The "Sexe" section shows two radio buttons: "Actuellement enceinte", "Récemment accouchée", and "Ni enceinte, ni récemment accouchée". The "Cas particulier" section is also visible. At the bottom, there is a red button labeled "SUITE".

On entre les mots ou groupes de mot un par un dans le search (le search actuel n'est pas capable de reconnaître au sein d'une phrase plusieurs mots pourtant connus du glossaire).

Les symptômes sélectionnés apparaissent dans le pavé de droite. Il n'y a pas de limite au nombre de symptômes sélectionnables. On peut désélectionner ceux qu'on veut.

### 1) Entrez vos symptômes un par un

Le symptôme retenu sera celui entre parenthèses

12 résultat(s)

diarrhée (Diarrhée)

fausse diarrhée (Diarrhée)

fausse diarrhée (Suintement par l'anus)

diarrhée liquide (Diarrhée)

diarrhée très sévère (Diarrhée)

diarrhée très sévère (plus des 10 fois par jour) (Diarrhée)

alternance diarrhée constipation (Diarrhée)

alternance diarrhée-constipation (Constipation)

### 2) Symptômes sélectionnés

Quel symptôme vous gêne ou vous inquiète le plus ?

☐ Mal au ventre

☐ Fièvre

☐ Vomissement

### 3) Informations complémentaires

SUITE

Dans cet exemple, les mots recherchés étaient les suivants :

- « Mal au bide » qui a activé le symptôme mal au ventre
- « Température » qui a activé le symptôme fièvre
- « Vomit son repas » qui a activé le symptôme vomissements et sa caractéristique vomit des aliments
- Etc.

Il faudra ensuite sélectionner le symptôme principal (celui qui gêne le plus), ce qui déclenche le début du questionnaire.

**Signes d'accompagnement**

Si vous avez d'autres symptômes, cochez-les parmi la liste suivante.  
Ne cochez que les cases qui correspondent à votre cas.

Douleurs	▼
Ecoulement	▼
Gonflement localisé	▼
Peau	▲
<input type="checkbox"/> Jaunisse	
<input type="checkbox"/> Pâleur	
<input type="checkbox"/> Sueurs	
Troubles	▼
Problème général	▼
Problème localisé	▼

Au terme de l'interrogatoire, des diagnostics sont posés avec leur probabilité respective, un niveau de gravité, des conseils médicaux adaptés à la situation et des renvois vers l'encyclopédie.

Finalement, l'objectif est de placer en amont du search actuel un champ de saisie qui est le NLP que l'on cherche à construire et qui constitue une sorte d'aide. Il ne peut pas être substitué au champs search actuel pour raisons réglementaires (Dispositif Médical de classe IIa)

### 3 EXEMPLES

Les phrases sont soit tapées par l'utilisateur (mission principale) soit dites dans un micro avec transformation en un texte par une application speech to text (mission secondaire).

Le texte peut comprendre des mots médicaux ou du langage courant.

Ci-dessous 2 exemples imaginaires de la même situation décrite par un patient et par un médecin :

Langage patient :

*J'ai **très mal au foie**, c'est comme des **spasmes**, ça me **tire dans le dos**, je dois avoir de la **fièvre** genre **grippe**, je **tremble** comme une feuille, et j'ai vomi tout mon repas. Quand je **respire** ça me fait encore plus **mal sous les côtes**, et je suis obligé de me mettre en **chien de fusil** pour avoir moins mal. Ca monte dans **l'épaule droite**, exactement comme quand j'ai eu un **problème de vésicule** il y a 2 ans, le jour de l'anniversaire de mes 54 ans, mais là je n'avais pas de fièvre. J'ai l'impression d'avoir*

**les yeux un peu jaunes, je pisse tout marron et mes selles sont très jaunes. On m'a fait des radios et des tas de prises de sang, j'avais le bras tout bleu. D'après eux j'aurais une cystite.**

Langage médecin

**Femme** de 70 ans, présentant une **douleur brutale** de **l'hypochondre droit**, **transfixiante**, de **type spasmodique**, **irradient dans le dos à droite et dans l'épaule droite**, avec **position antalgique recroquevillée**, et douleur **augmentant à l'inspiration profonde**. La douleur est accompagnée de **vomissements alimentaires**. Elle est **fébrile**, avec un **subictère conjonctival**, des **urines foncées**, des **selles décolorées**, et des **tremblements** pouvant faire craindre un **sepsis**. Elle a présenté des crises équivalentes de **colique hépatique non fébrile** il y a 2 ans, et l'échographie avait montré un **hydrocholécyste** résolutif spontanément. La **biologie** a montré [...]. Les résultats d'imagerie ont montré [...]. On évoque ici une **cholécystite aigue** avec risque potentiel **d'angiocholite**.

Ces deux textes expriment le même scénario, mais les mots employés ne sont pas les mêmes, de même que la façon d'organiser le discours est différente.

Il peut arriver d'ailleurs que dans certains scénarios, le médecin mêle à son discours des mots patients, et les patients des mots médicaux.

Tous les mots en gras italique sont des mots présents dans le glossaire.

Le problème est que notre Search actuel n'est capable de comprendre des mots ou groupes de mots que s'ils sont tapés de façon identique à ce qu'il y a dans le glossaire. Une fonction existant dans le search de AccelRare qui utilise Medvir permet d'être plus fin en termes de proximité (douleur [.....] de l'hypochondre droit).

## 4 STRUCTURE DE LA BASE DE DONNÉES EXISTANTE

La base de données est structurée en de nombreuses tables qui permettent des extractions par des requêtes sql :

Libellé du symptôme ou de la questions/synonymes/mot clé du symptôme ou N° de la question.

Exemple résumé :

- Symptôme/Mal au ventre/mal au bide, douleur abdominale, crampes dans le ventre..../MAVEN
  - Localisation/La douleur se situe/ sous les côtes à droite/douleur de l'hypochondre droit, douleur hépatique.../745

- o Début/La douleur a débuté/Brutalement/douleur syncopale, douleur brutale.../1271
- Symptôme/Vomissement/gerber, dégoûter, vomir.../VMS
  - o Aspect/Le vomissement contient/Des aliments/Vomissements alimentaires,vomi tout le repas,vomi tout mon repas/1505

Les éléments complets se trouvent dans le fichier excel joint qui montrent l'extract dans son intégralité. Copie AllSymptomsAndQuestions

---

## 5 PROPOSITION D'ANALYSE DES MOTS DU GLOSSAIRE

---

Les mots reconnaissables sont :

- Des mots isolés, médicaux ou non (fièvre, dysurie, ganglions, paresthésies...)
- Des expressions toutes faites (mal au bide, crise d'épilepsie, pété la cheville, colique hépatique...)
- Des associations de mots précédés, suivis ou entrecroisés avec des mots sans valeur pour le search (**douleur épouvantable sous les côtes, très mal dans le bas-ventre, ventre très douloureux**). Seuls les mots en gras ont du sens, les autres n'ont pas d'importance.

C'est sur ce dernier type de mot que le travail est important à réaliser.

---

## 6 PISTES ET MATERIEL DISPONIBLE

---

### 6.1 Création de classes ?

---

Nous pourrions structurer le glossaire en classes.

Pour cela nous avons déjà en base de données avec leurs synonymes :

- Les zones du corps avec leurs synonymes
- Les organes du corps avec leurs synonymes
- Les fonctions (respiration, digestion, etc.) avec leurs synonymes
- Les types de question (Localisation, intensité, irradiation, augmentation, diminution, etc.) mais dont le travail de synonymie est très incomplet.

Par contre, aucun travail de classification automatisée des mots du glossaire n'a été faite.

Vous pouvez proposer une nouvelle organisation de notre glossaire.

## 6.2 Notion de radicaux

---

L'observation des mots en médecine montre que la quasi-totalité des mots sont univoques au-delà d'un certain nombre de caractères. Par exemple :

\_DOUL\_ est suffisant pour reconnaître tous les mots relatifs à la douleur

\_SAIGN\_ est suffisant pour reconnaître tous les mots relatifs au saignement

\_DOIG\_ est suffisant pour reconnaître tous les mots relatifs aux doigts

\_ESTOM\_ est suffisant pour reconnaître le mot estomac.

Etc.

Donc une phrase du type : « je me suis coupé le doigt, je saigne beaucoup, et j'ai des douleurs à l'estomac depuis tout à l'heure » va se résumer en  
DOIG\_SAIGN\_DOUL\_ESTOM\_

Cette réduction à des radicaux a peut-être une utilité.

---

## 7 RÉSULTAT ATTENDU : CORRESPONDANCES MOTS-NUMÉROS

---

Au final, le NLP devra pouvoir être en mesure de transformer une série de mots en des numéros utilisables par l'API.

### 7.1 Amélioration Front office

---

Le champ de saisie du NLP permet de proposer à l'utilisateur les concepts qu'il a reconnus, comme dans notre search actuel.

L'utilisateur peut supprimer les éléments qui lui semblent ne pas avoir de sens pour lui. Cette liste est alors transférée à la page de search où le processus actuel peut se poursuivre.

Cette détermination de l'aspect du front office est encore un peu floue car il faudra voir dans les usages et avec les performances du NLP quelle UX mettre en place.