

# Documentation

Explication de termes technique

- Embedding model (représentations vectorielles)
- Embedding model dans une société

# Embedding model (représentations vectorielles)

## Qu'est-ce qu'un embedding ?

Un **embedding** est une représentation vectorielle d'un mot, d'une phrase ou même d'un texte complet dans un espace mathématique de haute dimension. L'objectif est de transformer des éléments de texte en vecteurs (des suites de nombres) qui capturent leur signification.

Par exemple, prenons les mots suivants :

- "chat"
- "chien"
- "voiture"

Un modèle d'embedding va attribuer à ces mots des **vecteurs numériques**. Ces vecteurs seront placés dans un espace multidimensionnel, et la distance entre ces vecteurs peut refléter la similarité sémantique entre les mots. Par exemple :

- Les vecteurs de "chat" et "chien" devraient être proches dans l'espace vectoriel car ces mots sont sémantiquement similaires (ce sont des animaux domestiques).
- Le vecteur de "voiture" serait éloigné de ceux de "chat" et "chien", car ce sont des concepts différents.

## Contexte d'un mot et fenêtre de tokens :

L'un des éléments clés d'un modèle d'**embedding** performant est sa capacité à comprendre le **contexte** dans lequel un mot apparaît. Par exemple, le mot "banc" peut signifier "un meuble" ou "une institution financière" en fonction du contexte.

Voici deux exemples avec la même **fenêtre de contexte** :

- **Phrase 1** : "Je suis allé au banc de touche."
- **Phrase 2** : "J'ai ouvert un compte au banc."

Un modèle avec une **fenêtre de contexte large** va prendre en compte non seulement le mot "banc" lui-même, mais aussi les mots qui l'entourent, comme "de touche" dans le premier cas et

"compte" dans le second. Grâce à une grande fenêtre de contexte, il comprendra que "banc" se réfère à un **siège** dans le premier exemple, et à une **institution financière** dans le deuxième.

## Pourquoi une grande fenêtre de contexte est importante ?

Un modèle avec une **fenêtre de contexte large** peut prendre en compte plus de mots autour du mot cible pour mieux comprendre son sens. Par exemple :

- Dans la phrase : "Le chat mange une souris."
- La fenêtre de contexte pour le mot "mange" pourrait inclure les mots "chat" et "souris", ce qui aide le modèle à comprendre que "mange" fait référence à une action de nourriture, et non à une autre signification possible (comme "manger un repas" dans un autre contexte).

Une **fenêtre de contexte large** signifie que le modèle peut analyser des sections plus longues du texte, ce qui améliore la compréhension du sens d'un mot, même dans des phrases complexes.

## Exemples concrets avec "nomic-embed-text" :

Supposons que nous utilisons un modèle "**nomic-embed-text**" pour générer des embeddings pour ces deux phrases :

1. **Phrase 1** : "Le chat dort sur le canapé."
2. **Phrase 2** : "Le chat a attrapé une souris."

Le modèle va générer des embeddings pour chaque mot en prenant en compte le contexte autour d'eux.

- Pour le mot "chat", le modèle pourrait produire un vecteur similaire dans les deux phrases, car il est utilisé dans des contextes relativement similaires (un animal domestique). Cependant, le modèle prendra aussi en compte des mots comme "dort" dans la première phrase et "attrapé" dans la deuxième, ajustant l'embedding de "chat" pour capturer les différences de contexte.
- Le mot "dort" dans "Le chat dort sur le canapé" aura un contexte avec "chat" et "canapé", ce qui le placera près d'autres mots associés à des actions de repos ou de sommeil, comme "dormir" ou "repos".
- Le mot "attrapé" dans "Le chat a attrapé une souris" aura un contexte avec "chat" et "souris", plaçant ce mot plus près de mots liés à l'action de chasser ou de capturer.

## Conclusion :

Le modèle "**nomic-embed-text**" est un modèle performant qui utilise une **fenêtre de contexte large** pour analyser les relations entre les mots dans un texte. Cela lui permet de produire des **embeddings** plus précis et contextuellement adaptés, ce qui est essentiel pour des tâches telles que la recherche sémantique, la traduction automatique ou l'analyse de texte, où il est important de comprendre le sens global d'un mot en fonction de son contexte spécifique.

En résumé, une grande fenêtre de contexte permet au modèle de mieux "comprendre" le texte dans son ensemble et de produire des embeddings qui reflètent correctement le sens des mots dans chaque situation.

# Embedding model dans une société

L'utilisation d'un modèle d'**embedding** comme "**nomic-embed-text**" dans une société qui a une grande quantité de documentation (manuels, guides, FAQ, rapports, etc.) peut être extrêmement utile pour répondre efficacement à des demandes internes ou externes. Voici comment un tel modèle pourrait être utile pour automatiser et améliorer la gestion des demandes et l'accès à la documentation dans votre entreprise :

## Améliorer la recherche de réponses pertinentes dans la documentation :

Un modèle d'**embedding** permet de transformer la documentation de la société en vecteurs de mots ou de phrases, qui capturent le sens global du texte plutôt que de se concentrer uniquement sur des mots-clés exacts. Cela signifie qu'il peut trouver des réponses même si les termes de la question de l'utilisateur ne correspondent pas exactement aux mots de la documentation.

### Exemple :

Supposons qu'un employé pose une question dans un chatbot comme :

“Comment installer le logiciel sur un PC Windows ?”

Un modèle d'**embedding** avec une **fenêtre de contexte large** va analyser non seulement les mots exacts dans la question (comme "installer", "logiciel", "Windows"), mais aussi les relations sémantiques, ce qui lui permettra de trouver la réponse appropriée dans la documentation, même si celle-ci n'utilise pas exactement les mêmes termes.

Par exemple, dans la documentation, il pourrait y avoir une section avec des phrases comme :

“Guide d'installation du programme sur un système Windows" ou "Procédure d'installation sur une machine sous Windows".

Un modèle avec des embeddings pourra comprendre la relation entre ces phrases et la question de l'utilisateur, même si les mots ne sont pas exactement identiques.

## Automatiser les réponses aux demandes fréquentes (FAQ) :

Dans de nombreuses entreprises, il y a un grand nombre de questions récurrentes. Un modèle d'**embedding** peut être utilisé pour automatiser les réponses à ces questions fréquemment posées (FAQ).

### **Exemple :**

Supposons que plusieurs employés demandent des informations sur le processus de demande de congés ou sur la procédure à suivre en cas de panne technique. Plutôt que de faire une recherche manuelle dans la documentation à chaque fois, le modèle d'**embedding** peut rapidement trouver la réponse correspondante et la fournir de manière autonome.

## Améliorer l'assistance interne avec un chatbot ou une interface de recherche intelligente :

Un modèle comme "**nomic-embed-text**" peut être intégré dans un chatbot ou une interface de recherche interne pour aider les employés à trouver rapidement des informations pertinentes dans la documentation de l'entreprise.

- **Chatbot intelligent** : Lorsqu'un employé pose une question via le chatbot, ce dernier peut utiliser l'IA pour analyser la demande, rechercher les informations pertinentes dans la documentation et y répondre en temps réel.
- **Recherche contextuelle** : Lorsqu'un employé utilise une barre de recherche dans la documentation interne, un modèle d'**embedding** peut fournir des résultats non seulement basés sur la correspondance exacte des mots-clés, mais aussi en fonction du contexte, ce qui améliore la qualité des résultats et rend la recherche plus intuitive.

## Réduction du temps de réponse et des erreurs humaines :

Les employés qui utilisent des systèmes manuels ou des moteurs de recherche traditionnels peuvent perdre du temps à trouver des informations dans une documentation longue et complexe. Un modèle d'**embedding** améliore l'efficacité de ce processus, en réduisant les erreurs humaines et en fournissant des réponses plus rapidement, sans avoir à passer par un processus de recherche fastidieux.

### **Exemple :**

Si un employé a une question sur une procédure technique et que la documentation est très détaillée, un moteur de recherche traditionnel pourrait renvoyer plusieurs résultats, dont certains sont peu pertinents. Un système d'IA basé sur des embeddings trouvera directement les passages les plus pertinents en comprenant le contexte de la question et en analysant la documentation en profondeur.

## Personnalisation des réponses selon les besoins :

Un modèle d'**embedding** peut également s'adapter aux besoins spécifiques d'un utilisateur ou d'un département. Par exemple, si un employé fait une demande sur un sujet spécifique (comme des informations sur le département IT), le modèle peut filtrer les réponses et se concentrer uniquement sur les sections de la documentation pertinentes à ce département.

## Maintenance et mise à jour de la documentation :

Un modèle d'**embedding** peut aussi être utilisé pour identifier les lacunes ou les incohérences dans la documentation existante. Par exemple, si des questions fréquemment posées ne trouvent pas de réponse directe dans la documentation, cela peut indiquer que la documentation doit être mise à jour ou enrichie. L'IA pourrait aussi suggérer de nouvelles sections à ajouter.

---

## Exemple concret dans une entreprise :

Imaginez une entreprise de services informatiques qui fournit une documentation détaillée sur l'utilisation de ses logiciels, ainsi que des procédures internes pour ses employés. Supposons que :

- Un employé du service client pose la question suivante : "**Comment résoudre un problème de connexion à notre réseau VPN ?**"
- Grâce à un modèle d'**embedding** intégré à un chatbot, la réponse pourrait être trouvée dans la documentation, même si la question est formulée différemment de la manière dont le problème est décrit dans le guide. Par exemple, le chatbot pourrait répondre par :

“ Pour résoudre un problème de connexion VPN, veuillez vérifier si votre connexion internet est stable et assurez-vous que votre logiciel VPN est bien à jour. Vous trouverez des instructions détaillées dans le guide d'utilisation de votre VPN dans le chapitre 5."

Le chatbot pourrait aussi proposer une liste de solutions possibles en fonction du contexte de l'entreprise et de l'historique des demandes.

## Conclusion :

En résumé, un modèle d'**embedding** comme "**nomic-embed-text**" serait extrêmement utile pour une entreprise disposant de grandes quantités de documentation et cherchant à répondre automatiquement aux demandes internes ou externes. Il permettrait d'améliorer la recherche de réponses, d'automatiser des tâches récurrentes, de réduire le temps de réponse et d'assurer une meilleure gestion des connaissances au sein de l'entreprise.