

# NLP DÉCODÉ : CRÉATION ET COMPLÉTION D'UNE BASE DE DONNÉES STRUCTURÉE EN ANALYSANT DES DOCUMENTS TECHNIQUES

Sur le marché actuel, les entreprises sont confrontées à une abondance d'informations provenant de diverses sources telles que les fournisseurs, les clients, les prospects et les bases de données publiques. Cependant, ces informations sont souvent dispersées et non structurées, rendant leur exploitation difficile pour les activités commerciales, marketing et relationnelles. C'est là que la création et la complétion d'une base de données structurée, en analysant des documents techniques tels que des fiches descriptives, deviennent essentielles. Cette approche permet de simplifier l'accès et l'exploitation des informations, facilitant ainsi la prise de décisions basées sur des données fiables et pertinentes.

## Méthodologie

La méthodologie utilisée pour créer et compléter la base de données structurée repose sur l'**identification des informations pertinentes** et l'**adaptation de l'extraction en fonction du niveau de normalisation des documents**.

### TROIS TYPES DE DOCUMENTS PEUVENT ÊTRE PRIS EN COMPTE :

1

Documents normés  
avec repères clairs

2

Documents normés  
avec parties en texte libre

3

Documents  
peu normés

L'utilisation de mots clés, de balises et d'expressions régulières permet d'extraire la bonne séquence d'informations, qui est ensuite normalisée et intégrée dans la base de données.

## Principales difficultés

Les principales difficultés rencontrées lors de cette démarche sont **la variabilité de l'expression de l'information** et **le taux d'exploitation de celle-ci**. En effet, les documents provenant de différentes sources peuvent présenter **des formats et des structures variés**, ce qui rend l'extraction des informations plus complexe. De plus, certaines informations peuvent être **incomplètes** ou **contradictoires**, ce qui nécessite **une validation et une vérification rigoureuses**.

## Bénéfices

La création et la complétion d'une base de données structurées offrent de nombreux bénéfices pour les entreprises :

- > Exploitation de documents insuffisamment valorisés :**  
En extrayant les informations pertinentes des documents techniques, les entreprises peuvent exploiter des sources d'informations sous-utilisées, augmentant ainsi leur connaissance des produits, des clients et du marché.
- > Automatisation des tâches d'extraction manuelles :**  
La mise en place d'une base de données structurées permet d'automatiser les tâches d'extraction d'informations, ce qui réduit les erreurs humaines et libère du temps pour des activités à plus forte valeur ajoutée.
- > Identification de documents incomplets et d'informations contradictoires :**  
La mise en évidence des informations manquantes ou contradictoires dans les documents permet aux entreprises de prendre des mesures correctives, améliorant ainsi la qualité des données utilisées pour la prise de décisions.
- > Utilisation des informations pour le calcul de KPI et le dashboarding :**  
Une base de données structurée permet d'extraire des informations pour le calcul des indicateurs clés de performance (KPI) et la création de tableaux de bord dynamiques, offrant une vision globale et en temps réel de la performance de l'entreprise. On peut également réutiliser ces données dans des segmentations, modèles prédictifs, recommandations, ....

En conclusion, la création et la complétion d'une base de données structurée à partir de l'analyse de documents techniques représentent une étape cruciale pour simplifier l'accès et l'exploitation des informations dans le contexte des activités commerciales, marketing et relationnelles. En normalisant et en intégrant les informations extraites, **les entreprises peuvent exploiter pleinement les données disponibles, automatiser des tâches manuelles, identifier des problématiques et utiliser les informations pour la prise de décisions éclairées.** Cela ouvre la voie à **une utilisation optimale des outils de BI et de Data Science**, favorisant ainsi la **croissance et la compétitivité des entreprises** dans un environnement concurrentiel en constante évolution.

Si vous avez la moindre question concernant la problématique de création et complétion d'une base de données structurées à partir de l'analyse de documents techniques ou sur le NLP en général, n'hésitez pas à nous contacter : [contact@aid.fr](mailto:contact@aid.fr)