L'exploration de l'Encyclopédie par l'intelligence artificielle : problèmes, méthodes, premiers résultats.

Morgan Blangeois, Aurelia Vasile, Naïs Sabatier, Henri Galinon

Intervention à la journée « Qu'est-ce que l'IA peut faire pour vous ? » du 20 novembre 2024, MSH Clermont-Ferrand

L'exploration de l'*Encyclopédie* par l'intelligence artificielle : problèmes, méthodes, premiers résultats.

Présentation du projet Encycloped·IA, porté par la MSH de Clermont-Ferrand et du laboratoire du PHIER, par
Morgan BLANGEOIS (CLERMA), Henri GALINON (PHIER),
Naïs SABATIER (PHIER), et Aurélia VASILE (MSH).

*Les illustrations sont extraites de la présentation, intitulée « L'exploration de l'*Encyclopédie *par l'intelligence artificielle : problèmes, méthodes, premiers résultats », présentée lors de la journée d'études.*

0. Introduction.

Le projet Encycloped·IA consiste dans la réunion de différent·es chercheur·euses autour de l'exploration du texte de l'*Encyclopédie* de Diderot et D'Alembert grâce aux outils d'intelligence artificielle, en particulier les LLM (*large language model*), adaptés aux grands ensembles textuels. Son objet consiste dans l'ensemble des dix-sept volumes de textes, des onze volumes de planches[1] et du paratexte de l'*Encyclopédie*, publiée de 1749 à 1772. Dans ce grand ensemble, Encycloped·IA envisage d'interroger plus spécifiquement le discours économique présent d'une manière éclatée dans le dictionnaire. En effet, étant donné que l'économie politique apparait au mitan du siècle sans pour autant qu'une catégorie disciplinaire lui soit attribuée dans l'ordre encyclopédique, les textes qui parlent des échanges marchands, de la valeur politique et morale du commerce, et d'une manière générale de la production et de la consommation de biens, n'ont pas de domaine éprouvé au 18ème siècle.

Si la taille monumentale de l'ouvrage doit décevoir toute prétention d'une lecture exhaustive des articles, c'est bien plutôt la nature même du discours économique et de sa distribution encyclopédique qui est problématique pour apprécier le corpus car il n'existe ni groupe de contributeurs, ni domaine lexicographique qui autoriserait un découpage *a priori* des articles d'économie. Par conséquent, il ne suffit pas d'aller voir les définitions des concepts économiques pour être certains que l'*Encyclopédie* les considère uniquement comme tel : dit plus simplement, il y a un écart entre la définition d'un mot, comme « luxe » par exemple, et ce qu'on pense du *luxe*, dans d'autres articles du dictionnaire.

A partir de cet ensemble de problèmes, caractéristiques de l'objet à explorer et de la catégorie à examiner, quelles peuvent être les applications d'outils d'intelligence artificielle ?

[1] Textes descriptifs qui accompagnent les planches.

1. Usages de l'IA sur le corpus économique de l'*Encyclopédie.*

Il s'agit dans un premier temps de définir les attentes qu'un·e chercheur·euse en SHS pourraient attendre de l'intelligence artificielle, que la méthode classique de la lecture et de l'interprétation des textes ne pourrait atteindre.

1.1 Approche qualitative des concepts économiques.

Avant même tout traitement par des logiciels de traitement automatique du langage, l'océrisation des textes dans des éditions numériques, comme celle de l'ENCCRE[2], permet de faire des recherches d'occurrences et de co-occurrences. Cependant, ces recherches restent limitées à une approche quantitative, car elles ne distinguent pas les significations et usages contextuels des mots. Pour saisir le sens précis de certains concepts dont la signification est en tension au 18ème siècle, il faut lire la phrase, voire l'article entier, afin de comprendre le sens.

Par exemple, le terme « *commerce* », un concept économique aux sens variables, apparaît plus de 5000 fois. Si on veut en comprendre la signification précise, il est nécessaire de lire chaque occurrence dans le contexte de l'article. Ainsi, « commerce » dans le contexte suivant : « *le commerce des femmes* » signifie « relation sexuelle », mais dans la phrase : « *le commerce avec les Anglois* » il signifie « relation économique », alors que l'expression : « *hors du commerce des hommes* » signifie « en l'absence de relation sociale ».

L'IA, puisqu'elle permet une approche qualitative du texte pourrait nous faciliter les traitements de cette nature, c'est-à-dire le traitement sémantique des concepts.

1.2 Explorations de nouveaux textes.

Par ailleurs, nous pourrions espérer que les outils de type LLM révèlent dans le Dictionnaire des textes sur l'économie que la recherche classique n'aurait pas explorés ou découverts, en raison de la fragilité de la catégorie de l'économie politique dans les dictionnaires du 18ème siècle. Au siècle de l'*Encyclopédie*, il est question d'une véritable anthropologie de l'échange, ce qui laisse penser que des textes sur l'économie, la circulation et l'échange de biens peuvent apparaitre dans des domaines variés tels que la géographie, la médecine ou l'histoire naturelle, qui examinent chacun des relations entre individus depuis des catégories scientifiques distinctes.[3]

Le traitement proprement sémantique des textes pourrait nous permettre de « traquer » la pensée économique partout où elle se trouve, en partant non pas des auteurs ou des domaines mais bien des représentations linguistiques construites par les LLM.

1.3 Usages internes et conflit des définitions du Dictionnaire.

Enfin, comme l'objet est une encyclopédie, son intérêt est de *fixer des significations* dans un espace polémique où les représentations s'affrontent, notamment sur la question économique

[2] Edition numérique collaborative et critique de l'*Encyclopédie* : http://enccre.academie-sciences.fr/encyclopedie/
[3] Voir section n°2. Premières explorations techniques de l'*Encyclopédie* et premiers résultats obtenus.

qui est un point chaud du 18ème siècle en particulier sur la question de la libéralisation des échanges et la querelle du luxe. L'intérêt d'une I.A, dans ce contexte, serait de pouvoir interagir avec un *bot* capable non seulement de fournir la définition du mot en question, mais aussi d'expliquer son usage général, voire ses emplois contradictoires dans l'ouvrage.

Par exemple, à la question « *Comment l'Encyclopédie définit-elle le luxe* », l'outil pourrait à la fois diriger vers l'article de Saint-Lambert, dont l'article LUXE est une tentative de neutralisation de la critique morale du luxe comme vice, et en même temps pointer vers d'autres usages du terme dans d'autres articles pour affiner la « place du luxe » dans l'univers moral des encyclopédistes par exemple.

2. Premières explorations techniques de l'*Encyclopédie* et premiers résultats obtenus.

Forts de ces attentes, nous avons commencé à tester des méthodes et des modèles issues de différentes branches de l'intelligence artificielle : dans un premier temps, l'objectif du projet Encycloped·IA a été de comprendre le fonctionnement de ces méthodes et les résultats auxquels on pouvait aboutir : l'*Encyclopédie* de Diderot et D'Alembert fut l'objet qui nous a réuni : elle a permis la jonction entre une démarche réflexive en philosophie et en histoire des idées, avec une démarche technique.

2.1 Vectorisation avec Word2vec et modèle de représentation des mots.

La première méthode que nous avons explorée s'appelle **Word2Vec**. C'est une technique d'intelligence artificielle qui permet de créer un modèle de représentation des mots d'un corpus en se basant sur le contexte d'apparition de ces mots dans la phrase. Ce modèle de représentation s'appelle *word embedding* (plongement lexical) : il peut être produit par d'autres méthodes que le groupe de modèles de Word2vec, et se présente concrètement sous la forme d'une liste de vecteurs numériques. C'est le contexte d'apparition du mot dans le texte qui constitue l'élément clé de Word2Vec : les mots qui partagent des contextes similaires auront des vecteurs numériques proches. Cela conduit à dire qu'il y a une *proximité sémantique* entre ces mots-là, dans un contexte langagier déterminé (un texte ou un ensemble de textes). Par exemple : dans un texte qui présente des animaux domestiques, « chat » et « chien » peuvent être des mots plus proches sémantiquement que les mots « chat » et « félin ».

Grace à Word2Vec nous avons généré un modèle de vecteurs pour tous les mots de l'*Encyclopédie*. A partir de ce modèle on peut maintenant repérer des similarités sémantiques entre les mots selon la 'vision' du dictionnaire.

Représentations des vecteurs : Word2vec

- La production d'un modèle de vecteurs pour tous les mots de l'*Encyclopédie*.

  « *word embedding* »

- *Ex*. les vecteurs du mot « **révolution** » :

```
print(model.wv['révolution'])
```

```
[0.31884596 -0.5381088 0.34372813 -1.0202217 -1.6776458 0.9757068 -1.4085875 -1.3355446
-0.7059597 0.24516176 -1.6001616 -1.2312181 -1.7969168 -0.6510634 2.3834803 -0.18950042 -
0.9099625 2.1977813 0.32053995 0.61372584 -1.5292628 0.06976119 0.31420055 0.8349429 -
0.56572527 -0.27992362 -1.6315123 0.5707176 -0.27478182 -2.261454 0.6054682 -2.7852979
1.4380549 -0.45856252 -0.03459271 1.2076389 0.64418477 -1.5595444 -1.9762436 -0.7825642 -
1.6353741 1.3280544 -0.4463582 -1.9173613 -2.0474184 -0.750122 -0.7650286 -0.10337191 -
0.34081945 -0.4005983 -1.0439534 0.83680725 -0.5914661 1.8397526 -0.697791 -1.1686355
0.20033859 -2.1497142 -2.4227314 -1.4242734 -1.4121526 2.5846162 1.2210379 0.07510211
0.06433702 -0.26146853 -0.99041116 0.33987406 0.0567982 2.2613618 1.7509662 1.2902793
2.9852033 -2.728478 -0.31732777 1.5251579 -1.9090866 -3.5765302 -0.2703839 0.8660372 -
0.9409886 -0.8489094 -2.0763397 1.8497517 -1.1066146 1.8537375 -1.5390269 -0.1658735 -
0.7378688 0.24992715 -0.8513995 -0.04721072 -1.1341677 -0.13458997 0.05696878 1.0442783 -
1.3501227 1.5403774 -1.7299869 -0.11297539]
```

Application encyclopédique : ESCLAVAGE

```
[('servitude', 0.7714689373970032),
('despotisme', 0.7502161860466003),
('tyrannie', 0.7018572688102722),
('avilissement', 0.6787083745002747),
('indépendance', 0.6602827310562134),
('anarchie', 0.654937744140625),
('barbare', 0.6460627913475037), …]
```

score de similarité à « **esclavage** »

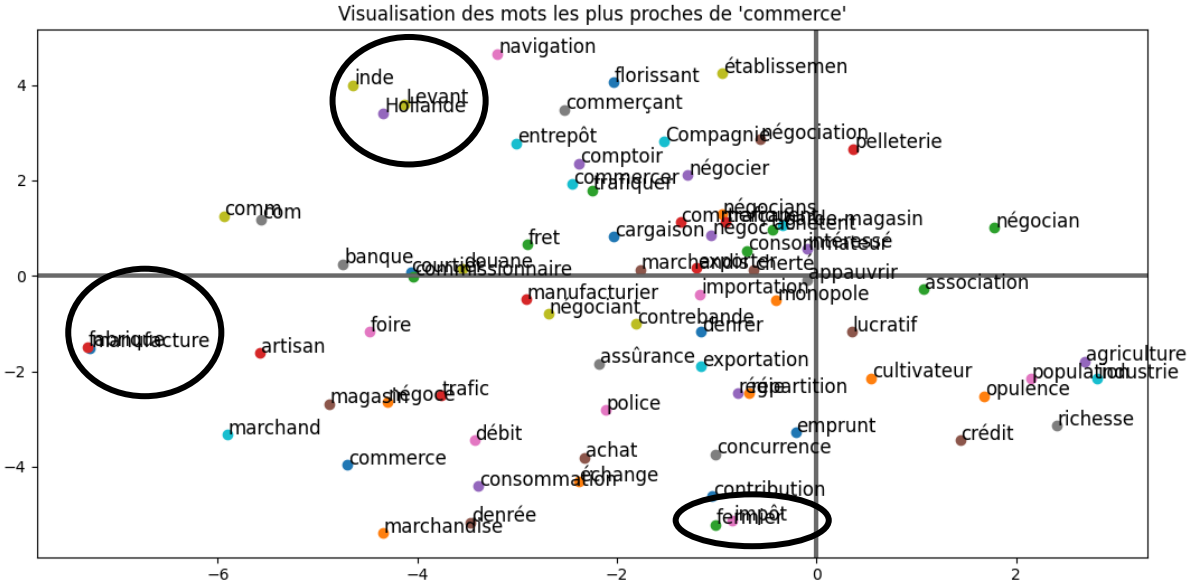
Application encyclopédique : COMMERCE

```
[('négoce', 0.7906147241592407),
('trafic', 0.7573156952857971),
('banque', 0.6853881478309631),
('négociant', 0.678597092628479),
('débit', 0.6777104735374451),
('denrée', 0.6698488593101501),
('exportation', 0.6679540276527405),
('consommation', 0.6577104926109314),
('marchandise', 0.6559945940971375)]
```

score de similarité à « **commerce** »

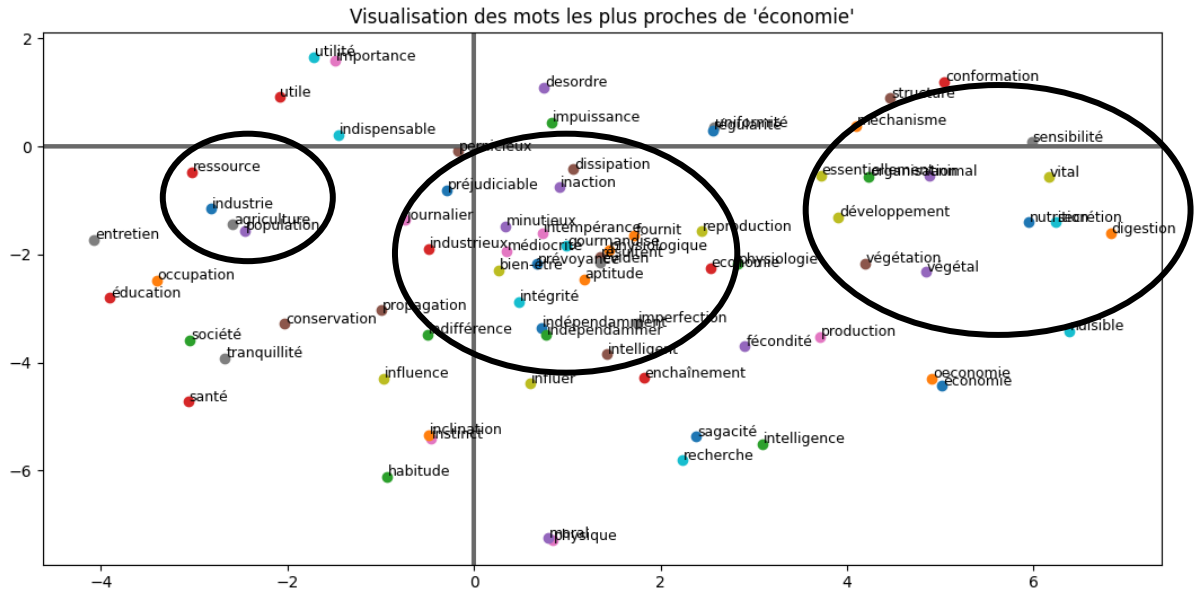
La représentation de la similarité est plus parlante si on utilise une visualisation en 2D. Elle permet de repérer non seulement la proximité par rapport au mot d'origine, mais le regroupement des mots proches entre eux formant une sorte de *cluster* thématique.

Voici par exemple une visualisation 2D des vecteurs de « commerce » :



Pour faire une analyse rapide et non exhaustive du résultat, l'*Encyclopédie* utilise le mot « commerce » dans plusieurs contextes : un contexte de description des espaces économiques dynamiques au 18ème siècle, d'où le cluster des mots « Inde, Levant, Hollande » (qui sont également proches les uns des autres). Mais nous voyons également des quasi-synonymes comme « fabrique » et « manufacture », ou la très grande proximité des termes « Ferme » et « Impôt » dans le discours sur le commerce, ce qui est intéressant puisque cela montre que le problème fiscal dans l'*Encyclopédie*, est corrélé aux débats sur la Ferme générale et donc sur le mode de prélèvement de l'impôt.

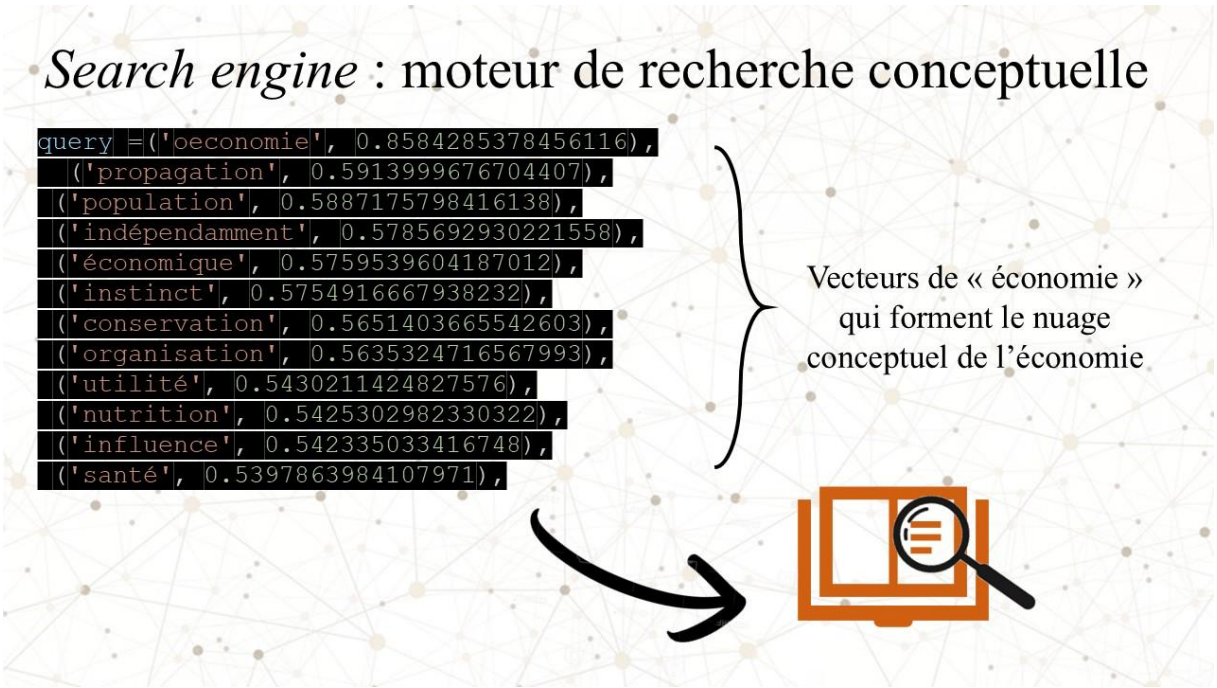
Autre exemple, avec le mot « économie » :



Pour les vecteurs du mot « économie », nous voyons que le contexte d'utilisation du terme est très flottant dans l'*Encyclopédie* parce que le mot n'a pas la définition contemporaine de l'« échange marchand » comme sens principal. Aussi on voit que l'*Encyclopédie* parle d'économie au sens de l'économie politique c'est-à-dire l'organisation de la vie civile et des ressources économiques ('ressource' 'industrie' 'agriculture' 'population', etc.). Il est possible de voir également que l'économie est un mot utilisé dans un contexte de discours moraux puisque nous constatons dans le nuage une concentration de termes éthiques relatifs aux passions. Enfin, l'*Encyclopédie* parle beaucoup d'économie au sens de l'économie animale, c'est-à-dire l'organisation et la génération des êtres vivants, végétaux et animaux (l'économie animale est l'ancêtre de la biologie recoupait des discours de médecine).

2.2 *Search engine* : « moteur de recherche conceptuelle ».

A partir de ces vecteurs et de ce premier modèle de représentation du texte encyclopédique, il est possible d'aller plus loin et de fabriquer un moteur de recherche *ad hoc*, pour trouver dans l'*Encyclopédie* des documents « cachés », relatifs à un sujet donné. Si l'on considère la liste des mots similaires générés par Word2Vec comme un ensemble de mots qui définirait le périmètre sémantique d'un autre mot, alors nous pouvons considérer cet ensemble comme un « concept ». L'objectif du moteur de recherche serait alors de trouver tous les textes qui évoquent ce concept, sans forcément utiliser le mot-clé lui-même. Par exemple, trouver les articles qui évoquent la question du commerce sans utiliser dans la recherche le mot « commerce » mais uniquement les mots de son champ sémantique de proximité, c'est-à-dire la liste générée par Word2vec grâce au plongement lexical.



Pour cela nous avons utilisé une méthode statistique qui calcule l'importance d'un mot dans un document par rapport à l'ensemble de documents du corpus : **la méthode TF-IDF**. Elle produit de nouveaux scores numériques qui indiquent l'importance des mots dans le document.

En combinant les deux scores pour chaque mot du champ sémantique du concept : le score de proximité généré par word2vec et le score TF-IDF pour chaque document, on obtient une liste de documents qui sont censés « parler » de l'économie, du commerce, du luxe, etc.

Par ailleurs, la notion de « *document* » dans le cadre de cette première exploration technique est elle-même particulière, car nous avons pris le texte de l'*Encyclopédie* comme un tout et nous l'avons décomposé en 26 000 fichiers d'environ 1000 mots qui constituent les documents du corpus (« *chunks* »). Aussi, un document ce n'est pas une entrée du dictionnaire mais un morceau de texte (qui peut donc recouper plusieurs articles ou seulement un segment d'un article plus long). De ce fait, c'est une méthode insatisfaisante actuellement, mais elle nous permet néanmoins d'explorer et de découvrir des textes inattendus.

Quels sont donc les résultats du *search engine*, si nous lui demandons de nous publier la liste de tous les documents qui parlent d'« *économie* » ?

Si nous analysons les premiers ***chunks*** vers lesquels pointe la machine, nous constatons d'abord que l'*Encyclopédie* parle d'économie majoritairement dans les articles d'économie animale, c'est-à-dire des articles de biologies et de médecine. Le modèle sort des textes qui parlent de la génération, de digestion, de la conservation de la santé et de la vie par exemple, ou de l'organisation des corps. Cependant, nous constatons aussi que le modèle trouve quelques textes, relativement haut dans la liste, qui parlent plus précisément d'économie politique. Il est possible d'avancer l'hypothèse que ces textes d'économie politique partagent conceptuellement des éléments avec l'économie animale (l'idée de la propagation par exemple, ou de la conservation, dont parlent les articles POPULATION et EDUCATION par exemple). Par contre, il y a un absent de cette liste : les articles – pourtant majoritaires en nombre dans le dictionnaire – qui parlent d'économie rustique (c'est-à-dire de la gestion du domaine agricole) : nous nous interrogeons sur les raisons de cette absence, qui confirme le caractère exploratoire et partiel de la mise en place de ce premier outil.

Search engine : moteur de recherche conceptuelle

Résultats avec le nuage conceptuel de « ***économie*** » :

Place	Article(s) correspondant(s) [*Chunks*]	Économie ?
1	ŒCONOMIE ŒCONOMIE (*critique sacrée*) ŒCONOMIE ANIMALE (*médecine*)	Économie animale (lois de l'organisation des corps vivants)
2	ŒCONOMIE ANIMALE (*médecine*)	Économie animale
3	INDUSTRIE (*métaphysique*) INDUSTRIE (*droit politique et commerce*)	Économie politique
4	GÉNÉRATION (*physiologique*)	Économie animale
5	ŒCONOMIE ANIMALE (*médecine*)	Économie animale
6, 7, 8	GÉNÉRATION (*physiologie*)	Économie animale
9	RÉGIME (*médecine, hygiène*)	Économie animale
10	ETAT (plusieurs articles de *géographie*) et ETAT (*médecine*)	
11	ŒCONOMIE ANIMALE (*médecine*)	Économie animale
12	ÉDUCATION	Économie politique et animale

Search engine : moteur de recherche conceptuelle

Résultats avec le nuage conceptuel de « ***économie*** » :

Place	Article(s) correspondant(s) [*Chunks*]	Économie ?
13	SANTÉ (*médecine*)	Économie animale
14	NUTRITION (*économie animale*) NUTRITION (*jardinage*)	Économie animale
15	INCLINAISON (*philosophie morale*) INCLINAISON, PENCHANT	Morale
16	POPULATION (*physique, politique, morale*)	Économie politique
17	SENSIBILITÉ, SENTIMENT (*médecine*)	Économie animale
18	GÉNÉRATION (*physiologie*)	Économie animale
19, 20, 21, 22	*Discours préliminaire*	Division des sciences
23	HYGIENE	Économie animale
24	OECONOMAT (*jurisprudence*) ŒCONOME – ŒCONOMIE	Économie politique
26	ÉDUCATION	Économie politique
…35	SOCIÉTÉ (*jurisprudence*)	Économie politique

Enrichis par ces premiers modèles d'exploration, nous faisons néanmoins le constat que les textes sources avec lesquels nous avons « nourris » Word2Vec et le *Search engine* (les « *chunks* »), constituent une limite pour la pertinence des résultats. En effet, les modèles ne pointent pas directement des articles de l'*Encyclopédie*, mais bien des morceaux de textes qui confondent parfois les thèmes (puisque les articles se suivent dans l'ordre alphabétique) et même les contributeurs. Aussi, il est possible d'envisager des méthodes de traitement du texte qui diffèrent par le seul fait qu'elles envisagent de pointer plus précisément les textes en question, ce qui permettrait de dégager plus facilement les sources (articles et auteurs concernés).

3. L'Encycloscope : un assistant conversationnel pour interroger l'*Encyclopédie*

Afin de prolonger les travaux décrits jusqu'ici et de faciliter l'exploration du texte monumental de l'*Encyclopédie*, le projet Encycloped·IA a entrepris la conception d'un outil interactif dénommé l'**Encycloscope**. Cet assistant conversationnel, basé sur des approches récentes d'intelligence artificielle, se fonde en particulier sur l'architecture du **Retrieval-Augmented Generation** (RAG). Il vise à proposer un accès dynamique au texte encyclopédique, à la fois comme démonstrateur technique et comme support herméneutique pour analyser les multiples usages des termes, les divergences entre contributeurs ou encore la répartition des concepts dans l'*Encyclopédie*.

3.1 Contexte et principes méthodologiques

Le recours à un assistant conversationnel spécialisé résulte d'une double nécessité : la masse considérable de l'*Encyclopédie* (plus de 70 000 articles) rend difficile toute lecture intégrale ou

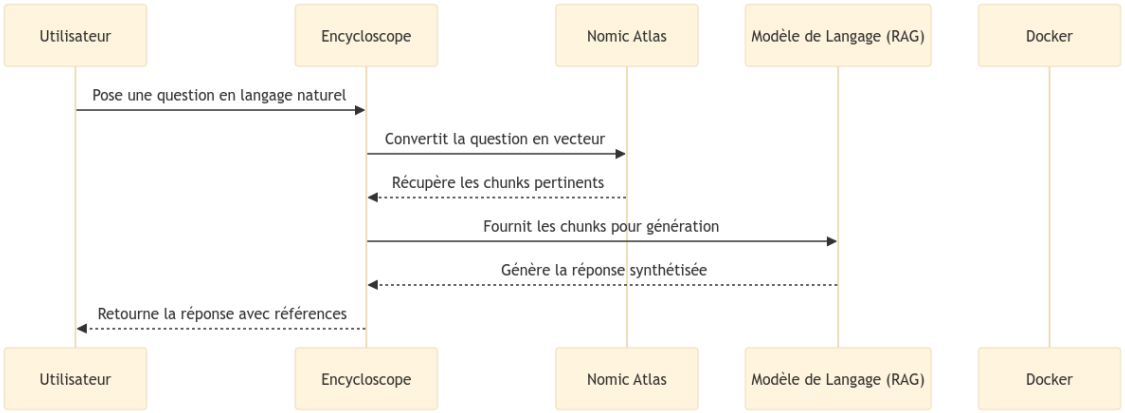
requête lexicométrique globale, tandis que les informations économiques, scientifiques ou philosophiques y sont disséminées de façon hétérogène. Dans ce contexte, les modèles de langage de grande taille (LLM) représentent un moyen prometteur d'automatiser une partie du travail de recherche, notamment pour l'identification des passages pertinents en contexte. Plusieurs solutions ont été examinées. Un affinage (*finetuning*) d'un LLM sur l'ensemble des volumes de l'*Encyclopédie* présentait de fortes contraintes techniques (coût de calcul, risque d'hallucination important) et ne garantissait pas une amélioration de la « découverte » de nouveaux articles. En revanche, l'approche RAG, qui adjoint un moteur de recherche à un LLM, permet de limiter ces hallucinations en faisant explicitement appel aux textes sources.

Le principe est donc d'interroger le modèle avec une requête conversationnelle, puis d'augmenter ce dernier par la recherche d'articles pertinents dans une base de données. Les réponses sont ainsi accompagnées d'une traçabilité explicite des références : un·e chercheur·euse a la possibilité de vérifier la correspondance entre l'article source et le résumé ou la synthèse proposée par l'assistant conversationnel. Cela s'inscrit dans une démarche à la fois expérimentale et critique, car le projet entend déterminer dans quelle mesure un modèle de langage peut contribuer à la compréhension d'un texte historique sans en trahir les nuances.

3.2 Architecture de l'Encycloscope

L'Encycloscope repose sur une suite de traitements qui commence par la segmentation du texte de l'*Encyclopédie* en unités (ou *chunks*) d'environ 8192 *tokens* (un token étant environ égal à un mot) pour la génération de la carte sémantique Atlas Nomic. Chaque *chunk* est alors converti en vecteur numérique (*embedding*) à l'aide d'un modèle spécialisé, ici *gte-multilingual-base*, dont l'intérêt est de produire des représentations sémantiques efficaces et de supporter la langue française, bien qu'ancienne dans notre corpus. Les vecteurs obtenus alimentent une base de type *vector database* permettant de repérer, via des calculs de similarité, les passages les plus voisins d'une requête.

Dans le cadre du projet Encycloped·IA, cette indexation vectorielle a été opérée avec l'outil *Nomic Atlas*, sollicité depuis un code Python via une clé d'API. *Nomic Atlas* a pour vocation de structurer et d'organiser des données non structurées à grande échelle, notamment des corpus textuels volumineux ; il fournit, entre autres, la possibilité de construire des *maps* qui visualisent les regroupements thématiques ou lexicaux. Cette visualisation éclaire les grands ensembles sémantiques et permet des analyses exploratoires. L'Encycloscope tire parti d'Atlas pour effectuer des recherches sémantiques : dès qu'un·e utilisateur·rice pose une question, celle-ci est convertie en vecteur par le modèle *gte-multilingual-base*, puis comparée aux vecteurs de la base indexée. Les segments jugés les plus pertinents sont remontés, pour être ensuite présentés au modèle de langage chargé de générer une synthèse ou une réponse argumentée.



Pipeline RAG.

Lorsque le LLM reçoit les articles, il les intègre dans un *prompt* étendu, ce qui lui permet de façonner sa réponse en s'appuyant sur les contenus réels de l'*Encyclopédie* et d'en citer les sources. Contrairement à une approche « classique », l'Encycloscope peut alors référencer avec exactitude le texte concerné, tel qu'il figure dans le corpus, ainsi que son auteur (lorsqu'il est identifié). Le dispositif global, assemblé dans un *pipeline* RAG, associe donc la richesse sémantique des embeddings vectoriels (grâce au moteur Nomic Atlas) à la puissance d'un grand modèle de langage, dans le but de fournir des réponses fidèles et contextualisées.

Afin d'illustrer ce travail, une carte interactive illustrant le corpus est disponible à l'adresse suivante : https://atlas.nomic.ai/data/encyclopedia-uca/encycloscope-2/map.

3.3 Usages et démonstration

L'Encycloscope se présente comme un outil conversationnel où l'utilisateur peut poser des questions en langage naturel. Une fois la question reçue, le module de recherche sémantique interroge la base vectorielle (les plus de 70 000 articles de l'*Encyclopédie*), puis transmet au LLM les articles jugés pertinents. La réponse générée, fondée sur cette sélection, est ensuite livrée à l'utilisateur avec des liens explicites vers les passages cités. Ce système offre plusieurs avantages. D'abord, il facilite une lecture thématique ou conceptuelle, puisqu'il repère et assemble rapidement des fragments sur un sujet précis qui, autrement, resteraient dispersés dans de nombreux articles. Il permet également de formuler des requêtes complexes, par exemple : « Comment l'*Encyclopédie* définit-elle le luxe, et quelles polémiques y sont associées ? » ou « Comment distinguer l'usage de "commerce" au sens moral et son usage au sens économique ? ». L'assistant peut apporter des éléments de réponse et guider vers les articles adéquats pour vérifier ou approfondir l'interprétation.

Les retours d'expérience suggèrent que ce mode conversationnel encourage une exploration heuristique : au lieu de suivre l'ordre alphabétique ou de parcourir la table des matières des dix-sept volumes, le chercheur navigue librement parmi les notions. De plus, l'intégration des métadonnées (quand elles sont disponibles) favorise la mise en évidence des attributs de l'article (auteur, volume, date, etc.). Sur le plan pédagogique, un tel dispositif simplifie l'accès à l'*Encyclopédie* pour des étudiant·es ou des publics moins familiers, tout en conservant la possibilité de consulter l'intégralité du texte primaire. Certains ajustements demeurent envisageables, comme l'amélioration du découpage en *chunks* et l'ajout de fonctionnalités plus

avancées (par exemple, la stylométrie pour l'attribution d'articles anonymes). Toutefois, l'idée même de combiner un moteur de recherche conceptuelle, un modèle d'*embedding* multilingue et une interface conversationnelle illustre la manière dont les techniques d'IA récentes peuvent ouvrir la voie à une relecture approfondie d'une œuvre historique telle que l'*Encyclopédie*.

Pour sa mise en production, l'application est déployée sur un serveur virtuel interne de l'Université Clermont Auvergne, accessible via le nom de domaine https://encycloscope.msh.uca.fr/.