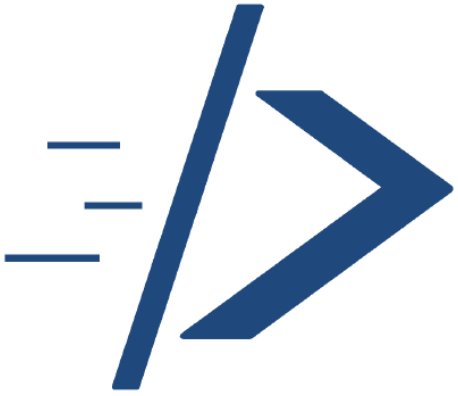
INSTITUTO SUPERIOR DE ENGENHARIA DE LISBOA

Área Departamental de Engenharia de Electrónica e Telecomunicações e de Computadores



Domain Specific Language generation based on a XML Schema

LUÍS CARLOS DA SILVA DUARTE

Licenciado em Engenharia Informática e de Computadores

Dissertação para obtenção do Grau de Mestre
em Engenharia Informática e de Computadores

Orientador : Doutor Fernando Miguel Gamboa de Carvalho

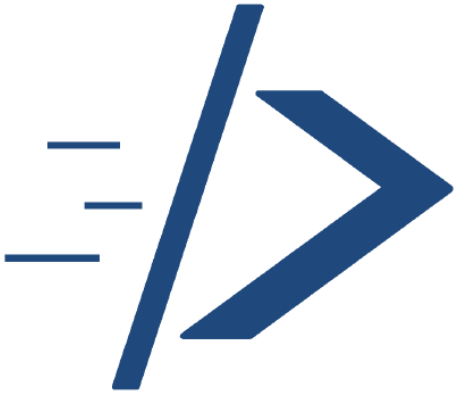
Setembro, 2018

Júri:

Presidente:	[Grau e Nome do presidente do júri]
Vogais:	[Grau e Nome do primeiro vogal]
	[Grau e Nome do segundo vogal]
	[Grau e Nome do terceiro vogal]
	[Grau e Nome do quarto vogal]

INSTITUTO SUPERIOR DE ENGENHARIA DE LISBOA

Área Departamental de Engenharia de Electrónica e Telecomunicações e de Computadores



Domain Specific Language generation based on a XML Schema

LUÍS CARLOS DA SILVA DUARTE

Licenciado em Engenharia Informática e de Computadores

Dissertação para obtenção do Grau de Mestre
em Engenharia Informática e de Computadores

Orientador : Doutor Fernando Miguel Gamboa de Carvalho

Setembro, 2018

Júri:

Presidente:	[Grau e Nome do presidente do júri]
Vogais:	[Grau e Nome do primeiro vogal]
	[Grau e Nome do segundo vogal]
	[Grau e Nome do terceiro vogal]
	[Grau e Nome do quarto vogal]

Aos meus pais.

Acknowledgments

Ao meu orientador, por todo o apoio que me deu ao longo da realização desta dissertação. A todos os meus amigos que me acompanharam, ajudaram e animaram nesta jornada. E em particular, um grande agradecimento aos meus pais, sem eles nada disto seria possível.

Acronyms and Abreviations

The list of acronyms and abbreviations are as follow.

API *Application Programming Interface* xi
DOM *Document Object Model* 28
DSL *Domain Specific Language* 1
HTML *HyperText Markup Language* xi
IDE *Integrated Development Environment* 54
JMH *Java Microbenchmark Harness* 61
JVM *Java Virtual Machine* 23
POM *Project Object Model* 52
SAX *Simple Application Programming Interface for eXtensive Markup Language* . . 28
URL *Uniform Resource Locator* 65
XHTML *eXtensive HyperText Markup Language* 28
XML *eXtensive Markup Language* 1
XSD *eXtensive Markup Language Schema Definition* xi
W3C *World Wide Web Consortium* 10

Abstract

The use of markup languages is recurrent in the world of technology today, with *HyperText Markup Language* (HTML) being the most prominent one due to its use in the Web. The need of tools that can automatically generate well formed documents with good performance is clear. Currently in order to tackle this problem the most used solution are template engines which base their solution on the usage of an external file, which doesn't ensure well formed documents and introduces the overhead of loading the template files to memory which degrades the overall performance.

Our objective is to create the required tools to generate fluent *Application Programming Interface* (API)s based on a language definition file, *eXtensive Markup Language Schema Definition* (XSD), while enforcing the restrictions of the given language. The generation of the APIs should be automated in order to avoid human error and expedite the coding process. By automating the API generation we also create a uniform approach to these languages.

To achieve our objectives we will use the Java language to extract the data from the language definition file. Based on the information provided by the language definition file we can then generate the adequate bytecodes to reflect the language definition to the Java language. To implement the language restrictions in Java we will always prioritize compile time validations, only performing run time validations of information that isn't available when the API is generated.

By comparing the developed solution to some existing solutions, including ten template engines and one other solution similar to the one we are proposing, we obtained very favorable results with the suggested solution being the best performance-wise in all the tests we performed. These results are important, specially considering that apart from being a more efficient solution it also introduces

validations of the language usage based on its syntax definition.

Keywords: XML, XSD, Automatic Code Generation, fluent API.

Resumo

Actualmente a utilização de linguagens de markup é recorrente no mundo da tecnologia, sendo o HTML a linguagem mais utilizada graças à sua utilização no mundo da Web. Tendo isso em conta é necessário que existam ferramentas capazes de escrever documentos bem formados de forma eficaz. Actualmente essa tarefa é realizada por template engines, tendo como base ficheiros externos com templates de resposta, o que não garante que estes sejam bem formados e acrescenta o overhead do carregamento do ficheiro para memória.

O nosso objectivo é criar as ferramentas necessárias para gerar APIs fluentes tendo em conta a sua definição sintática, expressa em XSD, garantindo que as restrições dessa mesma linguagem são verificadas. A geração de APIs deve ser automatizada de modo a evitar erro humano e tornar a geração de código mais rápida. Automatizando a geração de APIs cria-se também uma abordagem uniforme às diferentes linguagens utilizadas.

Para alcançar os nossos objectivos vai ser utilizada a linguagem Java para extrair informação sintática da linguagem do seu ficheiro de definição. Tendo essa informação em conta vão ser gerados bytecodes para refletir a definição da linguagem para a linguagem Java. Para implementar as restrições em Java é sempre prioritizada a validação de restrições em tempo de compilação, apenas validando em tempo de execução informação que não existe aquando da geração da API.

Comparando a solução desenvolvida com soluções semelhantes, incluindo dez template engines e outra solução semelhante à que é apresentada, obtemos resultados favoráveis, verificando que a solução sugerida é a mais eficiente em todos os testes feitos. Estes resultados são importantes, especialmente considerando que apesar de ser a solução mais eficiente introduz também a verificação das restrições da linguagem utilizada tendo em conta a sua definição sintática.

Palavras-chave: XML, XSD, Geração Automática de Código, API fluente.

Contents

List of Figures

List of Tables

List of Listings

1

Introduction

The research work that I describe in this dissertation is concerned with the implementation of a Java framework, named `xmlet`, which allows the automatic generation of a Java fluent interface[2] recreating a *Domain Specific Language* (DSL)[3] described on a *eXtensive Markup Language* (XML) schema. The approach described in this work can be further applied to any other strongly type environment. As an example of DSL generation we used `xmlet` to automatically create a Java DSL for HTML5, named HtmlFlow. A DSL for HTML can be used as a type safe template engine which results in an improvement on the numerous existing template engines. In this case, HtmlFlow is the most efficient template engine for Java in several benchmarks including the spring-comparing-template-engines where it is *** times faster than the second one.

1.1 Domain Specific Languages

In the programming world there are multiple well-know programming languages such as Java, C and C#. These language were created with the objective of being abstract, in the sense that they don't compromise with any specific problem. Using this generalist languages is usually enough to solve most problems but in some specific situations solving problems using exclusively those languages is counter-productive. A good example of that counter-productivity is thinking

about regular expressions. In the Martin Fowler DSL book[4] we have the following regular expression as an example, 1.1:

```
1 \d{3}-\d{3}-\d{4}
```

Listing 1.1: Regular Expression Example

By looking at the expression the programmer understands that it matches a String similar to `123-321-1234`. Even though the regular expression syntax might be hard to understand at first it becomes understandable after a while and it's easier to use and manipulate than implementing the same set of rules to verify a String using control instructions such as if/else and String operations. It also makes the communication easier when dealing with this concrete problem because there's a standard syntax with defined rules that are acknowledged by the members of the conversation. Regular expressions are only one of the many examples that show that creating an extra language to deal with a very specific problem simplifies it, other examples of DSLs are languages such as SQL, ant or make. The DSLs can be divided in two types, external or internal. External DSL are languages created without any affiliations to a concrete programming language, an example of that is the regular expressions DSL, since it defines its syntax without dependencies. An internal DSL on the other side is defined within a programming language, such as Java. An example of an internal DSL is the JMock[1] which is a Java library that provides tools for test-driven development. Internal DSLs can also be referred to as embedded DSLs since they are embedded in the language that they are using to define themselves. Other term to internal DSLs is fluent interfaces which are identified by the way that the usage of said DSL can be read fluently, almost as fluently as prose.

In this research what we are going to do can be summed up as a conversion of a external DSL into an internal DSL. On one side we have a XSD document. A XSD document defines a set of elements, attributes and rules that together define their own language. This language is qualified as an external DSL since it is defined in XML which is a markup language that doesn't depend on any programming language. This work has the objective of using all the information present in the XSD document and generating a Java fluent interface, which is an internal DSL since it will use the Java syntax to define the DSL.

In our use-case, the HTML5 language, what we will need to do is to generate the respective fluent interface, based on the information parsed from the HTML5

[1] JMock

XSD document. The result of this translation of XSD to Java will be the automatic code generation of Java classes and interfaces that will reflect all the information present in the XSD document. When we analyze the end result of this work, what we achieve is a Java interface to manipulate a DSL, in this case HTML, which can be used for anything related with HTML manipulation with the upside of having the guarantee that the rules of that language are verified. One of those usages is writing well-formed HTML documents and defining *dynamic views* that will be filled with information received in run-time.

1.2 Dynamic Views

A *dynamic view* is a type of textual document that defines a view to represent information. To represent the information the document has two distinct components, as shown in Listing 1.2, a static component, represented in blue, which defines the structure of the document and a dynamic part, represented in green, which is represented by placeholders which indicate where to place the received information. A simple example of a *dynamic view* can be an HTML page that presents the information of a given `Student` as shown in Listing 1.2.

```
1  <html>
2      <body>
3          <ul>
4          {{#student}}
5              <li>
6                  {{name}}
7              </li>
8              <li>
9                  {{number}}
10             </li>
11         {{/student}}
12         </ul>
13     </body>
14 </html>
```

Listing 1.2: Dynamic Student Info - Using Mustache Template Engine

To generate a complete view using the previous example we need external input, received in runtime, to replace the dynamic aspect of the view. In the previous example, Listing 1.2, the view needs to receive a value for the variable named `{{student}}`. The type that the `student` variable represents should be a type

that contains two fields, a `number` and a `name` field. An example of an object with that characteristics is presented in Figure 1.1.

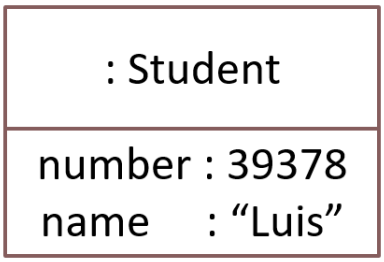


Figure 1.1: Student Object

Now that we have the two main ingredients for the proper construction of a dynamic view we can see the HTML document which results on the junction of these two aspects in Listing 1.3.

```
1  <html>
2      <body>
3          <ul>
4              <li>
5                  Luis
6              </li>
7              <li>
8                  39378
9              </li>
10         </ul>
11     </body>
12 </html>
```

Listing 1.3: Dynamic Student Info

1.3 Template Engines

The template engines are the entity which is responsible for generating the HTML document presented in Listing 1.3. Template engines are the most common method to manipulate *dynamic views*. Template engines are responsible for performing the combination between the *dynamic view*, also named *template*, and a data model, which contains all the information required to generate a complete document as shown in Figure 1.2.

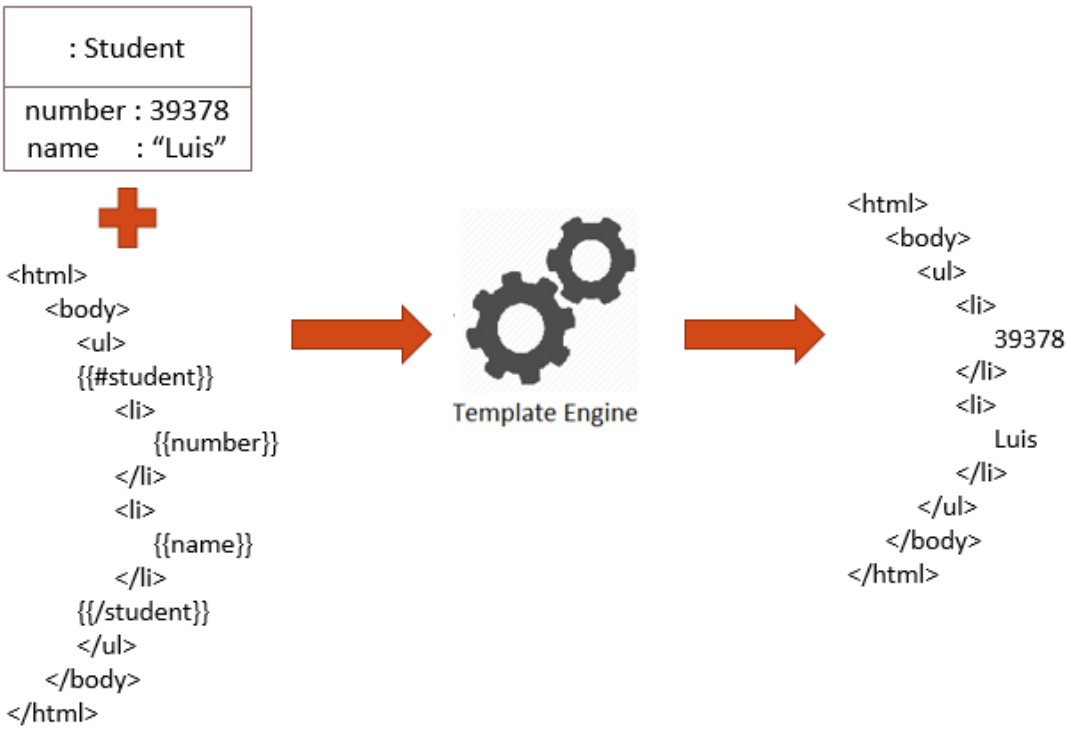


Figure 1.2: Template Engine Process

Since the Web appearance to this day there is a wide consensus around the usage of *template engines* to generate dynamic HTML documents. The consensus is such that there isn't a real alternative to the usage of template engines for dynamic generation of documents. The template engine scope is also wide, even that they are mostly associated with Web and its associated technologies they are also widely used to generate other types of documents such as emails and reports.

1.3.1 Template Engines Handicaps

Although there is a wide consensus in the usage of template engines this solutions still contain some handicaps, which we will further analyze.

- Language Compilation - There is no compilation of the language used in the templates nor the dynamic information. This can result in documents that don't follow the language rules.
- Performance - This aspect can be divided in two, one regarding the text files that are used as templates which have to be loaded and therefore slow the overall performance of the application and the heavy usage of string operations which are inherently slow.

- Flexibility - The syntax provided by the template engines is sometimes very limited which limits the operations that can be performed in the template files, often to if/else operations and a for each to loop data.

- Complexity - It introduces more languages to the programmer, for example a Java application using the Mustache[2] template engine to generate HTML forces the programmer to use three distinct languages, Java, the Mustache syntax in the template file and the HTML language.

These are the main problems that exist within the template engines solutions. To suppress these handicaps presented above we propose to use the solution implemented in this work, `xmlet`, which allows the automatic generation of a strongly typed and fluent interface for a DSL based on the rules expressed in the XSD document of that respective language, such as HTML. How will the `xmlet` solution address the handicaps of the template engines?

- Language Compilation - The generated Java DSL will guarantee the implementation of the language restrictions defined in the XSD file by reflecting those restrictions to the Java language and enforce them by using the Java compiler.

- Performance - The text files to contain templates are replaced by Java functions that represent templates, removing the need to load an additional textual file.

- Flexibility - The syntax to perform operations on templates is changed to the Java syntax which is much more flexible than any template engine syntax.

- Complexity - It removes the requirement to write three distinct languages, the programmer only needs to program in Java.

To understand how the generated API will work we will present a little example, Listing 1.4, that shows how the previous example in the Mustache idiom (Listing 1.2) will be recreated with the `xmlet` solution. The specific details on how the code presented in this example works will be provided in Chapter 4.

[2] Mustache Template Engine

```
1  Html<Element> root = new Html<>();
2
3  root.body()
4          .ul()
5              .<Student>binder((ulElem, student) ->
6                  ulElem.li()
7                          .text(Student::getNumber).°()
8                        .li()
9                          .text(Student::getName));
10
11 root.accept(new CustomVisitor<>(new Student("Luis", "39378")));
```

Listing 1.4: Xmlet Dynamic Hello

1.4 Thesis statement

This dissertation thesis is that it is possible to reduce the time spent by programmers by creating a process that automatizes the creation of DSLs based on a pre-existing DSL defined in a XSD document. The process encompasses three distinct aspects:

- XsdParser - Which parses the DSL described in a XSD document in order to extract information needed to generate the internal Java DSL.
- XsdAsm - Which uses XsdParser to gather the required information to generate the internal Java DSL.
- HtmlApi - A concrete Java DSL for the HTML5 language generated by XsdAsm using the HTML5 XSD document.

The use case used in this dissertation will be the HTML language but the process is designed to support any domain language that has its definition in the XSD syntax. This means that any XML language should be supported as long as it has its set of rules properly defined in a XSD file. To show that this solution is viable with other XSD files we used another XSD file that detailed the rules of the XML syntax used to generated Android[3] visual layouts.

[3] Android

1.5 Document Organization

This document will be separated in six distinct chapters. The first chapter, this one, introduces the concept that will be explored in this dissertation. The second chapter introduces the motivation for this dissertation. The third chapter presents existent technology that is relevant to this solution. The fourth chapter explains in detail the different components of the suggested solution. The fifth chapter approaches the deployment, testing and compares the `xmlet` solution to other existing solutions. The sixth and last chapter of this document contains some final remarks and description of future work.

2

Problem

2.1 Motivation

Text has evolved with the advance of technology resulting in the creation of *markup languages* [1]. Markup languages work by adding annotations to text, the annotations being also known as tags, that allow to add additional information to the text. Each markup language has its own tags and each of those tags add a different meaning to the text encapsulated within them. In order to use markup languages the users can write the text and add all the tags manually, either by fully writing them or by using some kind of text helpers such as text editors with IntelliSense[1] which can help diminish the errors caused by manually writing the tags. But even with text helpers the resulting document can violate the restrictions of the respective markup language because the editors don't actually enforce the language rules.

The most common markup language is the HTML language, which is heavily used in Web applications. Other uses of the HTML language are writing emails, writing reports, etc. The function of HTML in Web applications is to define the user-interface, which is also known as the view of the website. To generate the end view the most common solution is the usage of *template engine* solutions, which can be considered the *controller* which is responsible for joining the domain data with the views. This approach separates the *view* from the *controller*,

[1] Intellisense Definition

which allows both of the different layers of a project to be more independent. But, with the usage of *template engines*, there is another layers of complexity between the *view* and the *controller*, since both of these aspects of the project use different programming languages.

In the following examples we will show the Java code necessary to join a template, the `view`, that generates a table in HTML with a list of `TableElement` objects, Listing 2.1, which is the `domain data` which describes information about a company, its country and a name of a contact inside the company. We will present different examples in order to show how the complexity of the actions we want to perform affect the different solutions (the template engines and the `xmlet` solution).

```
1  class TableElement{
2      String company;
3      String contact;
4      String country;
5
6      TableElement(String company, String contact, String country){
7          this.company = company;
8          this.contact = contact;
9          this.country = country;
10     }
11 }
```

Listing 2.1: Company Information Table Element

The resulting table should be similar to the one presented in Figure 2.1, which is the example used in the Table HTML element page[2] of *World Wide Web Consortium* (W3C).

Company	Contact	Country
Alfreds Futterkiste	Maria Anders	Germany
Centro comercial Moctezuma	Francisco Chang	Mexico
Ernst Handel	Roland Mendel	Austria
Island Trading	Helen Bennett	UK
Laughing Bacchus Winecellars	Yoshi Tannamuri	Canada
Magazzini Alimentari Riuniti	Giovanni Rovelli	Italy

Figure 2.1: Table Element Page at W3C

[2] Table Element at W3C

To represent that table using the Mustache idiom we have the following template, Listing 2.2.

```
<table>
    <tr>
        <th>company</th>
        <th>contact</th>
        <th>country</th>
    </tr>
    {{#tableElements}}
    <tr>
        <td>{{company}}</td>
        <td>{{contact}}</td>
        <td>{{country}}</td>
    </tr>
    {{/tableElements}}
</table>
```

Listing 2.2: Company Information Table Element

In this situation it is possible to define the header rows directly in the template, adding the data rows of the table dynamically by receiving an `Iterable<TableElement>` named `tableElements` as shown in Listing 2.3. Since the names of the fields are known they can be directly referenced in the `td` elements.

```
Writer writer = new OutputStreamWriter(System.out);
MustacheFactory mf = new DefaultMustacheFactory();

// String template = Previous Listing

Mustache mustache = mf.compile(new StringReader(template), "");

mustache.execute(writer, new Object(){
    Iterable<TableElement> tableElements = getTableElements();
});

writer.flush();
```

Listing 2.3: Mustache With Iterable<TableElement>

Moving this example to the desired implementation of the `xmlet` solution we would have the following code present in Listing 2.4.

```
1  CustomVisitor<List<TableElement>> visitor =
2      new CustomVisitor<>(getTableElements());
3  Table<Element> table = new Table<>();
4
5  table.tr()
6          .th().text("company").°()
7          .th().text("contact").°()
8          .th().text("country").°().°()
9       .<List<TableElement>>binder((tableObj, tableElements) ->
10         tableElements.forEach(tableElement ->
11             tableObj.tr()
12                         .td().text(tableElement.company).°()
13                         .td().text(tableElement.contact).°()
14                         .td().text(tableElement.country)
15         )
16      );
17
18 table.accept(visitor);
19 System.out.println(visitor.getResult());
```

Listing 2.4: Mustache With Iterable<TableElement>

The complexity between the two previously presented seems similar, since the example is straightforward. But what if we wanted to create a template that instead of using an `Iterable<TableElement>` as the `domain data` used an `Iterable<T>`? This new solution would imply that the programmer doesn't have knowledge about the fields of the type to present on the table and therefore won't be able to use the same templates. To solve this new problem both solutions have to use reflection in order to obtain information about the concrete type represented by the `T` type. Since the `xmlet` solution is integrated in the Java environment it's easier to solve this issue. The solution uses reflection to obtain the field information of the type `T` and then uses the field information to generate the static part of the template represented by the header of the table and then it binds the `table` element to a function that receives the table data, a `List<T>` variable, and creates the table row elements, `tr`, accordingly. The final table is generated when the `table` element invokes the method `accept` receiving a concrete visitor implementation, i.e. variable `visitor`, which contains the concrete table data, i.e. variable `elementList`.

```
List<T> elementList = getTableElementsAsT();
List<Field> fields =
   elementList.isEmpty() ?
      new ArrayList<>() :
      Arrays.asList(elementList.get(0).getClass().getDeclaredFields());

CustomVisitor<List<T>> visitor = new CustomVisitor<>(elementList);
Table<Element> table = new Table<>();
Tr<Table<Element>> titleRow = table.tr();

fields.forEach(field -> titleRow.th().text(field.getName()));

table.<List<T>>binder((tableObj, tableElements) ->
        tableElements.forEach(tableElement -> {
           Tr<Table<Element>> tableRow = tableObj.tr();
           fields.forEach(field -> {
             try {
                tableRow.td().text(field.get(tableElement).toString());
             } catch (IllegalAccessException e) {
                e.printStackTrace();
             }
           });
     }));

table.accept(visitor);
System.out.println(visitor.getResult());
```

Listing 2.5: Mustache With Iterable<TableElement>

The Mustache solution to this new problem is a bit more complex, compared to the first example which used a `Iterable<TableElement>` as the `domain data`. The new template, Listing 2.6, is fairly similar, it now receives the `headerNames` variable which contains the names of the field of the given `T` type which should be inserted in the table headers and also receives the `data` variable. This `data` variable is a work around to the weaknesses of the syntax provided by the Mustache idiom, it's a `List<List<String»` object, with the inner list being the a list containing the values present in the fields of a single instance of type `T` and the outer list being the list that mimics a list of `T`. This solution was forced onto the programmer since Mustache doesn't provide enough tools to do this in other way. The first approach to this problem was to use a `Map<T, List<String»` which associated a `T` object with a list of `String` representing the names of the fields of `T`. This didn't work since Mustache doesn't allow multiple contexts which was needed in order to access the object and then access the

list to write the placeholders for the class fields, such as the `td` elements in Listing 2.2.

```
<table>
    <tr>
    {{#headerNames}}
        <th>{{.}}</th>
    {{/headerNames}}
    </tr>
    {{#data}}
    <tr>
       {{#.}}
       <td>{{.}}</td>
       {{/.}}
    </tr>
    {{/data}}
</table>
```

Listing 2.6: Company Information Table Element

This approach to the problem generated code that is more complex in the Java language. The code present in Listing 2.7 is the Java code needed to prepare the `Object` that has the context for the template, containing the `headerNames` and `data` variables.

```
Writer writer = new OutputStreamWriter(System.out);
MustacheFactory mf = new DefaultMustacheFactory();
Mustache mustache = mf.compile(new StringReader(template), "");

mustache.execute(writer, new Object(){
    List<String> headerNames = getNames();
    List<List<String>> data = getData();

    List<String> getNames() {
        return getFields().stream().map(Field::getName)
                          .collect(Collectors.toList());
    }

    List<List<String>> getData() {
        List<Field> fields = getFields();

        return getTableElementsAsT()
                .stream()
                .map(elem ->
                    fields.stream()
                          .map(field -> {
                              try {
                                  return field.get(elem).toString();
                              } catch (IllegalAccessException e) {
                                  e.printStackTrace();
                              }
                              return "";
                          }).collect(Collectors.toList()))
                .collect(Collectors.toList());
    }

    private List<Field> getFields() {
        List<Field> fields = new ArrayList<>();

        getTableElementsAsT()
           .stream().findFirst().ifPresent(elem ->
              fields.addAll(
                  Arrays.asList(elem.getClass().getDeclaredFields())));
        return fields;
    }
});
writer.flush();
```

Listing 2.7: Mustache With Iterable<TableElement>

With this comparison we observe that for simple problems both the template engines and the `xmlet` solution are viable but when the problems increase in complexity the template engine solutions start to present their weaknesses while the `xmlet` solution keeps the same complexity in its solution. We can also observe that when complexity escalates the extra coded needed to compensate the weaknesses of the template engines degrade the overall performance of the application.

2.2 Problem Statement

The problem that's being presented revolves around the handicaps of template engines, the lack of compilation of the language used within the template, the performance overhead that it introduces and the issues that it was when the complexity increases, as it was presented in the previous Section 1.3.1. To tackle those handicaps we suggested the automated generation of a strongly typed fluent interface. In order to show how that API will effectively work we will now present a small example which consists on the `html` element, Listing 2.8, described in XSD of the HTML5 language definition. The presented example is simplified for explanation purposes.

```
<xs:attributeGroup name="commonAttributeGroup">
    <xs:attribute name="someAttribute" type="xs:string">
</xs:attributeGroup>

<xs:element name="html">
    <xs:complexType>
        <xs:choice>
            <xs:element ref="body"/>
            <xs:element ref="head"/>
        </xs:choice>
        <xs:attributeGroup ref="commonAttributeGroup" />
        <xs:attribute name="manifest" type="xs:string" />
    </xs:complexType>
</xs:element>
```

Listing 2.8: Code Generation XSD Example

With this example there is a multitude of classes that need to be created, apart from the always present supporting infrastructure that will be presented in Section 4.2.1.

- Html Element - A class that represents the `Html` element (Listing 2.9), deriving from `AbstractElement`.
- Body and Head Methods - Both methods present in the `Html` class (Listing 2.9) that add `Body` (Listing 2.10) and `Head` (Listing 2.11) instances to `Html` children list.
- Manifest Method - A method present in `Html` class (Listing 2.9) that adds an instance of the `Manifest` attribute (Listing 2.12) to the `Html` attribute list.

```
1  class Html extends AbstractElement implements CommonAttributeGroup {
2      public Html() { }
3
4      public void accept(Visitor visitor){
5      visitor.visit(this);
6      }
7
8      public Html attrManifest(String attrManifest) {
9          return this.addAttr(new AttrManifest(attrManifest));
10     }
11
12     public Body body() { return this.addChild(new Body()); }
13
14     public Head head() { return this.addChild(new Head()); }
15 }
```

Listing 2.9: Html Element Class

- Body and Head classes - Classes for both `Body` (Listing 2.10) and `Head` (Listing 2.11) elements, similar to the generated `Html` class (Listing 2.9). The class contents will be dependent on the contents present in the concrete `xsd:element` nodes.

```
1 public class Body extends AbstractElement {
2     //Similar to Html, based on the contents of the respective
3     //xsd:element node.
4 }
```

Listing 2.10: Body Element Class

```
public class Head extends AbstractElement {
    //Similar to Html, based on the contents of the respective
    //xsd:element node.
}
```

Listing 2.11: Head Element Class

- Manifest Attribute - A class that represents the `Manifest` attribute (Listing 2.12), deriving from `BaseAttribute`.

```
public class AttrManifestString extends BaseAttribute<String> {
    public AttrManifestString(String attrValue) {
        super(attrValue);
    }
}
```

Listing 2.12: Manifest Attribute Class

- CommonAttributeGroup Interface - An interface with default methods that add the group attributes to the concrete element (Listing 2.13).

```
public interface CommonAttributeGroup extends Element {
    default Html attrSomeAttribute(String attributeValue) {
        this.addAttr(new SomeAttribute(attributeValue));
        return this;
    }
}
```

Listing 2.13: CommonAttributeGroup Interface

By analyzing this little example we can observe how the `xmlet` solution implements one of its most important features that was lacking in the template engine solutions, the user is only allowed to generate a tree of elements that follows the rules specified in the XSD file of the given language, e.g. the user can only add `Head` and `Body` elements as children to the `Html` element and the same goes for attributes as well, the user can only add a `Manifest` or `SomeAttribute` objects as attribute. This solution effectively uses the Java compiler to enforce the specific language restrictions, most of them at compile time. The other handicaps are also solved, the template can now be defined within the Java language eradicating the necessity of textual files that still need to be loaded into memory and

resolved by the template engine. The complexity and flexibility issues are also tackled by moving all the parts of the problem to the Java language, it removes the necessity of additional syntax and now the Java syntax can be used to create the templates.

2.3 Approach

The approach to achieve a solution was to divide the problem into three distinct aspects, as previously stated in the **??**.

The XsdParser project will be an utility project which is needed in order to parse all the external DSL rules present in the XSD document into structured Java classes.

The XsdAsm is the most important aspect of the `xmlet` solution, since it is the aspect which will deal with the generation of all the bytecodes that make up the classes of the Java fluent interface. This project should translate as many rules of the parsed language definition, its XSD file, into the Java language in order to make the resulting API as much similar as possible to the language definition.

The HtmlApi will be a representation of client aspect of `xmlet` solution. It's a concrete client of the XsdAsm project, it will use the HTML5 language definition file in order to request of XsdAsm a strongly typed fluent interface, named HtmlApi. This use case is meant to be used by the HtmlFlow API, which will implement its own Visitor in order to use the HtmlApi to write well formed HTML documents. At the moment HtmlFlow already supports a set of the HTML elements which were created manually and the rest of the library interacts with those elements in order to write the HTML files. By using the HtmlApi the HtmlFlow will support the whole HTML syntax.

3

State of Art

In this chapter we are going to introduce the the technologies used in the development of this work, such as the XSD language in order to provide a better understanding of the next chapters, and also introduce the latest solutions that try to achieve objectives similar to this work.

3.1 XSD Language

The XSD language is a description of a type of XML document. The XSD syntax allows for the definition of a set of rules, elements and attributes that together define an external DSL. This specific language defined in a XSD document aims to solve a specific issue, with its rules serving as a contract between application regarding the information contained in the XML files that represent information of that specific language. The XSD main usage is to validate XML documents, if the XML document follows the rules specified in the XSD document then the XML file is considered valid otherwise it's not. To describe the rules and restrictions for a given XML document the XSD language relies on two main types of data, elements and attributes. Elements are the most complex data type, they can contain other elements as children and can also have attributes. Attributes on the other hand are just pairs of information, defined by their name and their value. The value of a given attribute can be then restricted by multiple constraints existing on the XSD syntax. There are multiple elements and attributes present in the

XSD language, which are specified in the XSD Schema rules[6]. In this dissertation we will use the set of rules and restrictions of the provided XSD documents to build a fluent interface that will enforce the rules and restrictions specified by the given file.

3.2 Template Engines Evolution

We have already presented the idea behind template engines in the Section 1.3 and their handicaps in Section **??**, but here we are going to present some recent innovations that some template engines introduced in order to remove or minimize some of the problems listed previously. We are going to compare the features that each solution brings to the table and create a general landscape of the preexisting solutions similar to the use case that `xmlet` will use.

3.2.1 HtmlFlow Before Xmlet

The HtmlFlow[1] solution was the first to be approached in the developing process of the `xmlet` solution. The HtmlFlow motivation is to provide a library that allowed its users to write well formed type-safe HTML documents. The solution that existed prior to this project only supported a subset of the HTML language, whilst implementing some of the rules of the HTML language. This solution was a step in the right direction, it removed the requirement to have textual files to define templates by moving the template definition to the Java language. It also provided a very important aspect, it performed language validations at compile time which is great since it grants that those problems will be solved at compile time instead of run-time. The main downside of this solution was that it only supported a subset of the HTML language, since recreating all the HTML language rules manually would be very time consuming. This problem led to the requirement of creating an automated process to translate the language rules to the Java language.

3.2.2 J2html

The J2html[2] solution is a Java library used to write HTML. This solution does not verify the specification rules of the HTML language either at compile time

[1] HtmlFlow
[2] J2html

or at runtime, which is a major downside. But on the other hand it removes the requirement of having text files to define templates by defining the templates within the Java language. It also provides support for the usage of the whole HTML language, which is probably the reason why it has more garnered more attention than HtmlFlow. This library also shows that the issue we are trying to solve with the `xmlet` solution is relevant since this library has quite a few forks and watchers on their github page[3].

3.2.3 Rocker

The Rocker[4] solution is very different from the two solutions presented before. Its approach is at is core very similar to the classic template engine solution since it still has a textual file to define the template. But contrary to the classic template engines the template file isn't used at run-time. This solution uses the textual template file to automatically generate a Java class to replicate that specific template in the Java language. This means that instead of resorting to the loading of the template defined in a text file it uses the automatically generated class to generate the final document, by combining the static information present in the class with the received input. This is very important, by two distinct reasons. The first reason is that this solution can validate the type of the context objects used to create the template at compile time. The second reason is that this solution is very good performance wise due to having all the static parts of the template hardcoded into the Java class that defines a specific template. This was by far the best competitor with **xmlet!** (**xmlet!**) performance wise. The biggest downside of this solution is that it doesn't verify the HTML language rules or even well formed XML documents.

3.2.4 KotlinX

Kotlin[5] is a programming language that runs on the *Java Virtual Machine* (JVM). The language main objective is to create an inter-operative language between Java, Android and browser applications. Its syntax is not compatible to the standard Java syntax the JVM implementation of the Kotlin library allows interoperability between both languages. The main reasons to use this language is that it

[3] J2html Github Page
[4] Rocker Github
[5] Kotlin

heavily reduces the verbose needed to create code by using type inference and other techniques.

Kotlin is relevant to this project since one of his children projects, KotlinX, defines a DSL for the HTML language. The solution KotlinX provides is quite similar to what the `xmlet` will provide in its use case.

- Elements - The generated Kotlin DSL will guarantee that each element only contains the elements and attributes allowed as stated in the HTML5 XSD document. This is achieved by using type inference and the language compiler.
- Attributes - The possible values for restricted attributes values aren't verified.
- Template - The template is embedded within the Kotlin language, removing the textual template files.
- Flexibility - Allows the usage of the Kotlin syntax to define templates, which is richer than the regular template engine syntax.
- Complexity - Removes the three languages used in the process, the programmer only programs in Kotlin.

KotlinX[6] is probably the solution which resembles the `xmlet` solution the most. The only difference is that the `xmlet` solution takes advantage of the attributes restrictions present in the XSD document in order to increase the verifications that are performed on the HTML documents that are generated by the generated fluent interfaces. Both solutions also use the Visitor pattern in order to abstract themselves from the concrete usage of the DSL.

3.2.5 HtmlFlow With Xmlet

After developing the `xmlet` solution and adapting the HtmlFlow solution to use it the solutions characteristics changed. The positive aspects of the solution are kept since the general idea for the solution is kept with the usage of the `HtmlApi` generated by the `xmlet` solution. Regarding the negative aspects there were three main ones:

[6] KotlinX

- Small language subset - Solved by using the automatically generated `HtmlApi` which defines the whole HTML language within the Java language.
- Attribute value validation - The `HtmlApi` validates every attribute value based on the restrictions defined for that respective attribute in the HTML XSD document.
- Maintainability - Since it uses an automatically generated DSL if any change occurs in the HTML language specification the only change needed is to generate a new DSL based on the new language rules.

By using the `xmlet` solution the HtmlFlow solution was able to improve its performance. With the mechanics created by the usage of the `xmlet` solution it is now possible to replicate the performance improvements of the Rocker solution. Even though the template rendering using the HtmlFlow is made as the template is being defined it is possible to implement a caching strategy that caches the static parts of the template, which result in huge performance boosts when reusing a template.

3.2.6 Feature Comparison

	HtmlFlow1	J2Html	Rocker	KotlinX	HtmlFlow2
Template Within Language	✓	✓	+-	✓	✓
Elements Validations	✓	?	✗	✓	✓
Attribute Validations	✗	?	✗	✗	✓
Fully Supports HTML	✗	✓	✓	✓	✓
Well-Formed Documents	✓	✓	✗	✓	✓
Maintainability	✗	?	✓	✓	✓
Performance	✗	✓	✓	?	✓

Table 3.1: Template Engines Feature Comparison

4

Solution

This chapter will present the `xmlet` solution, its different components and how they interact between them. Generating a Java fluent interface based on a XSD file includes two distinct tasks:

1. Parsing the information from the XSD file;
2. Generating the fluent interface classes based on the resulting information of the previous task.

Those tasks are encompassed by two different projects, XsdParser and XsdAsm. In this case the XsdAsm has a dependency to XsdParser.

4.1 XsdParser

XsdParser is a library that parses a XSD file into a list of Java objects. Each different XSD tag has a corresponding Java class and the attributes of a given XSD type are represented as fields in Java. All these classes derive from the same abstract class, `XsdAbstractElement`. All Java representations of the XSD elements follow the schema definition for XSD elements, referred in Section 3.1. For example, the `xsd:annotation` tag only allows `xsd:appinfo` and `xsd:documentation` as children nodes, and can also have an attribute named `id`, therefore XsdParser has the following class as shown in Listing 4.1.

```java
public class XsdAnnotation extends XsdAbstractElement {

    private String id;
    private List<XsdAppInfo> appInfoList = new ArrayList<>();
    private List<XsdDocumentation> documentations = new ArrayList<>();

    // (...)
}
```

Listing 4.1: XsdAnnotation class (Simplified)

4.1.1 Parsing Strategy

The first step of this library is handling the XSD file. The Java language has no built in library that parses XSD files, so we needed to look for other options. The main libraries found that address this problem were *Document Object Model* (DOM) and *Simple Application Programming Interface for eXtensive Markup Language* (SAX). After evaluating the pros and cons of those libraries the choice ended up being DOM, since a XSD file is a tree of XML elements. This choice was based mostly on the fact that SAX is an event driven parser and DOM is a tree based parser, which is more adequate for the present issue. DOM is a library that maps HTML, *eXtensive HyperText Markup Language* (XHTML) and XML files into a tree structure composed by multiple elements, also named nodes. This is exactly what XsdParser requires to obtain all the information from the XSD files, which is described in XML.

This means that XsdParser uses DOM to parse the XSD file and obtain its root element, a `xs:schema` node, performing a single read on the XSD file, avoiding multiple reads which is less efficient (Listing 4.2).

```java
private Node getSchemaNode(String filePath)
    throws IOException, SAXException, ParserConfigurationException {
    DocumentBuilderFactory dbFactory =
                                DocumentBuilderFactory.newInstance();
    DocumentBuilder dBuilder = dbFactory.newDocumentBuilder();
    Document doc = dBuilder.parse(xsdFile);   //Parses the XSD file.

    //Obtains the first node, which is the xs:schema node.
    return doc.getFirstChild();
}
```

Listing 4.2: DOM Document Parsing

After obtaining the root node of the XSD file the XsdParser verifies if that node is a `XsdSchema` node as shown in Listing 4.3. If that is the case it proceeds by performing the `parse` function of the `XsdSchema` class.

```
1 Node schemaNode = getSchemaNode(filePath);
2
3 if (isXsdSchema(schemaNode)){
4     XsdSchema.parse(this, schemaNode);
5 }
```

Listing 4.3: XsdParser Node Parsing Process

The `XsdSchema` element `parse` function converts the `Node` attributes into a `Map` object, which `XsdSchema` receives in the constructor. Each class extracts their field information from that `Map` object in their `setFields` method (Listing 4.4). To guarantee that the information parsed by the classes is valid according to the rules in the XSD language standard there are multiple validations. To validate the possible values for any given attribute, e.g. the `formDefault` attribute from the `xsd:schema` element, we use `Enum` classes. Any parsed value that is meant to be assigned to one of this `Enum` variables has its content verified to assert if the received value belongs to the possible values for that attribute.

```
1  public class XsdSchema extends XsdAnnotatedElements {
2      private XsdSchema(XsdParser parser, Map<String, String> fieldsMap){
3          super(parser, fieldsMap);
4      }
5
6      @Override
7      public void setFields(Map<String, String> fieldsMap) {
8          super.setFields(fieldsMap);
9
10         this.attributeFormDefault =
11           EnumUtils.belongsToEnum(FormEnum.UNQUALIFIED,
12             elementFieldsMap.getOrDefault(ATTRIBUTE_FORM_DEFAULT,
13               FormEnum.UNQUALIFIED.getValue()));
14         this.elementFormDefault =
15           EnumUtils.belongsToEnum(FormEnum.UNQUALIFIED,
16             elementFieldsMap.getOrDefault(ELEMENT_FORM_DEFAULT,
17               FormEnum.UNQUALIFIED.getValue()));
18         this.blockDefault =
19           EnumUtils.belongsToEnum(BlockFinalEnum.DEFAULT,
20             elementFieldsMap.getOrDefault(BLOCK_DEFAULT,
21               BlockFinalEnum.DEFAULT.getValue()));
```

```
22          this.finalDefault =
23            EnumUtils.belongsToEnum(FinalDefaultEnum.DEFAULT,
24              elementFieldsMap.getOrDefault(FINAL_DEFAULT,
25                FinalDefaultEnum.DEFAULT.getValue()));
26          this.targetNamespace =
27            elementFieldsMap.getOrDefault(TARGET_NAMESPACE,
                targetNamespace);
28          this.version = elementFieldsMap.getOrDefault(VERSION, version);
29          this.xmlns = elementFieldsMap.getOrDefault(XMLNS, xmlns);
30      }
31
32      public static ReferenceBase parse(XsdParser parser, Node node) {
33          NamedNodeMap nodeAttributes = node.getAttributes();
34          Map<String, String> attrMap = convertNodeMap(nodeAttributes);
35
36          return xsdParseSkeleton(node, new XsdSchema(parser, attrMap));
37      }
38  }
```

Listing 4.4: XsdSchema Information Extraction (Simplified)

The parsing of the `XsdSchema` continues by parsing its children nodes. To parse children elements of any given `XsdAbstractElement` type we have the `xsdParseSkeleton` function present in the `XsdAbstractElement` class (Listing 4.5). This function will iterate in all the children of a given node, invoke the respective `parse` function of each children and then notify the parent element, using the Visitor pattern[5].

In the XsdParser the Visitor pattern is used to ensure that each concrete element defines different behaviours for different types of children. This provides good flexibility for implementing certain XSD syntax restrictions, e.g. the element A can reject the element B as his children if the element A doesn't support children of type B.

By using this strategy we ensure that the whole XSD file is parsed just by invoking the `parse` function of `XsdSchema`. This happens because the `XsdSchema` element is the top level element of all the XSD files and having all the concrete element types parsing their children results in the parsing of the whole XSD file.

```
ReferenceBase xsdParseSkeleton(Node node, XsdAbstractElement element){
    XsdParser parser = element.getParser();
    Node child = node.getFirstChild();

    while (child != null) { //Iterates in all children from node.
        //Only parses element nodes, ignoring comments and text nodes.
        if (child.getNodeType() == Node.ELEMENT_NODE) {
            String nodeName = child.getNodeName();

            //Searches on a mapper for a parsing functions
            //for the respective type.
            BiFunction<XsdParser, Node, ReferenceBase> parserFunction =
                XsdParser.getParseMappers().get(nodeName);

            //Applies the parsing functions, if any, and notifies
            //the parent objects Visitor to the newly created object.
            if (parserFunction != null){
                XsdAbstractElement childElement =
                  parserFunction.apply(parser, child).getElement();

                childElement.accept(element.getVisitor());
                childElement.validateSchemaRules();
            }
        }

        child = child.getNextSibling(); //Moves on to the next sibling.
    }

    ReferenceBase wrappedElement= ReferenceBase.createFromXsd(element);
    parser.addParsedElement(wrappedElement);
    return wrappedElement;
}
```

Listing 4.5: XsdParseSkeleton Parsing Children From a Node

Based on the explanation provided above, we will give a more detailed description about the parsing process made by `XsdParser` using a small concrete example extracted from the HTML XSD file, present in Listing 4.6.

```
1 <xs:schema>
2     <xs:element name="html">
3         <xs:complexType>
4             <!-- -->
5         </xs:complexType>
6     </xs:element>
7 </xs:schema>
```

Listing 4.6: Parsing Concrete Example

Step 1 - DOM parsing:

The parsing starts with the DOM library parsing the code (Listing 4.6), which returns the `xs:schema` node (i.e. schemaNode in Listing 4.3). `XsdParser` verifies if the node is in fact a `xs:schema` node and after verifying that in fact it is, it invokes the `XsdSchema parse` function (line 19 of Listing 4.4).

Step 2 - XsdSchema Attribute Parsing:

The `XsdSchema parse` function receives the `Node` object and converts it to a `Map` object (line 21 of Listing 4.4). The map object is then passed to the `XsdSchema` constructor which will result in the invocation of the `setFields` method (line 7 of Listing 4.4), which will extract the information from the `Map` object to the class fields.

Step 3 - XsdSchema Children:

To parse the `XsdSchema` children the `XsdAbstractElement xsdParseSkeleton` (Listing 4.5) function is called (line 24 of Listing 4.4) and starts to iterate the `xs:schema` node children, which, in this case, is a node list containing a single element, the `xs:element` node.

Step 4 - XsdElement Attribute Parsing:

The parsing of the `xs:element` node is similar to `xs:schema`, it extracts the attribute information from its respective node in its `setFields` function.

Step 5 - XsdSchema Visitor Notification:

After parsing the `xs:element` node the previously created `XsdSchema` object is notified. This notification informs the `XsdSchema` object that it contains the newly created `XsdElement` object using the Visitor pattern. The `XsdSchema` should then act accordingly based on the type of the object received as his children, since different types of objects should be treated differently.

4.1.2 Reference solving

After the parsing process described previously, there is still an issue to solve regarding the existing references in the XSD schema definition. In XSD files the usage of the `ref` attribute is frequent to avoid repetition of XML code. This generates two main problems when handling reference solving, the first one being existing elements with `ref` attributes referencing non existent elements and the other being the replacement of the reference object by the referenced object when present. In order to effectively help resolve the referencing problem some wrapper classes were added. These wrapper classes contain the wrapped element and serve as a classifier for the wrapped element. The existing wrapper classes are as follow:

- `UnsolvedElement` - Wrapper class to each element that has a `ref` attribute.
- `ConcreteElement` - Wrapper class to each element that is present in the file.
- `NamedConcreteElement` - Wrapper class to each element that is present in the file and has a `name` attribute present.
- `ReferenceBase` - A common interface between `UnsolvedReference` and `ConcreteElement`.

Having these wrappers on the elements allow for a detailed filtering, which is helpful in the reference solving process. That process starts by obtaining all the `NamedConcreteElement` objects since they may or may not be referenced by an existing `UnsolvedReference` object. The second step is to obtain all the `UnsolvedReference` objects and iterate them to perform a lookup search on the `NamedConcreteElement` objects obtained previously. This is achieved by comparing the value present in the `UnsolvedReference ref` attribute with the `NamedConcreteElement name` attribute. If a match is found then XsdParser performs a copy of the object wrapped by the `NamedConcreteElement` and replaces the element wrapped in the `UnsolvedReference` object that served as a placeholder. A concrete example of how this process works is in Listing 4.7.

```
<xsd:schema xmlns='http://schemas.microsoft.com/intellisense/html-5'
    xmlns:xsd='http://www.w3.org/2001/XMLSchema'>

    <!-- NamedConcreteType wrapping a XsdGroup -->
    <xsd:group id="replacement" name="flowContent">
        <!-- (...) -->
    </xsd:group>

    <!-- ConcreteElement wrapping a XsdChoice -->
    <xsd:choice>
        <!-- UnsolvedReference wrapping a XsdGroup -->
        <xsd:group id="toBeReplaced" ref="flowContent"/>
    </xsd:choice>
</xsd:schema>
```

Listing 4.7: Reference Solving Example

In this short example we have a `XsdChoice` element that contains a `XsdGroup` element with a `ref` attribute. When replacing the `UnsolvedReference` objects the `XsdGroup` with the `ref` attribute is going to be replaced by a copy of the already parsed `XsdGroup` with the `name` attribute. This is achieved by accessing the parent of the element, in this case accessing the parent of the `XsdGroup` with the `ref` attribute, in order to remove the element identified by "toBeReplaced" and adding the element identified by "replacement".

Having created these classes it is expected that at the end of a successful file parsing only `ConcreteElement` and/or `NamedConcreteElement` objects remain. In case there are any remainder `UnsolvedReference` objects the programmer can query the parser, using the function `getUnsolvedReferences` of the `XsdParser` class, to discover which elements are missing and where were they used. The programmer can then correct the missing elements by adding them to the XSD file and repeat the parsing process or just acknowledge that those elements are missing.

4.1.3 Validations

As it was already referred in Section 4.1.1 the parser uses some strategies to validate the rules of the XSD language. We already referred the usage of `Enum` classes for attribute values that have a set of possible values but there are more validations. This solution also validates the types of data received, e.g. validating if a given attribute is a positive `Integer` value. There are other more

intricate restrictions relating to the organization between elements, for example the `xsd:element` element is not allowed to have a `ref` attribute value if the `xsd:element` is a direct child of the top-level `xsd:schema` element. All those rules were extracted from the XSD language standard and each time a concrete element is create the respective rules are verified as seen with the `validateSchemaRules` method call in line 21 of Listing 4.5.

Each time any of these rules are violated a `ParsingException` is thrown containing a message detailing the rule that was violated, either being an attribute that doesn't match its type, an attribute that has a value that is not within the possible values for that attribute or the other more complex rules of the XSD language.

4.2 XsdAsm

XsdAsm is a library dedicated to generate a fluent Java interface based on a XSD file. It uses the previously introduced XsdParser library to parse the XSD file contents into a list of Java elements that XsdAsm will use to obtain the information needed to generate the correspondent classes.

To generate classes this library also uses the ASM[1] library, which is a library that provides a Java interface that allows bytecode manipulation providing methods for creating classes, methods, etc. There were other alternatives to the ASM library but most of them are simply libraries that were built on top of ASM to simplify its usage. It supports the creation of Java classes up until Java 9 and is still maintained, the most recent version, 6.2.1, was release in 5 august of 2018. ASM also has some tools to help the new programmers understand how the library works. These tools help the programmers to learn faster how the code generation works and allow to increase the complexity of the generated code. In Listing 4.8 we present a class that is the objective of our code generation, it's a simple class, with a field and a method. The ASM library provides a tool, `ASMifier`, that receives a `.class` file and returns the ASM code needed to generate it, as shown in Listing 4.9.

[1] ASM Website

```
1 public class SumExample {
2
3     private int sum;
4
5     void setSum(int a, int b){
6         sum = a + b;
7     }
8 }
```

Listing 4.8: ASM Example - Objective

```
1 ClassWriter classWriter = new ClassWriter(0);
2
3 classWriter.visit(V9, ACC_PUBLIC + ACC_FINAL + ACC_SUPER,
4   "Samples/HTML/SumExample", null, "java/lang/Object", null);
5
6 FieldVisitor fieldVisitor =
7   classWriter.visitField(ACC_PRIVATE, "sum", "I", null, null);
8 fieldVisitor.visitEnd();
9
10 MethodVisitor methodVisitor =
11   classWriter.visitMethod(0, "setSum", "(II)V", null, null);
12 methodVisitor.visitCode();
13 methodVisitor.visitVarInsn(ALOAD, 0);
14 methodVisitor.visitVarInsn(ILOAD, 1);
15 methodVisitor.visitVarInsn(ILOAD, 2);
16 methodVisitor.visitInsn(IADD);
17 methodVisitor.visitFieldInsn(PUTFIELD, "Samples/HTML/SumExample",
18   "sum", "I");
19 methodVisitor.visitInsn(RETURN);
20 methodVisitor.visitMaxs(3, 3);
21 methodVisitor.visitEnd();
22
23 classWriter.visitEnd();
24
25 writeByteArrayToFile(classWriter.toByteArray());
```

Listing 4.9: ASM Example - Required Code

The strategy while creating the `xmlet` solution was to have manually created classes that represent a certain type of class that XsdASm will need to generate, i.e. the concrete element, attribute classes. By using the `ASMifier` tool with those template-like classes the programming process was expedited.

4.2.1 Supporting Infrastructure

To support the foundations of the XSD language an infrastructure is created in every API generated by this project. This infrastructure is composed by a common set of classes. This supporting infrastructure is divided into three different groups of classes:

Element classes:

- Element - An interface that serves as a base to every parsed XSD element.
- AbstractElement - An abstract class from where all the XSD element derive. This class implements most of the methods present on the Element interface.

Attribute classes:

- Attribute - An interface that serves as a base to every parsed XSD attribute.
- BaseAttribute - A class that implements the Attribute interface and adds restriction verification to all the deriving classes. All the attributes that have restrictions should derive from this class.

Visitor class:

- ElementVisitor - An abstract class that defines methods for all the generated elements that can be visited with the Visitor pattern. All the implemented methods point to a single method. This behaviour aims to reduce the amount of code needed to create concrete implementations of Visitors.

Taking in consideration those classes, a very simplistic API could be represented with the class diagram (Figure 4.1). In this example we have an element, `Html`, that extends `AbstractElement` and an attribute, `AttrManifestString`, that extends `BaseAttribute`.

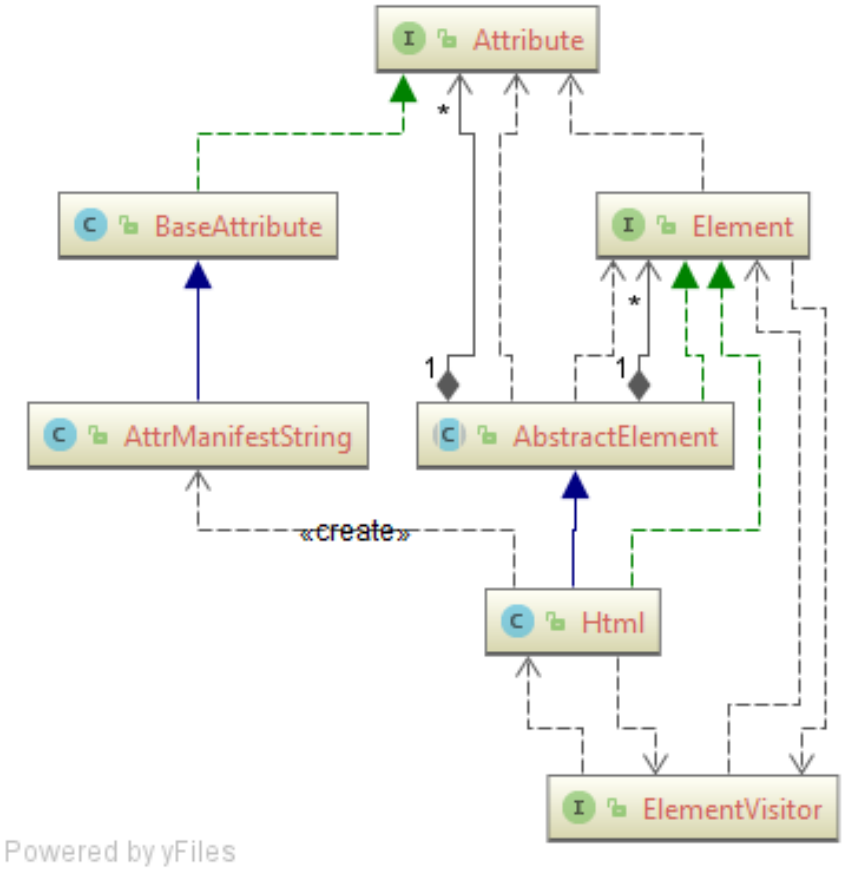


Figure 4.1: API - Supporting Infrastructure

4.2.2 Code Generation Strategy

As we already presented before in the Problem Statement section, Section 2.2, this solution focus on how the code is organized instead of making complex code. All the methods present in the generated classes have very low complexity, mainly adding information to the element children and attribute list. To reduce repeated code many interfaces with default methods are created so different classes can implement them and reuse the code. The complexity of the generated code is mostly present in the `AbstractElement` class, which implements most of the `Element` interface methods. Another very important aspect of the generated classes is the extensive use of type arguments which allows the navigation in the element tree while maintaining type information which is essential to guarantee the specific language restrictions.

4.2.3 Type Arguments

As this solution was designed an objective became clear, the generated API should be easily navigable. This is crucial to provide a good user experience while creating templates through the `xmlet` fluent interfaces. There are two main aspects, the fluent interface should be easily navigable and always implement the concrete language restrictions. To tackle this issue we rely on type arguments. Through type arguments we can always keep track of the tree structure of the elements that are being created and keep adding elements or moving up in the tree structure. In Listing 4.10 we can observe how the type arguments work.

```
1 Html<Element> html = new Html<>();
2 Body<Html<Element>> body = html.body();
3 Div<Body<Html<Element>>> div = body.div();
4 Div<Body<Html<Element>>> divAfterAttribute =
5                           div.attrClass("attrClassValue");
6 Body<Html<Element>> bodyAgain = divAfterAttribute.°();
7 Html<Element> htmlAgain = bodyAgain.°();
```

Listing 4.10: Explicit Example of Type Arguments

When we create the `Html` element we should indicate that he has a parent, for consistency. Then, as we add elements such as `Body` we automatically return the recently created `Body` element, but with parent information that indicates that this `Body` instance is descendant of an `Html` element. The same happens with the `Div` element. This behavior changes when adding attributes, in which case we return the object which had the attribute added to them, while maintaining all the parent information. The last two lines show how to move up in the element tree by using the method `°`. The method name is meant to be as short as possible and to avoid distractions while reading the template.

While in the example presented in Listing 4.10 the usage of the API might seem to have excessive verbose to define a simple HTML document that verbose isn't exactly needed. For specific purposes it might be needed to extract variables but the most common usage of the fluent interface should be more similar to Listing 4.11.

```
new Html<>()
      .body()
           .div().attrClass("attrClassValue").°()
      .°();
```

Listing 4.11: Simpler Example of Type Arguments

To provide a better understanding on how this is possible we need to showcase three distinct classes. First we have the `AbstractElement` class, Listing 4.14, from which all concrete elements derive. This class receives two type arguments:

- **T** - Represents the type of the concrete element;
- **Z** - Represents the type of the parent of the concrete element.

In the `°` method, which returns the parent of any concrete element, the type parameter is guaranteed by returning `Z`, which is the type of the parent of the current element, as shown in the last two lines of code in Listing 4.10.

```
class AbstractElement<T extends Element, Z extends Element>
  protected Z parent;

    protected AbstractElement(Z parent) {
        this.parent = parent;
    }

  public Z °() {
    return this.parent;
  }

  // (...)
}
```

Listing 4.12: Type Arguments - AbstractElement class

The second class is the `Html` class, which is representative of every concrete element of an `xmlet` API. It has a single type argument, Z which represents the type of its parent. It extends `AbstractElement` and therefore indicates that his type is `Html<Z>` and its parent type is `Z`. Any interface implemented by the concrete elements should receive the same type information as the `AbstractElement` class, as shown with `HtmlChoice0`. Regarding the `attrManifest` method, it indicates that returns the exact same instance as the one from which the method is called, keeping the type information by returning `Html<Z>`.

```
class Html<Z extends Element> extends AbstractElement<Html<Z>, Z>
                                implements HtmlChoice0<Html<Z>, Z>
    // (...)

    public Html<Z> attrManifest(String attrManifest) {
        return this.addAttr(new AttrManifestString(attrManifest));
    }
}
```

Listing 4.13: Type Arguments - AbstractElement class

As the third class we have the `HtmlChoice0` interface which is representative of most interfaces of an `xmlet` API. It should receive the same type arguments as `AbstractElement`:

- **T** - Represents the type of the concrete element;
- **Z** - Represents the type of the parent of the concrete element.

Since the `Html` is the only class implementing this interface its `T` type parameter is always `Html<Z>` which results in a return type of `Body<Html<Z>>` when the method `body` is called.

```
interface HtmlChoice0<T extends Element<T, Z>, Z extends Element>
                                                    extends Element<T, Z> {
    default Body<T> body() {
        return (Body)this.addChild(new Body(this.self()));
    }
}
```

Listing 4.14: Type Arguments - AbstractElement class

4.2.4 Restriction Validation

In the description of any given XSD file there are many restrictions in the way the elements are contained in each other and which attributes are allowed. To reflect those restrictions to Java language there are two alternatives, validation in runtime or in compile time. This library tries to validate most of the restrictions in compile time, as shown above by the way classes are created. But some restrictions can't be validated in compile time, an example of this is the following restriction (Listing 4.15):

```
<xs:schema>
    <xs:element name="testElement">
        <xs:complexType>
            <xs:attribute name="intList" type="valuelist"/>
        </xs:complexType>
    </xs:element>

    <xs:simpleType name="valuelist">
        <xs:restriction>
            <xs:maxLength value="5"/>
            <xs:minLength value="1"/>
        </xs:restriction>
        <xs:list itemType="xsd:int"/>
    </xs:simpleType>
</xs:schema>
```

Listing 4.15: Restrictions Example

In this example (Listing 4.15) we have an element (i.e. testElement) that has an attribute called `intList`. This attribute has some restrictions, it is represented by a `xs:list`, the list elements have the `xsd:int` type and its element count should be between 1 and 5. Transporting this example to the Java language will result in the following class (Listing 4.16):

```
public class AttrIntListObject extends BaseAttribute<List<Integer>> {
    public AttrIntListObject(List<Integer> list) {
        super(list);
    }
}
```

Listing 4.16: Attribute Class Example

But with this solution the `xs:maxLength` and `xs:minLength` values are ignored. To solve this problem the existing restrictions in any given attribute are hardcoded in the class constructor, which invokes methods present in `RestrictionValidator` that validate each type of restriction, e.g. `xs:maxLength` and `xs:minLength`. The values present in the restrictions on the XSD document are hardcoded in the bytecodes and help validate each attribute object that is created. This results in the generation of a constructor as shown in Listing 4.17.

```
1 public class AttrIntlistObject extends BaseAttribute<List<Integer>> {
2     public AttrIntlistObject(List attrValue) {
3         super(attrValue, "intlist");
4         RestrictionValidator.validateMaxLength(5, attrValue);
5         RestrictionValidator.validateMinLength(1, attrValue);
6     }
7 }
```

Listing 4.17: Attribute Static Constructor Restrictions

In total there are thirteen different restrictions on the XSD language. The `RestrictionValidator` class is a class with static methods that allow to validate most of those restrictions, the only restrictions that aren't validated by this class are `xsd:enumeration` which is already validated by the usage of `Enum` classes and `xsd:whitespace` since it represents an indication instead of an actual restriction on the language. In Listing 4.18 we can observe how simple is to validate the `xs:maxLength` and `xs:minLength` restrictions that were used in the previous example. All the methods work in the exact same way, a condition is verified and if the verification fails it will throw a `RestrictionViolationException` with a message indicating which restriction was violated.

```
1 public class RestrictionValidator {
2     public static void validateMaxLength(int maxLength, List list){
3         if (list.size() > maxLength){
4             throw new RestrictionViolationException("Violation of
                  maxLength restriction");
5         }
6     }
7
8     public static void validateMinLength(int minLength, List list){
9         if (list.size() < minLength){
10            throw new RestrictionViolationException("Violation of
                  minLength restriction");
11        }
12    }
13 }
```

Listing 4.18: RestrictionValidator Class (Simplified)

4.2.4.1 Enumerations

Regarding restrictions there is one that can be enforced at compile time, the `xs:enumeration`. To obtain that validation at compile time the XsdAsm library generates `Enum` classes that contain all the values indicated in the `xs:enumeration` elements. In the following example (Listing 4.19) we have an attribute with three possible values, command, checkbox and radio.

```
<xs:attribute name="type">
    <xs:simpleType>
        <xs:restriction base="xsd:string">
            <xs:enumeration value="command" />
            <xs:enumeration value="checkbox" />
            <xs:enumeration value="radio" />
        </xs:restriction>
    </xs:simpleType>
</xs:attribute>
```

Listing 4.19: Enumeration XSD Definition

This results in the creation of an `Enum`, `EnumTypeCommand` presented in Listing 4.20. The attribute class will then receive an instance of `EnumTypeCommand`, ensuring that only allowed values are used (Listing 4.21).

```
public enum EnumTypeCommand {
    COMMAND(String.valueOf("command")),
    CHECKBOX(String.valueOf("checkbox")),
    RADIO(String.valueOf("radio"))
}
```

Listing 4.20: Enumeration Class

```
public class AttrTypeEnumTypeCommand extends BaseAttribute<String> {
    public AttrTypeEnumTypeCommand(EnumTypeCommand attrValue) {
        super(attrValue.getValue());
    }
}
```

Listing 4.21: Attribute Receiving An Enumeration

4.2.5 Element Binding

To support repetitive tasks over an element the `Element` and `AbstractElement` classes were modified to support binders. This allows

programmers to define, for example, templates for a given element. An example is presented in Listing 4.22 using the HTML5 API.

```
1  public class BinderExample{
2      public void bindExample(){
3      Html<Element> root = new Html<>()
4              .body()
5                  .table()
6                      .tr()
7                          .th()
8                              .text("Title")
9                          .°()
10                     .°()
11                     .<List<String>>binder((elem, list) ->
12                         list.forEach(tdValue ->
13                                 elem.tr().td().text(tdValue)
14                         )
15                     )
16                 .°()
17             .°();
18     }
19 }
```

Listing 4.22: Binder Usage Example

In this example we use the HTML language to create a document that contains a table with a title in the first row as a title header (i.e. `th()`). In regard to the values present in the table instead of having them inserted right away it is possible delay that insertion by indicating behaviour to execute when the information is received. This is achieved by implementing an `ElementVisitor` that supports binding.

In Listing 4.23 we can observe how the `ElementVisitor` would work. It maintains the default behaviour on the elements that aren't bound (i.e. else clause). In the case that the element is bound to a function this implementation will clone the element and apply a model (i.e. a `List<String>` object following the example of Listing 4.22) to the clone, effectively executing the function supplied in the previously called binder method (i.e. Listing 4.22 line 8). This function call will generate new children on the cloned table element which will be iterated as if they belonged to the original element tree. This behaviour ensures that the original element tree isn't affected since all these changes are performed in a clone of the bound element, meaning that the template can be reused.

```
public class CustomVisitor<R> implements ElementVisitor<R> {

    private R model;

    public CustomVisitor(R model){
        this.model = model;
    }

    public <T extends Element> void sharedVisit(Element<T,?> element) {
        // ...
        if(element.isBound()) {
            List<Element> children = element.cloneElem()
                                            .bindTo(model)
                                            .getChildren();
            children.forEach( child -> child.accept(this));
        } else {
            element.getChildren().forEach(item -> item.accept(this));
        }
        // ...
    }
}
```

Listing 4.23: Visitor with binding support

4.2.6 Using the Visitor Pattern

In the previous sections we presented how the fluent interface is generated and how it implements the language restrictions, but what can the fluent interface actually be user for? That's strictly up to the user of the generated fluent interface. To achieve this we use the `Visitor` pattern[5]. There are multiple `visit` methods that are invoked by the generated classes and the user can define the behaviour that the `ElementVisitor` has when those methods are called. This way the generated code delegates the responsibility to define how the user wants to interact with the generated DSL. The generated `ElementVisitor` class defines four main `visit` methods, Listing 4.24:

- `sharedVisit(Element<T, ?> element)` - This method is called whenever a concrete element has its `accept` method called. By receiving the `Element` we have access to the element children and attributes.
- `visit(Text text)` - This method is called when the `accept` method of the special `Text` element is invoked.

- `visit(Comment comment)` - This method is called when the `accept` method of the special `Comment` element is invoked.
- `visit(TextFuction<R, U, ?> textFunction)` - This method is called when the `accept` method of the special `TextFunction` element is invoked.

```
1 public abstract class ElementVisitor<R> {
2    <T extends Element> void sharedVisit(Element<T, ?> element);
3
4    void visit(Text text);
5
6    void visit(Comment comment);
7
8    <U> void visit(TextFunction<R, U, ?> textFunction);
9 }
```

Listing 4.24: ElementVisitor generated by XsdAsm - 1

Apart from these four main method we also create specific methods, as shown in Listing 4.25. These methods default behaviour is to invoke the main `sharedVisit(Element<T, ?> element)` method, but they can be redefined to perform a different action, providing the concrete `ElementVisitor` to have a very simple implementation of only four methods or redefine all the methods for a concrete purpose for the respective DSL.

```
1 public abstract class ElementVisitor {
2    // (...)
3
4    default void visit(Html html) {
5       this.sharedVisit(html);
6    }
7 }
```

Listing 4.25: ElementVisitor generated by XsdAsm - 2

4.2.7 Performance - XsdAsmFaster

The `xmlet` developed two alternative solutions to generate fluent interfaces. The first solution that was implemented was XsdAsm, which generated a fluent interface that defined element and attribute classes. When interacting with those

elements it was possible to add children or attributes that were stored in a data structure as seen by the implementation of `AbstractElement` and the snippet of the `Html` code present in Listing 4.26 and 4.27, respectively.

```
1  abstract class AbstractElement<T extends Element, Z extends Element>
       implements Element<T, Z> {
2     protected List<Element> children = new ArrayList();
3     protected List<Attribute> attrs = new ArrayList();
4
5     // (...)
6
7     public <R extends Element> R addChild(R child) {
8        this.children.add(child);
9        return child;
10    }
11
12    public T addAttr(Attribute attribute) {
13       this.attrs.add(attribute);
14       return this.self();
15    }
16 }
```

Listing 4.26: AbstractElement generated by XsdAsm

```
1  class Html<Z extends Element> extends AbstractElement<Html<Z>, Z> {
2     public void accept(ElementVisitor visitor) {
3        visitor.visit(this);
4     }
5
6     public Html<Z> attrManifest(String attrManifest) {
7        return (Html)this.addAttr(new AttrManifestString(attrManifest));
8     }
9
10    public Body<T> body() {
11       return this.addChild(new Body(this));
12    }
13
14    public Head<T> head() {
15       return this.addChild(new Head(this));
16    }
17 }
```

Listing 4.27: Html class generated by XsdAsm

By using the XsdAsm generated solution we end up with a fluent interface that works in two steps basis:

- Creating the Element tree - We need to create the element tree by adding all elements and attributes (Listing 4.28);
- Visiting the Element tree - We need to invoke the `accept` method of the root of the tree in order for the whole tree to be visited (Listing 4.29).

```
1  Html<Html> root = new Html<>();
2
3  root.head()
4          .title()
5              .text("Title")
6          .°()
7      .°()
8      .body().attrClass("clear")
9          .div()
10           .h1()
11             .text("H1 text")
12           .°()
13         .°()
14     .°()
15 .°();
```

Listing 4.28: HTML5 tree creation - XsdAsm usage

```
1  CustomVisitor customVisitor = new CustomVisitor();
2
3  // root variable created in the previous Listing.
4  root.accept(customVisitor);
```

Listing 4.29: HTML5 tree visit - XsdAsm usage

Even though that this solution worked fine it had a performance issue. Why were we adding elements to a data structure just for it to be iterated at a later time? From this idea a new solution was born, XsdAsmFaster. This new solution aims to perform the same operations faster while providing a very similar user experience to the fluent interface generated by XsdAsm. To achieve that instead of storing information on a data structure we directly invoke the `ElementVisitor visit` method, this removes the need of storing and iterating information while maintaining all the expected behaviour. The two main moments that are affected

by this change are the moments when an element is added to the tree and when an attribute is added to a previously created element. The code generated by XsdAsmFaster to add elements is as shown in Listing 4.30.

```
1  public final class Html<Z extends Element> {
2     protected final Z parent;
3     protected final ElementVisitor visitor;
4
5     public Html(ElementVisitor visitor) {
6        this.visitor = visitor;
7        this.parent = null;
8        visitor.visitElementHtml();
9     }
10
11    public Html(Z parent) {
12       this.parent = parent;
13       this.visitor = parent.getVisitor();
14       this.visitor.visitElementHtml();
15    }
16
17    public final Html<Z> attrManifest(String attrManifest) {
18       this.visitor.visitAttributeManifest(attrManifest);
19       return this;
20    }
21
22    public Body<T> body() {
23       return new Body(this);
24    }
25
26    public Head<T> head() {
27       return new Head(this);
28    }
29 }
```

Listing 4.30: Html class generated by XsdAsmFaster

As we can see in the previously Listing we can invoke the `visit` method in the constructor of the concrete element classes, in this case the `Html` class, since the `ElementVisitor` object is passed to all the elements on the tree. Since adding elements results in the creation of new objects, such as `Body` and `Head` in the previous example, it results in the invocation of their respective `visit` method due to the `visit` method being called in each concrete element constructor. The attributes have a very similar behaviour, although they don't have instances created their restrictions are validated, if present, and if all restrictions are validated

the respective `visit` method is called and the method ends returning the object `this` to continue with the fluent tree creation.

The XsdAsmFaster solution also adds many other performance improvements. The `ElementVisitor` methods were changed to receive `String` objects instead of `Element`, `Attribute` or `Text` types. Changing this removes the requirement to instantiate attribute concrete classes since we can directly pass the `name` of the attribute and its `value` as shown in `attrManifest` method in Listing 4.30. The `ElementVisitor` class of XsdAsmFaster is as present in Listing 4.31.

```
1  public abstract class ElementVisitor {
2     public abstract void visitElement(String elementName);
3
4     public abstract void visitAttribute(String attributeName, String
          attributeValue);
5
6     public abstract void visitParent(String elementName);
7
8     public abstract <R> void visitText(R text);
9
10    public abstract <R> void visitComment(R comment);
11
12    public void visitParentHtml() {
13       this.visitParent("html");
14    }
15
16    public void visitElementHtml() {
17       this.visitElement("html");
18    }
19
20    // (...)
21 }
```

Listing 4.31: ElementVisitor generated by XsdAsmFaster

4.3 Client

To use and test both XsdAsm and XsdParser we need to implement a client for XsdAsm. Two different clients were implemented, one using the HTML5 specification and another using the specification for Android visual layouts. In this section we are going to explore how the HTML5 fluent interface is generated using the XsdAsm library and how to use it.

4.3.1 HtmlApi

To generate the HTML5 fluent interface we need to obtain its XSD file. After that there are two options, the first one is to create a Java project that invokes the XsdAsm `main' method directly by passing the path of the specification file and the desired fluent interface name that will be used to create a custom package name (Listing 4.32).

```
1 void generateApi(String filePath, String apiName){
2     XsdAsmMain.main(new String[] {filePath, apiName} );
3 }
```

Listing 4.32: API creation

The second option is using the Maven[2] build lifecycle[3] to make that same invocation by adding an extra execution to the *Project Object Model* (POM) file (Listing 4.33) to execute a batch file that invokes the XsdAsm main method (Listing 4.34).

[2]Maven
[3]Maven Build Lifecycle

```
<plugin>
    <artifactId>exec-maven-plugin</artifactId>
    <groupId>org.codehaus.mojo</groupId>
    <version>1.6.0</version>
    <executions>
        <execution>
            <id>create_classes1</id>
            <phase>validate</phase>
            <goals>
                <goal>exec</goal>
            </goals>
            <configuration>
                <executable>
                    ${basedir}/create_class_binaries.bat
                </executable>
            </configuration>
        </execution>
    </executions>
</plugin>
```

Listing 4.33: Maven API compile classes plugin

```
if exist "./src/main/java" rmdir "./src/main/java" /s /q

if not exist "./target/classes/org/xmlet/htmlapi"
    mkdir "./target/classes/org/xmlet/htmlapi"

call
    mvn exec:java -D"exec.mainClass"="org.xmlet.xsdasm.main.XsdAsmMain"
    -D"exec.args"="./src/main/resources/html_5.xsd htmlapi"
```

Listing 4.34: Maven API creation batch file (create_class_binaries.bat)

This client uses the Maven lifecycle option by adding an execution at the `validate` phase (Listing 4.33, line 8) which invokes XsdAsm main method to create the API. This invocation of XsdAsm creates all the classes in the target folder of the HtmlApi project. Following these steps would be enough to allow any other Maven project to add a dependency to the HtmlApi project and use its generated classes as if they were manually created. But this way the source files and Java documentation files are not created since XsdAsm only generates the class binaries. To tackle this issue we added another execution to the POM. This

execution uses the Fernflower[4] decompiler, the Java decompiler used by Intellij[5] *Integrated Development Environment* (IDE), to decompile the classes that were automatically generated (Listing 4.35, 4.36).

```
<!-- Plugin Information -->
<execution>
    <id>decompile_classes</id>
    <phase>validate</phase>
    <goals>
        <goal>
            exec
        </goal>
    </goals>
    <configuration>
        <executable>
            ${basedir}/decompile_class_binaries.bat
        </executable>
    </configuration>
</execution>
```

Listing 4.35: Maven API decompile classes plugin

```
if not exist "./src/main/java/org/xmlet/htmlapi" mkdir "./src/main/java
    /org/xmlet/htmlapi"

call
    mvn exec:java
    -D"exec.mainClass"="org.jetbrains.java.decompiler.main.decompiler.
       ConsoleDecompiler"
    -D"exec.args"="-dgs=true ./target/classes/org/xmlet/htmlapi ./src/
       main/java/org/xmlet/htmlapi"

if exist "./target/classes/org" rmdir "./target/classes/org" /s /q
```

Listing 4.36: Maven API decompile batch file (decompile_class_binaries.bat)

By decompiling those classes we obtain the source code which allows us to delete the automatic generated classes and allow the Maven build process to perform the normal compiling process, which generates the Java documentation files and the class binaries, along with the source files obtained from the decompilation process. This process, apart from generating more information to the programmer that will use the fluent interface in the future, also allows to find any problem

[4]Fernflower Decompiler
[5]Intellij IDE

with the generated code since it forces the compilation of all the classes previously generated.

4.3.1.1 Using the HtmlApi

After the previously described compilation process of the HtmlApi project we are ready to use the generated fluent interface. To start using it the first step is to implement the `ElementVisitor` class, which defines what to do when the created element tree is visited. A very simple example is presented in Listing 4.37 which writes the HTML tags based on the name of the element visited and navigates in the element tree by accessing the children of the current element.

```
1  public class CustomVisitor<R> implements ElementVisitor<R> {
2
3      private PrintStream printStream = System.out;
4
5      public CustomVisitor(){ }
6
7      public <T extends Element> void sharedVisit(Element<T,?> element) {
8          printStream.printf("<%s", element.getName());
9
10         element.getAttributes()
11                 .forEach(attribute ->
12                     printStream.printf(" %s=\"%s\"",
13                         attribute.getName(), attribute.getValue()));
14
15         printStream.print(">\n");
16
17         element.getChildren().forEach(item -> item.accept(this));
18
19         printStream.printf("</%s>\n", element.getName());
20     }
21 }
```

Listing 4.37: Custom Visitor

After creating the `CustomVisitor` presented in Listing 4.37 we can start to create the element tree that we want convert to text using the `CustomVisitor`. To start we should create a `Html` object, since all the HTML documents have it as a base element. Upon creating that root element we can start to add other elements or attributes that will appear as options based on the specification rules. To help with the navigation on the element tree a method was created to allow the

navigation to the parent of any given element. This method is named °, a short method name to keep the code as clean as possible. In Listing 4.38 we can see a code example that uses a good amount of the fluent interface features, including element creation and how they are added to the element tree, how to add attributes, attributes that receive enumerations as parameters and how to navigate in the element tree using the method °.

```
1  Html<Html> root = new Html<>();
2
3  root.head()
4          .meta().attrCharset("UTF-8").°()
5          .title()
6              .text("Title").°()
7          .link().attrType(EnumTypeContentType.TEXT_CSS)
8                  .attrHref("/assets/images/favicon.png").°()
9          .link().attrType(EnumTypeContentType.TEXT_CSS)
10                 .attrHref("/assets/styles/main.css").°().°()
11     .body().attrClass("clear")
12         .div()
13             .header()
14                 .section()
15                     .div()
16                         .img().attrId("brand")
17                               .attrSrc("./assets/images/logo.png").°()
18                         .aside()
19                             .em()
20                                 .text("Advertisement")
21                             .span()
22                                 .text("HtmlApi is great!");
23
24 CustomVisitor customVisitor = new CustomVisitor();
25
26 customVisitor.visit(root);
```

Listing 4.38: HtmlApi Element Tree

With this element tree presented (Listing 4.38) and the previously presented `CustomVisitor` (Listing 4.37) we obtain the following result (Listing 4.39). The indentation was added for readability purposes, since the `CustomVisitor` implementation in Listing 4.37 does not indent the resulting HTML.

```
<html>
    <head>
        <meta charset="UTF-8">
        </meta>
        <title>
            Title
        </title>
        <link type="text/css" href="/assets/images/favicon.png">
        </link>
        <link type="text/css" href="/assets/styles/main.css">
        </link>
    </head>
    <body class="clear">
        <div>
            <header>
                <section>
                    <div>
                        <img id="brand" src="./assets/images/logo.png">
                        </img>
                        <aside>
                            <em>
                                Advertisement
                                <span>
                                    HtmlApi is great!
                                </span>
                            </em>
                        </aside>
                    </div>
                </section>
            </header>
        </div>
    </body>
</html>
```

Listing 4.39: HtmlApi Visitor Result

The `CustomVisitor` of Listing 4.39 is a very minimalist implementation since it does not indent the resulting HTML, does not simplify elements with no children (i.e. the link/img elements) and other aspects that are particular to HTML syntax. That is where the HtmlFlow library comes in, it implements the particular aspects of the HTML syntax in its `ElementVisitor` implementation which deals with how and where the output is written.

5

Deployment

5.1 Github Organization

This project and all its components belong to a github organization called `xmlet`[1]. The aim of that organization is to contain all the related projects to this dissertation. All the generated APIs are also created as if they belong to this organization.

5.2 Maven

In order to manage the developed projects a tool for project organization and deployment was used, named Maven. Maven has the goal of organizing a project in many different ways, such as creating a standard of project building and managing project dependencies. Maven was also used to generate documentation and deploying the projects to a central code repository, Maven Central Repository[2]. All the releases of projects belonging to the `xmlet` Github organization can be found under the groupId, com.github.xmlet.

[1] xmlet Github
[2] Maven Central Repository

5.3 Sonarcloud

Code quality and its various implications such as security, low performance and bugs should always be an important issue to a programmer. With that in mind all the projects contained in the `xmlet` solution were evaluated in various metrics and the results made public for consultation. This way, either future users of those projects or developers trying to improve the projects can check the metrics as another way of validating the quality of the produced code. The tool to perform this evaluation was Sonarcloud[3], which provides free of charge evaluations and stores the results which are available for everyone. Sonarcloud also provides an API to show badges that allow to inform users of different metrics regarding a project. Those badges are presented in the `xmlet` modules Github pages, as shown in Figure 5.1.

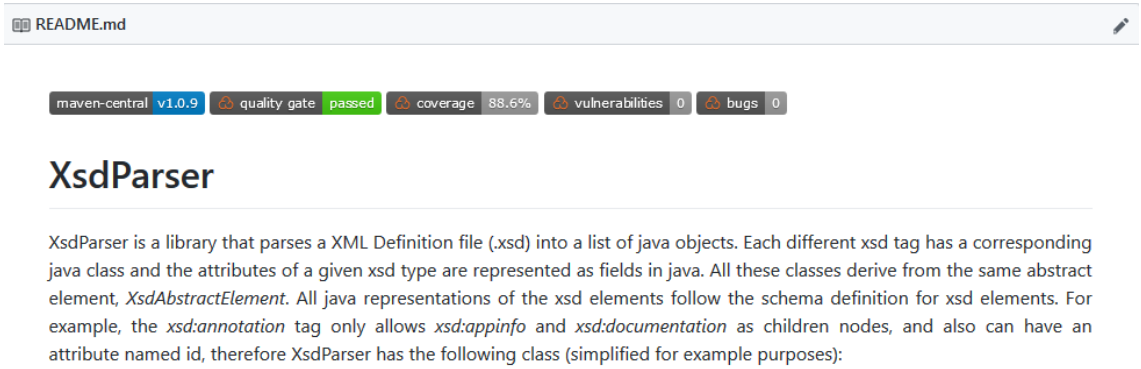


Figure 5.1: XsdParser Github badges

5.4 Testing metrics

To assert the performance of the `xmlet` solution we used the HTML5 use case to compare it against the J2html (Section 3.2.2) and Apache Velocity (Section **??**), presented earlier. This two libraries have a difference that is crucial when dealing with performance, J2html doesn't indent the generated HTML while Apache Velocity does it. In order to perform a fair comparison between solutions two Visitors were used with the HTML5 solution, one that indents the HTML to compare it with Velocity and other that doesn't indent it to compare it to J2html. The computer used to perform all the tests present in this section has the following specs:

[3] Sonarcloud xmlet page

Processor: Intel Core i3-3217U 1.80GHz
RAM: 4GB

The tests that are presented below consist in the generation of a simple HTML document, with a table and a variable amount of table entries as shown in Listing 5.1.

```
<html>
    <body>
        <table>
            <tr>
                <th>
                    Title
                </th>
            </tr>

            <!-- Repeated based on the number of elements -->
            <tr>
                <td>
                    ElemX
                </td>
            </tr>

        </table>
    </body>
</html>
```

Listing 5.1: Test HTML

The textual elements to feature in the table data rows are stored in the same data structure and each solution should iterate that structure when generating the expected HTML. To assert which solution is the fastest each one of them will generate the same HTML document and we will verify the number of iterations that are possible to perform in a certain unit of time. To achieve this objective we will use *Java Microbenchmark Harness* (JMH)[4], which is a tool used for benchmarking. The values gathered are result of the mean values of 8 testing iterations after 12 iterations of warm up.

[4]JMH Website

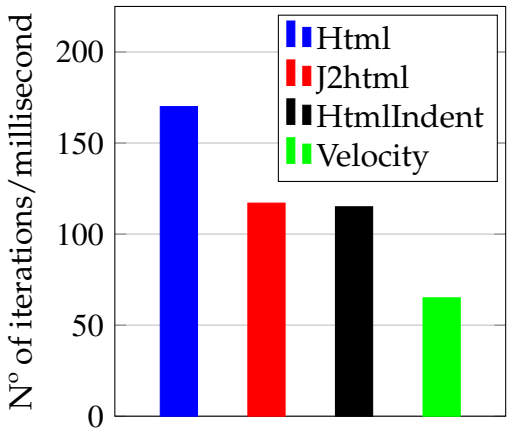


Figure 5.2: Benchmark: Generating an Html Table with 10 Elements

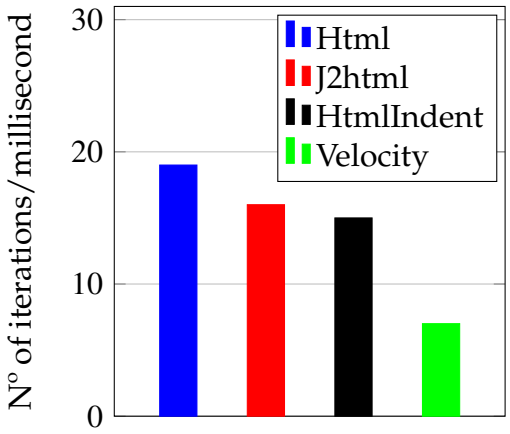


Figure 5.3: Benchmark: Generating an Html Table with 100 Elements

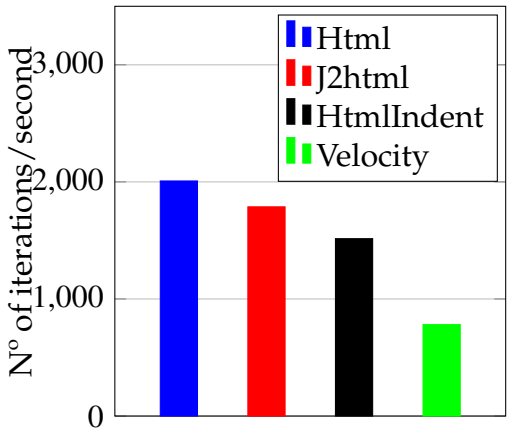


Figure 5.4: Benchmark: Generating an Html Table with 1.000 Elements

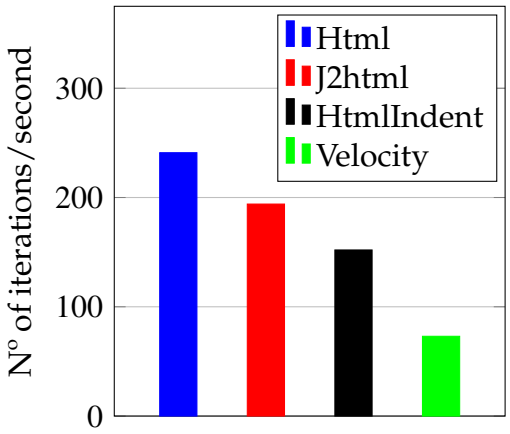


Figure 5.5: Benchmark: Generating an Html Table with 10.000 Elements

As we can see by the presented results the HtmlApi outperforms the other solutions, either with few elements or with a large number of elements.

5.4.1 Spring benchmark

To assert that the results obtained through the previous benchmark weren't in any way biased in favor of the `HtmlApi` we performed another benchmark. This second benchmark was performed by using a preexisting project that compares ten different template engines. To give some insight about each of the template engines used in this benchmark project we have Table 5.1 which enumerates the functionalities that each template engine provides. From this table we can observe that all these solutions being compared provide similar features.

Name	Variables	Functions	Loops	Includes	Exceptions	Hierarchy
Jade	✓		✓			✓
Mustache	✓	✓	✓	✓	✓	
Pebble	✓	✓	✓	✓	✓	✓
Velocity	✓	✓	✓	✓	✓	✗
Chunk	✓	✓	✓	✓	✓	✓
JSP	✓	✓	✓	✓	✓	
Jtwig						
FreeMarker	✓	✓	✓	✓	✓	✓
Thymeleaf	✓	✓	✓	✓	✓	✓
Handlebars	✓	✓	✓	✓	✓	✗

Table 5.1: Template Engines Functionality Comparison

The project used to perform this benchmark, named spring-comparing-template-engines [5], uses a tomcat server[6] to launch a WebApi based on the Java Spring[7] framework. This WebApi provides routes for each solution to test, e.g. http://localhost:8080/mustache, which returns a standard HTML page based in two distinct aspects.

- Template - Each solution under benchmark defines the same template, in their own syntax. An example is the Mustache template in Listing 5.2.

- Data - A `List` of `Presentation` objects (Listing 5.3), which all the solutions receive to complete their templates.

[5] Template Engines Benchmark
[6] Apache Tomcat
[7] Java Spring

```
<!DOCTYPE html>
<html>
    <head>
        <meta charset="utf-8"/>
        <meta name="viewport" content="width=device-width, initial-
            scale=1.0"/>
        <meta http-equiv="X-UA-Compatible" content="IE=Edge"/>
        <title>{{#i18n}}example.title{{/i18n}} - Mustache</title>
        <link rel="stylesheet" href="{{rc.contextPath}}/webjars/
            bootstrap/3.3.7-1/css/bootstrap.min.css"/>
    </head>
    <body>
        <div class="container">
            <div class="page-header">
                <h1>{{#i18n}}example.title{{/i18n}} - Mustache</h1>
            </div>
            {{#presentations}}
                <div class="panel panel-default">
                    <div class="panel-heading">
                        <h3 class="panel-title">{{title}} - {{
                            speakerName}}</h3>
                    </div>
                    <div class="panel-body">
                        {{{summary}}}
                    </div>
                </div>
            {{/presentations}}
        </div>
        <script src="{{rc.contextPath}}/webjars/jquery/3.1.1/jquery.min
            .js"></script>
        <script src="{{rc.contextPath}}/webjars/bootstrap/3.3.7-1/js/
            bootstrap.min.js"></script>
    </body>
</html>
```

Listing 5.2: Spring Benchmark Mustache Template

```
1 public class Presentation {
2     private Long id;
3     private String title;
4     private String speakerName;
5     private String summary;
6     private String room;
7     private Date startTime;
8     private Date endTime;
9 }
```

Listing 5.3: Spring Benchmark Presentation Object

Since all the different solutions used in this API generate the same HTML document using the same information we can evaluate which solution is faster. To incorporate the `HtmlApi` solution in this WebApi we registered another entry in Spring settings and defined the same template as the other solutions.

After setting the Spring project with our solution we need to introduce other tool, the ApacheBench[8]. This tool is used to flood a given *Uniform Resource Locator* (URL) with requests and measures how much time it takes for the server to respond to those requests. The goal is to flood every route with a set number of requests and evaluate the time that each solution needs to respond to the same number of requests, effectively measuring each solution performances. As a last step we also evaluated the overhead that the Spring framework introduces in this process by creating a route that returns an empty answer. By obtaining this overhead value we can subtract it to the results obtained in each benchmark route in order to exclusively compare the time that the respective solution needs to generate their response.

The benchmark values presented below are the result of a single iteration after running two warm up iterations. Multiple variants were used to test the template engines and the HtmlApi solution. The presented results are obtained after running the specified benchmark and *removing* the Spring overhead.

[8] ApacheBench

10.000 requests using 10 threads with a Spring overhead of 1.031 seconds:

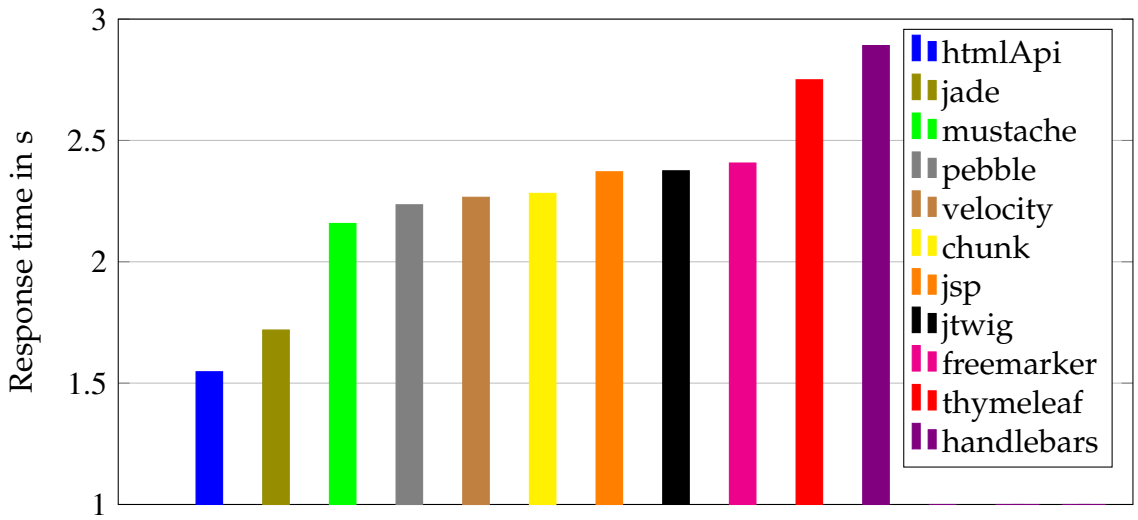


Figure 5.6: Benchmark: Spring Benchmark with 10.000 requests issued with 10 threads

30.000 requests using 10 threads with a Spring overhead of 3.285 seconds:

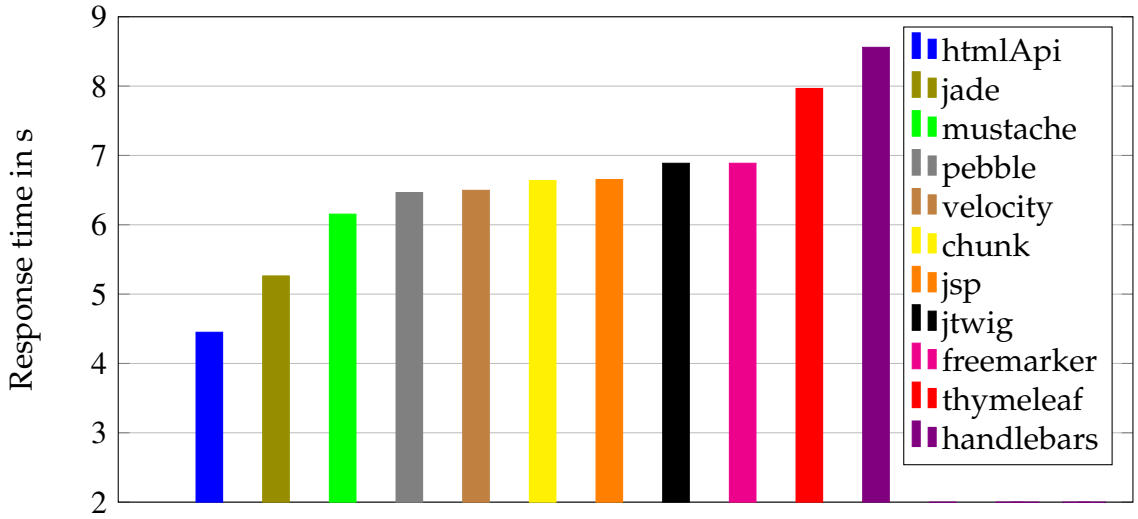


Figure 5.7: Benchmark: Spring Benchmark with 30.000 requests issued with 10 threads

30.000 requests using 25 threads with a Spring overhead of 3.787 seconds:

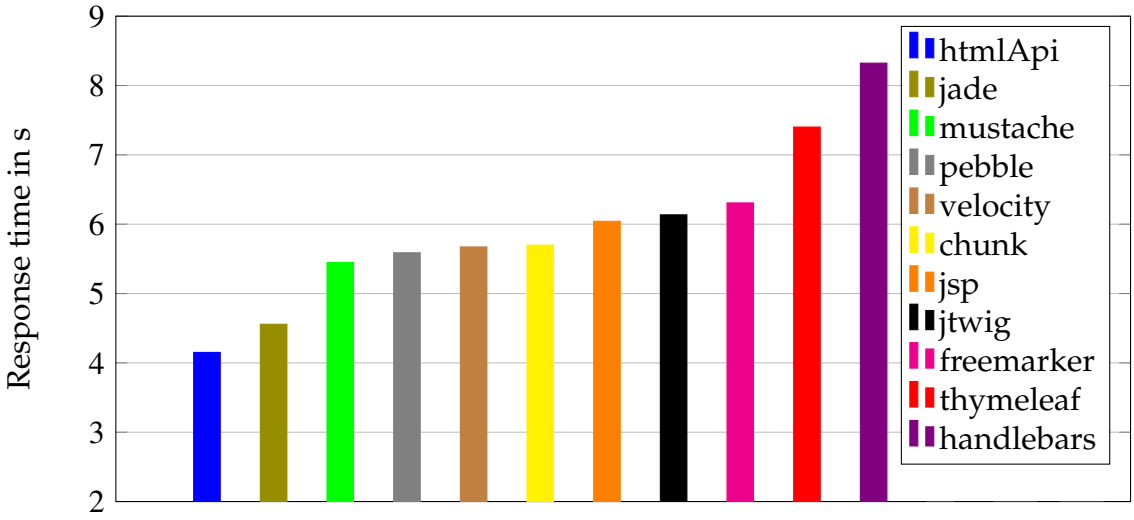


Figure 5.8: Benchmark: Spring Benchmark with 30.000 requests issued with 25 threads

100.000 requests using 10 threads with a Spring overhead of 9.890 seconds:

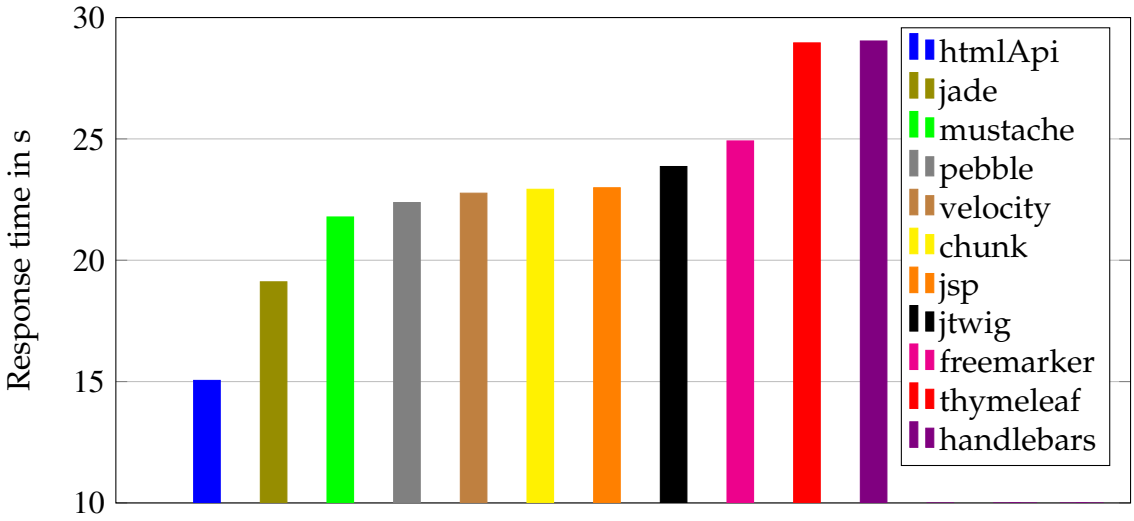


Figure 5.9: Benchmark: Spring Benchmark with 100.000 requests issued with 10 threads

As the results show, HtmlApi outperforms all the template engines in the examples presented with the gap in performance increasing as the number of elements grows. With the two examples presented with 30.000 requests we can observe that the results are approximately the same, even though the concurrent request

count increased from 10 to 25, with some template engines performing a bit better and other a bit worse, which may be due to invariable execution variations.

6

Conclusion

In this dissertation we developed a structure of projects that can interpret a XSD document and use its contents to generate a fluent Java API that allows to perform actions over the language defined in the XSD document while enforcing most of the rules that exist in the XSD syntax. The generated API only reflects the structure described in the XSD document, providing tools that allow any future usage to be defined according to the needs of the user. Upon testing the resulting solution we obtained better results than similar solutions while proving a solution with a fluent language, which should be intuitive even for people that never programmed in Java before, since it ends up being very similar to writing XML.

The main language definition used in order to test and develop this solution was the HTML5 syntax, which generated the HtmlApi project, containing a set of classes reflecting all the elements and attributes present in the HTML language. This HtmlApi project was then used by the HtmlFlow library in order to provide an API that writes well formed HTML documents. Other XSD files were used to test the solution, such as the Android layouts definition file, which defines the existing XML elements used to create visual layouts for the Android operating system and the attributes that each element contains.

6.1 Future work

In the future there are already changes planed to the way that the APIs are generated. The current APIs generated by the present solution are built on a two step basis, first the user has to create the element tree and after finishing its creation the user can then visit the whole tree. Even though this solution provides security to the user and already provides better results than similar solutions there is an improved solution that should increase its performance in a significant way. The improved solution removes the two steps nature of the solution, avoiding storing elements in a data structure and then having to iterate that structure. An example of this new solution usage is shown in Listing 6.1.

```
1 Visitor visitor = new Visitor();
2 Html<Html> documentRoot = new Html<>(visitor);
3 documentRoot.body();
```

Listing 6.1: Future Solution Usage 1

With the root receiving the Visitor instance its possible to start calling the respective visit methods right away, as shown in class Html in Listing 6.2 and Body in Listing 6.3.

```
1  public class Html{
2      private Visitor visitor;
3
4      public Html(Visitor visitor){
5          this.visitor = visitor;
6          visitor.visit(this);
7      }
8
9      public Body body(){
10         return new Body(this.self());
11     }
12
13     public Html someAttribute(String value){
14         visitor.visit(new SomeAttribute(value));
15         return this;
16     }
17 }
```

Listing 6.2: Future Solution Html Class (Simplified)

```
1 public class Body{
2     private Visitor visitor;
3
4     public Body(Element parent){
5         this.visitor = parent.getVisitor();
6         this.visitor.visit(this);
7     }
8 }
```

Listing 6.3: Future Solution Body Class (Simplified)

With this solution the overhead of using this solution is minimized, because as we can see the overhead of the code structure almost resumes itself to the overhead of instantiating classes. All the rules of the language are still enforced due to the way that the classes are organized and the usage keeps its fluent nature since the existent methods and its return type are unchanged. This also removes all the overhead of storing and iterating data from a data structure. Using this solution in the web world should improve the response times of servers in a very significant way since the solution can write to the response stream as soon as it executes the method code for adding an element, opposed to having to wait that the whole tree is built and only then iterated.

6.1.1 Early Results

After implementing a first version of this improved solution, called HtmlApiFaster, we ran a few tests in order to evaluate its performance. The benchmark used was the same presented in Section 5.4 with the spring-comparing-template-engines project. All the results presented below are presented without the respective Spring overhead, which was previously explained in Section 5.4.

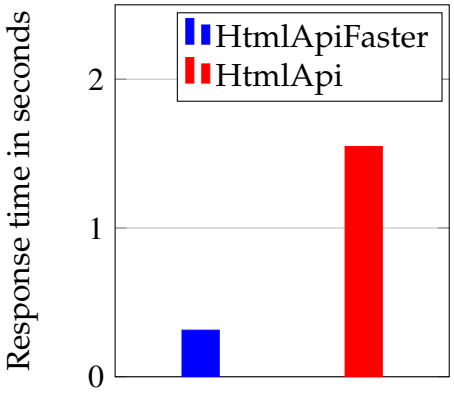


Figure 6.1: HtmlApiFaster Benchmark: Spring Benchmark with 10.000 requests issued with 10 threads

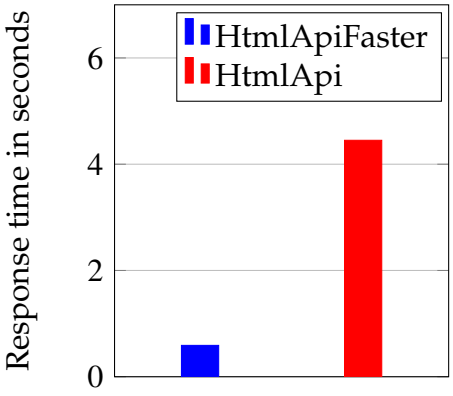


Figure 6.2: HtmlApiFaster Benchmark: Spring Benchmark with 30.000 requests issued with 10 threads

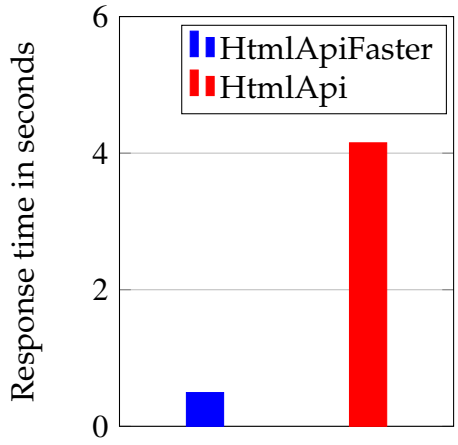


Figure 6.3: HtmlApiFaster Benchmark: Spring Benchmark with 30.000 requests issued with 25 threads

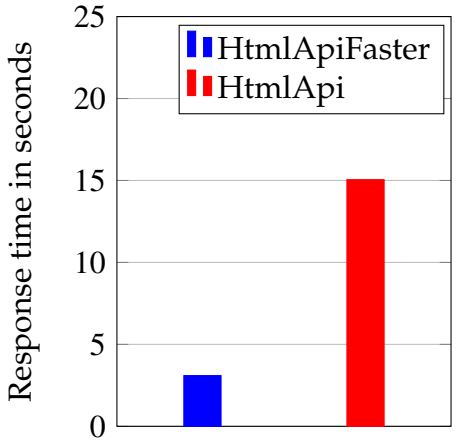


Figure 6.4: HtmlApiFaster Benchmark: Spring Benchmark with 100.000 requests issued with 10 threads

As we can see the performance gains are very good, with the lowest gap being with the example using 100.000 requests where HtmlApiFaster is 4 times faster than HtmlApi. On the other hand the biggest difference is present in the example using 30.000 requests and 25 concurrent threads, where HtmlApiFaster is 8 times faster than HtmlApi which was already the better option when comparing to the template engines. These results are promising and the HtmlApiFaster will be further developed in order to deploy it as an alternative to HtmlApi.

Bibliography

[1] Per. Christensson. Markup language definition., 2011. URL https://techterms.com/definition/markup_language. (p. 9)

[2] Martin Fowler. Fluent interfaces, 2005. URL https://martinfowler.com/bliki/FluentInterface.html. (p. 1)

[3] Martin Fowler. Domain specific languages, May 2008. URL https://martinfowler.com/bliki/DomainSpecificLanguage.html. (p. 1)

[4] Martin Fowler. *Domain Specific Languages*. Addison-Wesley Professional, 1st edition, 2010. ISBN 0321712943, 9780321712943. (p. 2)

[5] Erich Gamma, Richard Helm, Ralph Johnson, and John M. Vlissides. *Design Patterns: Elements of Reusable Object-Oriented Software*. Addison-Wesley Professional, 1 edition, 1994. ISBN 0201633612. URL http://www.amazon.com/Design-Patterns-Elements-Reusable-Object-Oriented/dp/0201633612/ref=ntt_at_ep_dpi_1. (pp. 30 and 46)

[6] Priscilla Walmsley. Xml schema, January 2004. URL http://www.datypic.com/sc/xsd/s-xmlschema.xsd.html. (p. 22)