**Département Informatique - Rapport Projet BD S5**

**Les sites UNESCO situés dans des pays à faible revenu sont-ils plus exposés à des menaces de conservation que ceux des pays à revenu élevé ?**

**Amine AMMI & Mickaël CERALINE – TP5A**

**Année universitaire 2023-2024**

**1. Introduction**

**Contexte du projet**

Dans le cadre du cours sur les Nouveaux Paradigmes de Bases de Données (NPBD), l’objectif de ce projet était de concevoir une application web sémantique permettant de répondre à une question sociétale pertinente en utilisant des données tabulaires et un graphe de connaissances. La question choisie pour ce projet est la suivante : **"Les sites UNESCO situés dans des pays à faible revenu sont-ils plus exposés à des menaces de conservation que ceux des pays à revenu élevé ?"**

Cette question soulève des problématiques liées à la conservation du patrimoine mondial et à la manière dont les ressources financières des pays influencent la protection de ces sites. Pour y répondre, il était nécessaire d'intégrer des données sur les sites UNESCO et les pays, puis d'interroger un graphe de connaissances pour obtenir des informations complémentaires.



**Palais de Versailles, premier site à avoir été inscrit au patrimoine mondial de l’Unesco**

**2. Données Sources et Conception des Ontologies**

**Fichiers CSV sources**

Pour répondre à la question de recherche, deux fichiers CSV ont été collectés et préparés :

* **CSV des Sites UNESCO** : Ce fichier contient des informations sur les sites inscrits au patrimoine mondial de l'UNESCO. Les colonnes incluent :
  + CodeSite : Identifiant unique du site.
  + NomSite : Nom officiel du site.
  + CodePays : Code ISO du pays où se trouve le site.
  + EtatSite : État de conservation du site (par exemple, "En danger", "Bon état").
* **CSV des Pays** : Ce fichier contient des informations sur les pays du monde, avec les colonnes suivantes :
  + CodePays : Code ISO du pays.
  + NomPays : Nom officiel du pays.
  + Localisation : Coordonnées du pays

Ces fichiers CSV ont servi de base pour créer des ontologies RDF, permettant une représentation sémantique des données et leur intégration dans un graphe de connaissances.

**Création des Ontologies**



Les ontologies ont été générées à partir des fichiers CSV grâce à l'outil **OntoRefine**. Ce processus a permis de transformer les données tabulaires en triples RDF tout en respectant les standards du web sémantique.

* **Fichier Pays.ttl** : Généré à partir du CSV des pays, il représente chaque pays comme une entité avec des propriétés telles que le nom, le code ISO et la catégorie de revenu.
* **Fichier Site.ttl** : Généré à partir du CSV des sites, il représente chaque site UNESCO comme une entité liée à son pays et à son état de conservation.

**3. Exploitation des données avec SPARQL**

Pour répondre à la question posée, plusieurs requêtes SPARQL ont été formulées afin d'analyser les relations entre les sites UNESCO, leur état de conservation, et les catégories de revenu des pays.



Résultats des analyses

Les résultats des requêtes ont permis d'établir que :

* Les revenus des pays n’influençaient pas tant que ça la dégradation des sites.
* La dégradation des sites naturel est souvent due aux dérèglement climatique.
* La dégradation des sites culturel ont pour la plupart été due à des évènements historiques (guerres, incendies,…).



**Incendie de la cathédrale Notre-Dame de Paris, Site inscrit au patrimoine mondial de l’Unesco, 2019**

**4. Développement de l'application web sémantique**

**Technologies utilisées**

L'application web a été développée pour permettre la visualisation des résultats des analyses SPARQL. Les technologies suivantes ont été utilisées :

**Fonctionnalités principales**

**5. Conclusion**

Ce projet a permis d'explorer l'utilisation des technologies du web sémantique pour analyser les données relatives aux sites UNESCO et leur état de conservation. Contrairement à l'hypothèse initiale, les résultats ont montré que les revenus des pays n'influencent pas de manière significative la dégradation des sites.

L'analyse a également mis en évidence des causes spécifiques en fonction du type de site :

* Pour les **sites naturels**, la dégradation est majoritairement liée au dérèglement climatique, qui représente une menace globale dépassant les frontières économiques.
* Pour les **sites culturels**, les principaux facteurs de dégradation sont des événements historiques tels que les guerres ou les incendies, qui reflètent des contextes locaux ou régionaux.

Ces observations soulignent l'importance d'adopter des approches différenciées et globales pour la préservation du patrimoine mondial, en tenant compte des spécificités de chaque type de site. Ce projet illustre également la puissance des outils sémantiques pour répondre à des questions complexes en intégrant et en analysant des données.