

RDF Cyling tour

Alban FALCOZ, Alexandre GRIPARI

24 Octobre 2025

Sommaire

1. Introduction	1
2. Modélisation du Graphe de Connaissances	1
2.1 Architecture Générale	1
2.2 Vocabulaire RDFS	2
2.3 Intégration avec le Linked Open Data	2
2.4 Contraintes SHACL	2
3. Données Instanciées	2
3.1 Flotte de Vélos	2
3.2 Clients et Guides	3
3.3 Parcours et Géographie	3
3.4 Organisation Commerciale	3
4. Requêtes SPARQL : Analyse et Complexité	3
4.1 Requêtes d'Analyse Tarifaire et Maintenance	3
4.2 Requêtes d'Agrégation Complexe	3
4.3 Requêtes de Gestion Opérationnelle	3
4.4 Requêtes d'Analyse Géographique	4
4.5 Requêtes de Planification et Capacité	4
4.6 Requêtes de Recommandation	4
4.7 Requêtes Temporelles	4
5. Choix de Conception et Justifications	4
5.1 Modélisation Multi-Types de Vélos	4
5.2 Séparation Stage/Package	4
5.3 Gestion Temporelle	4
5.4 Liaison DBpedia	5
6. Utilisation de l'IA	5

1. Introduction

Ce projet vise à concevoir et implémenter un graphe de connaissances RDF pour un opérateur de cyclotourisme. Le système modélise l'ensemble des entités et relations nécessaires à la gestion d'une entreprise de tour opérateur cycliste : vélos, clients, guides, parcours, réservations, et points d'eau.

2. Modélisation du Graphe de Connaissances

2.1 Architecture Générale

Le graphe de connaissances est structuré autour de plusieurs fichiers complémentaires :

- **cto_schema.ttl** : Définition du vocabulaire RDFS avec classes et propriétés
- **cto_shapes.ttl** : Contraintes SHACL validant la structure des données
- **cto_data_*.ttl** : Instances de données (vélos, clients, guides, parcours, réservations, etc.)
- **cto_queries.txt** : 15 requêtes SPARQL pour explorer et analyser le knowledge graph

2.2 Vocabulaire RDFS

Le schéma définit une hiérarchie de classes cohérente :

Classes principales :

- `cs:Bike` (classe parent) avec ses sous-classes : `ElectricBike`, `RoadBike`, `MountainBike`, `GravelBike`, `BeginnerBike`
- `cs:Client` et `cs:Guide` (sous-classes de `foaf:Person`)
- `cs:Path` et `cs:Mountain` pour la modélisation géographique
- `cs:TourStage` et `cs:TourPackage` pour l'organisation des circuits
- `cs:Booking` avec ses spécialisations `BikeBooking` et `TourBooking`
- `cs:WaterPoint` pour les infrastructures

Propriétés clés :

- Propriétés de tarification : `pricePerDayBike`, `pricePerDayTour`
- Propriétés temporelles : `availableFrom`, `bookingDate`, `endDate`, `duration`
- Propriétés de relation : `bike`, `guideAssigned`, `includesStage`, `stagePath`
- Propriétés géographiques : `hasStart`, `hasEnd`, `location`, `locatedInRegion`
- Propriétés descriptives : `difficulty`, `elevationGain`, `maintenanceStatus`

2.3 Intégration avec le Linked Open Data

Le projet intègre des références du Tour de France 2022 de DBpedia pour les localisations géographiques (villes, régions, montagnes), ce qui permet d'enrichir le graphe avec des données externes.

Exemples d'entités liées :

- Montagnes : `dbp:Col_du_Granon`, `dbp:Alpe_dHuez`, `dbp:Col_du_Galibier`
- Régions : `dbp:Hautes-Alpes`, `dbp:Isère`, `dbp:Savoie`, `dbp:Hautes-Pyrénées`
- Villes : `dbp:Nice`, `dbp:Cannes`, `dbp:Briançon`, `dbp:Copenhagen`

2.4 Contraintes SHACL

Les shapes définissent des contraintes garantissant la conformité des données :

Contraintes de cardinalité :

- Exactement 1 label par entité (`sh:minCount 1`, `sh:maxCount 1`)
- Exactement 1 guide assigné par stage/package
- Au moins 1 étape par package (`sh:minCount 1`)

Contraintes de type de données :

- Décimaux positifs pour prix, distances, gradients (`sh:minInclusive 0.0`)
- Gradients limités à 100% maximum (`sh:maxInclusive 100.0`)
- Formats email validés par regex pour les guides

Contraintes de valeurs :

- Status de maintenance : `"ok"`, `"needs_service"`, `"under_maintenance"`
- Difficulté : `"easy"`, `"moderate"`, `"hard"`, `"very hard"`
- Statut client : `"Finished"`, `"In Progress"`, `"Not Started"`

3. Données Instanciées

3.1 Flotte de Vélos

9 vélos de types variés avec états de maintenance:

- 4 vélos électriques (prix : 28-35€/jour)
- 1 vélo de route (22€/jour)
- 1 vélo gravel (26€/jour)
- 1 VTT (30€/jour)
- 2 vélos débutants (18-20€/jour)

Répartition des statuts : 5 “ok”, 4 “needs_service”.

3.2 Clients et Guides

9 clients avec profils diversifiés :

- Statuts : 5 “Finished”, 2 “In Progress”, 2 “Not Started”
- Types de vélos variés (analyse des préférences possible)
- Coordonnées complètes (email, téléphone)

4 guides professionnels avec informations de contact.

3.3 Parcours et Géographie

22 parcours cyclistes inspirés du Tour de France 2022 et de la Côte d’Azur :

- Distances : 12,2 à 220 km
- Dénivelés : 50 à 2300m
- Scores de difficulté : 1,0 à 5,8
- 15 cols/montagnes référencés avec caractéristiques détaillées

3.4 Organisation Commerciale

10 étapes de tour avec capacités variables (6-12 participants) et durées de 8-10h.

3 packages complets :

- “Coastal Easy Week” : 3 étapes côtières, 625€/jour
- “Mountain Adventure Week” : 3 étapes moyennes montagnes, 750€/jour
- “Alps Challenge Week” : 4 étapes alpines exigeantes, 975€/jour

10 réservations (4 vélos, 6 tours) avec dates réalistes permettant d’analyser les délais de réservation.

4. Requêtes SPARQL : Analyse et Complexité

Les 15 requêtes implémentent des cas d’usage métier variés.

4.1 Requêtes d’Analyse Tarifaire et Maintenance

Q1 - Quel est le prix moyen par jour pour chaque type de vélo, et quel type offre le meilleur rapport qualité-prix en tenant compte de l’état de maintenance ?

Q7 - Calculer le coût total d’un forfait touristique incluant la location de vélo pour une semaine (7 jours).

4.2 Requêtes d’Agrégation Complexe

Q2 - Trouver tous les forfaits touristiques avec leur distance totale, leur dénivelé positif total et leur score de difficulté global, classés par niveau de difficulté.

Q9 - Trouver le parcours multi-montagnes le plus difficile ainsi que toutes les montagnes incluses avec leurs statistiques cumulées.

4.3 Requêtes de Gestion Opérationnelle

Q3 - Quels clients ont effectué des tours sur des vélos nécessitant un entretien, et quels sont leurs coordonnées pour un suivi ?

Q4 - Lister les clients « En cours » avec leurs détails de vélo et de contact pour un suivi en temps réel.

Q10 - Pour chaque étape et forfait touristique, trouver la prochaine date disponible en tenant compte des affectations de guides et des réservations existantes.

4.4 Requêtes d'Analyse Géographique

Q5 - Trouver toutes les ascensions de montagne avec un gradient supérieur à 8 % faisant partie de parcours touristiques, ainsi que les régions dans lesquelles elles se trouvent.

Q12 - Trouver les parcours avec des points d'eau et calculer la disponibilité en eau par kilomètre de cyclisme.

Q15 - Requête complexe de type fédéré trouvant les parcours qui relient des régions avec des montagnes, en analysant le regroupement géographique des défis.

4.5 Requêtes de Planification et Capacité

Q6 - Quels forfaits touristiques incluent au moins une étape « très difficile », et quel est le ratio d'étapes très difficiles par rapport au total des étapes ?

Q11 - Générer un rapport de charge de travail théorique des guides avec leurs tours assignés, leurs étapes et leur capacité client.

Q13 - Analyser la distribution des statuts clients et leurs préférences de type de vélo avec les taux de complétion.

4.6 Requêtes de Recommandation

Q14 - Trouver le forfait touristique optimal pour les débutants en fonction de la difficulté (< 3.0), du prix et de la disponibilité de vélos adaptés aux débutants.

4.7 Requêtes Temporelles

Q8 - Quelle est l'analyse des tendances de réservation montrant les dates de tours, les délais de réservation et les préférences de vélo ?

5. Choix de Conception et Justifications

5.1 Modélisation Multi-Types de Vélos

La hiérarchie de vélos (ElectricBike, RoadBike, MountainBike, GravelBike, BeginnerBike) permet :

- Une tarification différenciée par catégorie
- Des analyses statistiques par segment
- Une extensibilité future (ajout de nouveaux types)

5.2 Séparation Stage/Package

Cette séparation offre :

- Réutilisabilité des étapes dans différents packages
- Assignment de guides différents par étape
- Calcul précis de capacités et durées cumulées
- Flexibilité pour créer des packages personnalisés

5.3 Gestion Temporelle

Les dates sont typées `xsd:date` ou `xsd:dateTime` selon le contexte :

- `bookingDate` : `dateTime` (horodatage précis des transactions)
- `availableFrom`, `endDate` : `date` (planification journalière)
- `duration` : `xsd:duration` (format ISO 8601 : PT8H = 8 heures)

5.4 Liaison DBpedia

L'utilisation de références DBpedia pour les localisations :

- Enrichit le graphe avec des données géographiques externes
- Permet des requêtes fédérées potentielles
- Évite la duplication d'informations géographiques

6. Utilisation de l'IA

Nous nous sommes aidés de chat GPT pour améliorer la tournure des phrases de notre rapport.