

INSA LYON
DÉPARTEMENT INFORMATIQUE

Projet Web Sémantique Spezialsuchmaschine

Auteurs :

Hazim ASRI
Nihal BOUTADGHART
Malte CAMIER
Jassir HABBA
Junior NOUKAM
Simon PERRET

Encadrants :

M. BENTO
Mme. CALABRETTO

5 janvier 2025

Table des matières

1	Introduction	3
2	Technologies et Outils Utilisés	3
2.1	Technologies Principales	3
2.2	Outils de Développement	3
3	Architecture et Fonctionnement	4
3.1	Organisation des Fichiers	4
3.2	Organisation des Fichiers	4
3.3	Flux de Données	4
4	Fonctionnalités	5
4.1	Recherche Interactive	5
4.2	Affichage des Marques	6
4.3	Affichage d'une marque	6
4.4	Affichage des Modèles	7
5	Réflexion sur le Web Sémantique	8
5.1	Avantages du Web Sémantique	8
5.1.0.1	Interconnexion des données	8
5.1.0.2	Interopérabilité et Flexibilité	8
5.1.0.3	Amélioration de la recherche	8
5.2	Limites et Défis du Web Sémantique	8
5.2.0.1	Complexité de l'intégration	8
5.2.0.2	Performance et Scalabilité	8
5.2.0.3	Dépendance aux sources externes	8
5.3	Enseignements tirés du projet	9
5.3.0.1	Importance de la modélisation des données	9
5.3.0.2	L'expérience utilisateur	9
5.3.0.3	Potentiel d'évolutivité	9
5.4	Conclusion	9
6	Conclusion	9

1 Introduction

Le développement des technologies du Web Sémantique a permis une avancée majeure dans la manière dont les données sont connectées, interprétées et exploitées sur le web. Contrairement au Web traditionnel, le Web Sémantique repose sur des structures standardisées et interopérables qui permettent une meilleure compréhension des données par les machines. À travers ce projet, nous avons exploré ces technologies en concevant une application de recherche spécialisée, exploitant les concepts de données liées et les requêtes SPARQL.

Objectif principal : Créer un moteur de recherche interactif et intuitif, capable d’explorer un domaine spécifique – celui des marques automobiles allemandes et de leurs modèles – à partir des données RDF disponibles sur DBpedia.

Enjeux majeurs du projet :

- **Exploitation des données structurées :** Manipuler des données hétérogènes, mais interconnectées, en utilisant des outils adaptés comme SPARQL pour les interroger.
- **Accessibilité et visualisation :** Rendre ces données accessibles à l’utilisateur final à travers une interface claire, ergonomique et interactive.
- **Défis technologiques :** Relever les défis liés à la dépendance envers DBpedia, à la complexité des requêtes SPARQL et à l’intégration des résultats dans une application moderne.

Ce projet ne se limite pas à une simple implémentation technique ; il offre également une réflexion sur les avantages et les limites des technologies du Web Sémantique dans le cadre d’une application concrète. Le rapport qui suit détaille les aspects techniques, les fonctionnalités implémentées, ainsi que les défis rencontrés et les apprentissages réalisés.

2 Technologies et Outils Utilisés

2.1 Technologies Principales

- **SPARQL :** Langage de requête pour les bases de données RDF.
- **DBpedia :** Base de données sémantique utilisée pour les marques automobiles.
- **HTML/CSS :** Structure et design de l’interface utilisateur.
- **JavaScript :** Interaction dynamique avec les utilisateurs et génération de requêtes SPARQL.

2.2 Outils de Développement

- **Visual Studio Code :** IDE principal pour le développement.
- **GitHub :** Gestion de version pour la collaboration.
- **Firefox :** Débogage et optimisation de l’application.

3 Architecture et Fonctionnement

3.1 Organisation des Fichiers

Le projet est structuré de manière logique pour séparer le code, le style et les fonctionnalités. Voici l'organisation :

```
/-- docs/
  /-- presentation/
    /-- images/
      |-- ...
    |-- main.tex
    |-- main.pdf
    |-- logo.png
  /-- report/
    /-- images/
      |-- ...
    |-- main.tex
    |-- main.pdf
    |-- logo.png
/-- images/
  |-- favicon.ico
/-- scripts/
  |-- marque.js
  |-- marques.js
  |-- modele.js
  |-- recherche.js
  |-- requetes.js
/-- styles/
  |-- index.css
  |-- main.css
  |-- marque.css
  |-- modele.css
  |-- recherche.css
/-- index.html
/-- marque.html
/-- marques.html
/-- modele.html
/-- README.md
```

3.2 Organisation des Fichiers

- **docs/** : Contient les rapports et présentations.
- **images/** : Images utilisées dans l'application.
- **scripts/** : Scripts JavaScript pour chaque page.
- **styles/** : Feuilles de style CSS pour chaque page.
- **index.html** : Page d'accueil de l'application.
- **marque.html** : Page détaillée d'une marque.
- **marques.html** : Liste des marques automobiles allemandes.
- **modele.html** : Page détaillée d'un modèle de voiture.
- **README.md** : Description du projet et instructions

3.3 Flux de Données

- L'utilisateur interagit avec l'interface en entrant un mot-clé dans la barre de recherche.
- Le mot-clé saisi est traité par un script JavaScript, qui construit dynamiquement une requête SPARQL adaptée au contexte (recherche de marques, de modèles, ou de détails spécifiques).
- La requête SPARQL est envoyée à l'endpoint public de DBpedia via une requête HTTP.
- Les résultats au format JSON sont analysés et traités pour extraire les informations pertinentes (noms, descriptions, images, etc.).
- Les données sont intégrées à l'interface sous forme de cartes interactives, permettant à l'utilisateur de visualiser et de naviguer facilement parmi les résultats.

4 Fonctionnalités

Le choix du sujet a été une étape cruciale dans notre projet. Nous avons voulu nous concentrer sur un domaine à la fois captivant et structuré, afin de pouvoir exploiter pleinement les liens sémantiques entre les données. Nous avons en premier temps opté pour un moteur de recherche sur les voitures en générale, cependant nous avons dans notre groupe un étudiant Erasmus venant d'Allemagne, nous avons finalement choisi le thème des automobiles allemandes. DBpedia, avec ses données bien organisées, nous a permis de tisser des connexions entre les modèles, les marques, et leurs caractéristiques respectives.

Nos fonctionnalités incluent les éléments suivants : pour chaque marque, nous affichons son nom, son logo, sa date de fondation et un lien vers son site officiel. Concernant les modèles, nous fournissons une description détaillée, l'année de production, le type de moteur, le design et des images illustrant chaque véhicule. Ces informations sont présentées sous forme de cartes interactives, offrant une navigation claire et intuitive.

Pour rendre l'expérience utilisateur plus fluide, nous avons intégré une barre de recherche avec une auto-complétion qui permet de saisir facilement le nom d'une marque ou d'un modèle. Cette fonctionnalité guide également l'utilisateur en proposant des suggestions pertinentes, même en cas de saisie partielle. De plus, nous avons implémenté une normalisation des chaînes de caractères pour éviter les erreurs dues à la casse, permettant ainsi une recherche sans contrainte sur la forme du texte.

Enfin, notre application organise les marques de manière alphabétique, offrant une vue d'ensemble intuitive et un accès rapide aux informations. Cette approche exploite pleinement les données sémantiques pour offrir une expérience enrichissante et interactive à l'utilisateur.

4.1 Recherche Interactive

La recherche interactive permet à l'utilisateur de rechercher des marques et des modèles automobiles allemands en temps réel. Les résultats sont affichés sous forme de pages interactives, avec des images et des informations détaillées.

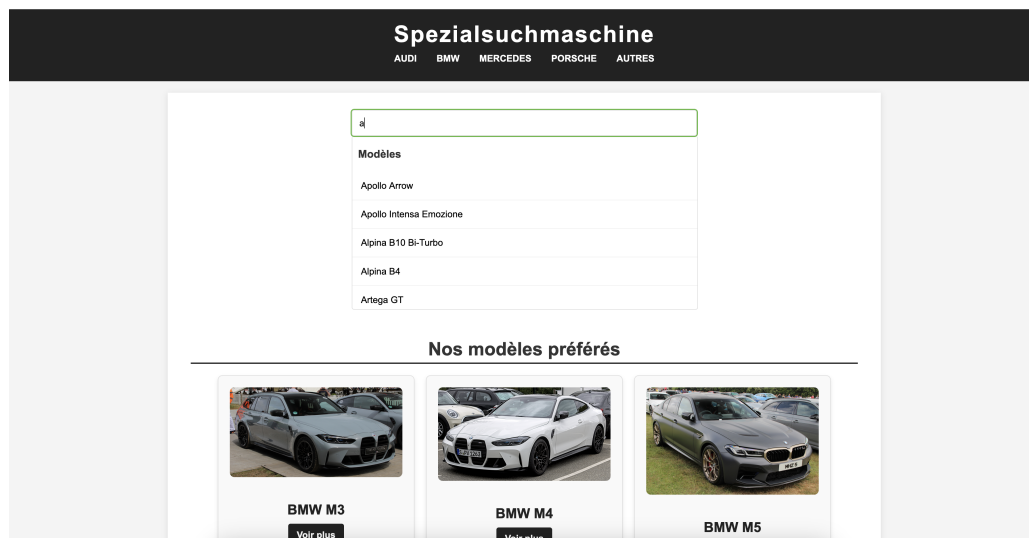


FIGURE 1 – Recherche interactive de marques et modèles

Listing 1 – Requête SPARQL pour rechercher une marques allemandes

```
SELECT DISTINCT ?model ?modelLabel ?manufacturer ?manufacturerLabel
WHERE {
    ?manufacturer dct:subject dbc:Car_manufacturers_of_Germany .

    ?model (dbo:manufacturer|dbp:manufacturer) ?manufacturer .
    ?model rdf:type dbo:Automobile .

    ?model rdfs:label ?modelLabel .
    FILTER(LANG(?modelLabel) = "fr") .
    ?manufacturer rdfs:label ?manufacturerLabel .
    FILTER(LANG(?manufacturerLabel) = "fr") .
    FILTER(REGEX(STR(?modelLabel), "^a", "i")) .
```

```
}
LIMIT 5
```

Listing 2 – Requête SPARQL pour rechercher un modèle allemand

```
SELECT DISTINCT ?manufacturer ?manufacturerLabel
WHERE {
  ?manufacturer dct:subject dbc:Car_manufacturers_of_Germany .
  ?manufacturer rdfs:label ?manufacturerLabel .
  FILTER(LANG(?manufacturerLabel) = "en") .
  FILTER(REGEX(STR(?manufacturerLabel), "^a", "i")) .
}
ORDER BY ASC(?manufacturerLabel)
LIMIT 5
```

4.2 Affichage des Marques

- Liste des marques allemandes triées par ordre alphabétique.
- Possibilité de cliquer sur la marque pour obtenir plus d'informations.

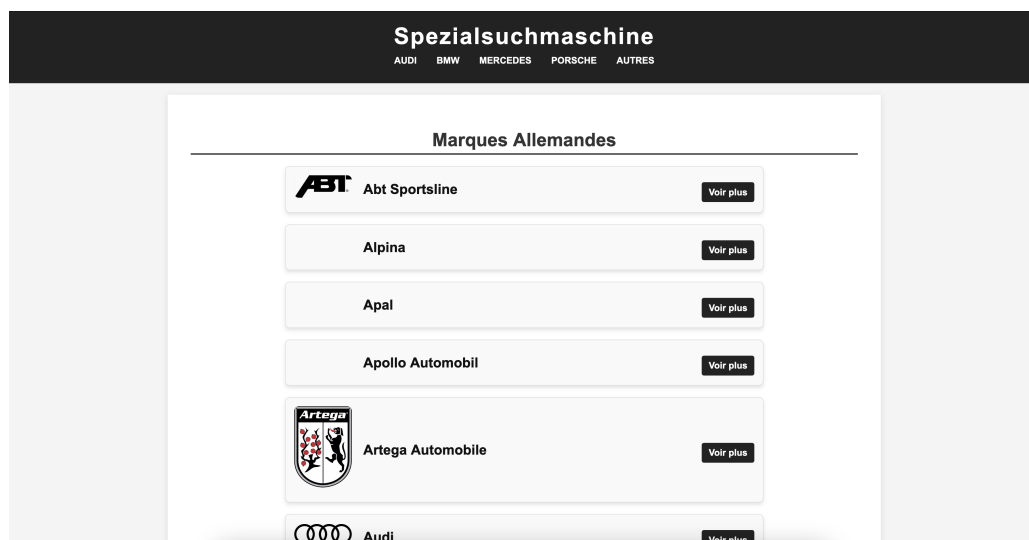


FIGURE 2 – Liste des marques automobiles allemandes

Listing 3 – Requête SPARQL pour rechercher toutes les marques allemandes

```
PREFIX dbo: <http://dbpedia.org/ontology/>
PREFIX dbc: <http://dbpedia.org/resource/Category:>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX dct: <http://purl.org/dc/terms/>

SELECT DISTINCT ?manufacturer ?manufacturerLabel ?thumbnail
WHERE {
  ?manufacturer dct:subject dbc:Car_manufacturers_of_Germany .
  ?manufacturer rdfs:label ?manufacturerLabel .
  FILTER(LANG(?manufacturerLabel) = "en") .
  ?manufacturer dbo:thumbnail ?thumbnail
}
ORDER BY ASC(?manufacturerLabel)
```

4.3 Affichage d'une marque

- Informations détaillées : description, site web, année de fondation, etc.
- Images pour illustrer chaque marque.
- Lorsque les informations sont manquantes, le container est rendu invisible.
- Nous avons fait le choix de rendre chaque requête indépendante pour éviter les erreurs de chargement.

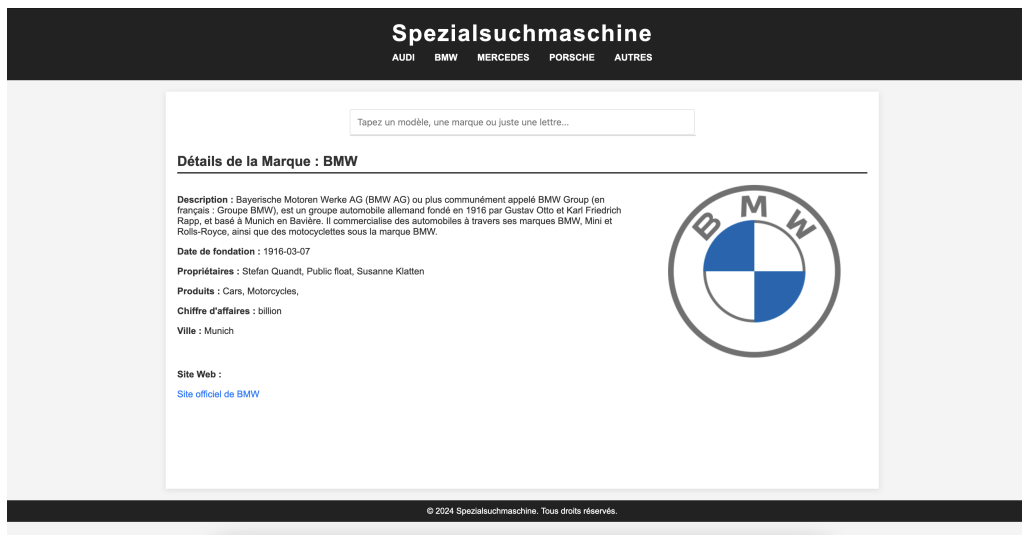


FIGURE 3 – Affichage des détails d’une marque

Listing 4 – Requêtes SPARQL pour rechercher la ville d’une marque

```
SELECT DISTINCT ?locationCity ?locationCityName
WHERE {
  OPTIONAL {
    <http://dbpedia.org/resource/BMW> dbp:locationCity ?locationCity .
    ?locationCity rdfs:label ?locationCityName .
    FILTER(LANG(?locationCityName) = "en") .
  }
  OPTIONAL {
    <http://dbpedia.org/resource/BMW> dbo:location ?locationCity .
    ?locationCity rdfs:label ?locationCityName .
    FILTER(LANG(?locationCityName) = "en") .
  }
}
```

4.4 Affichage des Modèles

- Informations détaillées : description, moteur, année de production, etc.
- Images pour illustrer chaque modèle.
- Lorsque les informations sont manquantes, le container est rendu invisible (idem que pour les marques)



FIGURE 4 – Affichage des détails d’un modèle

Listing 5 – Requêtes SPARQL pour rechercher la configuration d’un modèle

```
SELECT DISTINCT ?layoutLabel
WHERE {
  OPTIONAL {
    <http://dbpedia.org/resource/BMW_M3> dbp:layout ?layout .
    ?layout rdfs:label ?layoutLabel .
    FILTER(LANG(?layoutLabel) = "en") .
  }
}
```

5 Réflexion sur le Web Sémantique

Le Web Sémantique représente une évolution majeure du web, visant à rendre les données non seulement accessibles, mais aussi compréhensibles par les machines. À travers ce projet, nous avons pu constater à la fois les avantages et les défis liés à l'utilisation de ces technologies pour la construction d'un moteur de recherche spécialisé. Voici une réflexion sur ces aspects.

5.1 Avantages du Web Sémantique

5.1.0.1 Interconnexion des données L'un des principaux atouts du Web Sémantique est l'interconnexion des données, permettant à des informations provenant de sources diverses d'être intégrées et exploitées ensemble. Dans notre projet, l'utilisation de DBpedia comme source de données RDF a permis de lier de manière transparente des informations sur les marques, les modèles de voitures et leurs caractéristiques, créant ainsi un réseau riche et structuré de données liées.

5.1.0.2 Interopérabilité et Flexibilité Grâce aux standards ouverts comme RDF, SPARQL et OWL, le Web Sémantique offre une interopérabilité entre différents systèmes, indépendamment de leur plateforme ou de leur technologie sous-jacente. Ce projet a démontré la puissance de ces standards pour interroger et manipuler des bases de données distribuées et hétérogènes. Par exemple, les requêtes SPARQL que nous avons utilisées permettent de récupérer des informations très spécifiques tout en étant flexibles quant aux données recherchées.

5.1.0.3 Amélioration de la recherche Le Web Sémantique permet une recherche plus précise et intelligente grâce à l'exploitation des métadonnées et des ontologies. L'utilisation de SPARQL permet de poser des requêtes complexes qui vont au-delà des simples recherches par mots-clés. L'intégration de filtres linguistiques et de correspondances partielles dans les requêtes a facilité la recherche de marques et de modèles, ce qui est un pas en avant par rapport aux moteurs de recherche traditionnels.

5.2 Limites et Défis du Web Sémantique

5.2.0.1 Complexité de l'intégration L'un des défis majeurs rencontrés dans le cadre de ce projet a été l'intégration des données RDF dans une application web traditionnelle. Bien que le Web Sémantique propose une structure de données très puissante, il nécessite des connaissances spécifiques, notamment sur la modélisation RDF, la formulation de requêtes SPARQL et l'exploitation des ontologies. De plus, les résultats obtenus sont parfois complexes à manipuler et à afficher dans une interface utilisateur classique.

5.2.0.2 Performance et Scalabilité L'un des inconvénients évidents du Web Sémantique est la performance, surtout lorsque les requêtes deviennent plus complexes ou que la taille des données augmente. L'interrogation de grandes bases de données via SPARQL peut être lente, notamment si l'on travaille avec des endpoints publics comme DBpedia. Dans un contexte de production, il serait nécessaire d'optimiser les performances pour éviter les latences et garantir une expérience utilisateur fluide.

5.2.0.3 Dépendance aux sources externes Notre projet s'appuie sur DBpedia pour obtenir des données sur les marques et les modèles automobiles. Cette dépendance à une source externe a soulevé des questions liées à la fiabilité, la mise à jour des données, et la disponibilité du service. Le Web Sémantique repose sur la disponibilité et la qualité des données de source ouverte, ce qui peut poser problème si ces données deviennent obsolètes ou si les services d'accès ne sont plus disponibles.

5.3 Enseignements tirés du projet

5.3.0.1 Importance de la modélisation des données Ce projet a mis en lumière l'importance d'une bonne modélisation des données pour tirer pleinement parti des avantages du Web Sémantique. Bien que DBpedia offre une riche base de connaissances, une modélisation propre et une utilisation adéquate des ontologies peuvent permettre de construire des applications beaucoup plus puissantes, capables de raisonner sur les données et de les connecter de manière plus intelligente.

5.3.0.2 L'expérience utilisateur Il est également important de noter que l'expérience utilisateur joue un rôle crucial dans l'acceptation de technologies comme le Web Sémantique. Bien que les données sémantiques puissent être extrêmement riches, elles ne sont utiles que si elles sont accessibles de manière intuitive et pratique pour l'utilisateur. Ce projet a permis de mieux comprendre les défis liés à la conception d'interfaces utilisateur pour des applications sémantiques, en mettant l'accent sur la simplicité et l'interactivité.

5.3.0.3 Potentiel d'évolutivité Enfin, ce projet a montré que le Web Sémantique offre un énorme potentiel pour l'évolutivité des applications. En exploitant des données ouvertes et liées, il est possible de faire évoluer l'application pour inclure de nouveaux domaines ou de nouvelles sources de données sans devoir réécrire l'ensemble de l'infrastructure. L'application pourrait facilement être adaptée pour intégrer d'autres secteurs ou d'autres types de données en utilisant les mêmes principes de base.

5.4 Conclusion

Le Web Sémantique offre de nombreuses opportunités pour améliorer la recherche et la gestion des données sur le web. Cependant, sa mise en œuvre dans des applications concrètes nécessite une prise en compte des défis techniques et des exigences spécifiques liées à la manipulation de données liées. Ce projet nous a permis de mieux comprendre l'importance de la structuration des données, des requêtes complexes, et de l'interaction entre les utilisateurs et les données sémantiques.

6 Conclusion

Ce projet a permis d'explorer et d'appliquer les technologies du Web Sémantique dans le cadre de la création d'un moteur de recherche spécialisé. Grâce à ce projet, nous avons pu concevoir une application interactive permettant d'explorer les marques et modèles automobiles allemands, avec une interface utilisateur claire et intuitive.

Les technologies et outils employés, tels que DBpedia et SPARQL, ont joué un rôle essentiel dans la manipulation des données structurées, offrant une vue d'ensemble enrichissante et détaillée des informations. Le principal défi résidait dans l'intégration fluide de ces données sémantiques au sein d'une interface web moderne, ce que nous avons réussi à réaliser avec succès.

En conclusion, ce projet a non seulement renforcé nos compétences techniques en développement web et en Web Sémantique, mais a également montré l'énorme potentiel de ces technologies pour offrir des solutions intelligentes et performantes dans la gestion de données complexes et interconnectées.