

Générateur Dynamique de Contenu Statique

Projet de 1^{ère} Master réalisé par Maxime DE WOLF

Année académique 2017–2018

Directeur: Decan Alexandre

Service: Sciences Informatiques

Table des matières

| | | |
|---|---|----|
| 1 | Introduction | 2 |
| 2 | Cahier des charges | 3 |
| | 2.1 Générateur de contenu statique | 3 |
| | 2.2 Cas d'utilisations | 4 |
| 3 | Etat de l'art | 11 |
| 4 | Analyse fonctionnelle | 14 |
| | 4.1 Langage de programmation | 14 |
| | 4.2 Programmation modulaire | 14 |
| | 4.3 Règles | 14 |
| | 4.4 Système de fichiers <i>logs</i> | 14 |
| 5 | Analyse technique | 16 |
| | 5.1 Règles | 16 |
| | 5.2 Structure | 17 |
| | 5.3 Configuration | 18 |

1 Introduction

Dans le cadre du cours de projet et lecture et rédactions scientifiques, il nous est demandé de réaliser un projet d'envergure relativement importante. Dans notre cas, il s'agit d'implémenter un générateur de contenu statique. Ce rapport a pour but de vous présenter l'avancement de ce projet.

Il existe déjà énormément de générateurs de contenu statique disponibles sur Internet. StaticGen [1] établit un classement des générateurs de contenu statique *open-source* suivant leur popularité sur GitHub. Néanmoins, ces générateurs, sont souvent dédiés à une tâche spécifique ce qui limite donc leur utilisation.

L'objectif du générateur de contenu statique que nous développons est de pouvoir être utilisé pour diverses tâches. Il doit donc être paramétrable afin de s'adapter à un grand nombre de situations. Afin d'y parvenir, nous utilisons un système de règles relativement souple pour donner à l'utilisateur un contrôle total sur le comportement du générateur de contenu statique. Nous permettons également à l'utilisateur d'y ajouter de nouvelles fonctionnalités qu'il aura lui-même créées.

Dans un premier temps, nous dresserons le cahier des charges de ce projet. Cela consistera à lister les fonctionnalités nécessaires à respecter afin que ce projet soit mené à bien. Nous illustrerons également quelques cas d'utilisations afin de clarifier le cahier des charges.

Ensuite, nous listerons rapidement quelques générateurs de contenu statique existants. Nous exhiberons leurs différences ainsi que leurs fonctionnalités.

Troisièmement, nous dresserons une liste de l'ensemble des fonctionnalités que devra posséder notre générateur de contenu statique. Nous en profiterons également pour le comparer aux autres générateurs présentés en Section 3. Cela nous permettra donc de justifier la nécessité de ce projet.

Enfin, nous parlerons des techniques mises en place pour répondre aux besoins présentés en Section 4. En d'autres termes, cette section détaillera l'approche qui sera utilisée lors du développement de notre générateur de contenu statique.

2 Cahier des charges

Cette section a pour rôle de présenter le cahier des charges de ce projet. Nous introduisons également le concept de générateur de contenu statique. La seconde partie de cette section liste quelques cas d'utilisation afin d'illustrer le fonctionnement souhaité par ce projet. Notez que les fonctionnalités en elles-même seront détaillées dans la Section 4.

Le cahier des charges de ce projet est volontairement resté assez flou. En effet, les seules consignes données sont de concevoir un générateur de contenu statique qui donne à l'utilisateur une grande souplesse d'utilisation afin de pouvoir effectuer des tâches diverses. Ce générateur de contenu statique doit aussi être simple à utiliser pour le "grand public" et ne doit donc pas faire appel à des connaissances en programmation. Afin de préciser ce que sont les "tâches diverses" que ce générateur doit être capable d'accomplir, quelques cas d'utilisations sont illustrés dans cette section.

2.1 Générateur de contenu statique

Nous introduisons ici les notations que nous allons utiliser tout au long de ce rapport. Cette sous-section contient également notre définition d'un générateur de contenu statique.

Définition 1. *Un **générateur de contenu** sert à lier des données à un template (ou gabarit) afin de générer du contenu (document, site web, ...). Il sera qualifié de **statique** si ce contenu est identique peu importe le contexte dans lequel il est consulté. Les données, les templates et les contenus sont appelés **paramètres** d'un générateur de contenu statique dans le reste de ce rapport.*

Définition 2. *Un **item** est la brique de base d'un paramètre. Il s'agit d'une abstraction de ce qu'un paramètre représente comme valeur. En pratique un item peut être une page web, un fichier texte, une base de données, une structure de données, ...*

Définition 3. *La **multiplicité** des paramètres d'un générateur de contenu statique est le nombre d'items qu'ils représentent.*

Remarque 1. *Nous utilisons parfois les termes "générateur de contenu statique" et "générateur" de manière interchangeable afin d'éviter les répétitions et les lourdeurs d'écritures.*

A savoir que les générateurs de contenu statique sont très utilisés pour produire des sites web statiques. De ce cas, les *CMS* (ou *SGC*) peuvent être considérés comme des générateurs de contenu statique. Ils offrent toutefois plus de fonctionnalités qu'un générateur classique mais cela s'écarte du cadre de ce projet. Cela n'est vrai que pour les *CMS* qui ne génère que du contenu statique.

Notre générateur de contenu statique a une vocation un peu plus générale car il permettra de générer à peu près n'importe quel type de contenu statique. Leur utilisation est souvent liée au concept de *templates* qui permettent de séparer les informations de la mise en page en elle-même. La Figure 1 montre une utilisation type de ce genre de générateur. L'Exemple 1 illustre cela grâce à un cas simple et concret.

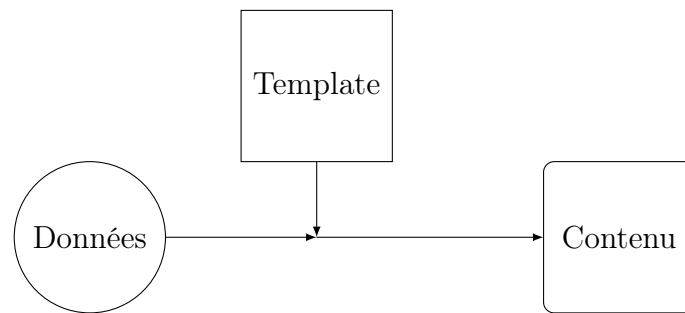


FIGURE 1 – Utilisation type d'un générateur de contenu statique

Exemple 1. Les Figures 2, 3, 4 représentent respectivement les données à utiliser, le template à appliquer et le contenu ainsi généré. Nous pouvons ainsi constater que le générateur de contenu statique génère ce dernier en remplaçant les champs du template par les valeurs contenues dans le fichier de données.

2.2 Cas d'utilisations

Nous distinguons principalement trois cas d'utilisations propres aux générateurs de contenu statique. Nous appelons ces trois cas comme suit : *OneToMany*, *ManyToOne* et *ManyToMany*. En pratique, ces trois cas se distinguent par la multiplicités des paramètres du générateur de contenu statique. Chacun d'entre eux est illustré à l'aide d'une figure explicitant leurs multiplicités.

```
{
  "players": [
    {
      "name": "Farar",
      "class": "rôdeur",
      "race": "humain",
      "level": "1"
    },
    {
      "name": "Börin",
      "class": "moine",
      "race": "nain",
      "level": "1"
    }
  ]
}
```

FIGURE 2 – Données au format JSON

Liste des joueurs :

```
{% for player in current.players %}
-> {{ player.name }}:
    {{player.class}} {{player.race}} de niveau {{ player.level }}
{% endfor %}
```

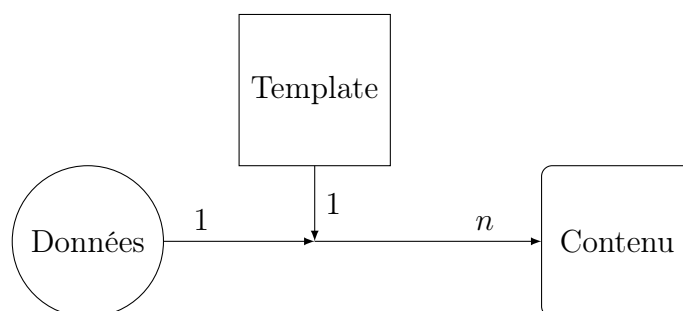
FIGURE 3 – Template au format Jinja2

Liste des joueurs :

```
-> Farar :
    rôdeur humain de niveau 1

-> Börin :
    moine nain de niveau 1
```

FIGURE 4 – Contenu au format *txt*

FIGURE 5 – Représentation du cas **OneToMany**.

Nous présentons également 2 cas d'utilisations supplémentaires qui ne sont pas directement réalisables grâce à notre générateur de contenu statique. Nous présentons donc aussi des techniques pour contourner ces limitations. Ces cas d'utilisations sont appelés *ManyTemplatesIn* et *ManyTemplatesOut*.

OneToMany

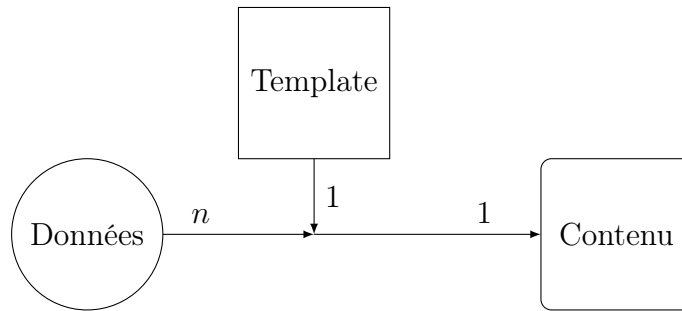
La cas *OneToMany* désigne l'utilisation d'un générateur de contenu statique en vue de générer plusieurs *items* de contenus à partir d'un seul *item* de données. Ce cas d'utilisation est illustré par la Figure 5 en considérant que **n** est la multiplicité des contenus à générer. L'Exemple 2 met en pratique ce cas d'utilisation.

Exemple 2. Par exemple, pour un site de vente, cela correspond à générer autant de page web qu'il y a de produits, chaque page ne contenant les caractéristiques que d'un unique produit (*items* de contenu). L'*item* de données est ici représenté par une base de données contenant les caractéristiques de tous les produits. Dans cet exemple, le template est appliqué à chaque produit individuellement pour générer une page web par produit.

ManyToOne

Le second cas, *ManyToOne*, est le cas inverse de *OneToMany*. En effet, ce cas désigne l'utilisation d'un générateur de contenu statique qui génère un seul *item* à partir de plusieurs. Cela donne la Figure 6 où **n** est la multiplicité des données. L'Exemple 3 met ce cas d'utilisation en pratique.

Exemple 3. Par exemple, pour un blog, cela correspond à générer une page web contenant l'ensemble de tous les posts (*item* de contenu). Celle-ci serait

FIGURE 6 – Représentation du cas **ManyToOne**.

générée à partir d'un ensemble de fichier contenant chacun un post (*items de données*). Dans ce cas, le template est appliqué une fois sur l'ensemble des posts dans le but de générer une unique page web.

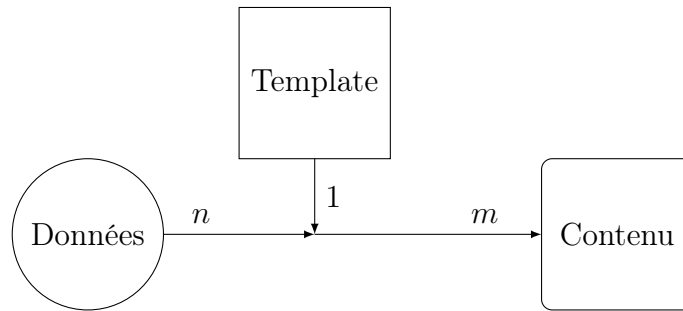
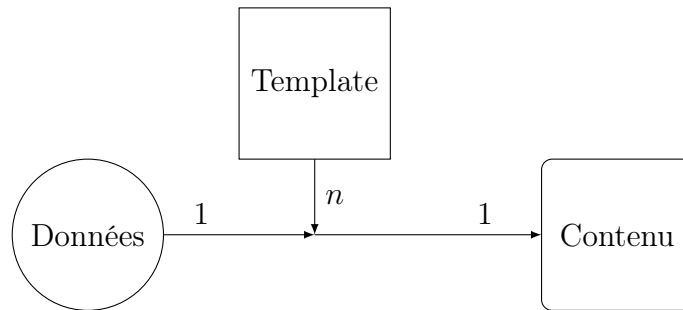
ManyToMany

Le dernier cas est celui que nous appellerons *ManyToMany*. Ce cas est en fait une généralisation des deux cas précédents. Effectivement, *ManyToMany* désigne l'utilisation d'un générateur de contenu statique en vue de générer plusieurs *items* de contenu à partir de plusieurs *items* de données. La Figure 7 permet une vision plus claire de ce cas où **n** désigne la multiplicité des données et **m** la multiplicité des contenus. L'Exemple 4 met en pratique ce cas d'utilisation.

Exemple 4. Par exemple, pour un blog, cela correspond à produire un certains nombres de pages web (disons 100 pages) listant les posts 10 par 10 (*items de contenus*) grâce à un ensemble de fichiers contenant chacun un post (*items de données*). Ici, le template est appliqué 100 fois sur des ensembles de 10 posts jusqu'à ce que tous les posts soient traités. Cela génère les 100 pages contenant chacune 10 posts.

ManyTemplatesIn

Ce cas d'utilisation implique d'utiliser un générateur de contenu statique dans le but de produire un fichier contenant plusieurs fois les mêmes données mais agencées par un *template* différent. Nous n'avons pas trouvé d'exemple pratique pour illustrer ce cas mais la Figure 9 permet de visualiser le format

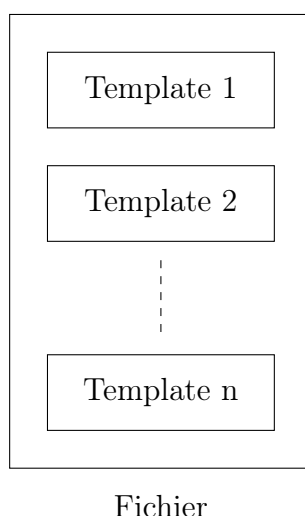
FIGURE 7 – Représentation du cas **ManyToMany**.FIGURE 8 – Représentation du cas **ManyTemplatesIn**.

de fichier générer dans ce cas. La Figure 8 exhibe plus clairement les multiplicités de ce cas, où n est le nombre de *templates* à appliquer sur le fichier de données.

Remarque 2. Notre générateur de contenu statique ne permet pas de gérer ce cas directement. Néanmoins, il existe un moyen de contourner cette limitation. En effet, suivant le moteur de template utilisé, il est possible de créer un template qui en inclut d'autres. Cela permet donc de produire un comportement similaire à *ManyTemplatesIn*.

ManyTemplatesOut

Ce cas d'utilisation est similaire au précédent, il implique également l'application de plusieurs *templates* sur un même fichier de données. La différence vient du fait que *ManyTemplatesOut* produit un fichier en sortie par *template* appliqué. L'Exemple 5 met ce cas d'utilisation en pratique. Cela se traduit par la Figure 10 où m est le nombre de *templates* à appliquer sur le fichier.

FIGURE 9 – Représentation d’un fichier de sorti du cas **ManyTemplatesIn**.

Exemple 5. *En pratique, cela revient à dire que pour un fichier contenant les informations propres à un CV, on veut générer un fichier \LaTeX et une page web. Ces deux contenus contiendraient les mêmes informations mais l’un pourra être imprimé (une fois le PDF produit) et l’autre sera mis à disposition sur Internet.*

Remarque 3. *Notre générateur de contenu statique ne donne pas à l’utilisateur le contrôle sur la multiplicité des templates. Cependant, comme pour le cas `ManyTemplatesIn`, il existe un moyen de contourner cette limitation afin de pouvoir réaliser le cas `ManyTemplatesOut`. En effet, intuitivement, ce cas d’utilisation correspond à m cas d’utilisation `ManyToOne` où n vaut 1. Ce cas d’utilisation est donc réalisable mais cela demande un travail supplémentaire de la part de l’utilisateur qui devra réaliser plusieurs cas `ManyToOne` afin d’y parvenir.*

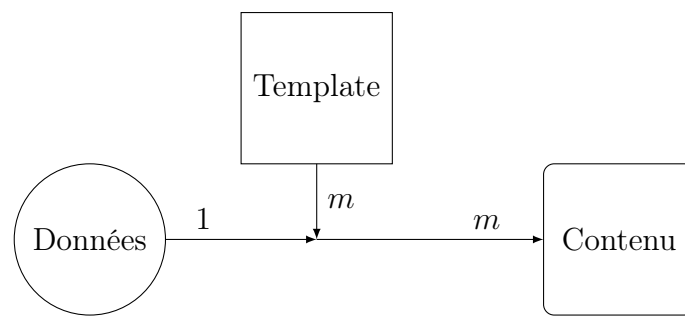


FIGURE 10 – Représentation du cas **ManyTemplatesOut**.

3 Etat de l'art

Cette section détaille, de manière non-exhaustive, l'existence d'autres générateurs de contenu statique similaires à ce projet. Leurs fonctionnalités principales seront présentées afin d'être comparées avec notre propre générateur.

Ces générateurs de contenu statique ont été choisis grâce à StaticGen [1] qui établit un classement d'un grand nombre de ces générateurs sur base de leur popularité sur GitHub. Parmi tous ceux qui y sont listés, nous en avons choisis trois comme étalons pour notre projet. Il s'agit de Jekyll, Pelican et Lektor.

Nous pouvons expliquer ces choix comme suit : Jekyll est le générateur *open-source* le plus populaire. Il s'agit donc d'un bon point de comparaison. Ensuite, Pelican est le générateur le plus populaire écrit en Python. Comme notre projet est également réalisé en Python, nous trouvons qu'il s'agit également d'une comparaison intéressante. Enfin, Lektor est le quatrième générateur le plus populaire réalisé en Python. Il offre cependant une plus grande souplesse d'utilisation que Pelican. Comme la souplesse d'utilisation est un critère important pour notre projet, nous avons jugé bon de l'ajouter à notre comparaison.

Pelican

Pelican [2] est un générateur de site web statique dont les principales fonctionnalités sont :

- Le support de *pages* (e.g. "Contact", ...)
- Le support d'*articles* (e.g. posts d'un blog)
- La régénération rapide de fichiers grâce à un système de caches et d'écriture sélective.
- La gestion de thèmes créés à partir de *templates Jinja2*
- La publication d'articles dans plusieurs langues

Pelican est un générateur de contenu statique spécialisé pour la création de blogs. Cela le rend plus facile à utiliser pour créer de tels contenus mais cela restreint également sa souplesse d'utilisation. En outre, Pelican permet de réaliser les cas d'utilisations suivants :

- **OneToOne** : permet de créer une page de blog à partir d'un fichier de données.

- **ManyToOne** : permet de générer une page d'accueil affichant plusieurs posts à partir d'un même nombre de fichiers de données.
- **OneToMany** : certains plugins permettent de générer plusieurs fichiers de contenu à partir d'un fichier de données. C'est par exemple le cas du plugin *pdf*.
- **ManyToMany** : pagination des différentes publications dans les archives.

Nous voyons ici que Pelican est capable de réaliser les trois principaux cas d'utilisations que nous avons exposé dans la section précédente. Néanmoins, Pelican n'est pas un générateur de contenu statique aussi flexible que l'on pourrait le vouloir. En effet, ces cas d'utilisations sont en fait imposés à l'utilisateur. Par exemple, le seul moyen de réaliser un cas **ManyToMany** est de paginer les publications dans les archives. L'utilisation de Pelican est ainsi limitée à la création de blogs.

Lektor

Lektor [4] est également un générateur de site web statique. Ses principales fonctionnalités sont les suivantes :

- Un système de construction intelligent qui ne reconstruit que le contenu qui a été modifié
- Un outil graphique qui permet la modification de pages sans toucher au code source
- Utilisation de système de *templates Jinja2* pour le rendu du contenu
- Un outil permettant la création relativement simple de sites web multilingues

On peut donc en conclure que c'est un outil assez similaire à Pelican sauf que Lektor propose un outil graphique pour l'édition de contenu.

Jekyll

Enfin, Jekyll [5] est le plus connu des générateurs de sites web statiques *open source*. Il est écrit en *Ruby* à la différence de Pelican et Lektor, écrits en *Python*. Voici ses principales fonctionnalités :

- Utilisation de *templates Liquid*
- Support de contenu de type *pages* (e.g "Contact", "Accueil", ...) et *articles* (e.g. posts d'un blog)
- Lancement d'un serveur local pour observer le rendu graphique des fichiers générés

Malgré leurs différences, chacun de ces générateurs de sites web statiques fonctionnent selon le cas d'utilisation *AllToOne* (voir Figure 6). Ce cas est très utilisé dans le cas de sites web de type blog d'où sa popularité.

4 Analyse fonctionnelle

Cette section est destinée à justifier les choix de conception de ce projet afin de remplir le cahier des charges. Certains choix seront inspirés directement de solutions existantes listées dans la section précédente.

La section suivante, sera directement liée à celle-ci car elle expliquera comment ces choix seront techniquement mis en œuvre lors de la phase d'implémentation.

4.1 Langage de programmation

Afin de garantir la facilité d'utilisation de notre générateur de contenu statique, notre choix de langage de programmation s'est tourné vers *Python 3*. En effet, ce langage possède une syntaxe suffisamment flexible pour nous permettre de proposer à notre tour une syntaxe intuitive pour l'utilisateur.

4.2 Programmation modulaire

Pour ce qui est de la souplesse d'utilisation, nous appliquerons les concepts de la programmation modulaire. Cela permettra à l'utilisateur de modifier ou de créer facilement de nouvelles fonctionnalités pour notre générateur de contenu statique. Nous fournirons tout de même les fonctionnalités de bases nécessaires au bon fonctionnement du projet.

4.3 Règles

Pour ce qui est de l'utilisation en elle-même de notre générateur de contenu statique, nous avons pensé à un système de règles relativement simple. Ces règles proposeront un système de variables pour permettre à l'utilisateur de réutiliser la même valeur dans plusieurs champs de la même règle. Ces règles seront contenues dans un fichier qui sera passé en paramètre à notre programme qui les exécutera une par une. Chacune de ces règles symbolisera un des trois cas d'utilisation exposé en Section 2 : *AllToOne*, *OneToAll* et *AlltoMany*. De plus amples informations sur la syntaxe de ces règles seront détaillées dans la section suivante.

4.4 Système de fichiers *logs*

Un système de fichiers *logs* permettrait de faciliter la correction de certaines règles lorsque les fichiers en sorties ne correspondent pas à ce que l'utilisateur attendait. Ces fichiers *logs* n'auront comme utilité que de lister

l'historique des règles exécutées pas à pas. Ce système viserait à atteindre l'objectif de simplicité d'utilisation -et d'aide à l'utilisation dans ce cas-ci- mentionné dans le cahier des charges en Section 2.

5 Analyse technique

Cette section est destinée à préciser les détails d’implémentation des principales fonctionnalités présentées dans la section précédente. Cette section n’est pas complète car l’implémentation de certaines fonctionnalités n’a pas encore été étudiée.

5.1 Règles

Les règles nécessaires au bon fonctionnement de notre générateur de contenu statique possèdent quatre champs obligatoires : *Target*, *Template*, *Data* et *Output*. Chacun de ces champs désigne un ou plusieurs fichiers grâce aux expressions régulières. La signification de chacun de ces quatre champs est expliquée ci-dessous.

Target

Les fichiers passés dans ce champ serviront de données pour le *template*. Chaque fichier produira un fichier en sortie. C’est donc ce champ qui permet de gérer la multiplicité des fichiers de sortie.

Template

Ce champ désigne le fichier *template* à utiliser sur les données chargées grâce aux champs *Target* et *Data*.

Concernant le cas *MultiTemplatesIn* mentionné dans la Section 3, nous pouvons solutionner ce problème en utilisant l’héritage de *template* de notre moteur de *template*. Par défaut nous utiliserons le moteur *Jinja2* [3] qui supporte cette fonctionnalité. Le problème qui en découle est donc que ce cas d’utilisation ne sera réalisable que si on utilise un moteur de *template* qui supporte une telle fonctionnalité.

L’autre cas non-couvert, également exposé en Section 3, pouvant être solutionné est *MultiTemplatesOut*. Dans l’état actuel des choses, il faudrait écrire la même règle **n** fois juste en changeant à chaque fois le champ *template* par le fichier correspondant où **n** est le nombre de *templates* à appliquer sur le fichier. Cette méthode deviendrait donc extrêmement lourde si le nombre de *templates* à appliquer est important.

Data

Les fichiers passés dans le champ *Data* serviront également de données pour le *template* comme *Target*. La différence est que ce champ n’influence

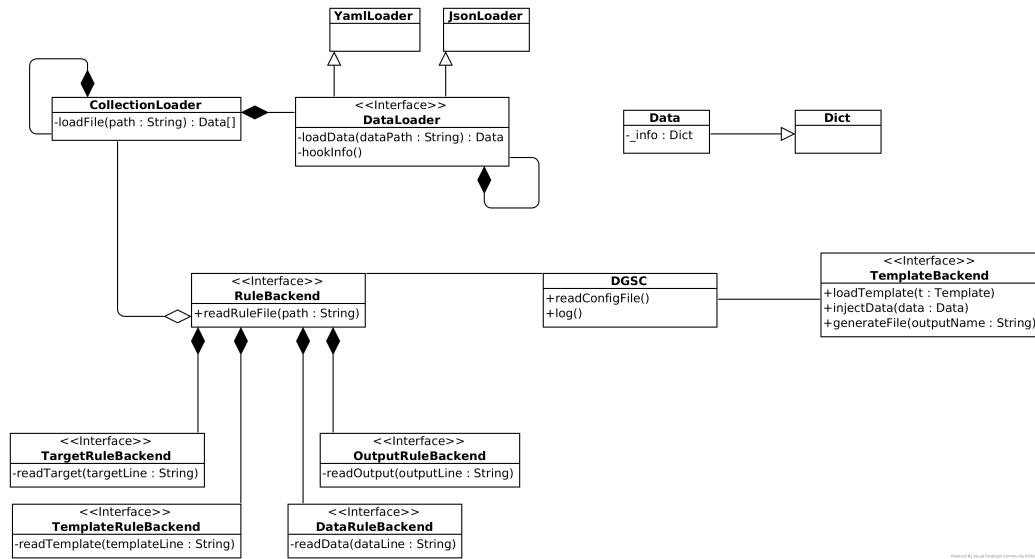


FIGURE 11 – Diagramme de classes de notre générateur de contenu statique

pas le nombre de fichiers de sortie. Les champs *Target* et *Data* permettent donc de gérer les multiplicités des fichiers d'entrée.

Output

Ce champ permet de spécifier le nom des fichiers de sortie. L'utilisation de variables de champs mentionnées dans la Section 4 permettra par exemple de nommer chaque fichier de sortie en fonction de son "homonyme" passé au champ *Target*, et cela, peut importe la multiplicité en sortie.

5.2 Structure

Comme mentionné dans la Section 4, nous appliquons la programmation modulaire afin de permettre une grande flexibilité à notre générateur de contenu statique. Nous avons donc conçu l'architecture dans cette optique comme le montre la Figure 11. Un utilisateur pourra donc modifier le comportement de notre générateur en implémentant lui même l'une de ces interfaces. Il pourra, entre autres, modifier la syntaxe des règles mentionnées ci-dessus en implémentant l'interface *RuleBackend*.

Comme expliqué dans la section précédente, nous fournirons par défaut une implémentation de ces interfaces qui permettra à l'utilisateur de réaliser les cas d'utilisations listés en Section 2.

5.3 Configuration

Un fichier de configuration permettra à l'utilisateur de spécifier quelques options à notre générateur de contenu statique. Parmi ces options, nous retrouverons entre autres : le répertoire de travail en entrée, le répertoire de travail en sortie, une liste de classes implémentant les interfaces du générateur, l'emplacement du fichier de *logs* (le cas échéant), ... Chacune de ses options aura une valeur par défaut, l'utilisateur ne devra donc spécifier que les éléments qui diffèrent de la configuration par défaut.

Bibliographie

- [1] Classement des générateurs de contenu statique *open-source*. <https://www.staticgen.com/>. (consulté le 19/08/18).
- [2] Documentation de pelican. <http://docs.getpelican.com/en/stable/>. (consulté le 14/11/17).
- [3] Page d'accueil de jinja2. <http://jinja.pocoo.org/>. (consulté le 31/12/17).
- [4] Page d'accueil de lektor. <https://www.getlektor.com/>. (consulté le 14/11/17).
- [5] Présentation de jekyll. <https://jekyllrb.com/>. (consulté le 14/11/17).