

# Favoriser l'adoption du numérique en histoire

## Générer des synthèses documentaires

### Contexte de l'étude

#### Surabondance informationnelle

- Les **ressources** n'ont jamais été aussi **nombreuses**, mais les historien.nes peinent à **trouver rapidement** les **informations pertinentes**.
- Articles, jeux de données et outils** dispersés dans des formats **hétérogènes** constituent un **verrou disciplinaire**
- Aucun **outil** ne permet de les **interroger** selon leurs **besoins** réels : croiser des méthodes, des époques, des terrains.

#### Synthèse de corpus

L'objectif de cette **thèse** est de produire des **synthèses** capables de répondre à des **questions d'historien.nes** en agrégeant l'**information pertinente** à travers un **corpus entier**.

Le corpus **Le Médiéviste et l'Ordinateur (1979-2003)** sert de terrain d'étude, il capture 25 ans d'évolution des pratiques numériques en histoire médiévale.

733 articles

25 ans

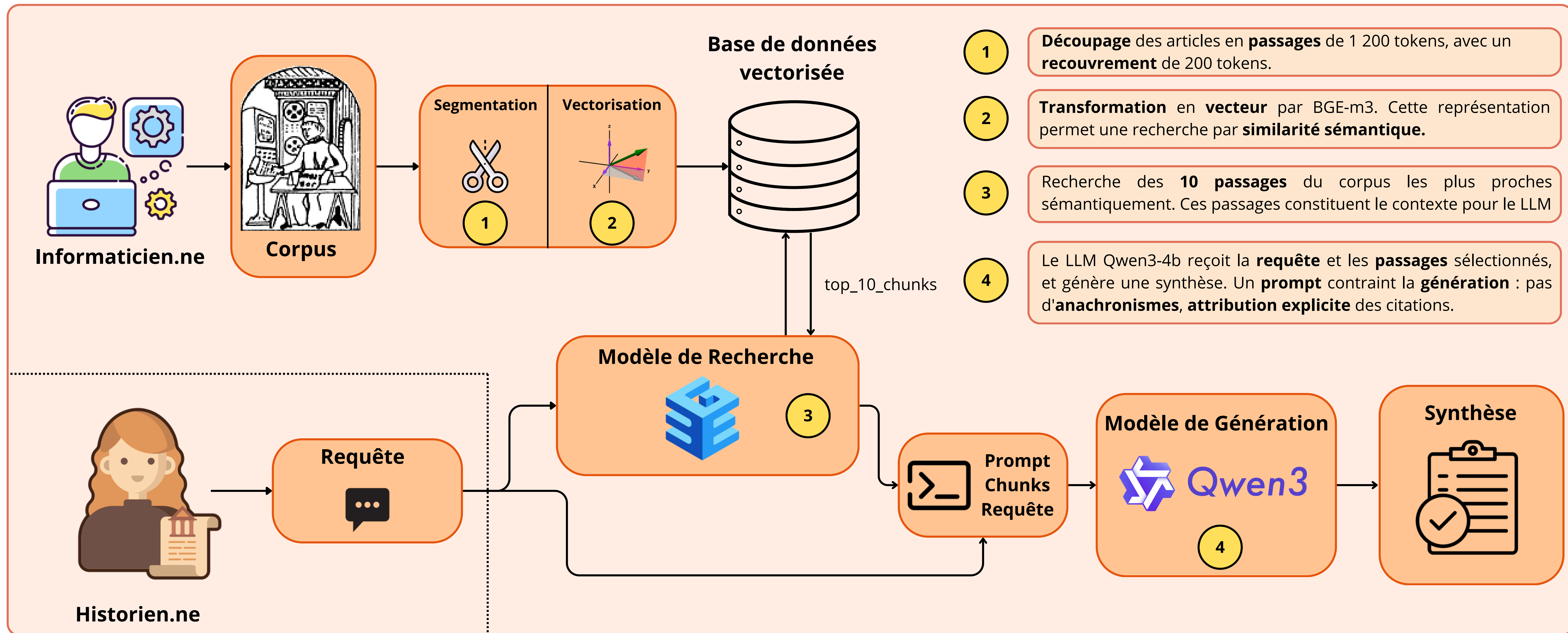
#### Reconstruire les requêtes historiennes

Une requête **historienne** croise plusieurs **axes** :

Objet d'étude	Terrains géographiques
Époques historiques	Méthodologies
Instruments numériques	Jeux de données

45 requêtes

### Architecture de génération de synthèses



### Premiers résultats critiques

#### Généralisation abusive

**Généralisation** des **opinions** d'individus comme une **vérité générale**.

Exemple tiré d'une synthèse :

« **Le discours sur l'utilisation de l'ordinateur en histoire est militant et optimiste dès les premières années.** »

**Contredit** par d'autres **articles** et l'**historiographie** de la discipline.

#### Dérive sémantique

Usage de **vocabulaire moderne** pour décrire des pratiques **anciennes**, création d'**anachronismes**.

Exemple tiré d'une synthèse :

« **La gestion automatisée de fichiers documentaires** »

« **Gérer** » des fichiers en 1979 impliquait un codage manuel rigide et des traitements par lots différés.

#### Confusion temporelle

Citations **sémantiquement pertinentes** mais **temporellement invalides**.

Exemple tiré d'une synthèse :

« **Les auteurs voient dans l'informatique un outil de libération du chercheur (Bourlet, 1993)** »

Attribuer le **sentiment** d'un article de **1993** aux pionniers de **1979** est un **contresens** majeur.

### Améliorer les synthèses documentaires

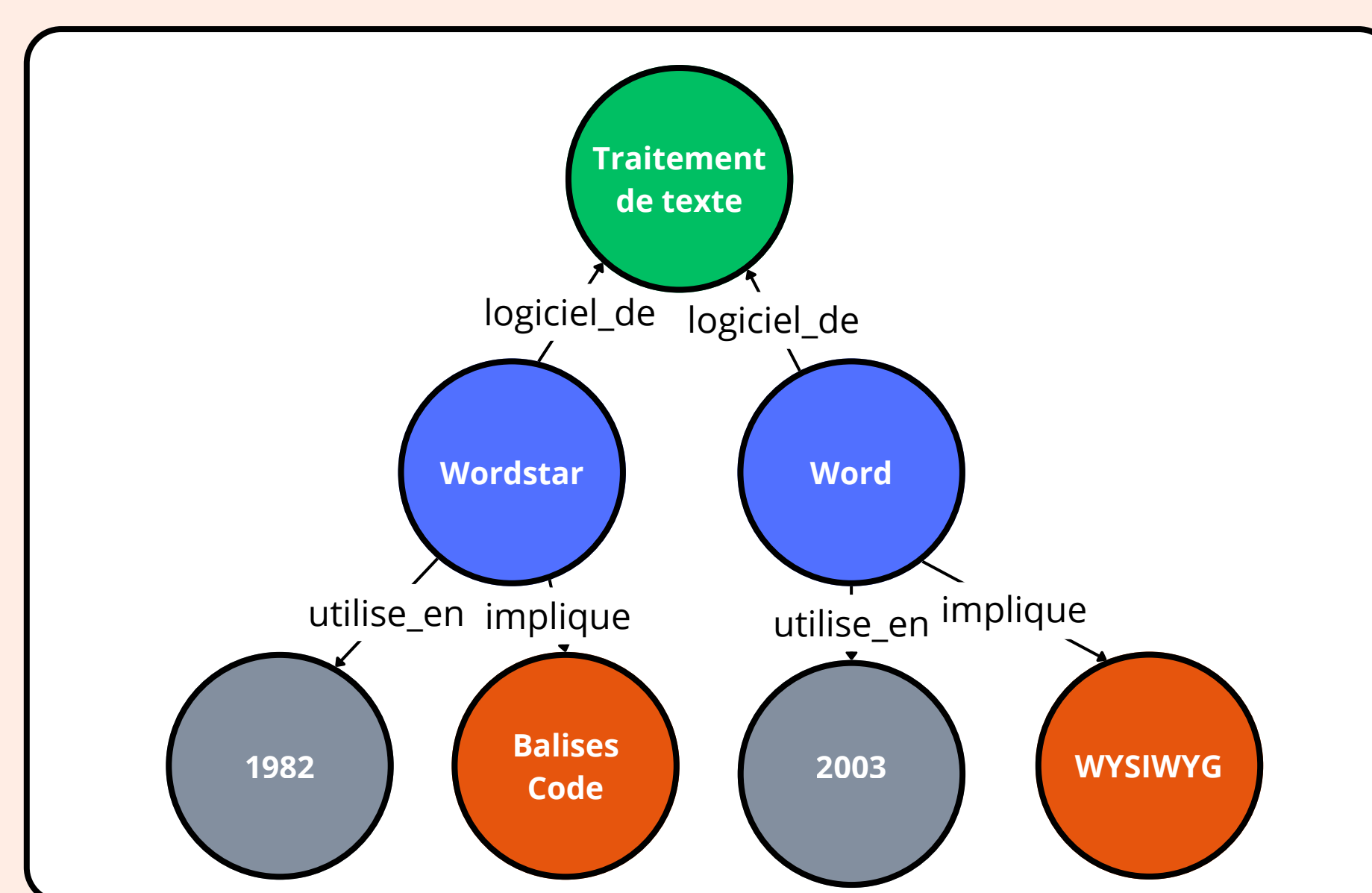
#### Grille d'évaluation multi-dimensionnelle

**Protocole hybride** : évaluation experte (Likert 4 pts) corrélée à des métriques automatiques.

- Temporalisation**  
Ancrage chronologique explicite et exact
- Couverture dimensionnelle**  
Les 6 axes couverts selon la requête
- Discours**  
Adéquation avec la rigueur historique
- Ancrage source**  
Citations vérifiables dans le corpus

#### Graphes de connaissances temporels

Intégration de **thésaurus** et **ontologies** à forte composante temporelle pour **contraindre** le RAG.



**Thibault Gautheron**  
Doctorant en Informatique  
thibault.gautheron@irit.fr

