

CDA 2023 - 2024 - Java Web Scraping

1. Introduction

1.1 Objectif

Le projet vise à développer un programme de scraping d'informations sur des sites internet de ventes immobilières. L'objectif principal est de créer une application permettant de réaliser des recherches à partir d'un client lourd basé sur JavaFX, scraper des informations sur différents sites internet, les stocker dans une base de données, et les afficher ensuite depuis un site web dédié.

1.2 Portée

Le projet couvre les fonctionnalités suivantes :

- Affichage d'une Interface Homme-Machine (IHM) avec JavaFX.
- Recherche sur différents sites internet en fonction de critères spécifiés.
- Affichage des résultats de la recherche.
- Stockage des informations dans un fichier texte.
- Envoi du fichier texte par courriel à un destinataire.
- Stockage des informations dans une base de données SQL.
- Affichage du contenu de la recherche depuis le site web.
- Réalisation de tests unitaires.

1.3 Aperçu

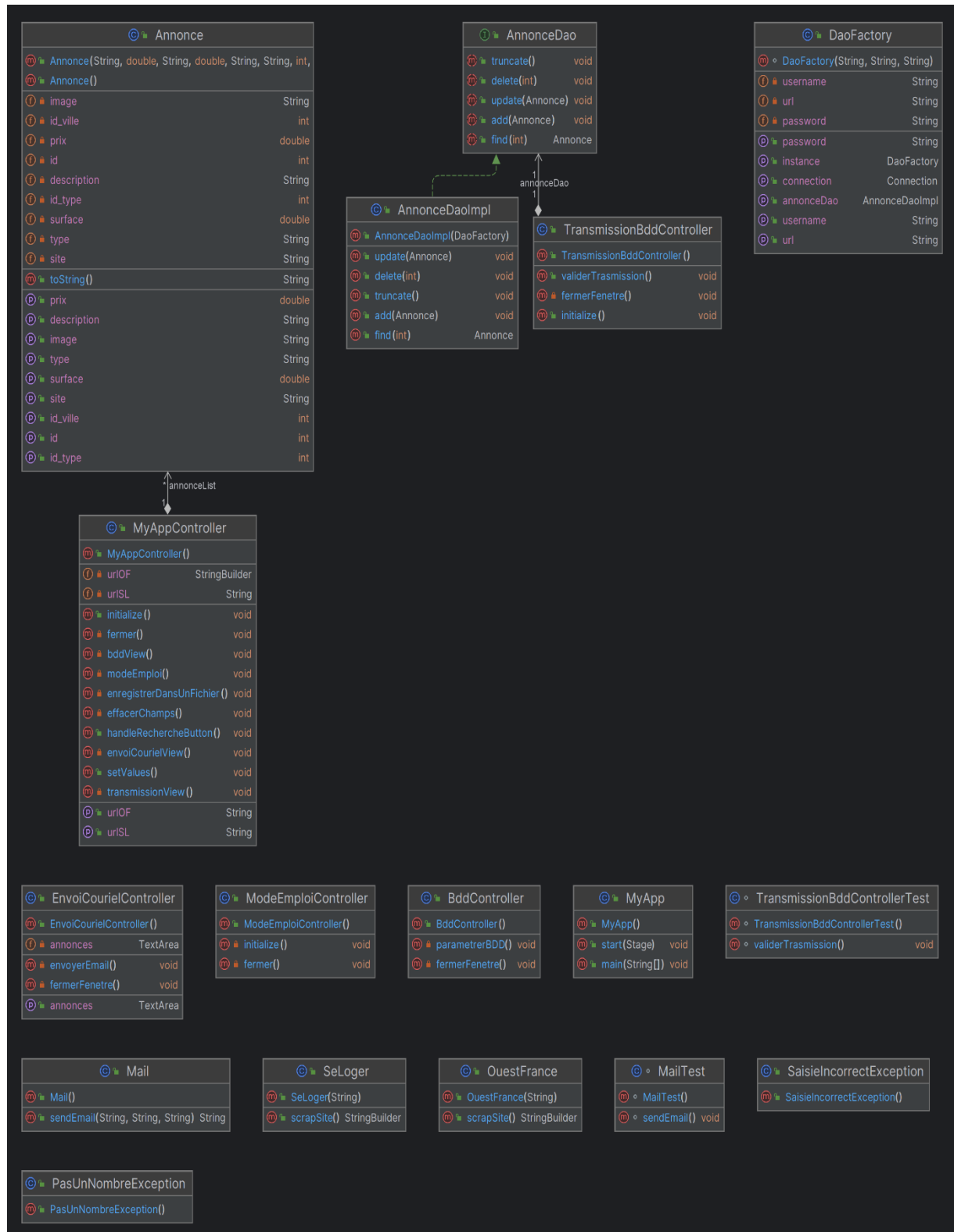
Une vue d'ensemble du projet consiste en trois étapes distinctes :

1. **Étape 1:** Réalisation des trois premières fonctionnalités, notamment l'affichage de l'IHM avec JavaFX, la recherche sur les sites internet, et l'affichage des résultats.
2. **Étape 2:** Mise en œuvre des fonctionnalités 4, 5 et 6, incluant le stockage des informations dans un fichier texte, l'envoi du fichier par courriel, et le stockage des données dans une base de données.
3. **Étape 3:** Finalisation du projet avec la réalisation des fonctionnalités 7 et 8, consistant à afficher le contenu de la recherche depuis le site web, et la réalisation des tests unitaires.

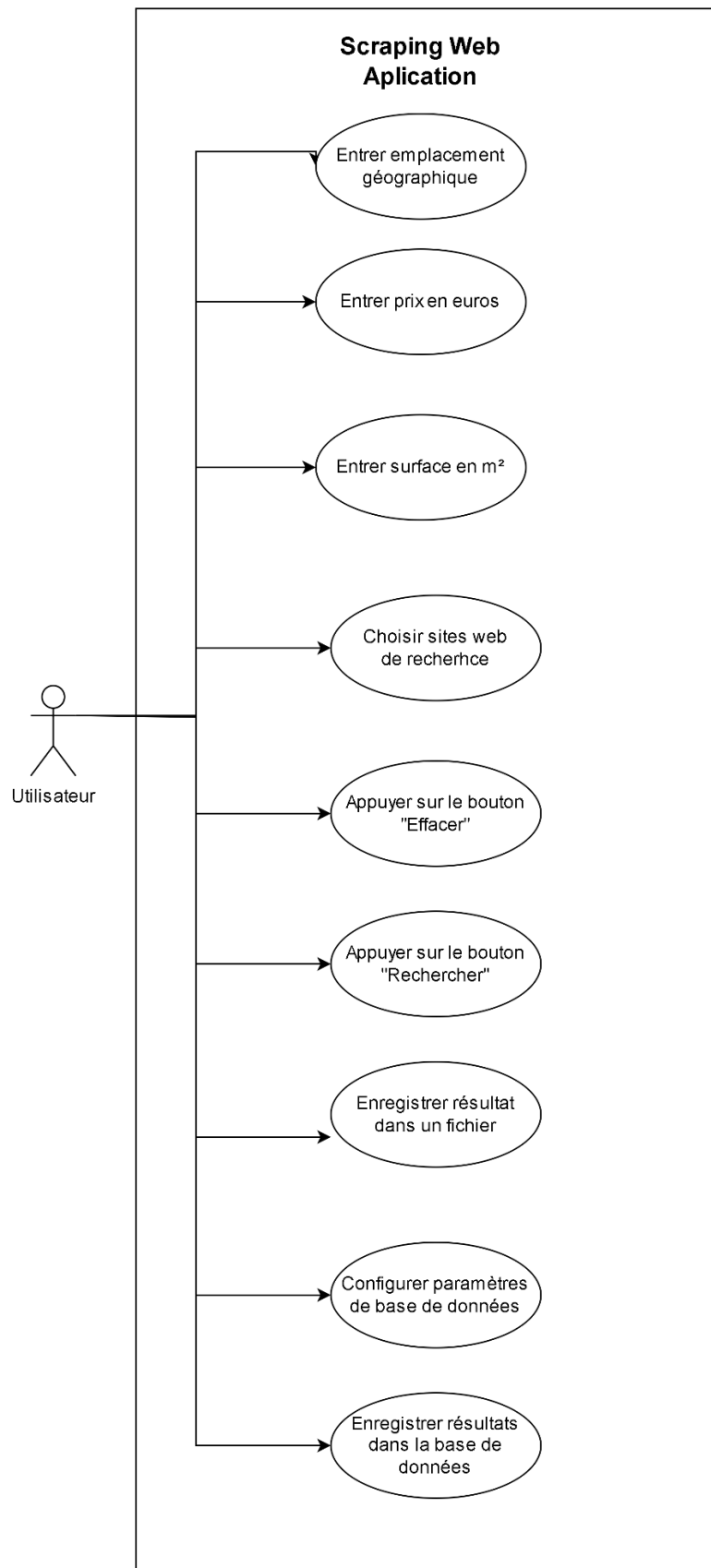
2.Diagramme

2.1 Diagramme de classe

Le diagramme de classe représente la structure statique du système, montrant les classes, les attributs, les méthodes, et les relations entre les classes. Les principales classes incluent Type, Ville, et Annonce.

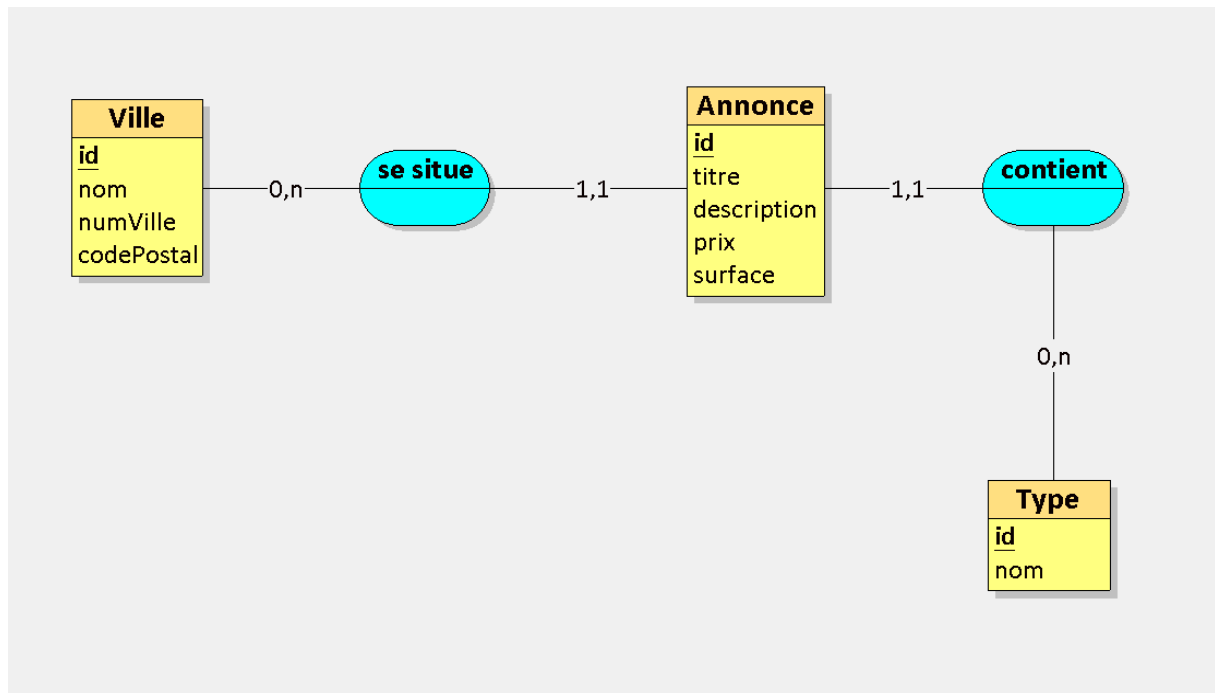


2.2 Diagramme de cas d'utilisation



2.3 Modèle Conceptuel de Données (MCD)

Le terme "MCD" fait référence à un Modèle Conceptuel de Données. Il s'agit d'un outil de modélisation utilisé dans le domaine de la gestion de bases de données pour représenter de manière abstraite la structure et les relations des données d'un système d'information. Le MCD est une étape préliminaire dans le processus de conception de bases de données et il se situe à un niveau conceptuel.



2.4 Modèle Logique de Données (MLD)

Le terme "MLD" fait référence au "Modèle Logique de Données". Il s'agit d'un modèle de données qui représente la structure organisationnelle des données dans une base de données, indépendamment des détails de mise en œuvre physiques. Le MLD se situe entre le modèle conceptuel de données (MCD) et le modèle physique de données (MPD) dans le processus de conception de bases de données.

Type = (id INT, nom VARCHAR(50));

Ville = (id INT, nom VARCHAR(50), numVille VARCHAR(50), codePostal VARCHAR(50));

Annonce = (id INT, titre VARCHAR(50), description VARCHAR(50), prix CURRENCY, surface DECIMAL(15,2), #id_1, #id_2);

2.5 Script SQL

SQL, ou Structured Query Language, est un langage de programmation standardisé utilisé pour gérer et manipuler des bases de données relationnelles. Il permet d'effectuer diverses opérations sur les données, telles que la création, la modification, la récupération et la suppression d'informations stockées dans une base de données.

```
CREATE TABLE Type(  
    id INT,  
    nom VARCHAR(50),  
    PRIMARY KEY(id)  
);
```

```
CREATE TABLE Ville(  
    id INT,  
    nom VARCHAR(50),  
    numVille VARCHAR(50),  
    codePostal VARCHAR(50),  
    PRIMARY KEY(id)  
);
```

```
CREATE TABLE Annonce(  
    id INT,  
    titre VARCHAR(50),  
    description VARCHAR(50),  
    prix CURRENCY,  
    surface DECIMAL(15,2),  
    id_1 INT NOT NULL,  
    id_2 INT NOT NULL,  
    PRIMARY KEY(id),  
    FOREIGN KEY(id_1) REFERENCES Ville(id),  
    FOREIGN KEY(id_2) REFERENCES Type(id)
```

3. Conclusion

En conclusion, cette documentation technique fournit une vue approfondie de l'architecture, des modèles de données, et des composants clés du projet de scraping immobilier. L'objectif fondamental de cette documentation est de servir de guide exhaustif pour l'équipe de développement, en offrant une compréhension détaillée de chaque aspect du projet.

L'architecture du projet, décomposée à travers le diagramme de classe, offre une représentation visuelle claire des entités, de leurs relations, et des interactions au sein du système. Cela facilite la communication au sein de l'équipe de développement et fournit une base solide pour les futures évolutions du projet.

Les modèles conceptuel de données (MCD) et logique de données (MLD) offrent une approche structurée de la gestion des données, garantissant une intégrité et une cohérence appropriées dans la base de données. Les requêtes SQL fournies contribuent à la mise en œuvre précise de la base de données, alignée sur la conception logique du système.

En adoptant une approche en trois étapes, le projet est conçu pour être évolutif, permettant une intégration continue de nouvelles fonctionnalités. Les tests unitaires ajoutent une couche de fiabilité au code, assurant la robustesse de l'application et facilitant la détection précoce des erreurs potentielles.

En somme, cette documentation technique représente une ressource essentielle pour l'équipe de développement, offrant une compréhension approfondie du projet, de ses composants, et des décisions techniques prises. Elle constituera un outil précieux pour la maintenance, l'optimisation, et l'expansion future de l'application de scraping immobilier.