

**MOHAMED
NAOUAR**

OPTIMISATION DE LA GESTION DES DONNÉES MÉDICALES

2023

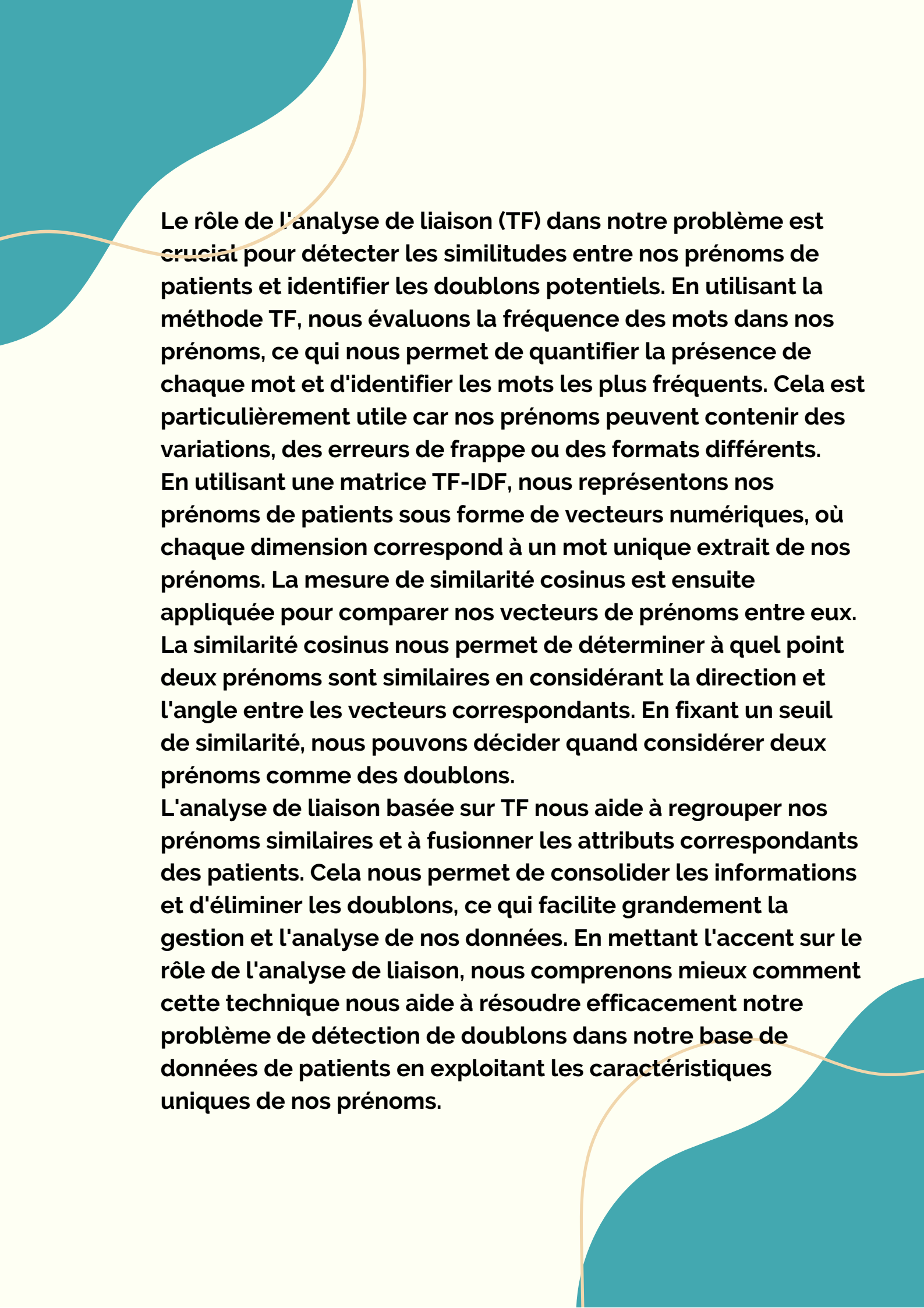
Introduction

Notre problème actuel concerne la gestion des données médicales au sein de notre système. Nous avons constaté qu'il existe des doublons et des incohérences dans les enregistrements des patients, ce qui entraîne une fragmentation des données et rend difficile l'obtention d'une vue d'ensemble précise.

Ces doublons et incohérences sont principalement dus à des erreurs de saisie, des variations dans l'enregistrement des noms et des informations incomplètes. Cela peut conduire à des difficultés dans l'identification des patients, dans l'analyse des opérations médicales effectuées et dans l'attribution des résultats des tests de laboratoire.

Pour résoudre ce problème, nous avons mis en place une approche basée sur la technique de liaison (linkage) et l'analyse de similarité des noms complets des patients. Nous utilisons le calcul de similarité cosinus et un seuil prédéfini pour détecter les doublons potentiels. Ensuite, nous fusionnons les enregistrements similaires en un seul enregistrement cohérent, en combinant les informations pertinentes.

Cette approche de fusion et d'analyse de liaison nous permet de nettoyer et d'harmoniser nos données médicales, ce qui facilite leur utilisation et leur analyse ultérieure. En consolidant les enregistrements des patients, nous obtenons une vue plus complète et précise de leur historique médical, ce qui contribue à améliorer la qualité des soins et la prise de décision médicale.



Le rôle de l'analyse de liaison (TF) dans notre problème est crucial pour détecter les similitudes entre nos prénoms de patients et identifier les doublons potentiels. En utilisant la méthode TF, nous évaluons la fréquence des mots dans nos prénoms, ce qui nous permet de quantifier la présence de chaque mot et d'identifier les mots les plus fréquents. Cela est particulièrement utile car nos prénoms peuvent contenir des variations, des erreurs de frappe ou des formats différents. En utilisant une matrice TF-IDF, nous représentons nos prénoms de patients sous forme de vecteurs numériques, où chaque dimension correspond à un mot unique extrait de nos prénoms. La mesure de similarité cosinus est ensuite appliquée pour comparer nos vecteurs de prénoms entre eux. La similarité cosinus nous permet de déterminer à quel point deux prénoms sont similaires en considérant la direction et l'angle entre les vecteurs correspondants. En fixant un seuil de similarité, nous pouvons décider quand considérer deux prénoms comme des doublons.

L'analyse de liaison basée sur TF nous aide à regrouper nos prénoms similaires et à fusionner les attributs correspondants des patients. Cela nous permet de consolider les informations et d'éliminer les doublons, ce qui facilite grandement la gestion et l'analyse de nos données. En mettant l'accent sur le rôle de l'analyse de liaison, nous comprenons mieux comment cette technique nous aide à résoudre efficacement notre problème de détection de doublons dans notre base de données de patients en exploitant les caractéristiques uniques de nos prénoms.

Mohamed	Naouar	Opération A	Labo A
John	Smith	Opération B	Labo B
Naouar	Mohamed	Opération C	Labo A
Emma	Johnson	Opération A	Labo B
Liam	Brown	Opération B	Labo
Sophia	Smith	Opération C	Labo B
Ava	Wilson	Opération A	Labo A
Lucas	Clark	Opération B	Labo B
Isabella	Davis	Opération C	Labo A
Jackson	Thomas	Opération A	Labo B
Mohamed	Naouar	Opération B	Labo A
Noah	Lee	Opération A	Labo A
Olivia	Johnson	Opération B	Labo B
Ethan	Williams	Opération C	Labo A
Amelia	Clark	Opération A	Labo B
Michael	Wilson	Opération B	Labo A
Sophia	Smith	Opération C	Labo B
William	Johnson	Opération A	Labo A
Smith	Sophia	Opération vC	Labo vB
Smith	John	Opération c	Labo v

ERROR: Invalid index 27 of 15

ID	Prénom	Nom	Opération	Labo	Nom Complet
0	Naouar,Mohamed	Naouar,Mohamed	Opération C,Opération B,Opération A	Labo A	Naouar Naouar
1	John,Smith	John,Smith	Opération B,Opération c	Labo B,Labo v	John John
2	Emma	Johnson	Opération A	Labo B	Emma Johnson
3	Liam	Brown	Opération B	Labo	Liam Brown
4	Smith,Sophia	Smith,Sophia	Opération C,Opération vC	Labo B,Labo vB	Smith Smith
5	Ava	Wilson	Opération A	Labo A	Ava Wilson
6	Lucas	Clark	Opération B	Labo B	Lucas Clark
7	Isabella	Davis	Opération C	Labo A	Isabella Davis
8	Jackson	Thomas	Opération A	Labo B	Jackson Thomas
9	Noah	Lee	Opération A	Labo A	Noah Lee
10	Olivia	Johnson	Opération B	Labo B	Olivia Johnson
11	Ethan	Williams	Opération C	Labo A	Ethan Williams
12	Amelia	Clark	Opération A	Labo B	Amelia Clark
13	Michael	Wilson	Opération B	Labo A	Michael Wilson
14	William	Johnson	Opération A	Labo A	William Johnson

ID	Prénom	Nom	Operation	Labo	Nom Complet
0	Naouar,Mohamed	Naouar,Mohamed	Opération B,Opération C,Opération A	Labo A	Naouar Naouar
1	John,Smith	John,Smith	Opération c,Opération B	Labo B,Labo v	John John
2	Emma	Johnson	Opération A	Labo B	Emma Johnson
3	Liam	Brown	Opération B	Labo	Liam Brown
4	Sophia,Smith	Sophia,Smith	Opération C,Opération vC	Labo B,Labo vB	Sophia Sophia
5	Ava	Wilson	Opération A	Labo A	Ava Wilson
6	Lucas	Clark	Opération B	Labo B	Lucas Clark
7	Isabella	Davis	Opération C	Labo A	Isabella Davis
8	Jackson	Thomas	Opération A	Labo B	Jackson Thomas
9	Noah	Lee	Opération A	Labo A	Noah Lee
10	Olivia	Johnson	Opération B	Labo B	Olivia Johnson
11	Ethan	Williams	Opération C	Labo A	Ethan Williams
12	Amelia	Clark	Opération A	Labo B	Amelia Clark
13	Michael	Wilson	Opération B	Labo A	Michael Wilson
14	William	Johnson	Opération A	Labo A	William Johnson