

Approche NLP pour l'Extraction d'Informations Architecturales à partir de Descriptions de Monuments Historiques

Licence D'excellence En Intelligence Artificielle



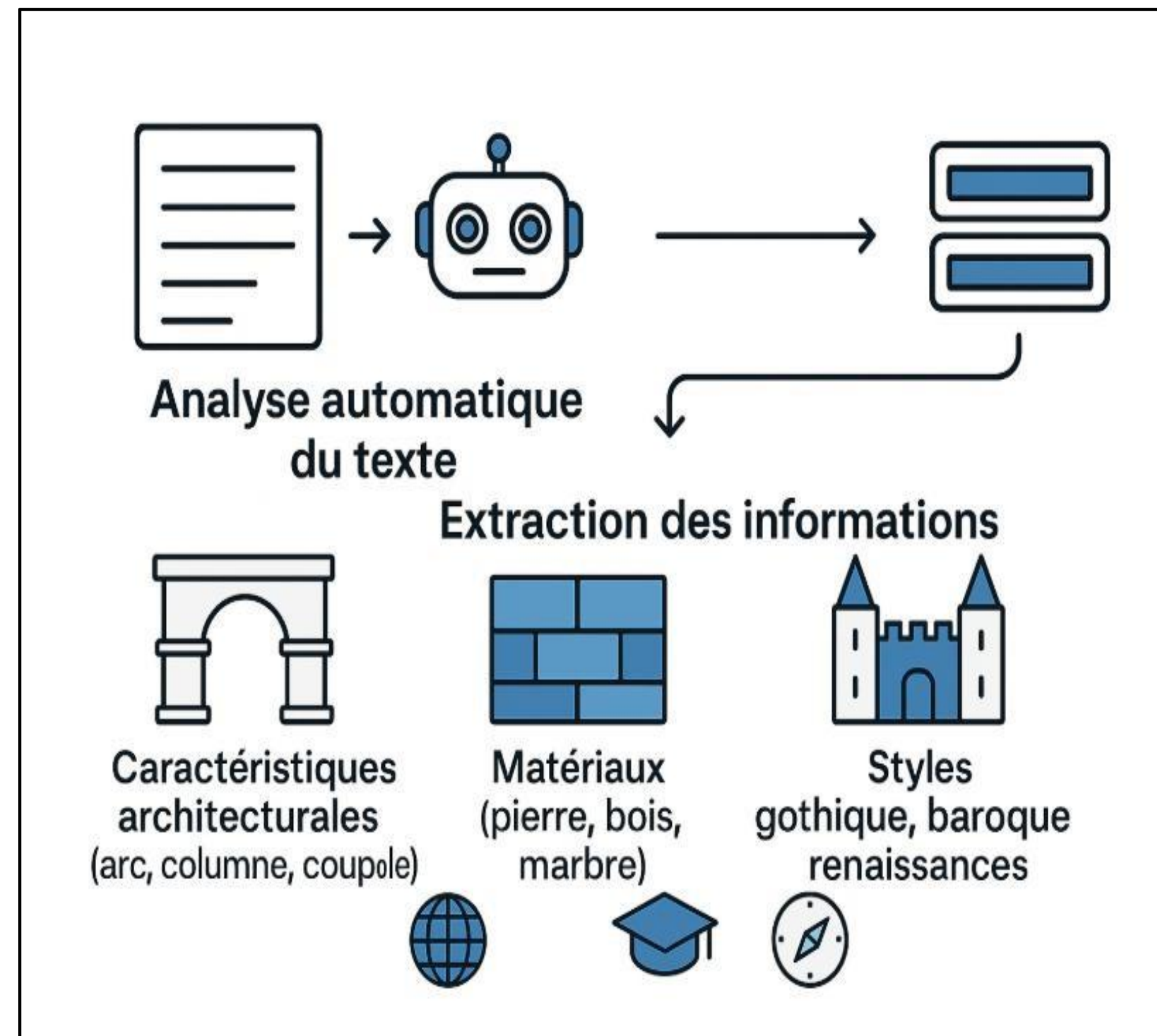
Introduction

Le patrimoine architectural constitue une ressource précieuse pour comprendre l'histoire, les savoir-faire et l'évolution des sociétés. Cependant, les informations qui le décrivent sont souvent contenues dans des textes narratifs non structurés, ce qui complique leur exploitation systématique.

Face à cette problématique, ce projet vise à développer un système intelligent capable d'extraire automatiquement des données clés (éléments architecturaux, matériaux, styles) à partir de descriptions textuelles, en s'appuyant sur les techniques de traitement automatique du langage naturel (NLP).

L'objectif est de structurer ces connaissances pour faciliter leur analyse, leur valorisation et leur intégration dans des outils destinés à la recherche, à la médiation culturelle et au tourisme patrimonial.

Objectif



Résumé

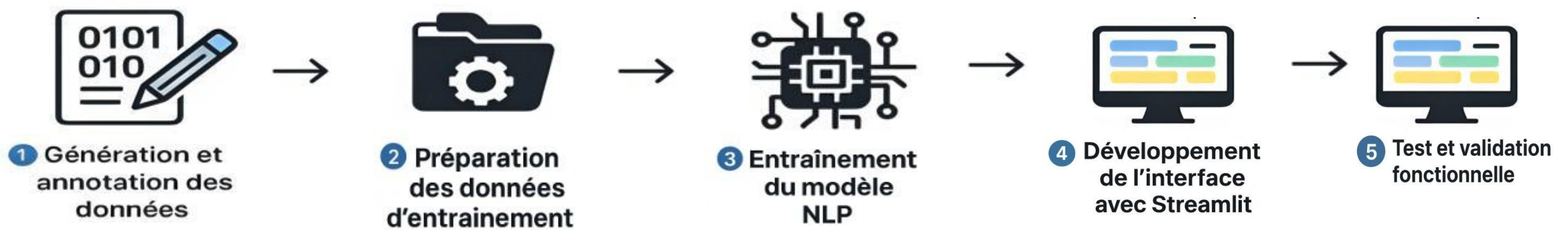
Ce projet mobilise des techniques avancées de traitement automatique du langage naturel (NLP) pour extraire et structurer des informations issues de descriptions textuelles de bâtiments historiques.

Un système a été conçu pour identifier automatiquement des éléments clés tels que les caractéristiques architecturales, les matériaux et les styles, à partir de textes non structurés.

Les données obtenues sont organisées de manière exploitable, afin de soutenir la recherche, la conservation du patrimoine, la médiation culturelle et le développement d'outils pour le tourisme intelligent.

L'approche permet de valoriser des corpus textuels existants, d'automatiser leur traitement à grande échelle, et d'ouvrir la voie à de nouvelles formes de recherche interdisciplinaire entre humanités et technologies.

METHODOLOGIE



La méthodologie globale adoptée dans ce projet suit une démarche progressive et structurée, articulée autour de cinq grandes étapes. Elle débute par la génération et l'annotation manuelle de textes décrivant des bâtiments, en l'absence de corpus annoté existant. Ces textes sont enrichis pour identifier des entités comme les matériaux, les styles ou les caractéristiques.

Une fois ces données constituées, elles sont préparées dans un format adapté à l'entraînement d'un modèle NER avec spaCy, permettant la reconnaissance automatique des entités dans de nouveaux textes.

La troisième étape consiste à entraîner le modèle NLP, en veillant à équilibrer performance et simplicité. Puis, une interface Web est développée avec Streamlit, afin d'offrir à l'utilisateur une analyse interactive avec surlignage des entités et affichage structuré.

Enfin, le système fait l'objet de tests et d'une validation fonctionnelle sur plusieurs exemples, garantissant la qualité des résultats et une expérience utilisateur fluide et efficace.

Résultat



Highlighted Text

Offer, a **Byzantine** marvel from the 20th century, graces the landscape of South Pacific. Crafted from **concrete** and **sandstone**, the structure incorporates **balconies** and elegantly rendered **arches**. It played a vital role in local rites and symbolized cosmic alignment through its orientation and form. The monument remains a testament to the community's devotion to spiritual continuity and environmental resonance.

Feature **Material** **Style**

Extracted Information

	Category	Extracted
0	Features	balconies, arches
1	Materials	sandstone, concrete
2	Styles	Byzantine

Conclusion

Ce projet met en lumière le potentiel du traitement automatique du langage naturel pour extraire et structurer des informations à partir de textes patrimoniaux.

Le système développé facilite l'analyse à grande échelle de descriptions architecturales et contribue à la valorisation du patrimoine dans des domaines comme la recherche, la médiation et le tourisme culturel.

Cette approche ouvre des perspectives pour enrichir des bases de données, créer des outils interactifs, et renforcer les liens entre sciences humaines et intelligence artificielle.

Elle peut également s'adapter à d'autres sources patrimoniales (archives, inventaires) et s'intégrer à des plateformes numériques pour améliorer l'accessibilité des savoirs.

Réalisé par :

- Mouad Radouani
- Abdelkarim Narjiss

Encadré par :

- Pr. EL HABIB Ben Lahmar
- Pr. Oussama Kaich

Année universitaire : 2024-2025