

Modernizing the WebDSL Front-End: A Case Study in SDF3 and Statix

Version of December 5, 2022

Max M. de Krieger

Modernizing the WebDSL Front-End: A Case Study in SDF3 and Statix

THESIS

submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE

in

COMPUTER SCIENCE

by

Max M. de Krieger
born in Delft, the Netherlands



Programming Languages Group
Department of Software Technology
Faculty EEMCS, Delft University of Technology
Delft, the Netherlands
www.ewi.tudelft.nl

© 2022 Max M. de Krieger.

The source code for this thesis can be found online:

<https://github.com/webdsl/webdsl>

<https://github.com/webdsl/webdsl-statix>

Modernizing the WebDSL Front-End: A Case Study in SDF3 and Statix

Author: Max M. de Krieger
Student id: 4705483
Email: maxdekrieger@gmail.com

Abstract

The front-end of a compiler reads the source program and performs analyses such as type checking. The goal of the front-end is to check for the presence of syntactic and semantic errors before the program is passed to the back-end of the compiler for tasks such as optimization and code generation.

WebDSL is a domain-specific language for web programming that is being used for over 15 years. With WebDSL, many applications have been developed which have thousands of daily users. While the language has evolved over the years, the core of its implementation remains unchanged and is starting to show signs of a legacy system. The current WebDSL syntax is defined in SDF2 and all other parts of the compiler are implemented in Stratego.

This thesis presents a modernized front-end of the WebDSL compiler, utilizing the meta-languages of the Spoofax language workbench. Specifically, we introduce a syntax definition of WebDSL in SDF3 that is implemented without the use of post-parse filters. Additionally, the modernized front-end features an executable declarative specification of the WebDSL static semantics in Statix.

We use the modernized front-end as the largest case study to date for the meta-languages SDF3 and Statix, in order to evaluate their expressiveness, performance, and elegance when they are used to implement a real world language.

Thesis Committee:

Chair:	Prof. dr. A. Van Deursen, Faculty EEMCS, TU Delft
Committee Member:	Dr. J. G. H. Cockx, Faculty EEMCS, TU Delft
University Supervisor:	Ir. D. M. Groenewegen, Faculty EEMCS, TU Delft

Thesis Advisor:	Prof. dr. E. Visser, Faculty EEMCS, TU Delft
-----------------	--

Preface

Almost three years. That is how long it took me to finish this thesis. During this period:

- Startup companies were joined and left;
- A full-time employment was started;
- Many hobby projects were started;
- Not a single hobby project was finished.

Although anecdotal, the list above reflects my tendency to distract myself with somewhat useful things when I doubt my ability to finish the main objective. Regardless of the time taken, I am happy with the result in the form of this thesis, which would not have been the same without the help and support of a lot of people. I am sure I will forget to mention some people, but I will attempt to list most of them.

First and foremost, I want to thank my supervisor Danny Groenewegen for his patience, for sharing his experience and for his practical help. In the same vein, I want to thank Jasper Cockx and Arie van Deursen who were able to give me the final push needed to finish this thesis. Next, I want to thank my colleagues Elmer and Daniël, as well as all other members of the Programming Languages group for sharing their experience and motivating me to keep going. Special thanks go to my parents, brother, and girlfriend for their unconditional support and trust.

Finally, this thesis would not exist without Eelco Visser. Although he is no longer with us, his ideas, contagious positivity, and continued trust in me are the foundations on which this thesis is built, and I am grateful to have experienced his mentorship.

Max M. de Krieger
Delft, the Netherlands
December 5, 2022

Contents

Preface	iii
Contents	v
List of Figures	vii
List of Tables	xi
1 Introduction	1
1.1 Case Study Description	2
1.2 Contributions	3
1.3 Outline	3
2 WebDSL	5
2.1 User Interfaces	5
2.2 Data Model	11
2.3 Access Control	13
2.4 Functions	14
2.5 Advanced WebDSL Features	14
2.6 Current Implementation of WebDSL	15
2.7 Modernization Goal	17
3 WebDSL in SDF3	19
3.1 Introduction to LR Parsing	19
3.2 WebDSL Grammar Specification	24
3.3 Migration from SDF2 to SDF3	25
3.4 Preparation for Statix	28
3.5 Disambiguation	31
4 WebDSL in Statix	37
4.1 Introduction to Statix	37
4.2 Encoding the WebDSL Basics	41
4.3 Advanced Entity Features	56
4.4 Function and Template Overloading	61
4.5 Placeholders, Actions and Submitting Forms	63
4.6 Type Extension	63
4.7 Module system	64
4.8 Pre-analyzed built-in library	71
4.9 String Manipulation in Statix	73

5	Evaluation	75
5.1	Evaluating the WebDSL SDF3 Specification	75
5.2	Statix	78
6	Related work	83
6.1	WebDSL	83
6.2	Statix	83
6.3	Alternative Approaches	84
7	Conclusion	87
7.1	Research Questions	87
7.2	Future work	89
	Bibliography	91
	Acronyms	95

List of Figures

2.1	An example of a simple WebDSL application with two pages, navigation and reusable templates.	6
2.2	An example of a WebDSL application that uses a form for navigation.	8
2.3	An example of template overloading in WebDSL with inputs and descriptions.	9
2.4	An example of placeholders and ajax actions in WebDSL.	10
2.5	An example of template overriding in WebDSL.	11
2.6	An example of template overriding in WebDSL.	12
2.7	Access control in WebDSL	13
2.8	Functions in WebDSL	14
2.9	Definition of the WebDSL type checking framework in Stratego. Some rules are omitted for brevity.	16
2.10	Definition of Stratego rule that shows an error message when two principals are declared in the access control rules of WebDSL.	17
3.1	Example context-free grammar with its parse table	20
3.2	An example of a parse tree for the input " $x * x$ "	20
3.3	An SDF3 specification of a language that allows addition and multiplication of integers.	21
3.4	The AST and ATerm representation of the expression $0 + 1$ according to the SDF3 specification of Figure 3.3	22
3.5	An SDF3 specification of a language that allows addition and multiplication of integers, showcasing injections, repetition and optional sorts.	23
3.6	The ATerm of input $0 + 0; 1 * -1$ according to the SDF3 specification of Figure 3.5	23
3.7	An SDF3 specification of a simple language with disambiguation rules.	24
3.8	An example of productions in SDF2. This example specifies the syntax of variable declaration in WebDSL.	24
3.9	WebDSL Syntax and desugaring for sections and definitions	25
3.10	Follow restrictions are used to for example ensure the longest-match of an identifier and enforcing whitespace after keywords.	26
3.11	SDF2 and generated SDF3 for the ascending and descending keyword in WebDSL. Duplicate constructors within the same sort are not allowed in SDF3.	27
3.12	Difference between priority chains in SDF2 and SDF3. The transformation tool does not perform the analysis to properly transform this.	27
3.13	An example of an SDF2 production, the generated SDF3 and the manually adjusted SDF3 for template productions.	28
3.14	An example of an expression language signature in Statix.	29
3.15	An example of how SDF3 injections are handled by the Statix signature generator.	30

3.16	An example of how to remove the optional sort to comply with the Statix signature generator.	30
3.17	An example of an ambiguous construct in WebDSL with an HQL query.	31
3.18	An example of rejecting keywords in template options.	32
3.19	SDF2 definitions of a WebDSL string.	33
3.20	SDF3 definitions of a WebDSL string.	33
3.21	Ambiguous WebDSL code using curly brackets for untyped set creation and passing template elements as argument.	33
3.22	Using non-transitive indexed priorities to disambiguate WebDSL syntax in SDF3.	34
3.23	Ambiguous WebDSL code using an optional typecase alias or a cast expression.	34
3.24	Using non-transitive indexed priorities to disambiguate WebDSL casts versus type aliases.	35
4.1	Language signature in Statix	37
4.2	Valid input terms for the described language	38
4.3	Statix signature for Boolean and integer types	38
4.4	Statix predicates and rules for typing booleans, integers and addition	39
4.5	Scope graph examples	39
4.6	Statix signature for let-bindings	40
4.7	A scope graph containing a single scope with two declared variables	40
4.8	Statix rules for let-bindings	41
4.9	Constructed scope graph after the example specification solved its constraints	41
4.10	Predicates that form the basis of the WebDSL Statix specification	42
4.11	Declaring built-in types in the project scope	42
4.12	WebDSL integer constant expression typing rules	42
4.14	WebDSL string typing rules	43
4.13	WebDSL string interpolation examples	43
4.15	WebDSL variable declaration and resolving	44
4.16	WebDSL requires declare-before-use of variables	44
4.19	Well-formedness predicate for variable paths	44
4.17	WebDSL statements use different scopes for querying and declaring data from the scope graph	45
4.18	Variable declarations example using a separate declaration scopes	45
4.20	The same variable identifier may only be declared once in an environment	46
4.21	Examples of type compatibility in WebDSL	46
4.22	WebDSL type compatibility predicate and general rules	46
4.23	Compatibility of the null expression encoded in Statix	47
4.24	Least-upper-bound rules for addition	47
4.25	Boolean computation results in Statix	48
4.26	Using boolean computation results in Statix for the equality expression	49
4.27	The Statix rules for declaring entities	49
4.28	An example of entity definition in WebDSL	50
4.29	declareType now shows an error when two types with the same name are declared and resolveType may optionally follow a DEF edge label	50
4.30	Statix rules for declaring the entity body	51
4.31	An example of an entity definition with multiple properties	51
4.32	Statix rules for instantiating an entity	52
4.33	Statix signature for pages and templates	52
4.34	Statix rules for declaring WebDSL pages and templates	53
4.35	An example of a module with a page <code>p</code> and a template <code>t</code>	53
4.36	Statix rules for type-checking a page reference	53
4.37	Statix rules for function parameters and variable shadowing	55

4.38	An example of a function with parameters	56
4.39	An example of access control in WebDSL. The entries related to entity <code>User</code> are omitted for brevity	56
4.40	Statix rules for declaring the access control principal	57
4.41	Entity inheritance Statix rules	57
4.42	An example of entity definition in WebDSL	58
4.43	The query that specifies what variables can be resolved, updated to reflect entity inheritance	58
4.44	Statix rules for allowing entity function calls to resolve to definitions in their ancestors	59
4.45	Statix rules for resolving entity functions that allow overriding	59
4.46	Statix rules for entity type compatibility that support inheritance	59
4.47	Examples of entity property annotations	60
4.48	An example of a derived property in an entity	60
4.49	An example of overriding the <code>name</code> property	61
4.50	Statix rules for overridable entity properties	61
4.51	An example of defining overloaded templates in WebDSL	61
4.52	Statix rules for allowing overloaded function definitions	62
4.53	An example of referencing overloaded templates in WebDSL	62
4.54	Statix rules for resolving overloaded function calls	63
4.55	An example of defining and referencing actions	64
4.56	Statix rules for declaring actions	65
4.57	An example of extending entity scopes	66
4.58	Statix rules for extending entities	67
4.59	An example of the current WebDSL module system	68
4.60	Statix rules for modelling the current module system	69
4.61	Run time of WebDSL Statix definition with the current module system.	70
4.62	Run time of WebDSL Statix definition with revised module systems.	71
4.64	Run time of WebDSL Statix definition with and without pre-analyzed standard library <code>built-in.app</code>	72
4.65	Statix rules for built-in type extension in Statix	73
4.66	An example of generated definitions and static code template expansion in WebDSL	74
5.3	Run time of WebDSL SDF2 definition vs. SDF3 definition	77
5.4	Defining a WebDSL for-loop in SDF2 and SDF3	78
5.8	Run time of the WebDSL static analysis in Stratego vs. Statix	82

List of Tables

4.63	Run time of WebDSL Statix definition with revised module systems on Reposearch and YellowGrass.	70
5.1	Results of parsing the syntax test suite with the WebDSL SDF3 specification. . . .	76
5.2	Data about the size of the parse tables generated from the WebDSL SDF2 and SDF3 grammar specifications.	76
5.5	Results of running the static analysis on Reposearch and Yellowgrass.	79
5.6	Results of the analysis test suite.	80
5.7	Results of the analysis test suite without taking exact error message expectation into account.	80

Chapter 1

Introduction

Computer programming is an essential skill that is increasingly important in diverse disciplines (Rafalski et al. 2019). To this end, many programming languages exist, each with different properties and advantages. Over time, the popularity of programming languages changes and developers tend to have preferences for one language over the other. In addition to preference, the implementation of a language and the tools that come with it can greatly boost the productivity of developers, if done well.

Another key to boost the productivity of software engineers is abstractions. Abstractions allow developers to think in terms closer to the domain rather than the implementation. In other words, the ideal level of abstraction increases the focus on the what, and steers away from the how. In this thesis, we will conduct a case study using *domain-specific languages* (DSLs). As opposed to a general-purpose language such as Java, C, or Python, a domain-specific language does not intend to provide solutions for problems from all domains, but instead focus on a single domain. This restriction allows for a high level of abstraction in the language itself, in an attempt to boost developer productivity. Examples of popular domain-specific languages are CSS for styling web pages and SQL for efficient database querying.

In this thesis, we use the domain-specific language *WebDSL* as a large case study for the domain-specific languages *SDF3* and *Statix*. *WebDSL* is a domain-specific language for developing web applications, developed and maintained by the Programming Languages research group of the Delft University of Technology.

When inspecting the implementation of a programming language, it can be split up in multiple parts such as parsing, static analysis, code generation and optimization. The parsing, desugaring and static analysis is often called the front-end of a programming language, and this is the part developers face directly. The code generation and code optimization is called the back-end, and is required to make the programming language operational. While the back-end of a programming language makes it work, the front-end plays a large role in how developers experience a programming language. Early feedback in the form of good error messages and hints are required to make the interaction with a programming language efficient (Becker et al. 2019).

Because of the language-based approach of *WebDSL* for encoding domain concepts, many features that would be a library or an external tool in a general purpose language, are linguistically integrated into *WebDSL*. Examples of such features are fuzzy search and defining the data model. The linguistic integration of these features allows for better consistency checking and more precise error descriptions.

Currently, the *WebDSL* implementation is composed of multiple definitions in meta-languages supported by the *Spoofax Language Workbench* (Kats and Visser 2010). *Spoofax* is an environment in which multiple meta-DSLs are used to declaratively specify a programming language. *WebDSL* is developed in *Spoofax*. In particular, the *WebDSL* syntax is defined in *SDF2* and the desugaring, type checking, optimization and code generation is defined in

the term transformation language Stratego. In the current Stratego implementation of the WebDSL language, the compilation steps are not clearly separated, which poses a threat to the readability and maintainability of the WebDSL language.

Continuous improvement of the Spoofax language workbench has introduced more meta-languages specialized in different parts of the language development chain. In this thesis, we will be modernizing the WebDSL front-end, by using the Spoofax meta-languages SDF3 to specify and disambiguate the WebDSL syntax from which a parser is generated, and Statix to declare the WebDSL static semantics from which a type checker is automatically generated.

1.1 Case Study Description

In this section we describe the design of the case study conducted in this thesis, according to the framework described by Runeson and Höst (2009).

Objective The objective of the case study is of exploratory nature, namely to gain insight into the meta-languages SDF3 and Statix. Specifically, we aim to gain insight into their expressiveness, performance and maintainability. With these insights, we hope to generate suggestions for improvements of the meta-languages.

The Case WebDSL is used as the case in this case study. Through implementing the WebDSL syntax in SDF3 and its static semantics in Statix we aim to gain insights as described in the objective.

Background Both WebDSL and the meta-languages of the Spoofax language workbench are created and maintained by the Programming Languages research group of the TU Delft. WebDSL is used in practice to make web applications with thousands of daily users, and its current front-end implementation is one of the largest projects in SDF2 and Stratego.

Research Questions In this thesis, we aim to answer the following research questions.

- **RQ1** Is it possible to define the WebDSL syntax in SDF3, without the use of post-parse filters?
- **RQ2** How does the run time efficiency of the parser generated from WebDSL in SDF3 compare to the current WebDSL parser generated from SDF2?
- **RQ3** How does the maintainability of WebDSL in SDF3 compare to the WebDSL syntax definition in SDF2?
- **RQ4** Is a Statix implementation of the WebDSL static semantics able to catch the same errors and warnings as the current implementation in Stratego?
- **RQ5** How does the run time efficiency of the WebDSL static semantics in Statix compare to the current WebDSL implementation?
- **RQ6** How does the maintainability of the WebDSL static semantics in Statix compare to the current implementation?

Methods To answer the research questions, we implement a modernized WebDSL front-end using SDF3 and Statix. We evaluate the generated parser and type-checker on multiple targets. The current WebDSL compiler contains large test suites that contain hundreds of test programs for the parser and static analysis. Most of these test programs stem from real life production bugs of the compiler. We use the test suites to evaluate the completeness and run time performance of the parser generated from SDF3, and the static analysis generated from Statix.

Why WebDSL? WebDSL is an interesting case study for SDF3 and Statix because of two main reasons. Firstly, WebDSL has a large amount of language features inspired by multiple paradigms of programming languages. As a consequence, the SDF3 and Statix specifications we present in this thesis are arguably the largest specifications to date. Accompanied by the large amount of publicly available source code for evaluation purposes, we aim to make observations about the elegance of the resulting specifications and the scalability of their performance. Since this thesis is a case study, we cannot use the result to make general claims about the performance of SDF3 and Statix, but only reveal and analyze results of the WebDSL specification. Secondly, WebDSL contains language features that have never been modelled in Statix before such as extension of built-in types, generated functions and classes and an unconventional module and scoping system.

1.2 Contributions

In this thesis, the following contributions are made.

- We present a modernized WebDSL front-end through an implementation of its grammar in SDF3 and its static analysis in Statix and document the challenges of this process.
- We assess the coverage of SDF3 and Statix by attempting to model all language features of WebDSL, evaluating the result on existing test suites.
- We give qualitative feedback on the meta-languages SDF3 and Statix and give suggestions for how to further improve their expressiveness and elegance.
- We assess the run time performance of SDF3 and Statix by benchmarking the new WebDSL front-end using test suites and large codebases of existing applications.

1.3 Outline

The rest of this thesis is structured as follows. In Chapter 2 we describe WebDSL, its features and its current implementation. Next, Chapter 3 and Section 4.1 go in detail about the new implementation of the WebDSL front-end in SDF3 and Statix respectively. The result of this implementation is evaluated in Chapter 5 and compared with related work in Chapter 6. Finally, Chapter 7 concludes this thesis.

Chapter 2

WebDSL

In this chapter we describe WebDSL. WebDSL is a domain-specific language for developing web applications. The language incorporates ideas from various web programming frameworks and produces code for all tiers in a web application (Groenewegen, E. v. Chastelet, and Visser 2020). Ever since its introduction over 15 years ago (Visser 2007), WebDSL has been a vehicle for research (cite some papers here) and on top of that, it is the programming language underpinning several applications used daily by thousands of users. Examples of WebDSL applications include:

- **WebLab**: An online learning management system, used by the Delft University of Technology.
- **conf.researchr.org**: A domain-specific content management system for conferences, used by most ACM SIGPLAN and SIGSOFT conferences.
- **researchr.org**: A platform for finding, collecting, sharing, and reviewing scientific computer science related publications.

In WebDSL, all aspects of web programming are defined in the same language which creates an opportunity for strong static consistency checking (Hemel, Groenewegen, et al. 2011). The data model is defined using entities, which are automatically translated to a database representation. These entities can be passed as arguments of pages without the need to explicitly fetch data from the database. Large amounts of boilerplate code are abstracted over by WebDSL, by providing built-in but customizable functionality for access control, relations between entities and data validation. This combination of features allows WebDSL developers to efficiently create functioning prototypes while providing the functionality to make it scale and customizable.

The rest of this chapter showcases the different aspects of WebDSL and zooms in on its non-trivial features. First, in Section 2.1 we will describe how WebDSL offers functionality for creating web user interfaces. Next, in Section 2.2 we illustrate how the language manages data models. Thirdly, Section 2.3 contains information about WebDSL's solution for access control and in Section 2.4 we highlight interesting aspects of its general-purpose object-oriented function code. We conclude this chapter by elaborating on the current implementation of the WebDSL static analysis in Section 2.6.

2.1 User Interfaces

The user interface of an application is the part of the system that the user interacts with. Through the user interface, users are able to navigate the system and manipulate its data.

```

application app
  page root {
    title { "Cookbook" }

    header { "Recipes" }

    linkToRecipe("Lasagne")
    linkToRecipe("Pancakes")
    linkToRecipe("Tomato Soup")
  }

  page recipe(s : String) {
    title { "Cookbook: ~s" }

    navigate root() {
      "Back to homepage"
    }

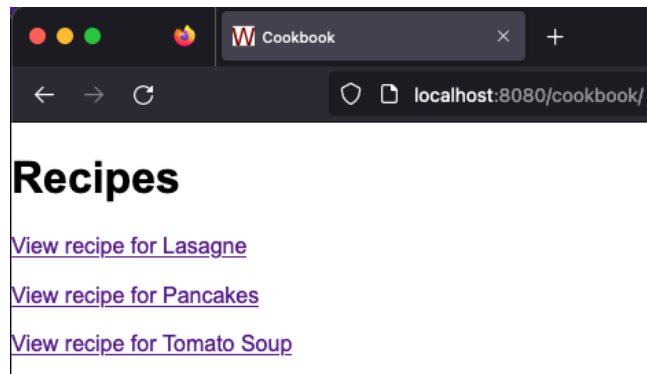
    header { "~s" }

    "Welcome to the recipe page for ~s!"
  }

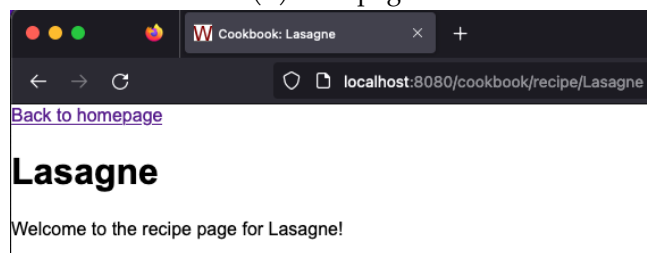
  template linkToRecipe(s : String) {
    par {
      navigate recipe(s) {
        "View recipe for ~s"
      }
    }
  }

```

(a) WebDSL code



(b) Root page



(c) Recipe page

Figure 2.1: An example of a simple WebDSL application with two pages, navigation and reusable templates.

2.1.1 Building Blocks: Pages, Templates and Template Elements

In a WebDSL application, the user interface is exposed to the user via pages that are accessible from the browser. In WebDSL, pages are defined by their unique name and can optionally take arguments that directly correspond with the path of the webpage.

Templates are reusable definitions of user interface code that can be included on pages and can be nested in other templates. Figure 2.1 shows an example of an application with two pages, the `root` page and the `recipe` page, with navigation between them and a reusable template called `linkToRecipe` that creates a link to the recipe page with an argument of type string. The recipe page also takes one argument of type string which is visible in the URL path.

The `navigate` calls are references to other pages in the application. Because all pages are known at compile time, WebDSL ensures that all navigation within an application refers to existing pages, and that the passed types are correct. Referring to a non-existing page or passing the wrong argument types would show an error in the WebDSL development environment and while compiling the application.

Other than pages and templates and navigation between pages, the application uses various calls to built-in template such as `title`, `header` and `par`. These built-in templates are translated to their respective HTML elements `<title></title>`, `<h1></h1>` and `<p></p>`. WebDSL also supports directly using HTML elements with their usual syntax such as `<div>"text"</div>`, but using template-style code such as `div { "text" }` is the convention in WebDSL.

Outputting text on a webpage in WebDSL can be done by typing the text between quotes in a template or page, as is shown in the example cookbook application of Figure 2.1.

Using the concepts described in this subsection, developers are able to create static websites with WebDSL. However, a powerful aspect of WebDSL is its ability to create dynamic pages without requiring much boilerplate code.

2.1.2 Request Processing and Action Code

HTML has a notion of forms and inputs that can be submitted to a web server, after which the web server returns a result. WebDSL code compiles to HTML forms and takes away the responsibility of manually linking the input data to variables. Figure 2.2 shows an example of the cookbook application where the user is able to enter a name in an input field and browse to that recipe.

Figure 2.2 no longer contains a fixed list with links to all possible recipes as in Figure 2.1 but now uses a form with an input and submits an action when the user presses a button. The root page contains a variable declaration as the first line, which defaults to an empty string. In the form later on the page, this variable is passed to an input element. After the input, the form has a button that send a request to the web server to execute the action `toRecipe` when it is pressed. When this request is sent to the server, the data entered in the input field is linked to variable `s` by WebDSL, and the action code is able to use this in its body. The return value of an action is the page that should be shown as response. In the example of Figure 2.2, the result of executing the action is to show the `recipe` page with the input that the user has entered in the form.

WebDSL also sanitizes the input that the user submits using the form. If the user was to submit the form with input `another/path/`, this input would be sanitized and rendered correctly on the recipe page, instead of navigation to an unknown path.

Note that WebDSL usually has *declare-before-use* semantics inside in page, templates or function bodies, but actions can be declared and resolved anywhere in a page or template definition.

2.1.3 Template Overloading

Templates in WebDSL must have a unique combination of name and argument types. This means that, similar to functions in Java, WebDSL supports overloading the same template with different arguments. The built-in `input` template that is used in previous subsection is an example of this. In Figure 2.2, a string argument is passed to the `input` template which rendered a text field on the resulting web page.

Figure 2.3 shows an example of template overloading. The `describe` template is defined multiple times for different argument types. The `input` template is provided by the WebDSL standard library and also defines different input elements for different types.

In WebDSL, multiple text-based types such as `String`, `Text`, `WikiText` and `Secret` are all compatible with the base type `String`. When the arguments are passed to a template, it resolves the most specific template. An example of this is that `describe(t)` in Figure 2.3 is compatible with `String` and `Text`, but it resolves to the definition with a `Text` argument.

When a template call has multiple arguments, it resolves to the template where all arguments are the most specific. If there is one template definition where the first argument

```

application app
  page root {
    var s : String

    title { "Cookbook" }

    header { "Recipes" }

    form {
      par {
        "Enter a recipe name:"
        input(s)
      }
      submit toRecipe() { "Go!" }
    }

    action toRecipe() {
      return recipe(s);
    }
  }

  page recipe(s : String) {
    title { "Cookbook: ~s" }

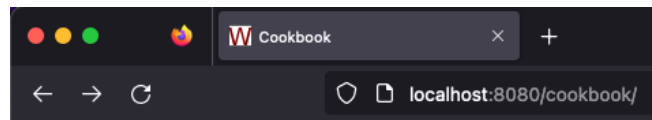
    navigate root() { "Back to homepage" }

    header { "~s" }

    "Welcome to the recipe page for ~s!"
  }

```

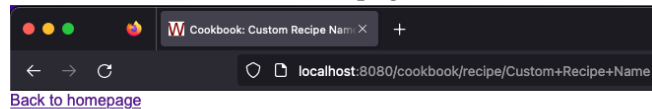
(a) WebDSL code



Recipes

Enter a recipe name:

(b) Root page



Custom Recipe Name

Welcome to the recipe page for Custom Recipe Name!

(c) Recipe page

Figure 2.2: An example of a WebDSL application that uses a form for navigation.

is more specific, and another definition where the second argument is more specific, this is incorrect code and the WebDSL development environment shows an error to the developer.

2.1.4 Ajax

To enable refreshing or replacing parts of a page with new content that is fetched from the server, WebDSL has support for placeholders that can be populated with content using *asynchronous JavaScript and XML* (ajax). Without ajax functionalities, refreshing data on a page would require a full page reload.

Figure 2.4 shows an example of a WebDSL application that defines a placeholder named `ph` which is initially empty. On the bottom of the page, a clickable link with an inline action is defined. When the action is executed, it will replace the content of the placeholder with the content of the ajax-enabled template `hello()`. The calling of the action does not incur a full page refresh, but instead fires an asynchronous request that returns the content of the ajax template, after which the placeholder content is replaced using JavaScript.

WebDSL automatically takes care of the necessary structure to enable this functionality, such as identifiers of DOM elements, JavaScript code to query the server and JavaScript code


```

application app
  page root {
    var s : String
    var b : Bool
    var t : Text
    var d : Date

    title { "Input Examples" }

    header { "Input Examples" }

    form {
      par {
        describe(s)
        input(s)
      }
      par {
        describe(b)
        input(b)
      }
      par {
        describe(t)
        input(t)
      }
      par {
        describe(d)
        input(d)
      }
    }
  }
}

```

```

template describe(s : String) { "String input:" }
template describe(b : Bool)   { "Bool input:"   }
template describe(t : Text)  { "Text input:"  }
template describe(d : Date)  { "Date input:"  }

```

(a) WebDSL code

Input Examples

String input: Bool input: ☒

Text input

Text input:

Date input:

December 2022						
Sun	Mon	Tue	Wed	Thu	Fri	Sat
27	28	29	30	1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	31
1	2	3	4	5	6	7

(b) Resulting page

Figure 2.3: An example of template overloading in WebDSL with inputs and descriptions.

```

application app
  page root {
    title { "Ajax Example" }

    header { "Ajax Example" }

    par {
      placeholder ph {}
    }

    submitlink action {
      replace(ph, hello());
    } { "Replace" }
  }

  ajax template hello {
    "Hello!"
  }

```

(a) WebDSL code

Ajax Example

Hello!

[Replace](#)

(b) Resulting page

Ajax Example

[Replace](#)

(c) After clicking the link

Figure 2.4: An example of placeholders and ajax actions in WebDSL.

to replace the correct elements.

As opposed to regular templates, ajax-enabled template can be requested on the server using a specific endpoint. The exact details of how this endpoint looks is only relevant for generated code of the WebDSL compiler, but WebDSL developers should pay attention to the fact that ajax-enabled content must be projected with proper access control rules (see Section 2.3) to prevent exposing sensitive content.

2.1.5 Dynamically Scoped Redefines

As last part of the user interface section, we give an example of dynamically scoped redefines in WebDSL. WebDSL templates can be redefined (i.e. overridden) inside another template. The redefined content will be shown when the redefined template is used within the template that redefines it. This also holds for calls to the redefined template, nested arbitrarily deep.

Figure 2.5 shows an example of a dynamically scoped redefinition of a template. In the figure, `templateB` redefines `templateA`, resulting in the overridden template being rendered when it is called within `templateB`. To prove that this is scoped, `templateA` is called before and after `templateB` and it renders the same (original) text. In the example of Figure 2.5, the redefined template is called directly from the template where it is redefined, but it has the same effect when it would be called in nested templates.

Apart from this redefinition of templates inside the context of other templates, WebDSL also has a notion of a one-time global redefinition. This is done using the keyword `override` on a template in the global scope. The WebDSL standard library contains many definitions that are able to be used by developers, such as an access-denied page, a login template or the input templates that are used in Figure 2.3. Using the `override` keyword, developers are able to customize the style and behavior of these templates.

<pre> application app page root { title { "Override Template" } header { "Override Template" } templateA templateB templateA } template templateA { par { "I'm Template A!" } } </pre>	<h1>Override Template</h1> <p>I'm Template A!</p> <p>I'm Template B!</p> <p>I'm Template A inside template B!</p> <p>I'm Template A!</p>
<pre> template templateB { // override template A template templateA { par { "I'm Template A inside template B!" } } par { "I'm Template B!" } templateA } </pre>	<p>(b) Resulting page</p>

(a) WebDSL code

Figure 2.5: An example of template overriding in WebDSL.

2.2 Data Model

The data model of a WebDSL application is defined using entities. Entities can have properties and relations with other entities. WebDSL translates the entities to database tables and columns such that any data stored in entities is persisted. The communication with the database is abstracted by WebDSL, allowing WebDSL developers to easily access, manipulate and pass around entity values.

Figure 2.6 shows an example of an application with stored recipes. A `Recipe` has two properties: a name and a duration. The application is similar to the earlier representation from Figure 2.1 where a recipe was simply defined by its name, encoded as a `String`. The representation using entities now defines the `recipe` page which takes a `Recipe` as argument, demonstrating how entities are represented in the WebDSL type system.

The `init{}` block in the global scope of Figure 2.6 is executed only once, and often used to seed the database.

Furthermore, the root page of Figure 2.6 uses a for-loop to iterate over all existing instances of an entity; Recipes in our case. It is uncommon to loop over *all* entities, thus WebDSL supports a filter in the for-loop similar to SQL syntax.

```
application cookbook
```

```
section data model
```

```
entity Recipe {
  name          : String
  minutesRequired : Int
}
```

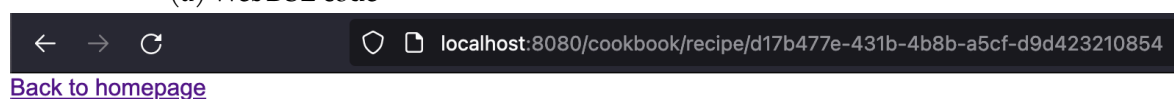
```
init {
  var r1 := Recipe {
    name := "Lasagne",
    minutesRequired := 60
  };
  var r2 := Recipe {
    name := "Pancakes",
    minutesRequired := 20
  };
  var r3 := Recipe {
    name := "Tomato Soup",
    minutesRequired := 40
  };
}
```

```
section user interface
```

```
page root {
  title { "Cookbook" }
  header { "Recipes" }
  for (r : Recipe) {
    par {
      navigate recipe(r) { "View recipe for ~r.name" }
    }
  }
}
```

```
page recipe(r : Recipe) {
  title { "Cookbook: ~r.name" }
  navigate root() { "Back to homepage" }
  header { "~r.name" }
  par {
    "Welcome to the recipe page for ~r.name!"
  }
  par {
    "Minutes required: ~r.minutesRequired"
  }
}
```

(a) WebDSL code



Lasagne

Welcome to the recipe page for Lasagne!
 12
 Minutes required: 60

(c) Recipe page



Recipes

[View recipe for Lasagne](#)

[View recipe for Pancakes](#)

[View recipe for Tomato Soup](#)

(b) Root page

2.2.1 Entity Inheritance

WebDSL allows entity inheritance, meaning that an entity can be a sub-entity of another entity. Similar to subclasses in Java, sub-entities inherit all properties from their parent and are able to define properties of their own.

2.2.2 Entity Extension

In the process of developing applications with WebDSL, entities tend to grow in size. For new features in an application, additional properties and functions are added to existing entities. To prevent having a single large entity declaration, WebDSL supports extending existing entities in multiple declarations, similar to partial classes in C#. Entity extensions are able to use all properties from other extensions and are not self-contained.

During compilation all entity extensions are merged, resulting in a large entity definition that is represented in the database in a single table.

2.3 Access Control

Almost all web applications use a form of access control. Examples of access control are using a log-in system, allowing users to only edit content of which they are the author, or the assigning of roles to groups of users in a large organization.

WebDSL recognizes the need for secure access control and provides a functionality where an entity is appointed as the user entity with some credentials. A policy can be assigned to all pages, templates and actions in WebDSL in a declarative way and separated from the implementation. Using the built-in access control functionality, WebDSL developers do not have to deal with communication and security details such as cookies and hashing.

Figure 2.7 shows an example of a WebDSL application with access control rules. The entity `User` is declared as the entity for access control using its properties `username` and `password`. Declaring the principal causes some properties to be exposed for handling access control such as `loggedIn` of boolean type and `principal` of type `User`. Multiple rules are declared for the root page, the templates and the action within a template.

In addition to these individual rules, WebDSL allows reuse of access control rules through predicates that can be called from individual rules or pointcuts to which multiple pages and templates are assigned. Furthermore, the matching of access control

```

application accesscontrol

section data model
  entity User {
    username : String
    password : Secret
    role : String
  }

section user interface
  page root {
    welcome
    secureContent
  }

  template welcome {
    if (loggedIn) { "Hello ~principal.username" }
    else           { "Hello unknown user" }
  }

  template secureContent {
    "This is only visible while logged in "
    "due to access control"

    submit a() { "Click!" }
    action a() {
      adminOnlyFunction();
    }
  }

access control rules
  principal is User
    with credentials username, password

  // pages are hidden by default
  rule page root() { true }

```

rules may contain wildcards in the name or arguments to prevent specifying the same rule multiple times for an overloaded template.

WebDSL infers the visibility of pages and elements on a page from the access control rules. For example, in case a page features a link to another page and the user does not have access to the other page according to the access control rules, the link will not be shown. If a user explicitly browses to the other page, it will show an access denied page that is provided by WebDSL by default but is customizable.

2.4 Functions

Most WebDSL constructs in previous sections are specified declaratively. However, WebDSL also supports imperative code using functions. Similar to Java functions, WebDSL functions can be scoped globally, or belong to a certain entity. These functions are able to take arguments and return a value of a certain type. Additionally, WebDSL generates many functions based on the declared entities and properties, some of which serve as hook and can be extended.

Figure 2.8 showcases some function behavior of WebDSL. It demonstrates extending a generated setter-function of the entity `Recipe`. The `setAuthor` function, including its extension, will be called when the author of a recipe is changed. Next, the entity `User` has an entity function `appendUsername`, of which the behavior is overridden in the `Admin` sub-entity. Lastly, a global function is defined, and a template demonstrates how the template body and function code can interact.

2.5 Advanced WebDSL Features

In the previous sections we listed the fundamental concepts of WebDSL that are necessary to create everyday systems. However, WebDSL enjoys more features which give the developer even more customization options.

```
application cookbook
entity Recipe {
  name : String
  minutesRequired : Int
  author : User

  // extend generated setter
  extend function setAuthor(u : User) {
    log("Set the author to ~u.username");
  }
}

entity User {
  username : String
  password : Secret

  function appendUsername(s : String) {
    this.username := this.username + s;
  }
}

// sub-entity Admin of with parent User
entity Admin : User {
  // override behaviour
  function appendUsername(s : String) {
    this.username := this.username + "admin" + s;
  }
}

function globalFunction() : Int {
  var x := 40;
  return x + 2;
}

template outputFortyTwo {
  var i := globalFunction();
}
```

Native Java Code WebDSL compiles to Java and abstracts over necessary boilerplate code to make a functioning web application with Java. However, these abstractions prevent some concepts from being expressible in WebDSL code. For this reason, it is possible to create Java files in a WebDSL project and provide an interface in the WebDSL code such that the definitions in the Java code are able to be typed and usable in WebDSL code. Such a definition of an interface in WebDSL is called a Native Java Class.

Built-in Type Extension In Section 2.2.2 we described the possibility of extending entities in WebDSL across multiple files. This same feature is available for extending built-in types such as `String`, `Bool` and the root of the type hierarchy `object`. The built-in types translate to their respective Java types, and with type extension it is possible to expose additional functions of the Java type to the WebDSL code, that are not exposed by default.

Search Van Chastelet (2013) extended WebDSL with a search-feature based on Lucene¹. For applications with a large quantity of data such as `researchr.org`² it is useful to make the data discoverable through search. As the amount data increases, implementing a search feature through querying the database becomes infeasible. The WebDSL search features allows entity properties to be annotated with a `searchable` annotation, declaring that they should be indexed. Through these annotations and advanced search mappings, it is possible to for example use fuzzy search on enormous amounts of data and search on multiple entity properties at the same time.

Routing By default, WebDSL translates pages to URLs as follows.

`https : // < domainname > / < pagename > / < arg_1 > / < arg_2 > / ... / < arg_n >`

For most cases this suffices, but it is possible to declare a `routing` block in WebDSL where the developer is able to intercept the requested URL, extract the page name and arguments, apply logic based on those and return the desired page in WebDSL.

2.6 Current Implementation of WebDSL

In the paper where Visser first introduced WebDSL (2007), it is presented as a large case study in domain-specific language engineering. Visser describes the implementation of the language with its syntax definition and code generation by term transformation using the Stratego/XT toolset.

The Stratego/XT toolset (Bravenboer, Kalleberg, et al. 2008) offers a set of features for syntax definition (with SDF2), program transformation (with Stratego) and various tools such as parser and pretty-printer generators. Stratego/XT is the predecessor of the Spoofox language workbench (Kats and Visser 2010). Spoofox offers an IDE for language engineers using an Eclipse plugin and features more meta-languages, each specialized for different components of a language definition.

¹<https://lucene.apache.org/>

²<https://researchr.org/>

```
1 strategies
2   constraint-error-all = bottomup(try(constraint-error))
3
4   constraint-error = constraint-error-ac
5   constraint-error = constraint-error-data
6   constraint-error = constraint-error-action
7   constraint-error = constraint-error-ui
8   constraint-error = constraint-error-search
9   // ...
```

Figure 2.9: Definition of the WebDSL type checking framework in Stratego. Some rules are omitted for brevity.

The core of the current WebDSL implementation is still the same, 15 years later. While the language has matured and is extended in many useful ways, the syntax is still defined in SDF2 and the rest of the compilation chain is implemented in Stratego, using the Stratego/XT toolset which is not actively developed anymore.

An application in WebDSL is compiled to a Java program, and a persisted data model through the use of the Hibernate Query Language (HQL) in the generated Java code. Before the correct target code can be generated, the WebDSL code has to pass the compilation chain that consists of components such as the parser, desugaring, type checking. In the rest of this section we discuss the implementation of the WebDSL static analysis in Stratego. In the next chapter (Chapter 3), we describe the parser generated from the WebDSL SDF2 definition.

2.6.1 WebDSL in Stratego

Apart from the syntax definition, every component of the WebDSL compiler is implemented in Stratego. For the scope of this thesis, we will focus on the implementation of the WebDSL static analysis in Stratego and we will not discuss the implementation of code generation.

Stratego (Bravenboer, Dam, et al. 2006; Visser, Benaissa, and A. P. Tolmach 1998) is a transformation language based on term rewriting. Stratego enables the definition of strategies that traverse an abstract syntax tree (AST) and applies rewrite rules. These rewrite rules may be conditional and are of the form $r : t_1 \rightarrow t_2 \text{ where } c$ with r being the name, t_1 and t_2 are terms and c is the condition. For a complete overview of the capabilities of Stratego, we refer to its documentation³.

The Stratego code for WebDSL consists of many Stratego rules that match an AST term with some conditions in the where clause that define the analysis. The static analysis is done through returning the same term as matched, but storing errors or warnings in the process. The rule that aggregates the individual rules is called `constraint-error` and this is applied in a bottom-up fashion. The bottom-up traversal strategy means that the rule will first be applied to subterms, and then to the term itself. The `constraint-error` rule is composed of many rules for specific WebDSL features. The implementation and usage of `constraint-error` is shown in Figure 2.9.

Stratego aggregates all definitions of the same rule from multiple files and combines them into one rule. The WebDSL implementation in Stratego relies on this feature by defining multiple rules with the same name, but matching on different elements with different conditions. An example of a WebDSL static analysis rule in Stratego is shown in Figure 2.10.

The `add-error` and `AddError` in Figure 2.10 utilize dynamic rules in Stratego using the `rules(...)` block. Similar to (normal) rules in Stratego, a dynamic rule matches a term,

³<https://www.spoofax.dev/references/stratego/>


```

1 rules
2   constraint-error-ac =
3     ?AccessControlPrincipal(_,_)
4     ; where(principals := <bagof-PrincipalDecl; uniq; length>)
5     ; where(<gt; (principals, 1))
6     ; add-error(["Only one access control principal can be defined."])
7
8   add-error(|msgs) =
9     rules(
10       FoundErrors := <inc> <FoundErrors <+ !0>
11     )
12     ; try(AddError(|msgs))
13
14   AddError(|msgs): node -> node
15   where rules(
16     AllErrors := (node, <error-to-string> msgs)
17   )

```

Figure 2.10: Definition of Stratego rule that shows an error message when two principals are declared in the access control rules of WebDSL.

returns a term and is able to have conditions. When dynamic rules are added during the execution of a Stratego program, it is added to the possible set of rules that can be applied. In the case of the WebDSL implementation, dynamic rules are (ab)used to store data, possibly limited to a certain scope, about the WebDSL program that is being compiled.

In the implementation, scoped dynamic rules are heavily utilized to store the internal state of the compiler. We argue that the extensive use of (scoped) dynamic rules obstructs the elegance, readability and maintainability of the WebDSL implementation in Stratego, because the effect of the dynamic rules are hard to predict before running the compiler without extensive experience with the codebase.

2.7 Modernization Goal

As discussed in the previous section, WebDSL is implemented in the Stratego/XT toolset which is not actively maintained anymore. Spoofox is the successor of the deprecated Stratego/XT toolset and features an extensive meta-languages, each specialized for defining different components of a language in a declarative way.

SDF3 (Vollebregt, Kats, and Visser 2012; Souza Amorim and Visser 2020) is the meta-language in Spoofox for syntax definition in a declarative way. From an SDF3 grammar, Spoofox generates a parser, a pretty-printer and syntactic code completions. Statix (Antwerpen, Bach Poulsen, et al. 2018) is a constraint-based declarative language for the specification of type systems in Spoofox, using the concept of scope graphs (Neron, A. Tolmach, et al. 2015) to model programs.

In this thesis we modernize the WebDSL front-end by defining the WebDSL grammar in SDF3 without post-parse filters and the WebDSL static semantics in Statix. Using this process and result, we hope to gain insight into the completeness, expressiveness and elegance of Statix and SDF3 when they are used to implement a real world language by answering the following research questions.

- **RQ1** Is it possible to define the WebDSL syntax in SDF3, without the use of post-parse

filters?

- **RQ2** How does the run time efficiency of the parser generated from WebDSL in SDF3 compare to the current WebDSL parser generated from SDF2?
- **RQ3** How does the maintainability of WebDSL in SDF3 compare to the WebDSL syntax definition in SDF2?
- **RQ4** Is a Statix implementation of the WebDSL static semantics able to catch the same errors and warnings as the current implementation in Stratego?
- **RQ5** How does the run time efficiency of the WebDSL static semantics in Statix compare to the current WebDSL implementation?
- **RQ6** How does the maintainability of the WebDSL static semantics in Statix compare to the current implementation?

Chapter 3

WebDSL in SDF3

In computer science, parsing is the process of analyzing a piece of text according to a grammar, and converting the textual representation to a more structured representation that is convenient for other processes such as a compiler or interpreter.

In this chapter we discuss the definition of the WebDSL grammar in SDF3, a meta-DSL in Spoofax for syntax definition. Currently, the WebDSL grammar is defined in SDF2, the predecessor of SDF3. The goal of defining the WebDSL grammar in SDF3 instead of SDF2 is to serve as a large case study for SDF3, while allowing the WebDSL parser to benefit from the regular updates of SDF3, compared to the deprecated SDF2.

We start by giving a brief introduction to parsing in general and introducing the SDF3 language. Then, we discuss the migration of the WebDSL syntax from SDF2 to SDF3, and we end this chapter by elaborating on the disambiguation of the WebDSL SDF3 grammar without the use of post-parse filters.

3.1 Introduction to LR Parsing

Every programming language can be described by a grammar that specifies what a syntactically correct program looks like. Given a specific program, this grammar can be used to analyze whether the program belongs to the language described by the grammar. A parser is a piece of software that is able to recognize whether a program belongs to the grammar described by the parser. Additionally, parsers create a structured representation of the input program, derived from the textual representation.

In this thesis, we will focus on *LR parsers*. LR stands for Left-to-right Rightmost-derivation, meaning that an LR parser reads the input from left to right and produces a rightmost (bottom-up) derivation. Knuth 1965 presents an LR parsing algorithm which is able to parse most languages that can be described by a context-free grammar.

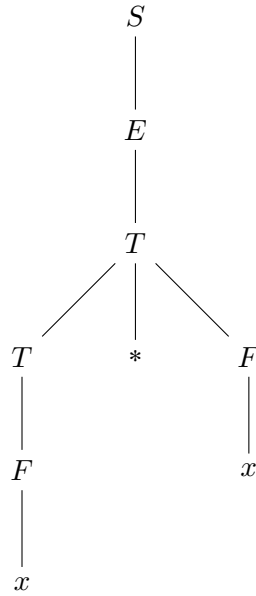
Before an LR parser can parse an input stream, it must also receive a parse table that describes a context-free grammar. In practice, the parse tables are usually generated from a syntax definition. Figure 3.1 shows an example of a context-free grammar describing a small language that features addition and multiplication. The parse table describes a push-down automaton that represents the LR parser. Using this parse table, an LR parser can build a parse tree as shown in Figure 3.2. A parse tree, or an Abstract Syntax Tree (AST) that we will discuss in Section 3.1.1 is a more structured way of representing a program, that can be used by other components of a compilation chain to analyze and transform the program.

LR parsers cannot handle ambiguous context-free grammars. The parse tables of ambiguous context-free grammars contain multiple state transitions for certain states and input tokens. The Generalized LR (GLR) parsing algorithm by Rekers (1992), is able to handle such parse tables and as a result is able to handle all context-free grammars. Visser 1997 introduced Scannerless GLR (SGLR) parsing. As opposed to LR and GLR parsers, SGLR

	State	Action				Goto			
		+	*	x	\$	S	E	T	F
	0			$S(4)$			1	2	3
$S \rightarrow E$	1	$S(5)$			<i>Accept</i>				
$E \rightarrow E + T$	2	$R(E_T)$	$S(6)$		$R(E_T)$				
$E \rightarrow T$	3	$R(T_F)$	$R(T_F)$		$R(T_F)$				
$T \rightarrow T * F$	4	$R(F_x)$	$R(F_x)$		$R(F_x)$				
$T \rightarrow F$	5			$S(4)$				7	3
$F \rightarrow x$	6			$S(4)$					8
	7	$R(E_{E+T})$	$S(6)$		$R(E_{E+T})$				
	8	$R(T_{T*F})$	$R(T_{T*F})$		$R(T_{T*F})$				

(a)
(b)

Figure 3.1: Example context-free grammar with its parse table

Figure 3.2: An example of a parse tree for the input " $x * x$ "

parsers do not have a separate lexing and parsing phase, but instead merge these through the use of grammars that are defined in terms of single characters.

3.1.1 SDF3

Syntax Definition Formalism 3 (SDF3) (Vollebregt, Kats, and Visser 2012; Souza Amorim and Visser 2020) is a meta-language in the Spoofox language workbench that makes use of the SGLR parsing algorithm. It is the latest version of the syntax definition formalism SDF (Heering et al. 1989; Visser 1997b). Language engineers are able to define context-free grammars in SDF3, which are transformed into parse tables and used by the JSGLR2¹ parsing algorithm. JSGLR2 (Denkers 2018) is the successor of the JSGLR algorithm; an implementation of the SGLR parsing algorithm in Java.

SDF3 is the successor of SDF2, in which the WebDSL syntax is currently defined. Souza Amorim and Visser 2020 argue that the SDF3 syntax is more similar to other grammar formalisms such as EBNF.

¹<https://github.com/metaborg/jsglr>

```

1 module ThesisTest
2
3 context-free start-symbols
4   Start
5
6 context-free sorts
7   Start Expr Lit
8
9 context-free syntax
10  Start.Expr = <<Expr>>
11
12  Expr.Add = <<Expr> + <Expr>>
13  Expr.Mul = <<Expr> * <Expr>>
14  Expr.Lit = <<Lit>>
15
16  Lit.Int = <<INT>>
17
18 lexical sorts
19   INT
20
21 lexical syntax
22   INT    = [0-9]+
23   LAYOUT = [\ \t\n\r]

```

Figure 3.3: An SDF3 specification of a language that allows addition and multiplication of integers.

A syntax definition in SDF3 is a declarative specification of syntactic sorts and their productions. Figure 3.3 shows an example grammar features a language that supports multiplication and addition of integers. The SDF3 definition must define a start-symbol that specifies what a syntactically valid program looks like. In our example, every program that can be derived from the `Start` sort, is a valid program. **Line 11** contains the first production of the specification. It states that the sort `Start` can be derived by deriving something of the sort `Expr`. The identifier behind the dot in the production (`Expr` in this case) declares the constructor in the AST. Figure 3.4 shows the AST of an example input, according to the SDF3 specification of Figure 3.3. **Line 12-14** defines the productions of the expression sort. It defines that an expression is either two expressions with a plus or asterisk in between, or it is something of the sort `Lit`. The production of literal sort on **line 16** references the lexical sort `INT`. Lexical sorts do not appear in the AST with constructors, but instead are parsed into a string. The lexical sort `INT` is defined on **line 22** with the regular expression `[0-9]+`, describing one or more characters in the range of 0 to 9. Finally, the production describing the built-in concept of `LAYOUT` on **line 23** states that a space, tab, carriage return and line feed character are layout (whitespace) characters and do not have to be parsed.

Figure 3.4 shows the abstract syntax tree of an example input for the SDF3 specification of Figure 3.3. In the Spoofox Language Workbench, abstract syntax trees are described in the Annotated Terms Format (ATerm) (Brand et al. 2000).

In addition to the basics as explained in the previous paragraph, SDF3 supports features such as injections, optional sorts and repetition to enhance the language engineers productivity. Figure 3.5 shows the SDF3 specification of Figure 3.3, but now enhanced with injections, repetition and optional sorts. Figure 3.6 shows the ATerm of the input `0 + 0; 1 * -1`. The

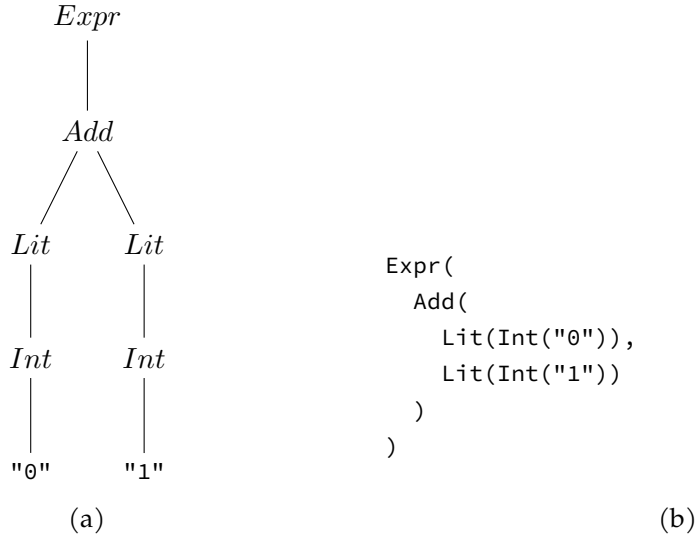


Figure 3.4: The AST and ATerm representation of the expression $0 + 1$ according to the SDF3 specification of Figure 3.3

repetition on **line x** of Figure 3.5 allows multiple expressions in a list, delimited by a semicolon. The injection on **line 14** allows a literal to be derived in the place of an expression, effectively omitting the `Lit(...)` constructor in the AST. Lastly, the optional minus sign of **line 16 and 18** allows negative integers to be parsed by this language. The AST contains `Some(...)` and `None()` constructors for the optional sorts.

Disambiguation

The simple grammar described by the SDF3 specification of Figure 3.3 is functional, but is ambiguous. For example, the input $1 + 2 + 3$ can be parsed as $(1 + 2) + 3$ or as $1 + (2 + 3)$. The process of altering and annotating the grammar such that there is only one way of parsing this, is called disambiguation. The example listed in the previous section (Figure 3.1) contains a grammar with expressions, terms and factors that is inherently unambiguous. However, disambiguating a grammar by introducing new sorts and productions is tedious and time-consuming. SDF3 provides multiple options for disambiguating a grammar.

First, SDF3 provides the `{bracket}` annotation, allowing the developer to disambiguate the program himself. To clarify, the input $1 + (2 + 3)$ is not valid according to our grammar, because the `"(" and ")"` symbols are not part of the grammar. If we add the production `Expr = "(" Expr ")" {bracket}`, SDF3 allows brackets around arbitrary expressions without it introducing new AST nodes.

Reject rules allow language engineers to filter derivations. A reject rule is simple a regular production, followed by the `{reject}` annotation. For example, appending our grammar with the rule `Expr = <<Expr> + <Expr>> {reject}` makes the set of valid derivations of `Expr` smaller, namely by disallowing any construction described by the right-hand side of the reject rule.

Another possibility for disambiguating SDF3 grammars is by indicating the associativity, either as annotation or using priority groups. If we add the annotation `{left}` to the production that specifies addition, $1 + 2 + 3$ is no longer ambiguous but instead will always be parsed as $(1 + 2) + 3$. It is also possible to declare the associativity in groups. Figure 3.7 contains such groups. **Line 8** shows that both addition and subtraction are left associative, and in the same group. The input $1 + 2 - 3 + 4$ will now be parsed as $((1 + 2) - 3) + 4$, whereas declaring the associativity annotation on both rules instead of the group would still make this input ambiguous.

```

1 module m
2
3 context-free start-symbols
4   Start
5
6 context-free sorts
7   Start Expr Lit Minus
8
9 context-free syntax
10  Start.Exprs = <<{Expr "; "}>>
11
12  Expr.Add = <<Expr> + <Expr>>
13  Expr.Mul = <<Expr> * <Expr>>
14  Expr      = Lit
15
16  Lit.Int = <<Minus?> <INT>>
17
18  Minus.Minus = <->
19
20 lexical sorts
21  INT
22
23 lexical syntax
24  INT    = [0-9]+
25  LAYOUT = [\ \t\n\r]

```

Figure 3.5: An SDF3 specification of a language that allows addition and multiplication of integers, showcasing injections, repetition and optional sorts.

```

Exprs([
  Add(
    Int(None(), "0"),
    Int(None(), "0")
  ),
  Mult(
    Int(None(), "1")
    Int(Some(Minus()), "1")
  )
])

```

Figure 3.6: The ATerm of input $0 + 0; 1 * -1$ according to the SDF3 specification of Figure 3.5

```

1 context-free syntax
2   Expr.Mul = <<Expr> * <Expr>>
3   Expr.Add = <<Expr> + <Expr>>
4   Expr.Sub = <<Expr> - <Expr>>
5
6 context-free priorities
7   {left: Expr.Mul} >
8   {left: Expr.Add Expr.Sub }

```

Figure 3.7: An SDF3 specification of a simple language with disambiguation rules.

```

1 "var" Id ":" Sort ";"          -> VarDeclStat {cons("VarDecl")}
2 "var" Id ":" Sort ":@" Exp ";" -> VarDeclStat {cons("VarDeclInit")}
3 "var" Id ":@" Exp ";"         -> VarDeclStat {cons("VarDeclInitInferred")}

```

Figure 3.8: An example of productions in SDF2. This example specifies the syntax of variable declaration in WebDSL.

Figure 3.7 also shows the declaration of priorities. In the example, multiplication has priority (e.g. binds tighter) over addition and subtraction. This results in the input $1 + 2 - 3 * 4$ being parsed as $(1 + 2) - (3 * 4)$.

SDF3 also provides the option of indexed priorities in the form of $p1 \ i \ .> \ p2$. This can be explained intuitively as the subterm with index i of production $p1$ may not be derived by production $p2$. In our example, the rule $\text{Expr.Add} \ \langle 0 \rangle \ .> \ \text{Expr.Mul}$ would imply that the left-hand side of an addition may never be a multiplication.

Lastly, SDF3 provides the ability to disambiguate through the use of post-parse filters annotated by `{prefer}` and `{avoid}`. When the parser encounters an ambiguous input, it will continue parsing and store all possibilities. At the end of parsing, it will use the annotations to prune the multiple ASTs according to these annotations. The `{prefer}` and `{avoid}` annotations are working but deprecated in the current version of SDF3 and will be removed in the future. Using the post-parse filters makes the disambiguation less transparent than using the other methods described in this section.

3.2 WebDSL Grammar Specification

The current grammar of WebDSL is specified in SDF2, the predecessor of SDF3. Similar to SDF3, a production in the grammar consists of terminals and terminals. In contrast to SDF3, it is optional to provide a constructor to appear in the abstract syntax tree. Figure 3.8 shows an example of productions in SDF2.

The WebDSL language SDF2 specification consists of 26 files that each describe a different part of the WebDSL syntax, ranging from the data model with entities to user interfaces with HTML and JavaScript. Additionally, the WebDSL syntax incorporates stand-alone grammars of other languages, such as HQL, Java and Stratego using mix syntax. Out of these mixed syntaxes, only the WebDSL and HQL syntax can be written by developers. The Java and Stratego syntax are used in other components of the WebDSL compiler, such as writing concrete WebDSL and Java syntax during desugaring, optimization and code generation.


```

1 context-free syntax
2 "section" SectionName Definition*
3   -> Section {cons("Section")}
4
5 "application" QId Definition+ Section*
6   -> Application {cons("ApplicationDefs")}
7
8 "application" QId Section*
9   -> Application {cons("Application")}

```

(a)

```

1 simplify-application-constructor :
2 ApplicationDefs(qid, defs, sections)
3   -> Application(qid, [Section("definitions", defs)|sections])

```

(b)

Figure 3.9: WebDSL Syntax and desugaring for sections and definitions

3.2.1 Desugaring WebDSL Syntax

In WebDSL, the root of the abstract syntax tree is the application or module definition. A project is allowed to have one application and the other files should be modules that are (transitively) imported by the main application file.

A file is divided in sections, which have no semantic meaning but are purely for code organization. The sections contain the top-level definitions such as pages, templates and entities. However, these sections are optional, and it is possible to list definitions without dividing them in sections.

To prevent an explosion of cases in the analysis and code generation, these constructs are normalized (desugared) to a core WebDSL syntax. Figure 3.9 shows an example of desugaring different definitions of a file.

3.3 Migration from SDF2 to SDF3

SDF2 is still supported by the Spoofax language workbench, but no longer actively developed. SDF3 is the most recent member of the Syntax Definitions Formalisms. In a sense, SDF3 is a high-level version of SDF2 with additional features. In the initial versions of SDF3, it compiled down to SDF2.

Spoofax contains tool to migrate SDF2 specifications to SDF3 specifications but not all constructs of the lower level SDF2 are expressible in SDF3. The WebDSL SDF2 specification requires manual adjustment before this transformation tool can be utilized. Additionally, the generated SDF3 definition is suboptimal, as not all features of SDF3 are utilized in the generated code.

3.3.1 Preparing the WebDSL SDF2 definition for migration

Not all SDF2 constructs can be expressed in SDF3. For this reason, the WebDSL SDF2 must be adjusted manually to be able to use the tool provided by Spoofax that transforms SDF2 specifications to SDF3 specifications.

```
1 lexical syntax
2  [a-zA-Z\_][a-zA-Z0-9\_]* -> ID
3 lexical restrictions
4  ID -/- [a-zA-Z0-9\_ ]
5  "in" -/- [a-zA-Z0-9\_ ]
```

Figure 3.10: Follow restrictions are used to for example ensure the longest-match of an identifier and enforcing whitespace after keywords.

Sorts Non-terminals are named *Sorts* in SDF2 and SDF3. A production in SDF2 reduces a sequence of terminals and non-terminals to a certain sort, as listed in the example of Figure 3.9. In SDF2 the sort does not have to be declared in a separate *sorts* section, while it does in SDF3. Additionally, SDF2 does not differentiate lexical sorts from context-free sorts. The result of this difference in design is that SDF2 sorts cannot be migrated to SDF3, since the transformation tool does not know if an SDF2 sort is used as lexical or context-free sort. To be able to translate the SDF2 specification to SDF3, we must remove the sort declarations.

Alternations Another manual adjustment we must make to the SDF2 specification is removing alternations. An alternation is a production of the form "a" | "b" -> s. This construct is not supported anymore in SDF3 and therefore the transformation tool is not able to handle this construct. Fortunately, the alternation can easily be split in two production rules. However, the transformation tool does not automate this process because the original production may contain a constructor annotation ("a" | "b" -> s {cons("AorB")}) and the duplicate constructors for the same sort are not allowed in SDF3. For this reason must manually remove alternations in SDF2 productions.

Follow Restrictions The transformation tool does also not include support for follow restrictions in the grammar. A follow restriction indicates what symbols are not allowed to be parsed after a production. A typical use case of this is ensuring the longest match for identifiers, or enforcing whitespace after a keyword. An example of follow restrictions are shown in Figure 3.10. For the migration to SDF3, these sections have to be manually copied.

3.3.2 Manual Adjustment of Generated WebDSL SDF3

The generated SDF3 specification of WebDSL according to the manually adjusted SDF2 is suboptimal SDF3 code. SDF2 can be seen as a more low-level version of SDF3 and not all constructs of SDF2 transform into the most elegant SDF3 code. Parts of the generated SDF3 code do not even adhere to the static semantics of SDF3. For this reason, the generated SDF3 code has to be manually adjusted.

Missing and duplicate constructors In SDF2, constructors are defined as an annotation on a production. An example of the different production syntax of SDF2 and SDF3 is shown in Figure 3.9. The constructor annotation is not required in SDF2, but in SDF3 it is necessary to provide a constructor for every production. In the WebDSL SDF2 definition, some constructors are missing and there were duplicate constructors that denoted alternative syntax for the same construct, essentially providing syntactic sugar. The SDF2 definition and generated SDF3 definition for the ascending and descending ordering is shown in Figure 3.11.

In the newly generated SDF3, duplicate constructors must be changed, in order for them to be unique. Additionally, missing constructors must be added for the pretty-printer to function correctly. Injections do typically not require a constructor, but preferably constructors

```

1 context-free syntax
2 "asc"          -> AscendingOrDescending {cons("Ascending")}
3 "ascending"    -> AscendingOrDescending {cons("Ascending")}
4
5 "desc"         -> AscendingOrDescending {cons("Descending")}
6 "descending"   -> AscendingOrDescending {cons("Descending")}

```

(a)

```

1 context-free syntax
2 HQLAscOrDescOpt.HQLAscending = <ascending>
3 HQLAscOrDescOpt.HQLAscending = <asc>    // Error: duplicate constructor
4
5 HQLAscOrDescOpt.HQLDescending = <descending>
6 HQLAscOrDescOpt.HQLDescending = <desc> // Error: duplicate constructor

```

(b)

Figure 3.11: SDF2 and generated SDF3 for the ascending and descending keyword in WebDSL. Duplicate constructors within the same sort are not allowed in SDF3.

```

1 context-free priorities
2 {left:
3   Expr "*" Expr -> Expr
4 } >
5 {left:
6   Expr "+" Expr -> Expr
7   Expr "-" Expr -> Expr
8 }

```

(a)

```

context-free priorities
2 {left: Expr.Mul} >
3 {left: Expr.Add Expr.Sub }

```

(b)

Figure 3.12: Difference between priority chains in SDF2 and SDF3. The transformation tool does not perform the analysis to properly transform this.

should be added there as well to reduce the length of the abstract syntax tree in Statix, as we will explain in Section 3.4.

Priority chains Priority chains are used to disambiguate grammars in SDF specifications. A typical example and an explanation is given in Figure 3.7. Both SDF2 and SDF3 use the concept of priority chains, but the SDF2 variant requires a repetition of the production inside the chain, whereas SDF3 uses a reference to the production. This causes the SDF2 priority chains to not be migrated to the SDF3 priority chains and requires them to be manually rebuilt. An example of the difference between priority chains in SDF2 and SDF3 is shown in Figure 3.12.

Transferring comments Comments are highly recommended and important for the readability and maintainability of large software projects. However, most grammars specified in SDF2 and SDF3 declare comments as part of the layout of a program and are not represented in the abstract syntax tree. For this reason, the transformation tool does not transfer

```

1 context-free syntax
2  "if" "(" Expr* ")" "{" Expr* "}" -> Expr {cons("If")}

```

(a)

```

1 context-free syntax
2  Expr.If = <if ( <Expr> ) { <Expr*> }>

```

(b)

```

1 context-free syntax
2  Expr.If = <
3    if ( <Expr> ) {
4      <Expr*>
5    }
6  >

```

(c)

Figure 3.13: An example of an SDF2 production, the generated SDF3 and the manually adjusted SDF3 for template productions.

comments from SDF2 grammars to the generated SDF3. Comments must be manually copied from the SDF3 and pasted in the correct parts of the SDF3 specification.

Template productions A major change in SDF3 compared to SDF2 are template productions, that allow for nice pretty printing and syntactic code completion. The productions in the generated SDF3 files are all template productions, but do not have the proper surrounding layout and indentation because there is no way to extract this information from the SDF2 source, as it is not present there. Adding the correct layout in template productions is tedious work but causes the pretty-printer to function properly. An example of SDF2, generated SDF3 and manually adjusted SDF3 is shown in Figure 3.13.

Deeply Embedding HQL Grammar in WebDSL The HQL syntax definition in SDF2 is a stand-alone definition, and is used in the WebDSL SDF2 using parameterized imports. SDF3 has no support for this feature, which means the syntax definition has to be transformed to be a part of the WebDSL grammar. Otherwise, WebDSL applications that use HQL syntax, which are all real world applications, would no longer parse correctly. Deeply embedding the HQL syntax in the WebDSL syntax causes sorts and constructors to overlap, and this must be adjusted manually.

3.4 Preparation for Statix

Statix is a constraint-based meta-language in which we implement the modernized WebDSL static analysis. In Chapter 4 Statix will be explained in detail. Even though the parsing and the static analysis are separate components of the compilation chain, the Spoofox language workbench contains a tool that extracts the sorts and constructors from an SDF3 definition, and generates these signatures in the Statix language to be able to use them while defining the static analysis. This tool is called the Statix signature generator. Using the Statix signature generator imposes additional constraints on the SDF3 definition.

3.4.1 Sorts and Constructors in Statix

To understand the role of the Statix signature generator, we will explain how sorts and constructors are used in Statix to define the static semantics of a language.

Unlike SDF3 and Stratego, Statix is strongly typed which requires all sorts and constructors to be declared before they can be used in the static semantics of a language. Figure 3.14 shows an example of declaration of the sorts and constructors of a simple expression language. These constructors and terms can be used in rules that form the static semantics.

The constructors in Statix do not support injections or overloaded constructors, which leaves some abstract syntax trees generated by the parser unable to serve as input for static analysis.

3.4.2 Statix Signature Generator

As mentioned and demonstrated in the previous subsection, Statix requires a definition of the constructors and sorts of the language to be analyzed. Defining these manually in Statix would essentially be code duplication because they are already defined in SDF3. However, Statix sorts and constructors do not support overloading and injections and is therefore stricter than the SDF3 type system.

To prevent manual redefinition of the sorts in Statix code, the Statix signature generator is part of Spoofax. This tool takes an SDF3 definition as input, and generates importable Statix files that contain the sorts and constructors from the syntax definition. For the Statix signature generator to work properly, Additional constraints are imposed on the SDF3 definition.

```

1 signature
2   sorts
3     Application
4     Expr
5
6   constructors
7     Application : Expr      -> Application
8     Add       : Expr * Expr -> Expr
9     Sub        : Expr * Expr -> Expr
10    Mul       : Expr * Expr -> Expr
11    Int       : string    -> Expr

```

Figure 3.14: An example of an expression language signature in Statix.

Explicitly Declare Sorts To define static semantics of all sorts in a language, these must be explicitly declared in Statix. In SDF3 this is not enforced. To be able to use the Statix signature generator, we have to explicitly declare all sorts in the WebDSL SDF3. Additionally, Statix views all lexical sorts as strings, so the SDF3 sorts have to be divided in lexical sorts and context-free sorts.

Injections With the semantics of sorts and constructors in Statix, it is not possible to model injections while the WebDSL SDF3 definition contains many injections. The Statix signature generator still functions with injections present in the SDF3, but it will generate an additional constructor and *explicate* the constructors of an abstract syntax tree before it is passed to Statix. Figure 3.15 shows an example of how injections are handled and explicated by the Statix signature generator. The abstract syntax tree is explicated after parsing and before static analysis. When injections are present in the SDF3, the Statix signature generator also generates Stratego rules that explicate the abstract syntax tree.

Optional Sorts As mentioned in Figure 3.5 as part of Section 3.1.1, SDF3 has built-in support for optional sorts, resulting in `Some(_)` and `None()` terms. These terms cannot be translated to Statix signatures, since it is not possible to overload the constructor for many sorts. To work around this limitation, the SDF3 definition must be altered make the `Some` and `None` constructors unique per sort. This leads to a much more verbose syntax definition. Figure 3.16 shows how the optional constructs can be altered to be unique per sort.

```

1 context-free sorts
2   S A B
3 context-free syntax
4   S.Start = A
5   A       = B // injection
6   B.Literal = "b"

```

(a) SDF3

```

signature
2 sorts
3   S A B
4
5 constructors
6   Start : A -> S
7   Literal :      B
8   // extra generated constructor
9   B2A : B -> A

```

(b) Generated Statix signatures

```

Start(
  Literal()
)

```

(c) Example of an AST

```

Start(
  B2A(
    Literal()
  )
)

```

(d) Example of an explicated AST

Figure 3.15: An example of how SDF3 injections are handled by the Statix signature generator.

```

1 context-free syntax
2   Definition.Function = <
3     function <ID> () <ReturnSort?> {
4       <Statement*>
5     }
6   >
7
8   ReturnSort.ReturnSort = <: <Sort>>

```

(a) SDF3 with optionals

```

1 context-free syntax
2   Definition.Function = <
3     function <ID> () <ReturnSortOpt> {
4       <Statement*>
5     }
6   >
7
8   ReturnSortOpt.ReturnSort = <: <Sort>>
9   ReturnSortOpt.ReturnSortNone = <>

```

(b) Rewritten to remove optionals

Figure 3.16: An example of how to remove the optional sort to comply with the Statix signature generator.

```

template t() {
  // from-clause is optional in this HQL expression
  // meaning that "select u.name" is a complete expression
  // and "from", "User", "as", and "u" can also be parsed
  // as template calls
  var names := select u.name from User as u

  // parentheses are optional in template calls
  callToOtherTemplate
}

```

Figure 3.17: An example of an ambiguous construct in WebDSL with an HQL query.

Disambiguation SDF3 provides many features to disambiguate a grammar as listed in Section 3.1.1. If an ambiguous code fragment is parsed, it will result in the `amb([...])` term in the AST, which contains a list of the possible interpretations. Similar to the `Some(_)` and `None()` constructors from the optionals as discussed in the previous paragraph, the `amb([...])` can not be expressed in Statix. The result of this is that ambiguous code fragments can not be used in Statix rules and therefore will never be analyzed, increasing the need for proper disambiguation even more. In the next section we will discuss the disambiguation of the WebDSL SDF3 grammar.

3.5 Disambiguation

When a program is ambiguous, it can be interpreted in multiple ways. Generally, ambiguous programs are undesired due to the undefined semantics. To prevent ambiguities in the syntax of a programming language, the syntax definition can be disambiguated.

SDF3 provides multiple language features to disambiguate a syntax, such as bracket rules, reject rules, associativity annotations, priority groups and indexed priorities. A complete list of ways to disambiguate is explained in Section 3.1.1.

Post-parse filters `{prefer}` and `{avoid}` provide a unique way to disambiguate the syntax, namely it keeps all the possible parse trees in memory and after the whole input is parsed, the parser prunes the resulting parse forest according to the `{prefer}` and `{avoid}` annotations.

The `{prefer}` and `{avoid}` post-parse filters are deprecated in the current version of SDF3. In this thesis we attempt to disambiguate the grammar of a WebDSL with the other disambiguation functionalities provided by SDF3.

3.5.1 Disambiguating Keywords

WebDSL template calls do not require parentheses or semicolons. This allows for clean code in templates, but makes the syntax ambiguous in the case of keywords that are able to parse as a template name.

Figure 3.17 shows an example of a WebDSL program where a template starts with a variable initialization with an HQL query that fetches all names of users from the database. In HQL, the `from`-clause is optional, meaning that a statement such as `select user.name` syntactically correct as an HQL expression. For this reason, the words after the `select` clause can also be parsed as template calls.

A simple yet effective solution is to disallow all keywords such as `from` and `as` from being valid template names. SDF3 automatically detects keywords from context-free productions,


```

1 lexical syntax
2   ID = [a-zA-Z]*
3
4 context-free syntax
5   Start.Hello = <hello <Name>>
6
7   Name.World = <world>
8   Name.Name = <<ID>>
9
10 template options
11   // "hello" and "world" are now no longer valid IDs
12   ID = keyword {reject}

```

Figure 3.18: An example of rejecting keywords in template options.

and all keywords can be rejected as identifiers at once using template options². An example of how keyword rejection in template options is configured, is shown in Figure 3.18.

Rejecting all keywords in the WebDSL language using the template options of SDF3 would be effective, but too rigorous. For example, the WebDSL grammar contains productions such as the following. Rejecting all keywords would imply also rejecting `Int` as identifier, meaning that it is now allowed as type anymore.

```
Statement.ForCountStmt = <for ( <VarId> : Int from <Exp> to <Exp> ) <Block>>
```

Another option is to add all the reject rules for keywords individually. This is tedious work, but allows for better precision. In the new WebDSL SDF3, all keywords are rejected individually.

3.5.2 Disambiguating Strings

WebDSL strings are defined in SDF2 using kernel syntax. Kernel syntax does not allow implicit layout (whitespace) between the different elements of a production. The SDF2 productions for a WebDSL string are shown in Figure 3.19. A string can be either a “simple” string or a string containing one or more expressions that have to be evaluated. These expressions must be prefixed by a tilde (`~`) to be parsed as interpolation.

To prevent a simple string from being parsed as a `StringInterp` with only one part, the SDF2 definition used a post-parse filter to avoid this construct in case of ambiguities.

Figure 3.20 shows the SDF3 definition of a WebDSL string. The “simple” string constructor was dropped, to ensure no ambiguity would arise. A simple string is now a string consisting of only one part.

3.5.3 Non-transitive Priority Rules

Another ambiguity in the WebDSL language is the overlap between passing template elements and the untyped set creation. Figure 3.21 demonstrates the ambiguity. The curly brackets are used for an untyped set creation, as shorthand for `Set<Int>(...)`. On the other hand, curly brackets are also used to pass template elements as argument to another template.

In SDF3, indexed non-transitive priorities are able to disambiguate this case. An indexed non-transitive priority rule is of the form `SortA.ConsA <n> .> SortB.ConsB`. Intuitively, the

²<https://www.spoofox.dev/references/syntax/templates/#reject>


```

1 %% Kernel syntax is required here since we do not want LAYOUT to be parsed
2 %% between the first QMLex and StringLex
3 syntax
4   %% string literal as expression
5   <QMLex-LEX> <StringLex-LEX> <QMLex-LEX>
6     -> <String-CF> {ast("String(<2>"))}
7
8   <QMLex-LEX> StringPart* <QMLex-LEX>
9     -> <String-CF> {ast("StringInterp(<2>"), avoid)}
10
11   SimpleStringPart      -> StringPart
12   <StringLexInterp-LEX> -> StringPart
13
14   "~" <SimpleExp-CF>    -> SimpleStringPart {cons("InterpExp")}
15   "~" "(" <Exp-CF> ")" -> StringPart {cons("InterpExp")}

```

Figure 3.19: SDF2 definitions of a WebDSL string.

```

1 // Kernel syntax is required here since we do not want LAYOUT to be parsed
2 // between the first quote and StringLex
3 syntax
4
5   String-CF.String = "\" StringPart-CF* "\""
6
7   StringPart-CF.StringValue = StringLex-LEX
8   StringPart-CF.InterpValue = "~" StringInterpExp-CF
9   StringPart-CF.InterpExp = "~" "(" Exp-CF ")"
10
11   StringInterpExp-CF.InterpSimpleExp = SimpleExp-CF

```

Figure 3.20: SDF3 definitions of a WebDSL string.

```

page p() {
  var mySet := { 1, 2, 3 } // untyped set creation

  t({ "Hello World" }) // pass template elements as argument
}

template t(elems : TemplateElements) {
  elems // this will render the content of elems
}

```

Figure 3.21: Ambiguous WebDSL code using curly brackets for untyped set creation and passing template elements as argument.

```

1 context-free syntax
2   Exp.UntypedSetCreation = <{ <{Exp ","}>+> }>
3
4   TemplateArg.TemplateArgExp = <<Exp>>
5   TemplateArg.TemplateArgElements = <{ <TemplateElements*> }>
6
7 context-free priorities
8   TemplateArg.TemplateArgExp <0> .> Exp.UntypedSetCreation

```

Figure 3.22: Using non-transitive indexed priorities to disambiguate WebDSL syntax in SDF3.

```

template t(e : Entity) {
  // example of a cast expression
  var x := e as User

  // switch-case statement on the type of a variable
  typecase(e) {
    User    { "e is a User"    showUser(e as User)    }
    Student { "e is a Student" showStudent(e as Student)}
    default { }
  }

  // switch-case statement on the type of a variable using an alias
  // the alias automatically casts the expression based on the case
  typecase(e as t) {
    User    { "e is a User"    showUser(t)    } // t := e as User
    Student { "e is a Student" showStudent(t)} // t := e as Student
    default { }
  }
}

```

Figure 3.23: Ambiguous WebDSL code using an optional typecase alias or a cast expression.

rule indicates that `SortB.ConsB` may never be the n -th subterm of `SortA.ConsA`. Figure 3.22 shows how the WebDSL syntax can be disambiguated using these priorities.

Another instance of an ambiguity that is corrected with indexed non-transitive priority rules is optional type aliases versus a cast expression. WebDSL allows a switch-case statement on the type of an expression. Figure 3.23 shows an example of how it is used in WebDSL code. The `typecase` takes an expression as subterm, and an optional type alias. The type alias automatically is the expression cast to whatever case is relevant. However, the expression given to the `typecase` can also be a cast expression, making `typecase(e as t)` ambiguous.

Figure 3.24 shows how the optional type alias versus a cast expression is disambiguated in the modernized WebDSL SDF3 definition. The index of the `<Exp>` subterm is 2 and not 0 because the index counts literals as well, `typecase` and `(` in this case.

```
1 context-free syntax
2 Exp.Cast = <<Exp> as <Sort>>
3 Statement.TypeCaseStmt = <
4   typecase ( <Exp> <OptTypeAlias> ) {
5     <TypeCaseAlt*>
6   }
7 >
8 OptTypeAlias.TypeAlias = <as <Id>>
9 OptTypeAlias.OptTypeAliasNone = <>
10
11 context-free priorities
12 Statement.TypeCaseStmt <2> .> Exp.Cast
```

Figure 3.24: Using non-transitive indexed priorities to disambiguate WebDSL casts versus type aliases.

Chapter 4

WebDSL in Statix

In this chapter, we elaborate on the implementation of the WebDSL static semantics in Statix, using the examples from Chapter 2 as a basis. We start this chapter by introducing the meta-DSL Statix. Once the goal and basics of Statix are stated, we describe the implementation of the type system that is the core of WebDSL. Next, we address and discuss the challenges faced while implementing non-trivial WebDSL features in Statix, and lastly we reflect on the developer experience of using Statix to implement static analyses.

4.1 Introduction to Statix

Statix is a constraint-based declarative language for the specification of type systems, introduced in 2018 (Antwerpen, Bach Poulsen, et al. 2018). Since then, the meta-DSL Statix has become a part of the Spoofox Language Workbench and allows language developers to implement static analyses to provide language-specific feedback to developers on written code.

A Statix specification consists of rules over terms that define constraints. Additionally, Statix rules build and query a *scope graph* (Neron, A. Tolmach, et al. 2015) that provides a language-agnostic representation of a program. A scope graph consists of nodes and edges that can be used to for example model the lexical scope of variables.

4.1.1 Language Signature

Consider a language consisting of booleans, integers and addition, for which we want to create a type-checker with Statix. First, Statix requires us to declare all types and sorts that we will be using in the rules. These Statix constructor names have to match the constructors of the input term (the AST). The Statix code that declares the sorts and constructors of our example language is shown in Figure 4.1. When writing a Statix specification for a language implemented in the Spoofox language workbench, it is a common practice to have the Statix signature generated from your SDF3 specification by the Statix signature generator (see Section 3.4.2), to prevent code duplication.

```
1 signature
2   sorts
3     Application
4     Exp
5
6   constructors
7     Application : Exp      -> Application
8     True       :           Exp
9     False      :           Exp
10    Int        : string   -> Exp
11    Add       : Exp * Exp -> Exp
```

Figure 4.1: Language signature in Statix

So far, our specification consists of two sorts. The `Application` sort defines the entry point of our language, it has one constructor with an identical name. Next, the sort `Exp` describes what expressions are allowed. It has four constructors: the Boolean values `True` and `False`, `Int` which requires an integer literal as subterm, and `Add` which takes two nested expressions as subterms. Examples of valid input according to our defined signature are shown in Figure 4.2.

```
Application(True())           // true
Application(Int("42"))       // 42
Application(Add(Int("40"), Int("2"))) // 40 + 2
Application(Add(Int("40"), False())) // 40 + false
```

Figure 4.2: Valid input terms for the described language

4.1.2 Semantic Types

Not all the valid input terms according to our signature are well-typed. For example, the last term shown in Figure 4.2 features an addition of the integer literal `40` and the Boolean value `False`. Using Statix' constraint solving capabilities, we would like to give feedback to the programmer that the input is ill-typed.

Given the code in Figure 4.1, our Statix specification does not yet generate any constraints. Constraints that we would like to generate using Statix rules are firstly that a program must be well-typed and secondly, in order for an addition expression to be well-typed, its two subterms must be of integer type.

To reason about the types of expressions and use them in constraints, we must first define them in our specification, as shown in Figure 4.3. To distinguish input sorts and constructors from semantic types that we will use in our constraints, those sorts and constructors are defined in upper-case. With the new `TYPE` sort that has two constructors: `BOOL` and `INT`, we can start generating constraints on input terms.

```
1 signature
2   sorts
3     TYPE
4
5   constructors
6     BOOL : TYPE
7     INT  : TYPE
```

Figure 4.3: Statix signature for Boolean and integer types

4.1.3 Predicates and Rules

Figure 4.4 lists the Statix predicates and rules required to generate the constraints we want to be satisfied in order for a program to be well-typed.

```

1 rules
2
3 applicationOk : Application
4 applicationOk(Application(e)) :- { T }
5   typeOfExp(e) == T.
6
7 typeOfExp : Exp -> TYPE
8 typeOfExp(True()) = BOOL().
9 typeOfExp(False()) = BOOL().
10 typeOfExp(Int(_)) = INT().
11 typeOfExp(Add(e1, e2)) = INT() :-
12   typeOfExp(e1) == INT(),
13   typeOfExp(e2) == INT().

```

Figure 4.4: Statix predicates and rules for typing booleans, integers and addition

The type of all Statix predicates must be explicitly declared, for example the `applicationOk` predicate on **line 3** specifies that all rules of `applicationOk` match exactly one constructor `Application`. An instantiation of the `applicationOk` predicate is on **line 4**. In prose English it would read “An application is well-typed, given that for some type τ , the expression e has type τ ”.

The other Statix rule in our small example specification is a *functional predicate*, meaning that it returns a value. All but the last rules of the `typeOfExp` predicate compute a `TYPE` for a given expression, without conditions. The last rule of the example does have two conditions, in prose English it would read “ e_1 plus e_2 is of type `INT`, given that e_1 is of type `INT` and e_2 is of type `INT`”.

4.1.4 Building and Querying Scope Graphs

When we expand our small example language with let-bindings and we want to add typing rules for this new construct, we come across a new feature in Statix. To facilitate typing rules for name binding, Statix uses *scope graphs* (Neron, A. Tolmach, et al. 2015). Scope graphs are built out of three components: scopes, edges and declarations.

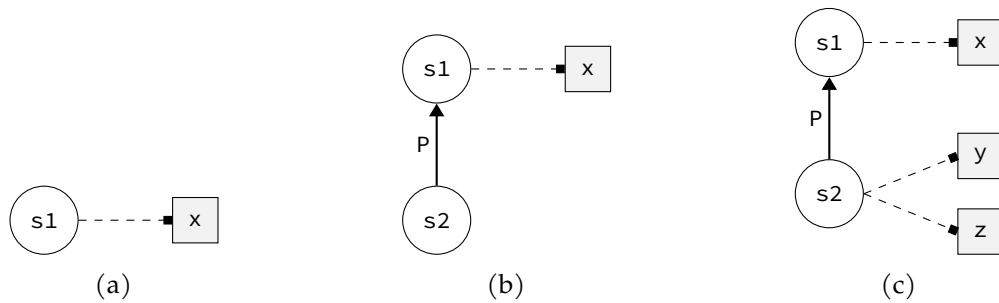


Figure 4.5: Scope graph examples

Figure 4.5 showcases three examples of scope graphs. Figure 4.5a consists of a single scope `s1` with declaration `x` that could be a model of a module with a single global variable `x` declared inside. The second example, Figure 4.5b, consists of two scopes: a root scope `s1` with again a declaration of `x`, and a scope `s2` with an outgoing edge to `s1` labeled `P`. The `P` label is often used to denote the relation of a lexical parent scope. In this example, `s2` could for example model an empty function declared in module `s1`. The last example again has two scopes, with one declaration in `s1` and two declarations in `s2`. This could model the same

program as described previously, but now with two local variable declarations inside the function body of `s2`.

The first step in implementing let-bindings in Statix is adding the signature. In addition to the new constructors on **line 3 and 4**, we now introduce an edge label `P` and the relation `var`. The edge labels defined in the constructor provide the set of allowed labels to use in rules later on. The relation `var` on **line 11** specifies that any declaration made under the `var` relation in a scope, maps an identifier to its type.

For illustration purposes, when we want to encode a single scope with two variable declarations, `x` of type `INT` and `b` of type `BOOL`, its scope graph would be as shown in Figure 4.7.

```

1 signature
2 constructors
3   Let : string * Exp * Exp -> Exp
4   Var : string                -> Exp
5
6 name-resolution
7   labels
8     P // to denote parent scope
9
10 relations
11   var : string * TYPE

```

Figure 4.6: Statix signature for let-bindings

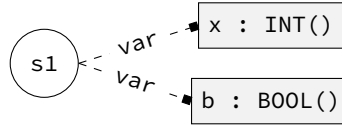


Figure 4.7: A scope graph containing a single scope with two declared variables

In Statix, scopes can be passed around as data. When we are evaluating an expression in our extended language, we now also want to pass the current scope. If the current input term that we are generating constraints for is a let-binding, we want to create a new scope, link it to the previous one, declare the variable in the new scope and evaluate the expression. To generate constraints for a variable expression, we want to query the scope graph and get its type. The Statix rules to reflect this are shown in Figure 4.8.

Figure 4.8 showcases various previously unexplained constructs:

- **Line 4** creates a new scope `s`. This scope is the root scope since it is created once at the start of an application and is not linked to any other scope.
- **Line 7** shows the new signature of the `typeOfExp` functional predicate. Given a scope and an expression, the rules of `typeOfExp` will compute the type of the expression.
- **Line 9-14** gives the typing rule of a let-binding. Given that the let-binding is of form `let x = e1 in e2`, the rule:
 - computes the type of `e1` on line 10;
 - creates a new scope `s_let` on line 11 for the body of the let to evaluate in;
 - declares variable `x` with associated type `τ1` in the newly created scope `s_let`;
 - computes the type of `e2` and this is the result of the rule.
- **Line 16-21** holds the implementation of the variable typing rule. It executes a query with the following properties:
 - It only returns entries in the `var` relation (line 17)
 - It may follow zero or more `P` edge labels to other scopes (line 17);
 - It only returns declarations under the same identifier as `x` (line 18);


```

1 rules
2 applicationOk : Application
3 applicationOk(Application(e)) :- { s T }
4   new s,
5   typeOfExp(s, e) == T.
6
7 typeOfExp : scope * Exp -> TYPE
8 // ... previous rules
9 typeOfExp(s, Let(x, e1, e2)) = T2 :- { s_let T1 }
10   typeOfExp(s, e1) == T1,
11   new s_let,
12   s_let -P-> s,
13   !var[x, T1] in s_let,
14   T2 == typeOfExp(s_let, e2).
15
16 typeOfExp(s, Var(x)) = T :-
17   query var filter P*
18     and { x' :- (x', _) == (x, _) }
19     min $ < P
20     and true
21     in s |-> [(_, (_, T))].

```

Figure 4.8: Statix rules for let-bindings

- It prefers local declarations over declarations for which P edges must be followed (line 19);
- Shadowing according to the shadowing rules of line 19 is enabled (line 20);
- The query starts in the passed scope s (line 21);
- The result may only be one declaration (line 21).

Figure 4.9 shows a possible input and the constructed scope graph after the constraints have been solved.

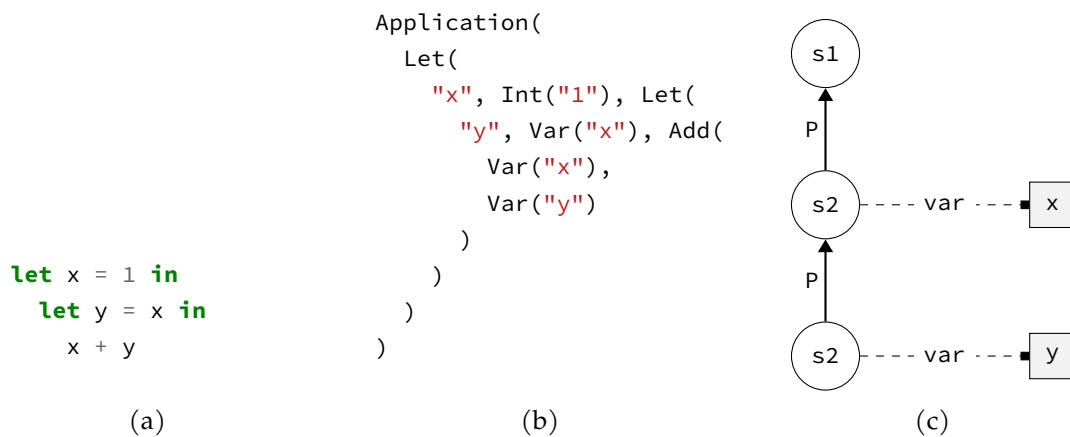


Figure 4.9: Constructed scope graph after the example specification solved its constraints

4.2 Encoding the WebDSL Basics

The WebDSL language adheres to a structure similar to many popular programming languages. A WebDSL application consists of multiple files. At the topmost level in a file, there is a module or *unit* declaration. Within a module, multiple *sections* of *definitions* exist, such as pages, templates, entities and functions. A function consists of consecutive *statements* such as variable assignment (`var n := 2`). At the innermost level, these statements contain *expressions* that form the basis the WebDSL type system.

To define well-typedness of the mentioned constructs, the Statix predicates as shown in Figure 4.10 form the backbone of the WebDSL Statix specification.

```

1 rules
2   projectOk : scope
3   unitOk    : scope * Unit
4   sectionOk : scope * Section
5   defOk     : scope * Definition
6   typeOfExp : scope * Exp -> TYPE

```

Figure 4.10: Predicates that form the basis of the WebDSL Statix specification

4.2.1 Built-in Types and Constant Expressions

Constant expressions such as strings, integers and booleans form the building blocks of more complication constructs. For reasons explained later (see Section 4.6), a built-in type such as string is not declared as `STRING : TYPE` but instead as `BUILTINTYPE : scope * string -> TYPE`, where the instantiation of the string type is as follows: `BUILTINTYPE(s, "String")`.

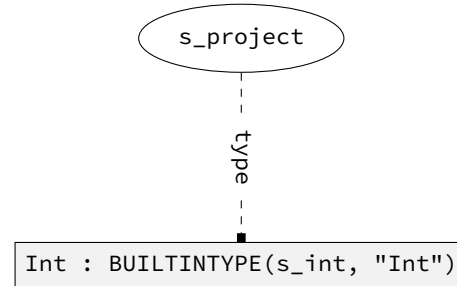
These built-in types are declared in a scope that is reachable from almost every location, the project scope, once per analysis. All WebDSL type declarations are made under the type relation, which associates the human-readable type name with a `TYPE` term: `type : string * TYPE`. The part of the Statix specification to achieve this, and the resulting scope graph are shown in Figure 4.11.

```

1 projectOk(s_project) :-
2   declareTypeBuiltIns(s_project).
3   // ...
4
5 declareTypeBuiltIns : scope
6 declareTypeBuiltIns(s) :-
7   declareType(s, "Int",
8     BUILTINTYPE(new, "Int")).
9   // ...
10
11 declareType : scope * string * TYPE
12 declareType(s, name, t) :-
13   !type[name, t] in s.

```

(a)



(b)

Figure 4.11: Declaring built-in types in the project scope

To retrieve a built-in type when evaluating a constant expression, we need to query the scope graph and resolve the type associated with the string representation. For example, the typing rules of an integer constant are listed in Figure 4.12. The integer typing rule introduces a constraint that the scope graph must contain

```

1 typeOfExp(s, Const(Int(_))) = t :-
2   resolveType(s, "Int") == [(_, (_, t))].
3
4 resolveType : scope * string
5   -> list((path * (string * TYPE)))
6 resolveType(s, name) = ts :-
7   query type filter P*
8     and { t' :- t' == (name, _) }
9     in s |-> ts.

```

Figure 4.12: WebDSL integer constant expression

```

1 typeOfExp(s, Const(StringConst(String(str)))) = t :-
2   resolveType(s, "String") == [(_, (_, t))],
3   stringPartsOk(s, str).
4
5 stringPartsOk maps stringPartOk(*, list(*))
6 stringPartOk : scope * StringPart
7 stringPartOk(s, StringValue(_)).
8 stringPartOk(s, InterpExp(exp)) :- typed(s, exp).
9 stringPartOk(s, InterpValue(InterpSimpleExp(simple_exp))) :- { T }
10  typeOfSimpleExp(s, simple_exp) == T.

```

Figure 4.14: WebDSL string typing rules

a single type declaration associated with "Int" under the type relation. The result of the `resolveType` functional predicate on **line 2** should be a list containing one entry, namely the pair that we declared in Figure 4.11. Other WebDSL constant expressions such as booleans, longs and floats have similar typing rules.

The typing of perhaps the most common constant expression, a string, has an additional condition to be well-typed. Because string interpolation is possible, the constructor of a WebDSL string contains multiple parts that may impose additional constraints. A demonstration of the different interpolated parts is shown in Figure 4.13 and the complete typing rules are shown in Figure 4.14. The parts can be a simple string value which imposes no additional constraints, they can be a complete interpolated expression which requires the expression to be typed, or lastly they can be a “simple” expression which is directly inlineable.

```

1 "Hello world" // value
2 "Hello ~( 1 + 2 )" // exp
3 "Hello ~x.y" // simple exp

```

Figure 4.13: WebDSL string interpolation examples

Now that all the typing rules for constants are implemented, typing rules for unary and binary operators are a step towards more complicated expressions. While it might seem trivial, we might require additional construct functional predicates for determining type compatibility or determining the resulting type of an expression.

4.2.2 Variables

Similar to other imperative languages, WebDSL allows the use of variables to store values. These variables can be defined on multiple levels, such as in the module, within a function or at the top of a page/template definition. Additionally, functions may be embedded in entities, allowing direct access to entity properties as variables without having to prefix it with the `this` keyword.

The basic variable declaration and resolving rules are shown in Figure 4.15. Given a scope `s`, the declaration rule will make a declaration in `s` of variable `x` with associated type `t`.

```

1 declareVar : scope * string * TYPE
2 declareVar(s, x, t) :-
3   !var[x, t] in s,
4   noDuplicateVarDefs(s, x)
5   | error $[A variable named [x] already exists in this scope].
6
7 resolveVar : scope * string -> list((path * (string * TYPE)))
8 resolveVar(s, x) = ps :-
9   query var filter P* /* The filter will be expanded throughout the chapter */
10      and { x' :- x' == (x, _) }
11      min $ < P
12      and true
13      in s |-> ps.

```

Figure 4.15: WebDSL variable declaration and resolving

The implementation of variable typing is similar to the example of let-bindings in Section 4.1.4. One difference between the let-bindings and WebDSL variables is that the introduction of consecutive statements in WebDSL requires a structure that defines declare-before-use semantics, to prevent backwards- or self-references such as shown in Figure 4.16.

```

function f() {
  var a := b;
  var b := b;
}

```

Figure 4.16: WebDSL requires declare-before-use of variables

Figure 4.17 shows how the scope graph is constructed when there are consecutive statements. To catch declare-before-use related errors, a new scope is created for each statement (**line 6 and 7**). When constraints are generated for a constraint (such as on **line 11**), it has access to two scopes. Scope *s* denotes the scope of the current statement. Any scope graph queries will be executed in this scope. Example: the type of this statement is queried starting in scope *s* on **line 12**). Scope *s_{decl}* denotes the scope of the next statement. Any scope graph declarations will be made this scope. Example: a variable declaration is being made in scope *s_{decl}* on **line 14**).

Using this tactic, a statement can never access declarations made by itself or by the next statements, it can only access declarations from previous statements.

An example of how this structure influences the building of scope graphs, a visualization of a function, accompanied by the scope graph of its body is shown in Figure 4.18.

Another difference between the let-binding rules from an earlier example and WebDSL variables is the complexity of the shadowing rules. The WebDSL variable shadowing rules which we reverse-engineered from the current compiler and static analysis implementation, state that the same variable identifier may be used multiple times, but never twice in the “environment”. Such environments are: module scope, entity properties, functions, templates, etc. If a variable reference has multiple declarations in reach, the closest one according to the shadowing rules will be picked. The regular expression that defines the reachability of variables (left out in **line 9** of Figure 4.15) is shown in Figure 4.19.

The edge label *P* as introduced in Figure 4.17 is the edge label used for linking consecutive statements together. The other edge labels such as complicate this regular expression, and will be explained in more

```

P*
F*
(
  (EXTEND? (INHERIT EXTEND?)*
  | (DEF? (IMPORT | IMPORTLIB?))
)

```

```

1 stmtOk : scope * scope * Statement
2
3 stmtsOk : scope * list(Statement)
4 stmtsOk(_, []).
5 stmtsOk(s, [stmt | tail]) :- {s_decl s_next}
6   new s_decl, s_decl -P-> s,
7   new s_next, s_next -P-> s_decl,
8   stmtOk(s, s_decl, stmt),
9   stmtsOk(s_next, tail).
10
11 stmtOk(s, s_decl, VarDecl(x, sort)) :- { t }
12   t == typeOfSort(s, sort),
13   inequalType(t, UNTYPED()) | error $[Unknown type [sort]] @sort,
14   declareVar(s_decl, x, t),
15   @x.type := t.

```

Figure 4.17: WebDSL statements use different scopes for querying and declaring data from the scope graph

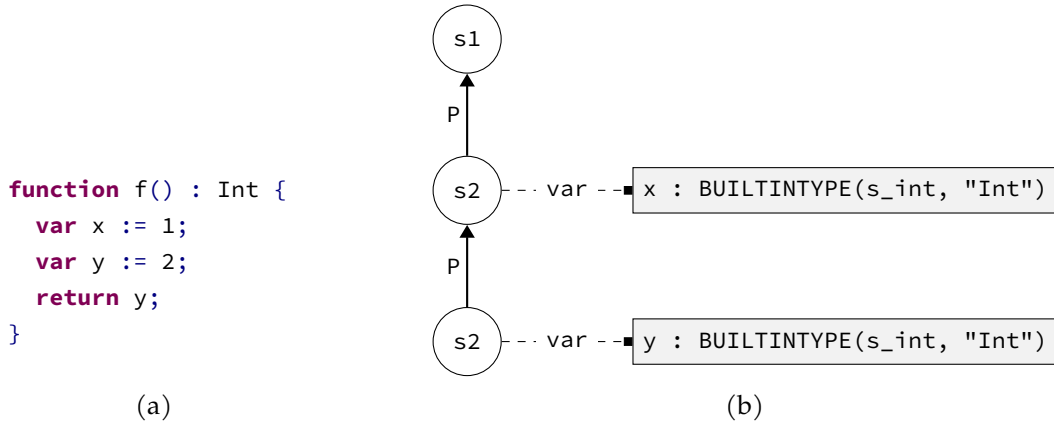


Figure 4.18: Variable declarations example using a separate declaration scopes

details in later sections when their use is discussed.

Figure 4.19 defines what data is reachable from any point in the scope graph, but we also want some restrictions of declarations. The same environment such as function body or an entity definition may never declare the same variable twice. To achieve this, **line 4** of Figure 4.15 uses the helper predicate `noDuplicateVarDefs`. The implementation of this predicate is straight-forward and shown in Figure 4.20. The predicate queries the current scope and checks whether all scopes reachable using only `P` edge labels, results in a list containing only one entry.

4.2.3 Type Compatibility

WebDSL has a notion of type compatibility. For example, the WebDSL superclass of all entities is conveniently called `Entity`. When assigning a value to a variable that requires type `Entity`, passing an instance of a user-defined entity such as `Person` or `Project` also suf-

```

1 noDuplicateVarDefs : scope * string
2 noDuplicateVarDefs(s, x) :-
3   query var filter P*
4     and { x' :- x' == (x, _) }
5     in s |-> [_].

```

Figure 4.20: The same variable identifier may only be declared once in an environment

```

entity Person {}

function f() {
  var e : Entity := Person{}; // all user-defined entities are compatible with Entity
  var d : Date := now(); // now() produces a value of type DateTime
  var p : Person := null; // null is compatible with many types
}

```

Figure 4.21: Examples of type compatibility in WebDSL

```

1 typeCompatible : TYPE * TYPE
2 // By default, two types are not compatible
3 typeCompatible(T1, T2).
4 // Same type is always compatible
5 typeCompatible(T, T).

```

Figure 4.22: WebDSL type compatibility predicate and general rules

fices. In this case, type `Person` is compatible with type `Entity`, but not the other way around. Type compatibility is not limited to entities. For instance, all WebDSL date types (`Date`, `Time`, `DateTime`) are compatible with each other. As a last example, `null` is compatible with many types. The examples given above are shown in Figure 4.21.

To encode the type compatibility as shown in Figure 4.21 in Statix, we need a predicate that tells us, given two types A and B , if type A is compatible with B . The signature and its general rules are shown in Figure 4.22.

With only the basic rules from Figure 4.22, we have created the equality (`==`) from Statix in predicate form. The advantage of listing it like this, is that we can now add rules to make it fit the WebDSL type system. To continue the example of `null` being compatible with every type, we can add the rules shown in Figure 4.23 to achieve this. An example of how to use our new `typeCompatible` predicate is also given on line 8 of Figure 4.23.

Typing of the addition expression

The typing rules for most binary operations such as conjunction is trivial: the resulting value is of boolean type, with the constraint that both operators must be of boolean type. However, the range of values in WebDSL is greater than only natural numbers and booleans. WebDSL supports other numeric types such as `Floats` and `Longs`, as well as string types and multiple subtypes of strings such as `Secret`, `Text` and `WikiText`. The addition operator supports most of these values, and the typing of this operator is not as trivial as boolean conjunction. For example: the addition of two strings results in a string, the addition of a string and an integer results in a string value and the addition of a boolean and a string is not supported.

To calculate the return type of addition, we introduce a functional rule that calculates

```

1 typeOfExp(_, Null()) = NULL().
2 typeCompatible(NULL(), _).
3
4 // example of usage:
5 stmtOk(s, VarDeclInit(x, sort, exp), _) :- { sortType expType }
6   sortType == typeOfSort(s, sort),
7   expType == typeOfExp(s, exp),
8   typeCompatible(expType, sortType)
9   | error $[Expression [exp] is not of type [sort], got type [expType]] @exp,
10  declareVar(s, x, sort),
11  @x.type := t.

```

Figure 4.23: Compatibility of the null expression encoded in Statix

```

1 lubForAdd : TYPE * TYPE -> TYPE
2 lubForAdd(T1, T2) = lubForAddNumeric(T1, T2).
3 lubForAdd(t@BUILTINTYPE("String", _), _) = t.
4 lubForAdd(_, t@BUILTINTYPE("String", _)) = t.
5
6 lubForAddNumeric : TYPE * TYPE -> TYPE
7 lubForAddNumeric(_, _) = UNTYPED().
8 lubForAddNumeric(t@BUILTINTYPE("Int", _), t) = t.
9 lubForAddNumeric(t@BUILTINTYPE("Long", _), t) = t.
10 lubForAddNumeric(t@BUILTINTYPE("Float", _), t) = t.
11 lubForAddNumeric(t@NATIVECLASS("Double", _), t) = t.
12
13 // implicit widening from int to long
14 lubForAddNumeric(BUILTINTYPE("Int", _), t@BUILTINTYPE("Long", _)) = t.
15 lubForAddNumeric(t@BUILTINTYPE("Long", _), BUILTINTYPE("Int", _)) = t.
16
17 // implicit widening from float to double
18 lubForAddNumeric(t@NATIVECLASS("Double", _), BUILTINTYPE("Float", _)) = t.
19 lubForAddNumeric(BUILTINTYPE("Float", _), t@NATIVECLASS("Double", _)) = t.

```

Figure 4.24: Least-upper-bound rules for addition

the least-upper-bound of two types: `lubForAdd : TYPE * TYPE -> TYPE`. The implementation of this rule is given in Figure 4.24. The functional rule `lubForAddNumeric` is reused in other contexts, in particular when generating the constraints for comparison with operators such as greater-than, to check if two types are comparable.

4.2.4 Boolean Logic in Statix

So far, most of the WebDSL's static semantics are expressible in Statix. However, the elegance of the Statix definition is sometimes lost due to code duplication. For example, logical negation and disjunction of predicates are not natively expressible in Statix, and require boilerplate code to function. To tackle this challenge, we introduced a notion of explicit boolean results for predicates that are reusable. The implementation in Statix is shown in Figure 4.25. The figure shows a predicate from before (`typeCompatible : TYPE * TYPE`) now changed to return an explicit result: `typeCompatibleB : TYPE * TYPE -> BOOL`. Additionally, we scope


```

1 signature
2   sorts
3     BOOL    // used as return values of functional rules
4
5   constructors
6     TRUE : BOOL
7     FALSE : BOOL
8
9   rules
10    // return a TRUE() or FALSE() value instead of failing/passing constraint
11    typeCompatibleB : TYPE * TYPE -> BOOL
12
13    // scope this explicit results in a predicate to avoid having to work
14    // with boolean computation everywhere
15    typeCompatible : TYPE * TYPE
16    typeCompatible(T1, T2) :- typeCompatibleB(T1, T2) == TRUE() .

```

Figure 4.25: Boolean computation results in Statix

this boolean result in a predicate `typeCompatible(T1, T2) :- typeCompatibleB(T1, T2) == TRUE()` ., such that existing references can be left unchanged.

As hinted before, the explicit return values of functional rules open up new possibilities for expressing constraints. One instance of where this is necessary, is expressing the semantics of an equality check in WebDSL. For the expression `A == B` to type check, the types have to be compatible. The naive implementation would be to define the constraint `typeCompatible(T_A, T_B)`. However, type compatibility is not symmetrical while the equality check should be: $A == B \iff B == A$. An example of type compatibility not being symmetrical is when dealing with entity inheritance (see Section 4.3.1). To properly define the static semantics for the equality expression in Statix, we need the newly defined boolean computation rules. The result is shown in Figure 4.26.

4.2.5 Entities and Properties

Entities form the basis of the type system and data structure in a WebDSL application. Using Hibernate as an object-relational mapping (ORM) tool, instances of entities can be persisted without explicit communication with a database management system. Entities typically have multiple properties which values are persisted, and functions that can be called and will be executed in the scope of the instantiated entity. Entity properties and entity functions together form the entity body declarations.

In the WebDSL type system, entities are declared in the scope of the module they are defined in. An entity is a type in the WebDSL type system, similar to built-in types such as `String` and `Int`. The Statix code to declare entities is shown in Figure 4.27 and an example of a simple program with entity definition plus its scope graph is shown in Figure 4.28.

The `declareType` and `resolveType` rules as introduced in Figure 4.11 need to be updated to work as intended for resolving and declaring entities. To prevent duplicate entity definitions, the `declareType` rule is extended with one additional rule as shown in Figure 4.29. **Line 4** was added to `declareType`, to make sure when you declare a new type or entity, its name is unique.

In addition to the added constraint to the `declareType` rule, we added an optional `DEF` edge label that may be followed when querying the scope graph for a type (**line 9** of Figure 4.29). The `DEF` (short for definition) is used to link the scope of top-level elements, such as entities and functions, to the module scope. This can be seen in **line 13** of Figure 4.27.


```

1  or  : BOOL * BOOL
2  orB : BOOL * BOOL -> BOOL
3
4  or(b1, b2) :- orB(b1, b2) == TRUE().
5
6  orB(_, _) = FALSE().
7  orB(TRUE(), _) = TRUE().
8  orB(FALSE(), TRUE()) = TRUE().
9
10 // (e1 == e2)
11 typeOfExp(s, Eq(e1, e2)) = t :- { T1 T2 }
12   t == bool(s),
13   typeOfExp(s, e1) == T1,
14   typeOfExp(s, e2) == T2,
15   or(
16     typeCompatibleB(T1, T2),
17     typeCompatibleB(T2, T1)
18   ).

```

Figure 4.26: Using boolean computation results in Statix for the equality expression

```

1 signature
2 constructors
3   // an entity constructor has two subterms:
4   // - the entity name
5   // - the scope of the entity where all the properties and
6   //   functions are declared
7   ENTITY : string * scope -> TYPE
8
9 rules
10 defOk(s_module, EntityNoSuper(entity_name, body)) :- { s_entity }
11   // a new scope for the entity is created and linked to the module scope
12   // using the 'DEF' (for definition) edge label
13   new s_entity, s_entity -DEF-> s_module,
14
15   // the new entity is declared as type in the module scope
16   declareType(s_module, entity_name, ENTITY(entity_name, s_entity)),
17
18   // finally a helper rule is called that properly handles
19   // the entity body definitions (properties, functions, etc.)
20   declEntityBody(s_entity, entity_name, body).

```

Figure 4.27: The Statix rules for declaring entities

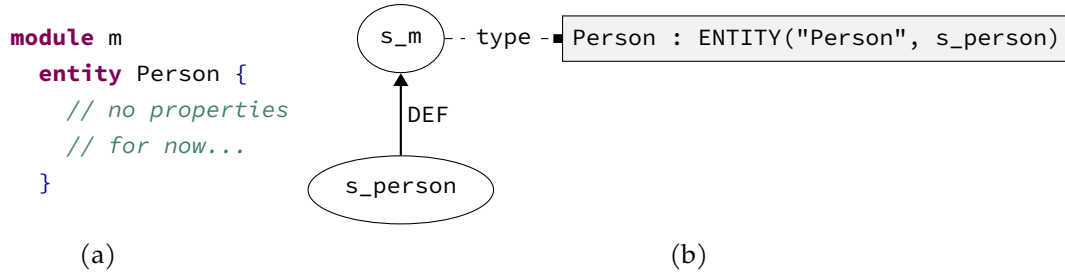


Figure 4.28: An example of entity definition in WebDSL

```

1 declareType : scope * string * TYPE
2 declareType(s, name, t) :-
3   !type[name, t] in s,
4   resolveType(s, name) == [(_, (_, t))]
5   | error $[Type [name] is defined multiple times] @name.
6
7 resolveType : scope * string -> list((path * (string * TYPE)))
8 resolveType(s, name) = typesOf(ts) :-
9   query type filter P* DEF? // resolving a type may
10                                // optionally follow DEF edge label
11   and { t' :- t' == (name, _) }
12   in s |-> ts.

```

Figure 4.29: declareType now shows an error when two types with the same name are declared and resolveType may optionally follow a DEF edge label

So far, there has been no reason to query for types inside the entity body because we have always worked with empty entities. In practice, entities are filled with properties and functions. **Line 20** of Figure 4.27 calls the `declEntityBody` predicate, of which the implementation is shown in Figure 4.30 and an example of an entity definition with two properties is shown in Figure 4.31.

Entity properties are declared under the variable relation inside the entity scope, such that functions inside entities can reference their own properties without using the `this` prefix. The `this` construct is supported, but not necessary. Declaring properties in this way, allows us to reuse the already existing rules such as those against duplicate definition, without duplicating the code for another relation.

When instantiating an entity, the properties declared in the entity body may be given a value in the instantiation expression. To express this in Statix, an entity instantiation first retrieves the scope of the entity. If the scope cannot be retrieved, it means that the entity is unknown at the position of the expression, so either the entity was never declared or it is not imported correctly. Secondly, all instantiated properties must be declared under the `var` relation of the entity scope. An example of the declaration and scope graph of an entity declaration is shown in Figure 4.31. A part of the Statix rules for instantiating entities is shown in Figure 4.32.

Even though the concepts, rules and approach mentioned in this subsection are present in the Statix specification of WebDSL, we had to simplify the examples and shown Statix rules to hide the extra complexity added concepts such as inheritance, property annotations and type extension. Those concepts will be explained in detail in section Section 4.3 and Section 4.6.

```

1 declEntityBody maps declEntityBodyDeclaration(*, *, list(*))
2 declEntityBodyDeclaration : scope * string * EntityBodyDeclaration
3
4 // entity property
5 declEntityBodyDeclaration(s, ent,
6   Property(x, propkind, sort, PropAnnos(annos))) :- { sortType }
7
8 // resolve the type of the property
9 sortType == typeOfSort(s, sort),
10
11 // there are some restrictions on property types
12 sortType != UNTYPED()
13 | error $[Cannot resolve type [sort]] @sort,
14 sortType != VOID()
15 | error $[Property type 'Void' not allowed] @sort,
16 sortType != REF(_)
17 | error $[Reference type is not allowed in property] @sort,
18 isValidTypeForPropKind(propkind, sort, sortType),
19
20 // declare the property as variable in the entity scope
21 declareVar(s, x, sortType),
22
23 // use a helper predicate to check for the uniqueness of
24 // the property name
25 resolveLocalProperty(s, x) == [_]
26 | error $[Property [x] of entity [ent] is defined multiple times] @x.

```

Figure 4.30: Statix rules for declaring the entity body

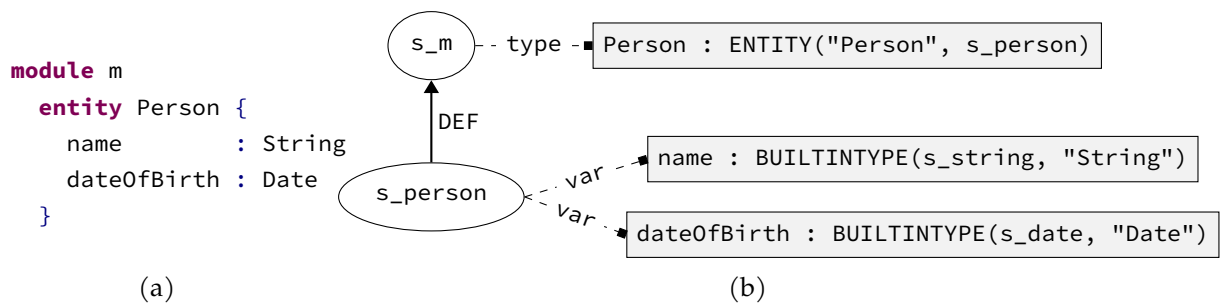


Figure 4.31: An example of an entity definition with multiple properties

```

1 typeOfExp(s, ObjectCreation(x, prop_assignments)) = e :-
2   definedType(s, x) == e,
3   e == ENTITY(_, _),
4   propAssignmentsOk(s, e, prop_assignments).
5
6 propAssignmentsOk maps propAssignmentOk(*, *, list(*))
7 propAssignmentOk : scope * TYPE * PropAssignment
8 propAssignmentOk(s, ent@ENTITY(e, s_ent),
9   PropAssignment(x, exp)) :- { propType expType }
10  typeOfProperty(s, ent, x) == propType,
11  typeOfExp(s, exp) == expType,
12  typeCompatible(expType, propType).

```

Figure 4.32: Statix rules for instantiating an entity

```

1 signature
2 constructors
3   PAGE      : string * list(TYPE) -> TYPE
4   TEMPLATE : string * list(TYPE) -> TYPE
5
6 relations
7   page      : string * TYPE
8   template  : string * TYPE

```

Figure 4.33: Statix signature for pages and templates

4.2.6 Pages and Templates

The user-interface of a WebDSL application is built out of *pages* and *templates*. A page defines a path that can be requested by the browser while a template is a reusable component that can be part of a page or nested in other templates.

The name of a page must be unique, while a template can be defined multiple times for different argument types (*overloading*), but never multiple times for the same argument types. The Statix rules to implement these checks can be found in Figure 4.34 and an example of a module with a page and a template definition is shown in Figure 4.35. In the latter image, the argument type of template *t* is shortened to `String`, instead of its full version `BUILTINTYPE(s_string, "String")`.

Type-checking a page reference is easier than the that of a template, since a page definition cannot be overloaded. In order for a page reference to be well-typed, the page must be defined exactly once, and the types of the passed arguments must be compatible with the parameter types of the page. The Statix rules that ticks those boxes is shown in Figure 4.36. The resolving of templates is similar to that of functions and will be explained later in Section 4.4.

The body of templates and pages consist of so-called *Template elements*. The simplest template element is simply a text to be printed on the page. Next to plain text, hyperlinks to other pages can be created using the `navigate` element. If we take Figure 4.35 as basis, an example of a valid `navigate` call would be `navigate p() { "Go to p" }`. To type-check this, the code from Figure 4.36 can be used. Other examples of template elements are forms, nested template calls, and at the top of a template, variables can be initialized, followed by a block of computational statements that get executed when the template is being loaded.

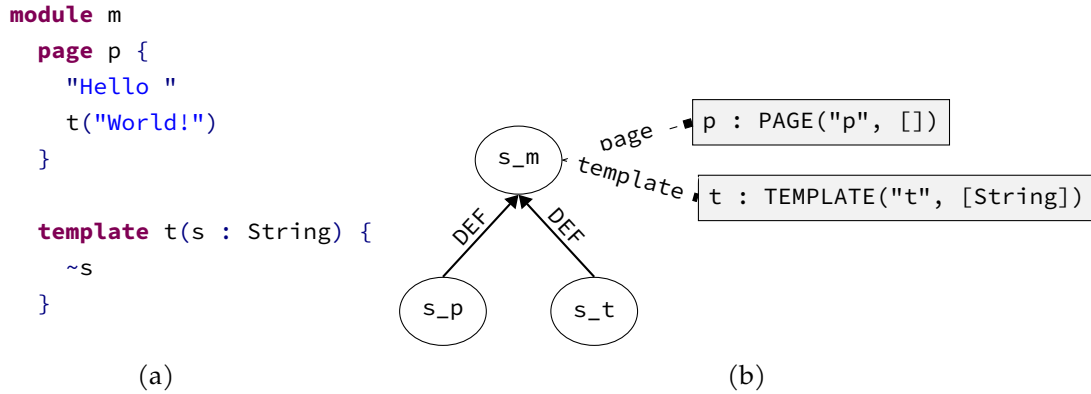
Apart from regular templates, WebDSL also has a notion of Ajax templates. Ajax tem-

```

1 declarePage : scope * string * list(TYPE)
2 declarePage(s, p, ts) :-
3   !page[p, PAGE(p, ts)] in s,
4   resolveTemplate(s, p) == []
5   | error $[Multiple page/template definitions with name [p]] @p,
6   resolvePage(s, p) == []
7   | error $[Multiple page/template definitions with name [p]] @p.
8
9 declareTemplate : scope * string * list(TYPE)
10 declareTemplate(s, t, ts) :-
11   !template[t, TEMPLATE(t, ts)] in s,
12   resolvePage(s, t) == []
13   | error $[Multiple page/template definitions with name [t]] @t,
14   filterTemplateResultsArgs(resolveTemplate(s, t), ts) == []
15   | error $[Multiple page/template definitions with name [t] and argument types [ts]] @t.

```

Figure 4.34: Statix rules for declaring WebDSL pages and templates

Figure 4.35: An example of a module with a page *p* and a template *t*

```

1 pageCallOk : scope * string * list(Exp)
2 pageCallOk(s, p, args) :- {argTypes ts}
3   pageType(s, p) == PAGE(_, ts)
4   | error $[There is no page with signature [p]] @p,
5   argTypes == typesOfExps(s, args),
6   typesCompatible(argTypes, ts)
7   | error $[Given argument types not compatible with page definition] @args.
8
9 // root page is always accessible from all locations
10 pageCallOk(_, "root", []).

```

Figure 4.36: Statix rules for type-checking a page reference

plates can be used as building blocks of the user interface, just like regular templates. Additionally, Ajax templates also have a possibility of being replaced on a rendered page without reloading the whole page, for example to refresh the results of a poll on a page. This addition makes Ajax templates useful for more interactive and modern web applications.

Certain action code such as the `replace` and `refresh` statements are only supposed to work on Ajax templates, and not on regular templates. To this end we need to differentiate between them in the scope graph. We have chosen the most trivial way of implementing this, namely adding an additional argument in the type constructor of a template: `TEMPLATE : string * list(TYPE) * BOOL -> TYPE`. The last boolean argument indicates whether the template is an Ajax template or not. When resolving templates we can now resolve only Ajax templates by adding a `TRUE()` to the filter statement of the query.

4.2.7 Functions

In WebDSL, a function is a sequence of statements that perform some sort of computation and can return a value. The type of the return value must be stated in the function header and is part of the signature. The implementation of the declaration and resolving of functions is similar to that of templates, as explained in Section 4.2.6, and therefore will not be repeated here.

An additional characteristic of functions that is similar to templates, is the use of parameters. The parameters with their corresponding types have to be declared statically. The parameters are readable from the function body, but never writable or overridable by a local variable. Additionally, the name of parameters may shadow the name of definitions outside the function such as entity properties or global definitions. To enforce these constraints, we introduce a new edge label `F` for embedding the function scope in their surrounding scope, which is either global or within an entity. Using this new edge label, the shadowing rules can be adjusted to properly check the listed semantics. The result is shown in Figure 4.37 and an example of a WebDSL snippet with the resulting scope graph is shown in Figure 4.38. In the example, parameter `x` of function `f` shadows the globally declared `x`.

Apart from globally declared functions, functions may also be part of an entity. In this case, functions can be called similar to how entity properties are referenced. Lastly, entity functions may have the `static` annotation, which is similar to static class functions in the Java programming language. Static functions may be called without having an instantiated entity.

4.2.8 Access Control

When developing any application that will be used in practice, access control is an important part of the system. It controls which user is allowed to see what data, what actions can be executed. Generally, this is implemented through a log-in system where different user accounts are given different rights. In all popular programming languages, developing a system access control is the responsibility of the developer, either through manual coding or using frameworks and libraries. In WebDSL however, access control is embedded in the language and all pages are protected by default.

Concretely, the developer is able to declare what entity represents a user in the system, and what data the user needs to show to log in. In the rest of the WebDSL code, the globally available security context is extended with the properties `principal` which references the logged-in user, and `loggedIn` which is true if the user has logged in. If the developer has not specified what entity represents a user, the security context is available but does not have these properties. An example of WebDSL code with resulting scope graph is shown in Figure 4.39 and the Statix rules used to achieve this are shown in Figure 4.40. In the code for

```

1 functionOk : scope * Function
2 functionOk(s_outer,
3     Function(name, FormalArgs(args), OptSortSome(returnSort), Block(stmts)))
4     :- { argTypes returnType s_function s_body }
5
6 // embed the function scope with edge label F
7 new s_function, s_function -F-> s_outer,
8
9 // declare parameters in function
10 argTypes == typesOfArgs(s_outer, args),
11 declareParameters(s_function, zipArgTypes(args, argTypes)),
12
13 // create the function body and generate constraints
14 new s_body, s_body -P-> s_function,
15 stmtsOk(s_body, stmts, returnType),
16
17 // declare the function in the outer scope
18 returnType == typeOfSort(s_outer, returnSort),
19 declFunction(s_outer, name, argTypes, returnType).
20
21 // resolve variables via P and F edges
22 resolveVar(s, x) = ps :-
23     query var filter P* F*
24         and { x' :- x' == (x, _) }
25         min $ < P, $ < F,
26             P < F
27         and true
28         in s |-> ps.
29
30 // a definition is only duplicate in a line of P edges
31 noDuplicateVarDefs : scope * string
32 noDuplicateVarDefs(s, x) :-
33     query var filter P*
34         and { x' :- x' == (x, _) }
35         in s |-> [_].

```

Figure 4.37: Statix rules for function parameters and variable shadowing

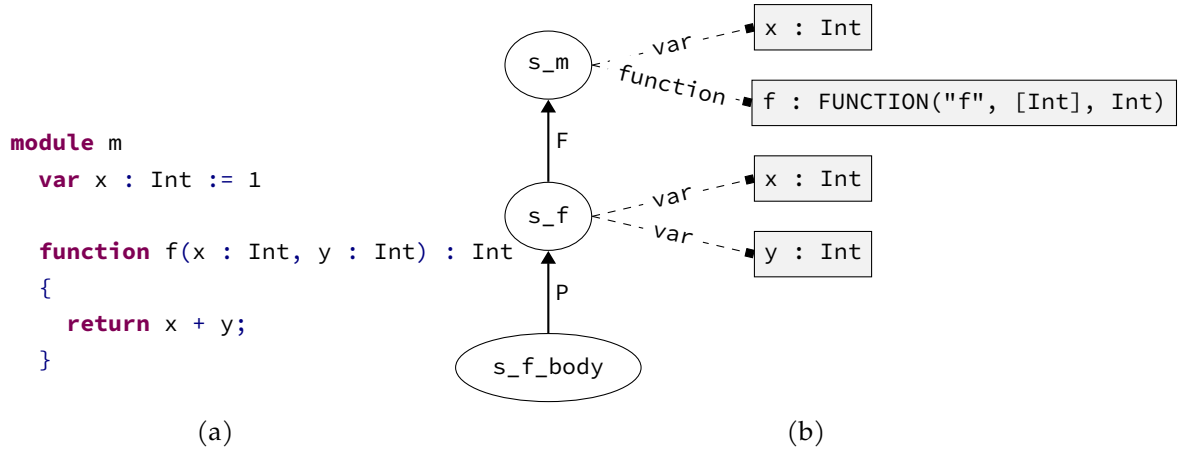


Figure 4.38: An example of a function with parameters

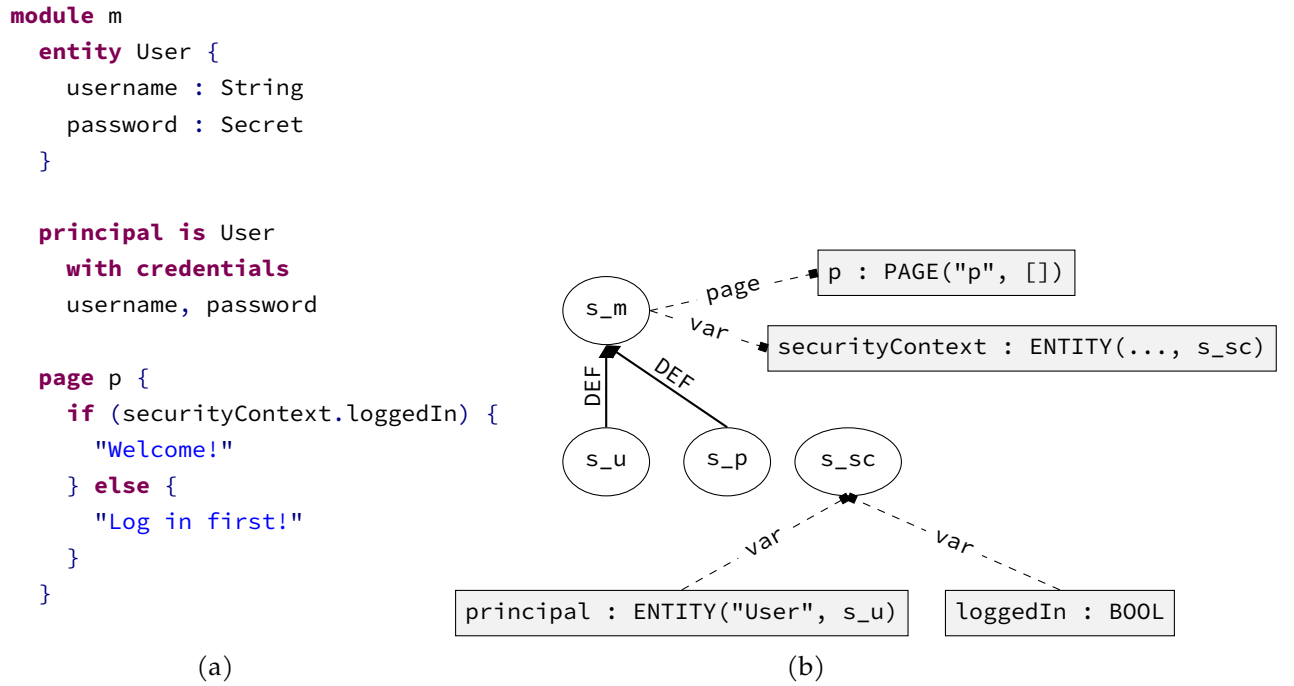


Figure 4.39: An example of access control in WebDSL. The entries related to entity User are omitted for brevity

extending the security context with two additional properties, the same mechanics are used as for entity and built-in type extension. An explanation can be found later in Section 4.6.

4.3 Advanced Entity Features

4.3.1 Inheritance

Linking the Scopes

The implementation of inheritance requires the scope of the sub- and super-entity to be connected such that Statix queries can resolve to declarations from the super-entity when necessary. To achieve this, we introduce an edge label `INHERIT` as shown in Figure 4.41.

First, the super-entity referred to in the declaration must refer to an existing entity in the scope graph. Secondly, the new scope belonging to the sub-entity `s_entity` is linked to the scope of the super class `s_super` via an `INHERIT` edge. Finally, some additional constraints


```

1 principalDefOk : scope * string * list(string)
2 principalDefOk(s, ent, properties) :-
3   { s_ent entityName credentialTypes t }
4   definedType(s, ent) == t@ENTITY(entityName, s_ent),
5   principalPropertyTypes(s_ent, properties, ent) == credentialTypes,
6   compatibleCredentialTypes(properties, credentialTypes),
7   declSecurityContext(s, t, credentialTypes).
8
9 compatibleCredentialTypes maps compatibleCredentialType(list(*), list(*))
10 compatibleCredentialType : string * TYPE
11 compatibleCredentialType(x, s) :-
12   isStringCompatibleType(s).
13
14 declSecurityContext : scope * TYPE * list(TYPE)
15 declSecurityContext(s, principalType, credentialTypes) :-
16   { s_extend_security_context }
17   new s_extend_security_context,
18   declProperty(s_extend_security_context, "securityContext"
19     , "principal", principalType),
20   declProperty(s_extend_security_context, "securityContext"
21     , "loggedIn", bool(s)),
22   declareExtendScope(s, "securityContext", s_extend_security_context),
23   extendScopes(resolveExtendScope(s, "securityContext")
24     , s_extend_security_context).

```

Figure 4.40: Statix rules for declaring the access control principal

```

1 signature
2   name-resolution
3   labels
4     INHERIT // inherit edge label for subclasses
5
6 rules
7   defOk(s_global, Entity(x, super, bodydecs)) :- {s_entity super' s_super}
8     resolveEntity(s_global, super) == [(_, (super', ENTITY(_, s_super)))],
9     new s_entity, s_entity -INHERIT-> s_super,
10    noCircularInheritance(s_entity),
11    declEntity(s_global, s_entity, x, bodydecs),
12    @super.ref := super'.

```

Figure 4.41: Entity inheritance Statix rules

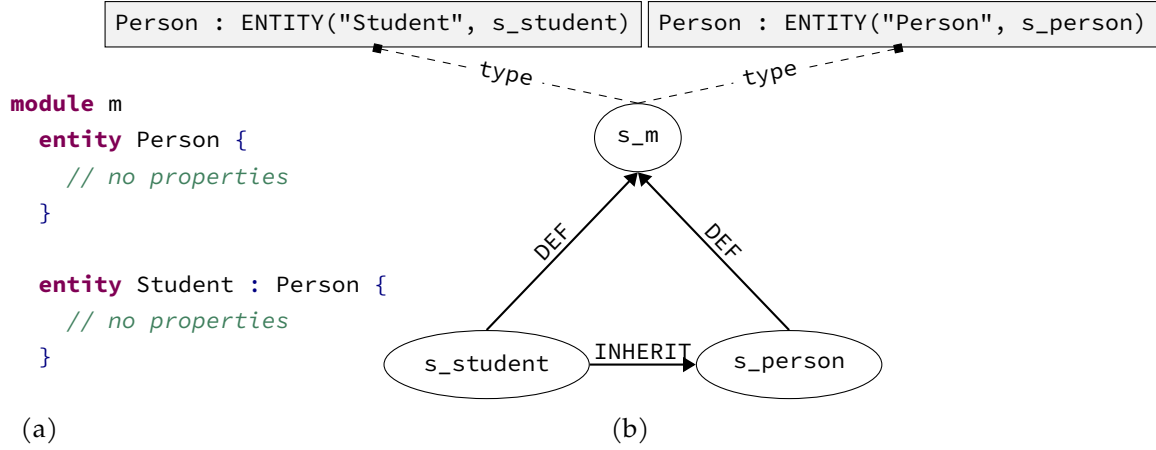


Figure 4.42: An example of entity definition in WebDSL

Figure 4.43: The query that specifies what variables can be resolved, updated to reflect entity inheritance

are generated to make sure no circular inheritance exists and constraints for the entity body declarations of the sub-entity are generated.

The variable resolving query as listed in Figure 4.43 reflects the addition of the `INHERIT` label. The addition of `INHERIT*` in the query filter makes all variables declared in ancestors reachable, but the shadowing rule as declared after the `min` keyword ensures correct shadowing behavior, namely that local variables are preferred over variables defined in ancestors.

Overwriting Functions

Generally, defining two functions with the same name and same argument types is not allowed in WebDSL. Entity functions are an exception to this such that entity function definitions shadow global function definitions. With the introduction of inheritance there comes another exception, namely that sub-entities are allowed to override function definitions of their ancestors.

Previously, the resolving of entity functions was done using a query that resolves within the entity scope only. With the introduction of entity inheritance, the path well-formedness over edge labels was changed such that functions from ancestors are also in scope. Changing filter `e` to filter `INHERIT*` accomplishes this. Both the previous and resulting queries are shown in Figure 4.44.

This query definition is adequate when sub-entities do not override functions. When a sub-entity does define a function that is already defined in one of its ancestors, resolving the entity function gives two results while the desired outcome is only one result, namely the overridden function defined in the sub-entity. To tackle this challenge, we defined a Statix anonymous shadowing rule combined with a label order. This ensures that when two functions with the same name and argument types exist, only the most specific (i.e. the least inheritance edges) is returned. This is implemented as shown in Figure 4.45.

Entity Type Compatibility

A perk of having the notion of inheritance in the WebDSL language, is that it allows for better abstraction and less code duplication. An example of this is a function definition, where the argument type is an entity. This function can be called with an argument of the entity type, or one of its sub-entities. To know if the given type is compatible with the required type, we require a predicate that defines this compatibility. We have created such

```

1  // previously (local entity scope only)
2  resolveEntityFunction(s, x) = ps :-
3      query function filter e
4          and { x' :- x' == (x, _) }
5          min
6          in s |-> ps.
7
8  // new (allow resolving to ancestors)
9  resolveEntityFunction(s, x) = ps :-
10     query function filter INHERIT*
11         and { x' :- x' == (x, _) }
12         min /* */
13         in s |-> ps.

```

Figure 4.44: Statix rules for allowing entity function calls to resolve to definitions in their ancestors

```

1  resolveEntityFunction(s, x) = ps :-
2      query function filter INHERIT*
3          and { x' :- x' == (x, _) }
4          /* prioritize local scope over inheritance */
5          min $ < INHERIT
6          /* shadow when function name and argument types match */
7          and {
8              (f, FUNCTION(args, _, _)),
9              (f, FUNCTION(args, _, _))
10         }
11         in s |-> ps.

```

Figure 4.45: Statix rules for resolving entity functions that allow overriding

Figure 4.46: Statix rules for entity type compatibility that support inheritance

a predicate while implementing general type compatibility in Section 4.2.3, in the form of `typeCompatibleB : TYPE * TYPE -> BOOL`.

With the addition of entity inheritance, we need to expand this definition. To this end, we added the rules as shown in listing Figure 4.46. Given two entity scopes, the `inherits(s_sub, s_super)` predicate returns true when the query has one result. The query in the `inherits` rule requests all paths from scope `s_sub` to scope `s_super` consisting of only `INHERIT` edges. Such a path exists if and only if the entity belonging to scope `s_sub` inherits the entity belonging to `s_super`. An example of a scope graph with entity inheritance is shown in Figure 4.42.

4.3.2 Property Annotations

So far, the Statix specification can validate entities, their properties and their functions. Since the goal is to never manually touch the database specification, we would like entity properties to be more expressive by for example specifying default values, or put a constraint on the possible values of a property. WebDSL uses *property annotations* for this. Figure 4.47 shows an WebDSL code of an entity with properties that have annotations.

Figure 4.47: Examples of entity property annotations

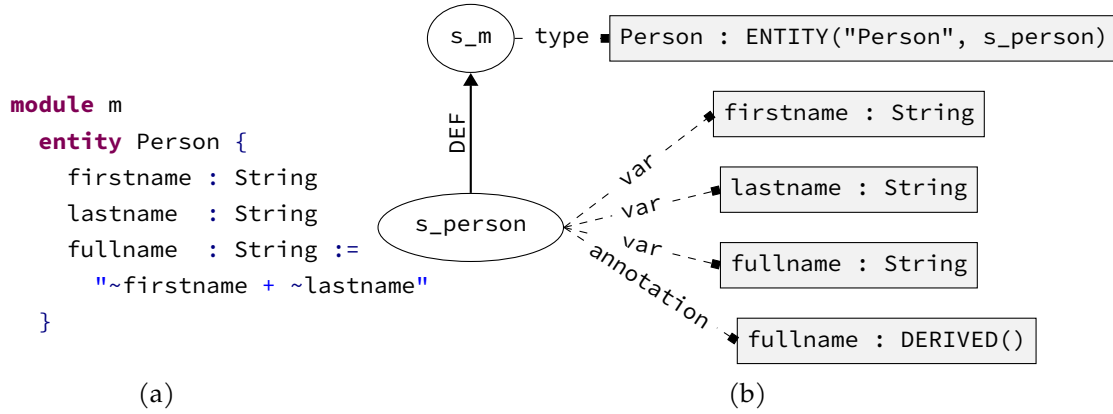


Figure 4.48: An example of a derived property in an entity

Many property annotations do not influence the scope graph. An example of this is the `default = <exp>` annotation, where Statix only needs to check whether the given expression is compatible with the property type. For the `length = <exp>` annotation, the same holds, except that the expression must now have type `Int`.

An interesting property to point out is the derived property, as shown in for property `fullname` of the `Teacher` entity in Figure 4.47. While this is not strictly an annotation, it does change something in the scope graph. A derived value can be calculated from other properties of the entity and does not have to be stored. Its value can also not be changed directly. The latter property is something we need to store in the scope graph, such that we can give an error when developers attempt to assign a value directly to a derived property. For this, a new relation is introduced in Statix, which allows us to declare annotations on properties in the scope of an entity. An example of this is shown in Figure 4.48. When assigning a variable, the left-hand side of the assignment can now be checked for mutability. The implementation of these checks whether the entity property that is referenced on the left-hand side has the `DERIVED()` annotation. To prevent code duplication, we chose to re-use the `DERIVED()` property for function parameters, which can only be referenced but never changed in the function body.

Another interesting annotation to mention, is the `inverse = <var>` annotation as shown for the `courses` property of the `Teacher` entity in Figure 4.47. The inverse annotation is introduced to prevent data duplication in the database. To continue with the example of `Teacher` and `Course` of Figure 4.47, the `Course` table saves the corresponding teacher, and when a teacher is fetched from the database, the `courses` property is instantiated according to the data in the `Course` table. When specifying `inverse=teacher`, the Statix specification has to validate that the entity mentioned in the property type (`Course` in this case) has the `teacher` property, and that the type of that `teacher` property is equal to the type of the entity that the inverse annotation was declared in (`Teacher` in this example). To prevent a situation where none of the two entities is responsible for saving the data, a double inverse annotation is not allowed. To enforce this, another constraint has to be added, namely that the `teacher` property of `Course` does not have an inverse annotation. The only way to reliably check this is to save the annotation `INVERSE()` in the scope graph.

4.3.3 Entity Hierarchy and the Name Property

Similar to `Object` in the Java programming language, WebDSL also has a root of the entity hierarchy, namely `Entity`. If a defined entity does not explicitly inherit from another entity,

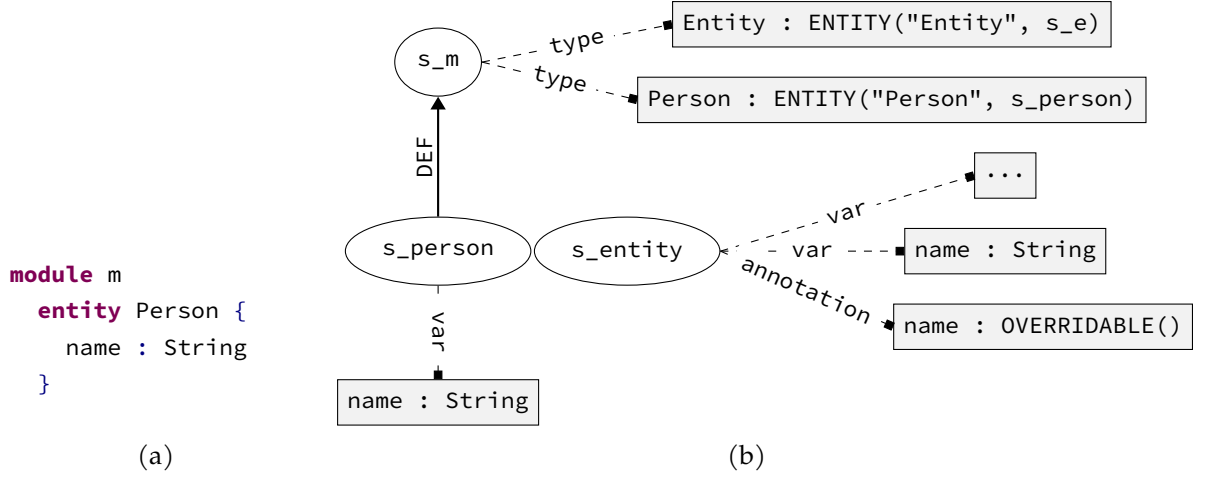
Figure 4.49: An example of overriding the `name` property

Figure 4.50: Statix rules for overridable entity properties

Figure 4.51: An example of defining overloaded templates in WebDSL

it will automatically inherit from `Entity`. This built-in superclass is convenient to store properties that all entities will have out of the box, such as the property `id` of type `UUID` and the property `created` of type `DateTime`. User-defined entities are not allowed to redefine such properties, but they may edit the values. An exception to this rule is the property `name` of type `String` that all entities have by default, but may be overridden once by sub-entities.

To achieve the overridability, we can re-use the property annotations in the scope graph as explained in previous section. The `name` property of the built-in entity `Entity` gets a `OVERRIDABLE()` annotation declared in the scope of `Entity`, and when attempting to catch duplicate property definitions, we must discard properties from parents that have the `OVERRIDABLE()` property. An example of the overriding is shown in Figure 4.49 and the Statix code to discard overridable properties is shown in Figure 4.50. In the Statix rules, there are now two predicates to prevent duplicates: one for within the entity, and another one to check inherited properties. The difference is rules allow for better error messages, but most importantly allows us to only discard properties with the `OVERRIDABLE()` annotation from inherited properties. At the point of writing this thesis, the built-in `name` property is the only use case for the `OVERRIDABLE()` annotation and it is not possible to mark a property as overridable by code.

4.4 Function and Template Overloading

Function overloading is the practice of defining multiple functions with the same name, that differ in argument types or in the amount of arguments. Before running the WebDSL program, the compiler determines what instance of the function will be run, based on the types of the function call arguments. WebDSL supports overloading for both functions and templates, and their implementation in the static analysis is the same. The concepts and solutions explained in this section are therefore applicable for both. An example of template overloading is shown in Figure 4.51. In the rest of this section we will talk about functions, but the concept is exactly the same for templates.

Overloading complicates the static analysis in two ways. The first and easiest to implement is that functions may now be defined multiple times with the same name, as long as the amount of arguments and argument types do not exactly match. The Statix rules that

Figure 4.52: Statix rules for allowing overloaded function definitions

```
module m

  entity Animal {
    name : String
  }
  entity Cat : Animal {
    breed : String
  }

  template description(a : Animal) {
    "~a.name"
  }
  template description(c : Cat) {
    "~c.name (Breed: ~c.breed)"
  }

  page p {
    var a := Animal{ name := "Alice" }
    var c := Cat{ name := "Charlie", breed := "Sphynx" }

    description(a) // will output "Alice"
    description(c) // wil output "Charlie (Breed: Sphynx)"
  }
```

Figure 4.53: An example of referencing overloaded templates in WebDSL

achieve this are shown in Figure 4.52. The essence of the rules is that all functions with the relevant name are retrieved, and the declarations with argument types exactly matching with the relevant types are counted. The resulting number should be 1, namely the newly declared function.

Now that the static analysis allows for overloaded functions and templates to be defined, the code that type checks function calls and template calls should be updated to reflect the new changes. The semantics of resolving the correct overloaded function or template are listed below. A practical example of how these rules work is shown in Figure 4.53.

1. Retrieve all function signatures with the matching name from the scope graph
2. Filter the result to end up with function signatures with matching arity (amount of arguments) and compatible argument types.
3. If the filtered result is exactly one signature: this is the function that will be called
4. If the filtered result is more than one signature: choose the signature with the "most specific" argument types:
 - If there are exactly matching types, always choose this one.
 - Otherwise; count the amount of `INHERIT` edges that have to be taken from the given expression types to the function argument types, and choose the signature with the least total edges taken.

Figure 4.54: Statix rules for resolving overloaded function calls

The implementation of resolving the correctly overloaded function according to the listed semantics is shown in Figure 4.54. The essence of the semantics is encoded in the Statix rules, but the brevity and elegance is lost due to many helper predicates being required to transform the data into the correct forms to perform the queries that calculate the inheritance edges, and filter the function signatures accordingly. Note that the `typeOfFunctionCall` predicate is not specific to functions, because the signature requires a string and a list of expressions, which causes the predicate to be re-usable for resolving template calls.

4.5 Placeholders, Actions and Submitting Forms

Forms are the basis of most information systems on the web. WebDSL has linguistically integrated forms through inputs and actions. The static analysis of referencing actions and placeholders is unlike variable referencing, due to their scoping semantics. Actions and placeholders are defined in pages or templates and may be referred to in the rest of the page or function body. Unlike variables, actions and placeholders do not follow the declare-before-use principle (explained in Section 4.2.2). Actions and placeholders may be defined anywhere in the body, and referenced from anywhere in the body, but there is a twist. Inside the action body, the statements and expressions may reference variables defined in the template or page body, and thus the scope of the action must be linked to the scope of the template or page in some way.

Summarized, actions and placeholders cannot be declared in the scope they are defined in, because that scope follows a declare-before-use regime, but the body of an action must have access to the scope it was declared in. As a solution to this challenge, the Statix rules for declaring a page or template were altered to not only create a scope for its body, but create an additional scope where placeholders and actions will be declared. This scope is passed along such that variables are queryable using the regular function body scope, and querying placeholders and actions is available using the additional scope. An example of an action and its representation in the scope graph is shown in Figure 4.55.

The updated Statix rules for type checking a page or template definition, and defining actions is shown in Figure 4.56. The Statix rules contain a separate predicate `templateActionOk` where the last boolean parameter indicates whether the action should be declared or not, and this is passed to `optionallyDeclareTemplate`. The declaration is optional because it is possible to create a form with a submit button that defines an action inline; an anonymous action in some sense. The anonymous inline action definition can now be type-checked by re-using the `templateActionOk` predicate.

The Statix rules which implement the placeholders and actions code behave correctly because of Statix' constraint scheduling algorithm as first described by in the original Statix paper (Antwerpen, Bach Poulsen, et al. 2018) and further improved by Rouvoet et al. (Rouvoet et al. 2020).

4.6 Type Extension

WebDSL allows the developer to define types such as entities or use existing built-in types such as strings and dates. Often times, when working on a new functionality for existing applications, developers want to extend existing entities with new properties or functions. To this end, WebDSL allows for entity extension across modules. In addition to entities, WebDSL also allows extension of built-in types. The built-in types can not be extended with arbitrary properties or functions. Since the WebDSL built-in types are based on Java types,

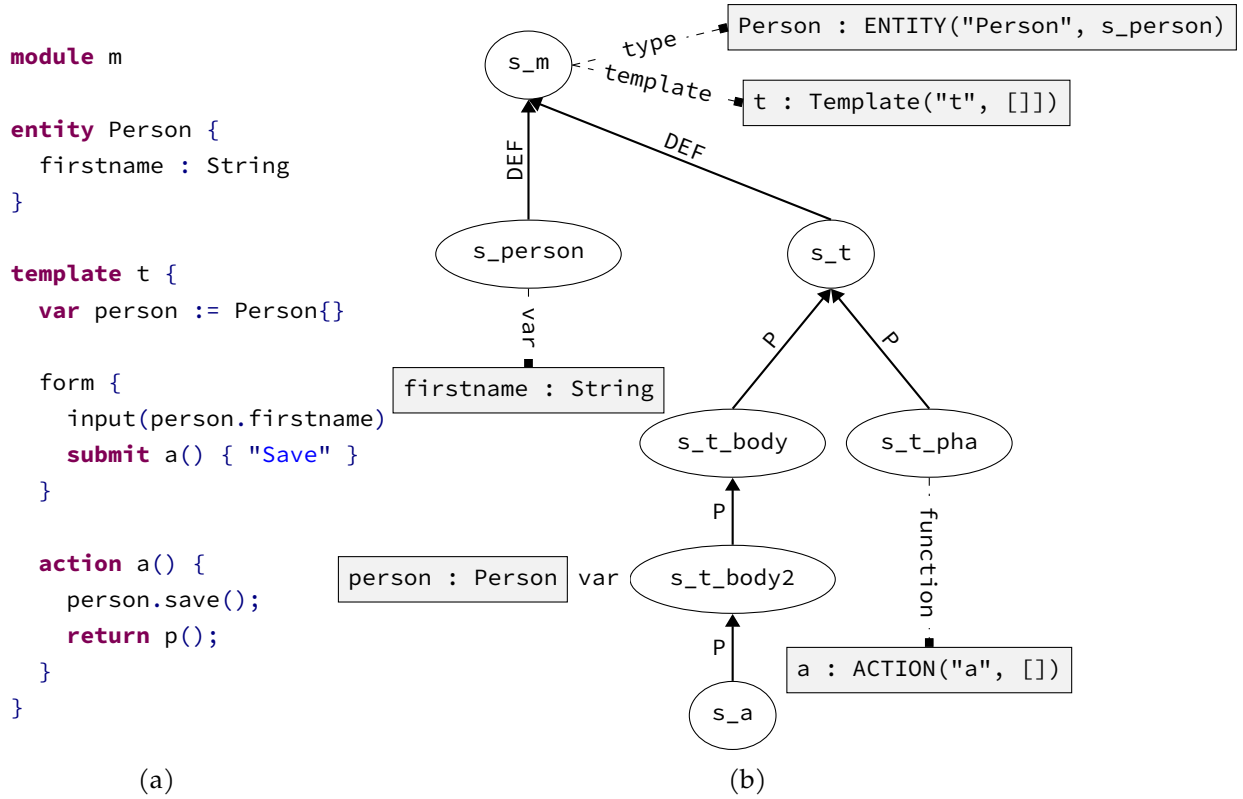


Figure 4.55: An example of defining and referencing actions

developers can expose properties of those Java types for use in WebDSL code. In this section we discuss how we implemented these semantics in Statix.

Figure 4.57a shows a WebDSL program containing the entity `Person` with one property and an extension of that entity with another property and a function. In Figure 4.57b, we show how we represent the program in a scope graph. When a (partial) entity is defined, a declaration is made in the scope graph under the `extendscope` relation that declares the scope linked to the name of the entity. When declaring the type in the module (such as `Entity("person", s_person)`), all scopes linked to the entity are retrieved from the scope graph and linked together using `EXTEND` edges.

The Statix rules that transform a WebDSL program with extended entities into the scope graph as shown in Figure 4.57 are listed in Figure 4.58. Note that the original entity does not need to import all modules in which it is extended. **Line 9** of Figure 4.58 shows the regular expression for well-formedness of the path to resolve declarations in the `extendscope` relation. It includes the optional `GLOBAL` edge, which is the edge from the global scope to every module, meaning that entity extensions in all modules, regardless of imports, are the result of the query. **Line 30** enforces that an `extend` entity declaration does always import the main entity declarations. The main entity declaration is what separates WebDSL entities main entity definition and explicit extensions, as opposed to C# partial classes that do not have a main definition.

4.7 Module system

The paper that introduced WebDSL (Visser 2007) contains the following description of its module system: “a very simple module system has been chosen that supports distributing functionality over files, without separate compilation”. We describe the current WebDSL module system as having transitive and symmetric imports. The implementation of the mod-


```

1 defineTemplateOk(s, DefineTemplate(mods, t
2   , FormalArgs(args), _, elements))
3 :- {fargTypes s_template s_phi s_body}
4 new s_template, s_template -DEF-> s,
5 argTypes == typesOfArgs(s, args),
6 declareParameters(s_template, zipArgTypes(fargs, argTypes)),
7 new s_phi, s_phi -P-> s_template, // scope for placeholders and actions
8 new s_body, s_body -P-> s_phi, // scope for template body
9 declareTemplate(s, t, argTypes, isAjaxTemplate(mods)),
10 overriddenElementExists(s, Template(), t, isAjaxTemplate(mods)),
11 templateElementsOk(s_body, s_phi, elements).
12
13 templateElementOk(s, _, s_phi
14   , Action2TemplateElement(Action(_, a, FormalArgs(args), Block(stmts))))
15 :- templateActionOk(s, s_phi, a, args, stmts, TRUE()).
16
17 templateActionOk : scope * scope * string
18   * list(FormalArg) * list(Statement) * BOOL
19 templateActionOk(s, s_phi, a, args, stmts, declare)
20 :- {s_fun s_fun_body argTypes}
21 new s_fun, s_fun -P-> s,
22 argTypes == typesOfArgs(s, args),
23 declareParameters(s_fun, zipArgTypes(args, argTypes)),
24 new s_fun_body, s_fun_body -P-> s_fun,
25 optionallyDeclareAction(s_phi, a, args, argTypes, declare),
26 stmtsOk(s_fun_body, stmts, PAGE(_, _)).
27
28 optionallyDeclareAction : scope * string * list(FormalArg) * list(TYPE) * BOOL
29 optionallyDeclareAction(_, _, _, _, FALSE()).
30 optionallyDeclareAction(s, a, args, ts, TRUE())
31 :- declareAction(s, a, args, ts).

```

Figure 4.56: Statix rules for declaring actions

ule system in the current compiler maintains a list of modules of which an application consists. When a module is imported by the main application file or transitively imported, the module is added to the list and its definitions are visible to all other modules in the list. An example application with these semantics and its scope graph is visualized in Figure 4.59. The implementation of these semantics are listed in Figure 4.60.

The modelling of this system causes an explosion in this amount of import edges between module scopes, as it is a fully connected graph for all modules that are (transitively) imported by the application root. The amount of import edges would scale quadratically with the number of imported modules:

$$\text{imported edges} = \#modules(\#modules - 1) + \#modules$$

The additional $+\#modules$ in this calculation is for the import edges to the global scope.

This representation of the WebDSL module system in scope graphs is functional, however, we found this to be extremely taxing on the run time performance of an analysis. To gather information about the performance of the current module system semantics in the WebDSL Statix specification, we conducted an experiment where the same program was an-

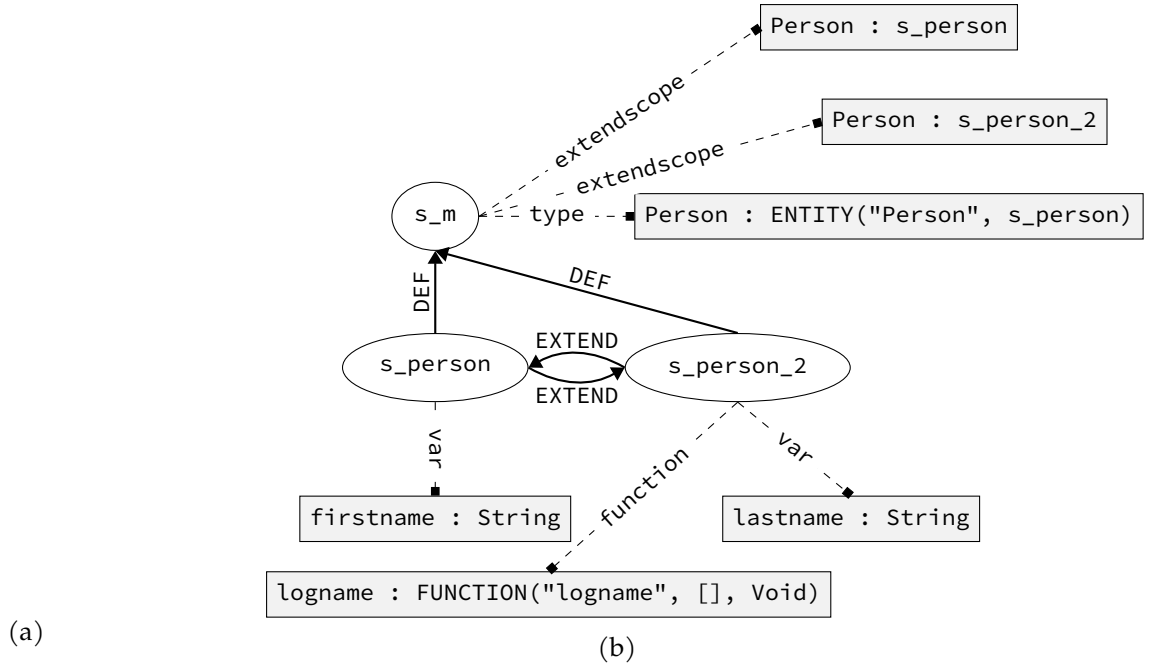


Figure 4.57: An example of extending entity scopes

alyzed multiple times, but with the addition of extra empty modules. The result as shown in Figure 4.61 confirms our suspicion.

To run the experiment, we used a 2019 MacBook Pro running macOS Monterey 12.2. The machine has a 2,3 GHz 8-core Intel Core i9 with 64 GB RAM available, of which 8 GB was dedicated to the evaluation scripts. The evaluation scripts¹ were configured to analyze the same application with different amounts of empty modules. Using the Java Microbenchmark Harness², we timed the run time of the Statix specification using 5 warm-up iterations and 20 regular iterations.

From the results shown in Figure 4.61, we argue that the implementation of the current WebDSL module system in Statix, approach does not scale to real world applications. We attempted to run the Statix specification on two open-source WebDSL applications.

- **Reposearch**³: A source code search engine that helps to find implementation details and example usages. Reposearch consists of 16 main files, 19 library files and 1 standard library file, totaling at 8,722 lines of code spread over 36 files.
- **YellowGrass**⁴: A tag based issue tracker similar to GitHub Issues, complete with access control and used daily by WebDSL developers. YellowGrass consists of 54 WebDSL files plus 20 WebDSL library files and 1 standard library file, coming to a total of 12,898 lines of code spread over 75 files.

Running the static analysis on Reposearch took roughly 25 minutes, and on YellowGrass it took more than 8 hours.

4.7.1 Revised WebDSL Module System

We designed a revised module system for WebDSL, in an attempt to improve the static analysis run time. In addition, the revised module system is designed to largely keep existing

¹<https://github.com/metaborg/statix-benchmark/>

²<https://github.com/openjdk/jmh>

³<https://codefinder.org/>, Source code: <https://github.com/webdsl/reposearch/>

⁴<https://yellowgrass.org/>, Source code: <https://github.com/webdsl/yellowgrass/>

```

1 declareExtendScope : scope * string * scope
2 declareExtendScope(s, e, s_extend) :-
3   !extendscope[e, s_extend] in s.
4
5 resolveExtendScope : scope * string -> list((path * (string * scope)))
6 resolveExtendScope(s, x) = ps :-
7   // extended entities do not have to be imported and
8   // can be resolved via the global scope edge:
9   query extendscope filter P* F* DEF? (IMPORT | IMPORTLIB)? GLOBAL?
10  and { e' :- e' == (x, _) }
11  min $ < P, $ < F, $ < DEF, $ < IMPORT, $ < IMPORTLIB, $ < GLOBAL,
12      P < F, P < DEF, P < IMPORT, P < IMPORTLIB, P < GLOBAL,
13      F < DEF, F < IMPORT, F < IMPORTLIB, F < GLOBAL,
14      DEF < IMPORT, DEF < IMPORTLIB, DEF < GLOBAL,
15      IMPORT < GLOBAL, IMPORTLIB < GLOBAL
16  and { (entity, entity_scope), (entity, entity_scope) }
17  in s |-> ps.
18
19 declEntity : scope * scope * string * list(EntityBodyDeclaration)
20 declEntity(s, s_entity, entity_name, bodydecs) :- { entityType }
21   // declare entity_scope to be linked to entity_name
22   declareExtendScope(s, entity_name, s_entity),
23   // link scopes using EXTEND edges
24   extendScopes(resolveExtendScope(s, entity_name), s_entity),
25   entityType == ENTITY(entity_name, s_entity),
26   declareType(s, entity_name, entityType).
27
28 declExtendEntity : scope * string * list(EntityBodyDeclaration)
29 declExtendEntity(s, entity_name, bodydecs) :- {s_extend_entity entity_scopes}
30   resolveType(s, entity_name) == [(_, (_, ENTITY(_, _)))]
31   | error $[Entity [entity_name] is not defined],
32   new s_extend_entity, s_extend_entity -DEF-> s,
33   declareExtendScope(s, entity_name, s_extend_entity),
34   extendScopes(resolveExtendScope(s, entity_name), s_extend_entity),
35   declEntityBody(s_extend_entity, entity_name, bodydecs).
36
37 extendScopes maps extendScope(list(*), *)
38 extendScope : (path * (string * scope)) * scope
39 extendScope((_, (_, s)), s). // Do not inherit own scope
40 extendScope((_, (_, s')), s) :-
41   s -EXTEND-> s'.

```

Figure 4.58: Statix rules for extending entities

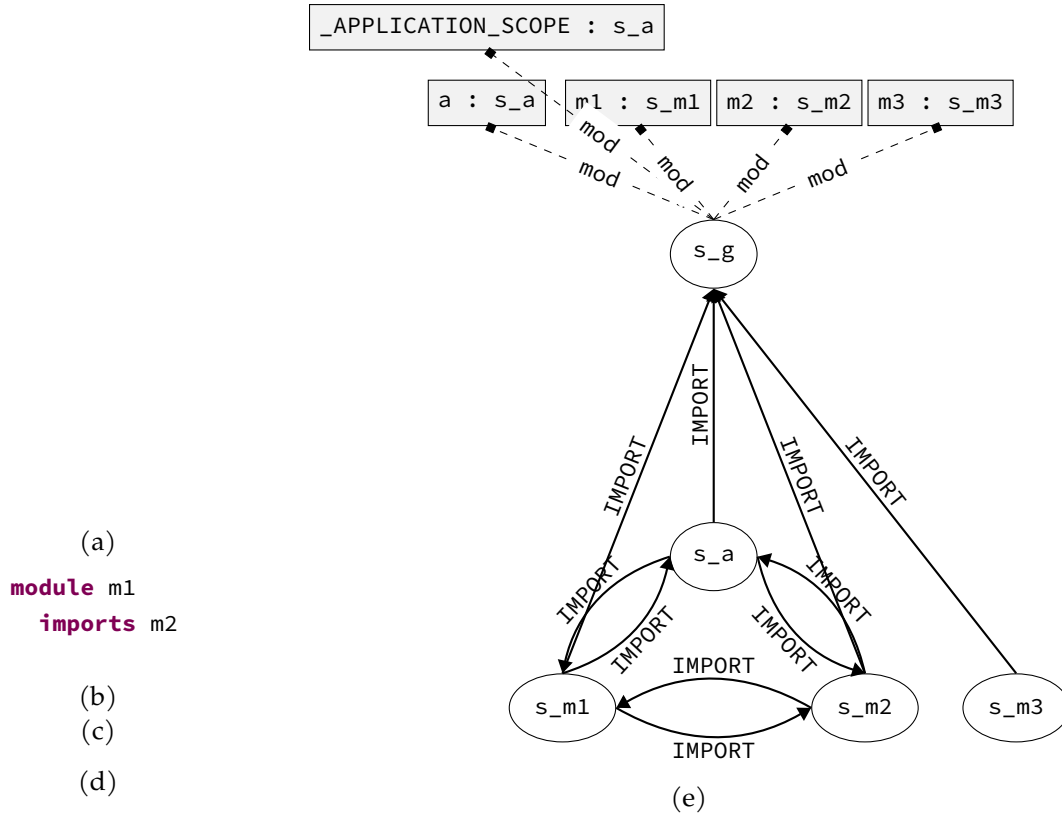


Figure 4.59: An example of the current WebDSL module system

WebDSL applications as they are, such that no substantial effort is required to adhere to the revised module system. We considered three options for the revised module system.

Everything is Global The first option boils down to getting rid of imports altogether and using one global scope. The upside of this approach is that existing WebDSL applications do not have to be changed at all, since the import statement will be deprecated and it does not change the behavior of programs. The downside is that, since all declarations and resolving takes place in the global scope, it could negate the possible performance boosts provided by the concurrent (Antwerpen and Visser 2021) or incremental (Zwaan, Antwerpen, and Visser 2022) Statix solver.

Strict Modules The second option is to get rid of transitive and symmetric imports, making the WebDSL module system similar to that of simple toy languages such as LMR: Language with Modules and Records (Neron, A. Tolmach, et al. 2015). Existing WebDSL applications have to be extended with more import statements to correctly resolve all the references that are now transitively or symmetrically imported. This module system allows for more transparent behavior; a reference to a definition from another module requires the importing of that module.

Strict Modules With Wildcard Imports The third and last option we consider is similar to the strict module system from previous paragraph, but allowing wildcard imports for the convenience of the developer. Assuming that existing WebDSL applications are structured correctly in terms of nesting of modules, using concepts defined in other modules can be imported in one statement. The downside of using wildcard imports is a potential increase in import edges of which some are unused.

```

1 unitOk(s_global, Application(app, sections)) :- {s_app}
2   new s_app, s_app -IMPORT-> s_global,
3   declareMod(s_global, app, s_app),
4   declareMod(s_global, "_APPLICATION_SCOPE", s_app),
5   importModules(s_global, s_app, app, sections),
6   rootPageDefined(s_app, app).
7
8 unitOk(s_global, Module(m, sections)) :- {s_mod}
9   new s_mod, s_mod -IMPORT-> s_global,
10  declareMod(s_global, m, s_mod),
11  importModules(s_global, s_mod, m, sections),
12  importedByApplicationRoot(s_global, s_mod, m).
13
14 importModules : scope * scope * string * list(Section)
15 importModules(s_global, s, m, sections) :-
16   declareImports(s_global, m, sections),
17   extendModuleScope(s_global, m, s).
18
19 extendModuleScope : scope * string * scope
20 extendModuleScope(s_global, m, s) :- {modules}
21   resolveImport(s_global, m) == modules,
22   importModulesInScope(s_global, s, modules).
23
24 importedByApplicationRoot : scope * scope * string
25 importedByApplicationRoot(s_global, s_mod, mod) :- {s_app}
26   resolveMod(s_global, "_APPLICATION_SCOPE") == [(_, (_, s_app))],
27   try { query () filter IMPORT*
28         and { s_app' :- s_app' == s_app }
29         in s_mod |-> [_|_]
30   } | warning $[Module is not imported by the application root] @mod.

```

Figure 4.60: Statix rules for modelling the current module system

We implemented the changes required in Reposearch and YellowGrass to apply the latter two module systems. Reposearch required 92 extra import statements, spread over 16 files⁵ and YellowGrass required 328 extra imports spread over 54 files⁶. In both systems, the strict import semantics required about 6 new imports per file. Using the wildcard import system, this was reduced to about 4 per file, requiring 75 instead of 92 for Reposearch⁷ and 193 instead of 328 for YellowGrass⁸.

Figure 4.62 shows the run time of the WebDSL Statix analysis with the different module systems described at the start of this subsection. The figure shows that all the revised module systems greatly speed up the analysis time. The run time of the strict and wildcard system are similar, while the system where everything is part of the global scope takes slightly longer. We suspect that this is due to negated effects of the concurrent Statix solver that we used for the analysis.

Table 4.63 lists the run time of the different module systems on Reposearch and Yellow-

⁵<https://github.com/webdsl/reposearch/tree/experiment/strict-imports>

⁶<https://github.com/webdsl/yellowgrass/tree/experiment/strict-imports>

⁷<https://github.com/webdsl/reposearch/tree/experiment/wildcard-imports>

⁸<https://github.com/webdsl/yellowgrass/tree/experiment/wildcard-imports>

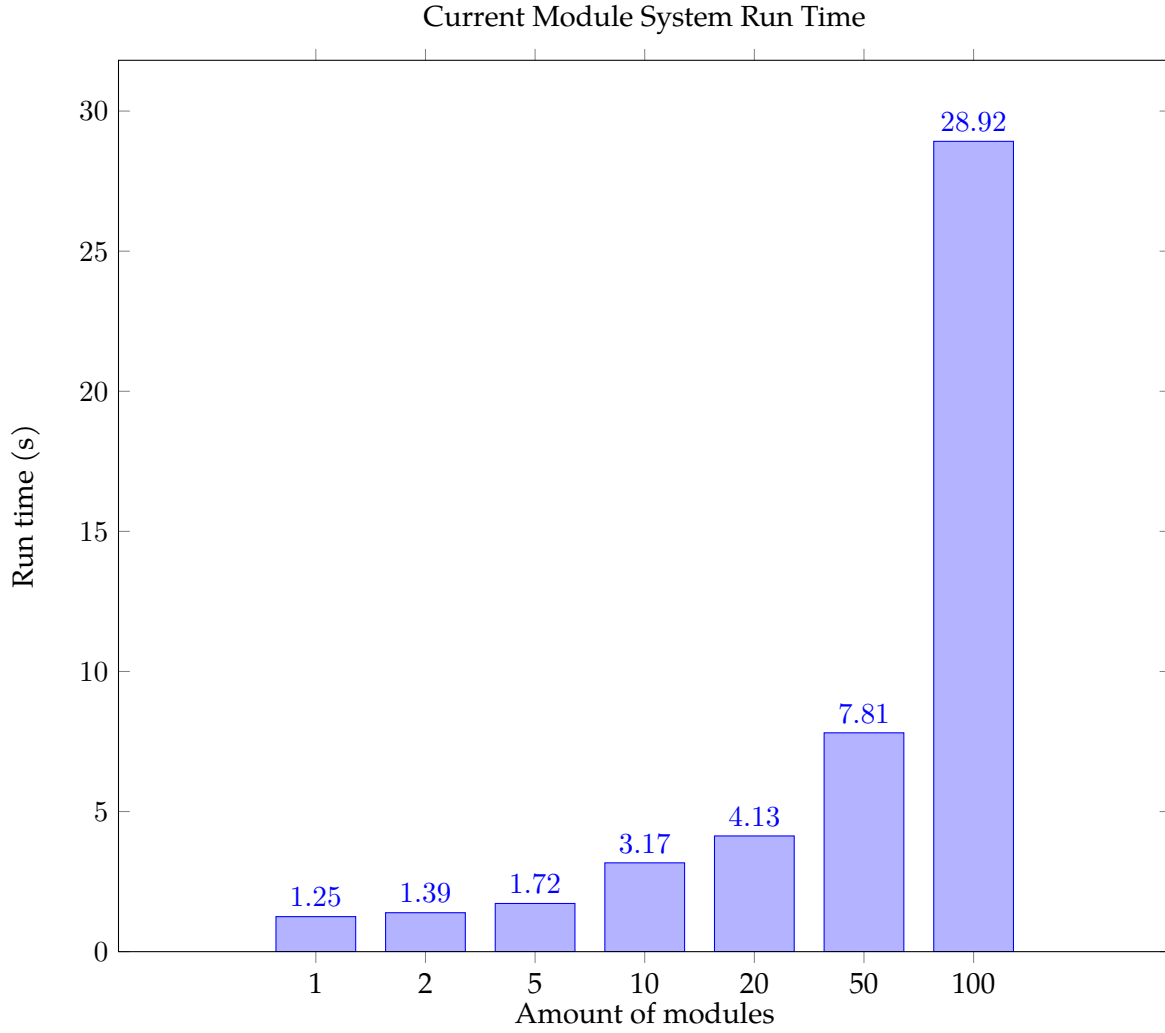


Figure 4.61: Run time of WebDSL Statix definition with the current module system.

Module System	Reposearch	YellowGrass
Current	1667s	>8 hours
Global	39s	40s
Strict	51s	53s
Wildcard	51s	51s

Table 4.63: Run time of WebDSL Statix definition with revised module systems on Reposearch and YellowGrass.

Grass. Similar to the test with the empty modules, there is a great performance boost from using any of the revised module systems. Surprisingly, the strict and wildcard module system take more time than global module system for YellowGrass. This is in contrast with Figure 4.62, so we suspect that having populated modules instead of empty modules impacts the performance of the strict- and wildcard imports by having longer (and therefore more costly) query paths and shadowing rules in a module system with imports.

Although all revised module systems would massively boost the performance of the WebDSL Statix specification, the run time of analyzing a real world application is still unfit for use in practice. To this end, we need to speed up the analysis time even more. For other performance boosts, as we explain in the next section, a scope graph representation with split modules was required. Since the wildcard import system is more convenient for

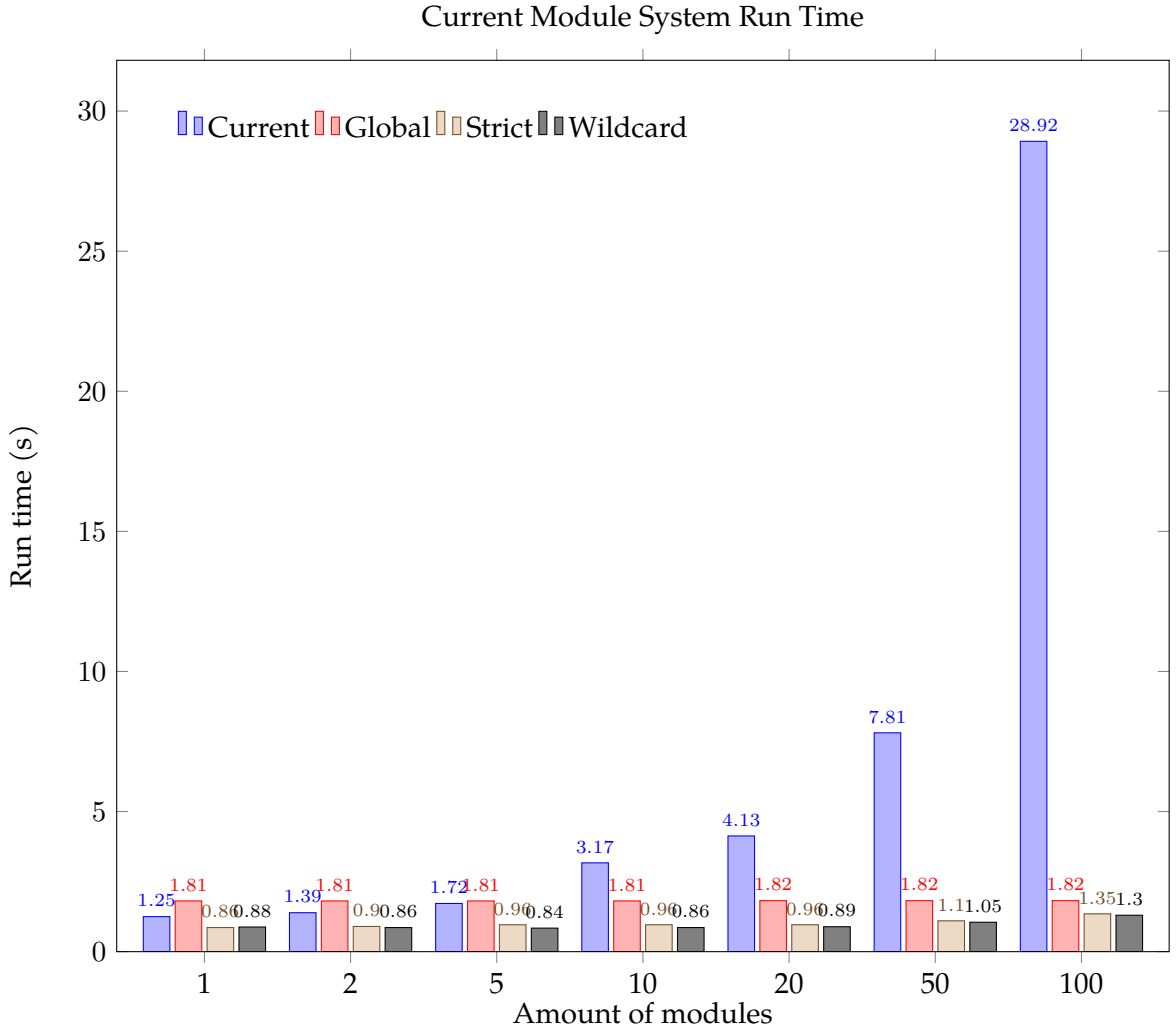


Figure 4.62: Run time of WebDSL Statix definition with revised module systems.

WebDSL developers, the WebDSL Statix specification is developed with the revised module system that uses wildcard imports.

4.8 Pre-analyzed built-in library

WebDSL applications heavily rely on the standard library named `built-in.app`. It contains all necessary type, page, template and function definitions and even the WebDSL compiler itself depends on this file being present. The `built-in.app` consists of 3,397 lines of WebDSL code and is by far the largest file in all WebDSL applications.

The concurrent Statix solver provides speed-ups by distributing the compilation units over the available processor cores on the development machine. However, the granularity of compilation units in the concurrent Statix solver is file-level granularity which results in the run time scaling with the largest file in the project (Antwerpen and Visser 2021). Van Antwerpen and Visser evaluated their concurrent Statix solver on multiple Java projects without dependencies besides the Java standard library (JRE). They indicate that the JRE dominated the run time of the concurrent Statix solver and decided to pre-compute the scope graph for the JRE and statically load it at the start of type-checking.

We investigated the approach of pre-computing the WebDSL `built-in.app` scope graph and statically loading it at the start of type checking. However, this requires the modules in

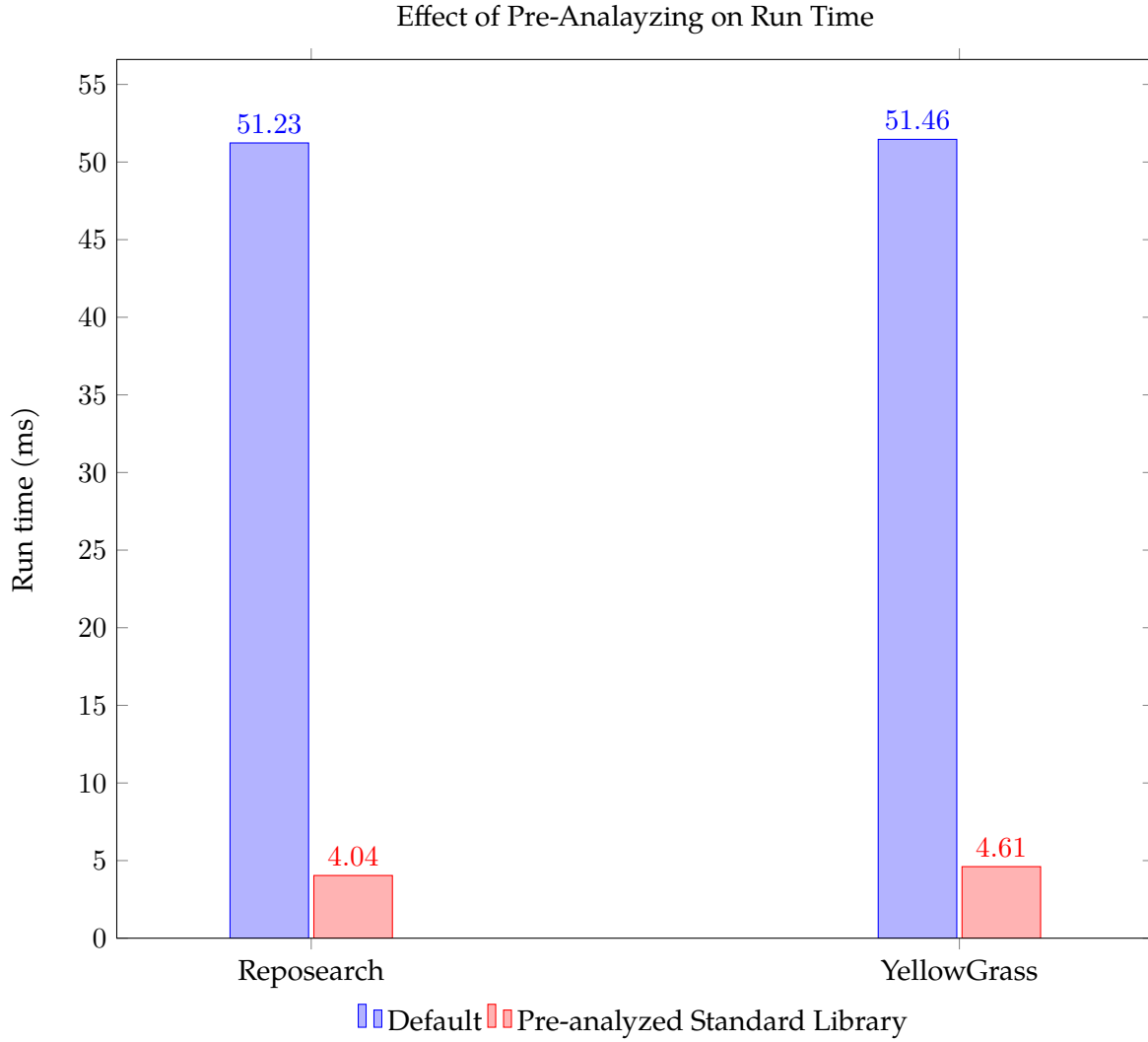


Figure 4.64: Run time of WebDSL Statix definition with and without pre-analyzed standard library `built-in.app`.

WebDSL to be clearly separated and this impacted our choice for a revised module system as explained in the previous section (Section 4.7.1).

Figure 4.64 shows a major improvement in run time of the WebDSL Statix specification when the analysis result of the standard library (`built-in.app`) is pre-computed, and the resulting scope graph is statically loaded at the start of type-checking other projects.

4.8.1 Built-in Type Extension

In the WebDSL Statix specification, built-in types are declared when processing the WebDSL standard library `built-in.app`. However, this clashes with our definition of extending types as explained in Section 4.6. The scopes of extended types are combined once, after which the type is declared. In Statix, existing scopes resolved via queries cannot be extended in order to keep the constraint scheduling sound. This does not impact types which are not declared in the standard library (such as all entities), but at first glance, this would nullify our attempt at implementing built-in type extension.

To combat this challenge, we changed the implementation of the query that resolves types in the WebDSL Statix specification, to the rules listed in Figure 4.65. **Lines 6, 7 and 8** show that the query for resolving now has a shadowing rule that allows shadowing types that are


```

1 resolveType : scope * string -> list((path * (string * TYPE)))
2 resolveType(s, name) = typesOf(ts) :-
3   query type filter P* F* ((EXTEND? INHERIT*) | (DEF? (IMPORT | IMPORTLIB)?))
4     and { e' :- e' == (name, _) }
5     // allow shadowing of types defined in an imported statix library
6     min $ < IMPORTLIB, P < IMPORTLIB, F < IMPORTLIB,
7       EXTEND < IMPORTLIB, INHERIT < IMPORTLIB,
8       DEF < IMPORTLIB, IMPORT < IMPORTLIB
9     and true
10    in s |-> ts.

```

Figure 4.65: Statix rules for built-in type extension in Statix

defined in an imported Statix library.

4.9 String Manipulation in Statix

String manipulation functions are the possibilities that a programming language provides to allow the developer to change or create new strings based on other data. In this section, we discuss the impact of the lack of string manipulation features in Statix.

Statix provides built-in functionalities for integer arithmetic⁹. This feature is useful to, for example, choose the most specific function call to resolve in function or template overloading (Section 4.4). Most other arithmetic such as boolean arithmetic can be encoded by the Statix developer, as we showed in Section 4.2.4.

The lack of the string manipulation features in Statix disallows certain WebDSL features, such as generated definitions or static code template expansion, to be encoded in Statix. An example of generated definitions and static code template expansion is shown in Figure 4.66. Generated definitions such as the `find<Entity>By<Property>` require the Statix specification to build a declaration in the scope graph with a name that is not present in the AST. Because it is impossible to concatenate two strings together, or to capitalize strings, this cannot be encoded in Statix.

Even though it is not possible to encode this in pure Statix, there is a workaround for the generated definitions. Because all information is present in the abstract syntax tree, the generation of the definitions can be seen as a desugaring of the syntax. Following this workaround, we extended the desugaring step of the WebDSL syntax with rules that generate the definitions. The result of the desugaring step is that the names of the generated functions are present in the analysis, and Statix is able to declare the functions.

The static code template expansion however, does not have such a workaround, because name-binding rules are required. Figure 4.66 shows the `expandtemplate ...` to `alert` definition to be declared in one definition, and then `alert` being referenced in the next definition. This name-binding logic can be encoded in Stratego, but it defeats the purpose of using the Statix definition as static analysis.

In addition to allowing new language concepts to be implemented, supporting string manipulation would also allow better error messages in Statix specifications. For example, the Statix specification could then pretty-print types, function signatures and template signatures to point developers to the error in their code with more accuracy.

In conclusion, the lack of string manipulation is preventing certain WebDSL language concepts from being encoded in Statix. Additionally, the introduction of string manipulation

⁹<https://www.spoofax.dev/references/statix/basic-constraints/#arithmetic-constraints>

application a

```
// generated definitions:
entity Person {
  address : String
}

function f() {
  // this function is generated by WebDSL:
  var p : Person := findPersonByAddress("");
}

// template expansion:
expand
  Success
  Info
  Warning
  Error
to alert

expandtemplate alert to Type{
  template alertType() {
    alert[class="alert-type", all attributes]{
      elements
    }
  }
}

template t {
  // templates are defined through the expansion above
  alertSuccess{}
  alertWarning{}
}
```

Figure 4.66: An example of generated definitions and static code template expansion in WebDSL

could increase the user-friendliness of error messages generated by Statix.

Chapter 5

Evaluation

In this chapter we evaluate the newly introduced SDF3 and Statix specifications for WebDSL. The new specifications have two concrete use-cases, namely serving as a case study for SDF3 and Statix, and being used on a daily basis by WebDSL developers. For both purposes it is useful to gather information about how the specifications behave in various situations. As a result of the case study, we want to show strengths and weaknesses of SDF3 and Statix based on information from the specifications, and for the WebDSL developers we would like to decide whether the new specifications are ready to be used in practice.

For both specifications, we will evaluate their correctness and performance on existing test suites, as well as WebDSL code that is used in practice. Then, we conclude this chapter by discussing the usability of the modernized implementation in practice.

5.1 Evaluating the WebDSL SDF3 Specification

Evaluating the SDF3 specification of the WebDSL grammar is done in two parts: its correctness and its performance in terms of generated parse tables and their run time. In this section we will use the current implementation of the WebDSL grammar in SDF2 as the reference grammar for correctness and performance.

5.1.1 Correctness

In this thesis we do not formally prove the correctness of the new grammar. Instead, we parse test suites that are intended for the current SDF2 specification and observe whether the files are parsed correctly and construct an AST without ambiguities. Additionally, we parse open source WebDSL applications that are used in practice and again observe whether the files parse correctly.

The test suite consists of 231 WebDSL snippets, ranging from single expressions to complete functioning applications. To re-use this test suite for the SDF3 specification, we converted the snippets into SPT tests. The existing syntax test suite is not a complete test suite of all syntax constructs but mostly contain syntax fragments which were problematic in the past to serve as a regression test suite. For the sake of completeness, we extended the SPT test suite, leading to a new total of 1118 SPT tests, where the newly added tests have an expected AST result, instead of only expecting the snippets to parse correctly. The result of running the WebDSL SDF3 specification on the syntax test suite is shown in Table 5.1.

In addition to the test suite, we used two open-source WebDSL applications for verifying that the new parser generated from the SDF3 specification does not suddenly fail or see ambiguities in existing applications:

	Parsed succesfully	Parsed with ambiguities	Failed to parse
Original test suite	231	0	0
Extension	887	0	0
Total	1118	0	0

Table 5.1: Results of parsing the syntax test suite with the WebDSL SDF3 specification.

- **Reposearch**¹: A source code search engine that helps to find implementation details and example usages. Reposearch consists of 16 main files, 19 library files and 1 standard library file, totalling at 8,722 lines of code spread over 36 files.
- **YellowGrass**²: A tag based issue tracker similar to GitHub Issues, complete with access control and used daily by WebDSL developers. YellowGrass consists of 54 WebDSL files plus 20 WebDSL library files and 1 standard library file, coming to a total of 12,898 lines of code spread over 75 files.

Using the parser generated from the WebDSL SDF3 specification, all files of both projects parsed successfully without ambiguities.

One thing to note in discussing correctness of the WebDSL SDF3 completeness is that, while the results are promising, the SDF3 specification has introduced many new sorts and constructors for disambiguation purposes, and to comply with the Statix signature generator expectations. The effect of this change is that we cannot automatically guarantee correctness of the disambiguation, because the resulting AST from the SDF3 definition is different compared to the SDF2 definition. Instead, we manually inspected the ASTs of handpicked snippets and no incorrect results were found.

5.1.2 WebDSL Parser Performance

The performance of a parser of a programming language is essential due to the rest of the compilation chain depending on its output. A requirement to use the parser generated by the new SDF3 specification in practice, is that its run time should not increase substantially.

Grammar specifications in SDF2 and SDF3 are not interpreted directly. Both formalisms generate a parse table, which is interpreted by the parser implementation JSGLR³. JSGLR is an implementation of SGLR parsing in Java, used within the Spoofax Language Workbench. Because of this architecture, it is insightful to inspect the generated parse tables and highlight the differences, as well as comparing the run times of both parsers on the test suite and existing applications.

Parse table from	States	Gotos	Max gotos per state	Actions	Max actions per state
SDF2	10,449	179,454	510	62,127	107
SDF3	12,866	244,688	821	525,728	2,491

Table 5.2: Data about the size of the parse tables generated from the WebDSL SDF2 and SDF3 grammar specifications.

The parse table generated from the SDF3 specification has more states, gotos and actions than the parse table from the SDF2 specification. Even though the described grammar did not change, it is implemented differently, leading to the increase in parse table size. Given

¹<https://codefinder.org/>, Source code: <https://github.com/webdsl/reposearch/>

²<https://yellowgrass.org/>, Source code: <https://github.com/webdsl/yellowgrass/>

³<https://github.com/metaborg/jsgrl>

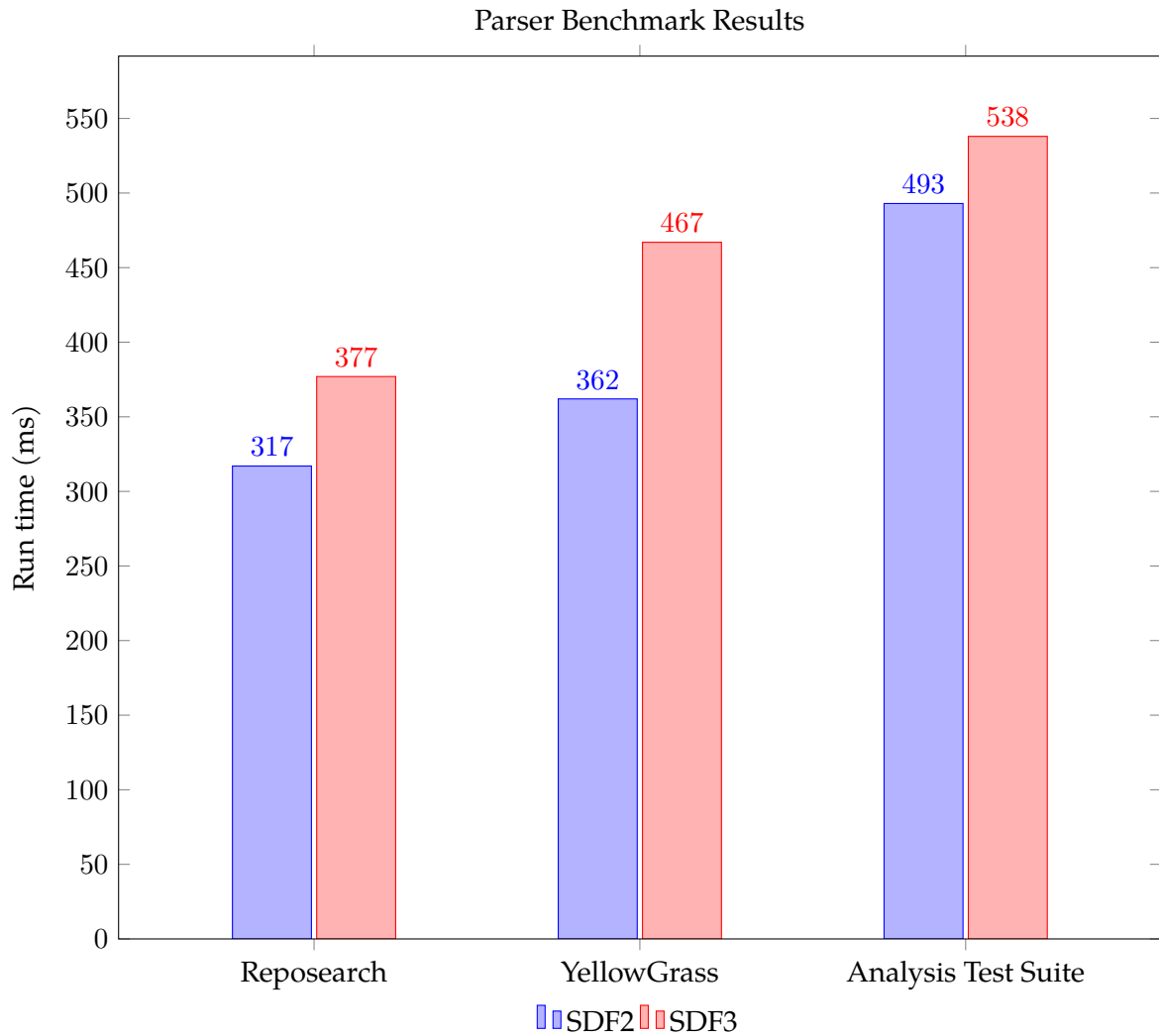


Figure 5.3: Run time of WebDSL SDF2 definition vs. SDF3 definition

that the parse table of the SDF3 specification is larger, we expect that this has a negative effect on the run time. To see the impact of the larger parse table on the run time of the parser, we executed the evaluation on Reposearch, YellowGrass and all files in the analysis test suite, which contains complete WebDSL programs as opposed to the syntax test suite. The analysis test suite consists of 521 small files, with a total of 19,644 lines of code.

To execute the evaluation, we used a 2019 MacBook Pro running macOS Monterey 12.2. The machine has a 2.3 GHz 8-core Intel Core i9 with 64 GB RAM available, of which 8 GB was dedicated to the evaluation scripts. The evaluation scripts⁴ were configured to parse the described files with the SDF2 parse table, as well as the SDF3 parse table using the JSGLR1 parser implementation. Using the Java Microbenchmark Harness⁵, we timed the run time of the parsers using 5 warm-up iterations and 10 regular iterations.

The result of benchmarking the run time of the syntax definitions is shown in Figure 5.3. Similar to the growth of the parse table generated from the SDF3 definition, the run time has also increased.

⁴<https://github.com/metaborg/jsglr2evaluation>

⁵<https://github.com/openjdk/jmh>

```

1 context-free syntax
2
3 "for" "(" Id ":" Sort "in" Exp OptFilter ")"
4   "{" TemplateElement* "}" ForSeparator -> TemplateElement {cons("For")}
5
6 "separated-by" "{" TemplateElement* "}" -> ForSeparator{cons("ForSeparator")}
7                                     -> ForSeparator{cons("None")}

```

(a)

```

1 context-free sorts
2
3 TemplateElement ForSeparator
4
5 context-free syntax
6
7 TemplateElement.For = <
8   for ( <VarId> : <Sort> in <Exp> <OptFilter> ) {
9     <TemplateElement*>
10   } <ForSeparator>
11 >
12
13 ForSeparator.ForSeparator = <separated-by { <TemplateElement*> }>
14 ForSeparator.ForSeparatorNone = <>

```

(b)

Figure 5.4: Defining a WebDSL for-loop in SDF2 and SDF3

5.1.3 Maintainability

Two of the goals of introducing SDF3 as successor of SDF2 in the syntax formalism family, were to support more declarative syntax definition and to make the syntax definitions more readable and understandable (Souza Amorim and Visser 2020). Figure 5.4 shows snippets of the SDF2 and SDF3 specifications that define a WebDSL for-loop. Souza Amorim and Visser argue that SDF3 syntax is more similar to other grammar formalisms such as EBNF (Backus et al. 1963) and for this reason we argue that the WebDSL syntax definition in SDF3 is easier to read and understand than its predecessor in SDF2.

However, being easier to read and understand does not automatically make the new syntax definition easier to maintain. The compliance with Statix and its Signature Generator⁶ imposes constraints on the grammar, such as disallowing optional sorts, which in the worst case causes the amount of sorts in the grammar to double as described in Section 3.4.2. Additionally, disambiguation without the `prefer` and `avoid` keywords, as described in Section 3.5, removes the post-parse filters but does create the need for more sorts which artificially complicate the grammar definition.

5.2 Statix

Static consistency checking through static analysis is one of the core aspects of WebDSL (Hemel, Groenewegen, et al. 2011). However, since no formal semantics of WebDSL are

⁶<https://www.spoofox.dev/howtos/statix/signature-generator/>

described, we rely on the current implementation of the static analysis in Stratego as the ground truth.

The evaluation of the Statix specification of WebDSL consists of three parts. First, we list results on the correctness; whether the static analysis allows well-formed programs to pass and whether it gives the correct feedback for erroneous programs. Next, we evaluate the performance in terms of run time on the applications Reposearch and YellowGrass. Lastly, we make qualitative observations on the maintainability of the new Statix specifications.

5.2.1 Correctness

One of the goals of analyzing source code before compiling and running it, is to provide early feedback to the developer regarding possible errors. The range of errors that can be caught early is extensive in WebDSL compared to other programming languages, because of the linguistic integration of user interfaces, request handling, access control and the data model.

For evaluating the correctness of the Statix specification, we first run the static analysis on Reposearch and YellowGrass. Both applications are well-formed programs without errors, therefore the desired result of the static analysis is to report no errors. The result of analyzing the programs with the Statix specifications is shown in Table 5.5. The Statix specification found a handful of errors that are caused by an unimplemented WebDSL feature in Statix, namely static code template expansion. This feature cannot be implemented in Statix due to the absence of String manipulation features, as we discussed in Section 4.9. Apart from this feature, the WebDSL Statix specification considers the applications to be well-typed which is the desired result.

Project	Files	Lines of code	Errors
Reposearch	36	8.722	4
Yellowgrass	75	12.898	6

Table 5.5: Results of running the static analysis on Reposearch and Yellowgrass.

It is trivial to write a program that does not analyze anything and therefore never give an error, and it would technically suit our goal of analysing Reposearch and YellowGrass without errors. To make sure the Statix specification gives feedback when it encounters an incorrect program, we run the Statix static analysis on the analysis test suite of the current implementation in Stratego. In total, the test suite consists of 521 small programs, testing different aspects of the WebDSL language. 273 files contain a correct program and expect the analysis to give no errors, while 248 programs contain in incorrect program where the static analysis must give specific feedback. The expectation as in those 248 files are listed as first lines in the file as comments, and can be one or more of the following.

- Should give an error containing the message s , denoted by `//s`.
- Should give an error containing the message s exactly x times, denoted by `//#x s`.
- Should not show an error with message s , denoted by `//^s`.

The results of running the Statix specification on the analysis test suite is shown in Table 5.6 below.

The result of the analysis test suite shows that the Statix specification of WebDSL is not yet on the level of the current static analysis in terms of correctness. While there are more passing tests than failing tests, there is room for improvement in both categories of the test suite, but particularly in the incorrect programs. The result of the analysis test suite scales with the engineering effort put in, and cannot be solely explained by Statix' shortcomings which we will highlight in the next subsection.

	Test succeeded	Test failed
Correct programs	231	42
Incorrect programs	71	177
Total	302	219

Table 5.6: Results of the analysis test suite.

Discussion of Failed Tests

Failing tests can be divided in categories. While some categories of failing tests can be solved by more engineering effort, others are inherent to using Statix as a language for implementing a static analysis.

Error messages that are not specific enough This is not something fundamentally wrong in the Statix language, as it can be improved by more development effort on the Statix specification, but at the cost of the brevity of the specification.

Error messages that Statix can not generate Some tests require an error along the lines of A function with Signature `f(string, int)` does not exist. Statix is unable to format strings to produce such an error. Statix is able to use string concatenation in error messages, but the given expression will always be surrounded by quotes. Apart from the quotes, a lack of string manipulation functions in Statix prevents users from properly formatting AST terms back to human-readable signatures.

Cascading errors This is mostly targeted by the test expectation where a certain error `s` should not be present (`//^s`). Statix as a language is not ideal for these test expectations, as errors are designed to show when the corresponding constraint fails. All generated constraints must either fail or succeed and there is no way to stop execution after a failed constraint, having the result of showing multiple errors. Consider the following expression:

```
var i : Int := nonexistentvariable + 2;
```

The error in this code snippet is that a non-existing variable is being referenced, but the expression cannot be typed properly, resulting in another error on the statement as a whole that says that the expression must be of boolean type.

When running the analysis test suite again, without taking the exact error message expectation into account, the outcome is as listed in Table 5.7. This simpler test suite expects a correct program to pass, and expects an incorrect program to show at least one error. The result is that more than a hundred tests now succeed, compared to the strict test suite from Table 5.6. Discarding the exact error message expectation lowers the accuracy of what is being tested, but it shows that the WebDSL Statix specification does throw errors for incorrect programs, but not always the correct amount or the expected message.

	Test succeeded	Test failed
Correct programs	231	42
Incorrect programs	185	63
Total	416	105

Table 5.7: Results of the analysis test suite without taking exact error message expectation into account.

Unimplemented features WebDSL has an extensive history and many features were added and deprecated over time. However, the deprecated features are still present in the analysis test suite, leading to failing tests. More development effort could make the tests pass.

5.2.2 Performance

Providing early feedback on written code to developers is an essential part of increasing productivity (Becker et al. 2019). For this reason, we want the static analysis to give quick and accurate results. To make claims about the run time of the WebDSL specification in Statix, we run the specification on two open-source applications: Reposearch and YellowGrass, and compare it to the run time of the current static analysis in Stratego.

To execute the evaluation, we used a 2019 MacBook Pro running macOS Monterey 12.2. The machine has a 2,3 GHz 8-core Intel Core i9 with 64 GB RAM available, of which 8 GB was dedicated to the evaluation scripts. The evaluation scripts⁷ were configured to analyze the two applications. Using the Java Microbenchmark Harness⁸, we timed the run time of the Statix specification using 5 warm-up iterations and 20 regular iterations. The run time of the current static analysis in Stratego was measured using a shell-script that executed the command-line tool shipped with the WebDSL compiler⁹.

Figure 5.8 shows the result of evaluating the WebDSL static analysis in terms of run time. As opposed to the Stratego implementation which always runs on a single core, the Statix specification utilizes all available cores on a machine. Regardless of the amount of cores, the Statix run time scales with the largest file in the project. Due to performance increases such as refactoring the WebDSL module system (Section 4.7) and pre-analyzing the WebDSL standard library (Section 4.8), the run time of the Statix WebDSL analysis is in the same order of magnitude as the Stratego WebDSL analysis from four cores onwards.

5.2.3 Maintainability

The current WebDSL compiler uses Stratego to implement the static analysis. As stated in the introduction of this thesis, we argue that the implementation of many compiler steps (e.g. desugaring, type checking, optimization and code generation) in the same Stratego project without clear intermediate representations poses a threat to the maintainability of WebDSL. While the results of the correctness- and performance evaluation look promising, we argue that there is still room for improvement in terms of the maintainability of the Statix specification.

Where Stratego is a more general term transformation language, Statix is developed specifically for implementing static analysis using scope graphs. This smaller domain allows the language engineer to express certain concepts with ease. Examples include declaration, resolving, shadowing and overriding. Additionally, the declarative nature of Statix increases focus on the what, instead of the how. Its syntax reads closer to inference rules, with a predicate being the conclusion, and the constraints being the premises.

A positive aspect of defining the static analysis in Statix is that, as opposed to the more imperative and handcrafted implementation in Stratego, we are able to profit from all new features introduced in Statix such as concurrency (Antwerpen and Visser 2021), incrementality (Zwaan, Antwerpen, and Visser 2022) or better editor services based on the Statix specification (Pelsmaeker, Antwerpen, and Visser 2019).

Unfortunately, the elegance and brevity of the WebDSL Statix specification is lost due to the amount of boilerplate code, helper functions and code duplication for specific error mes-

⁷<https://github.com/metaborg/statix-benchmark/>

⁸<https://github.com/openjdk/jmh>

⁹<https://github.com/webdsl/webdsl>

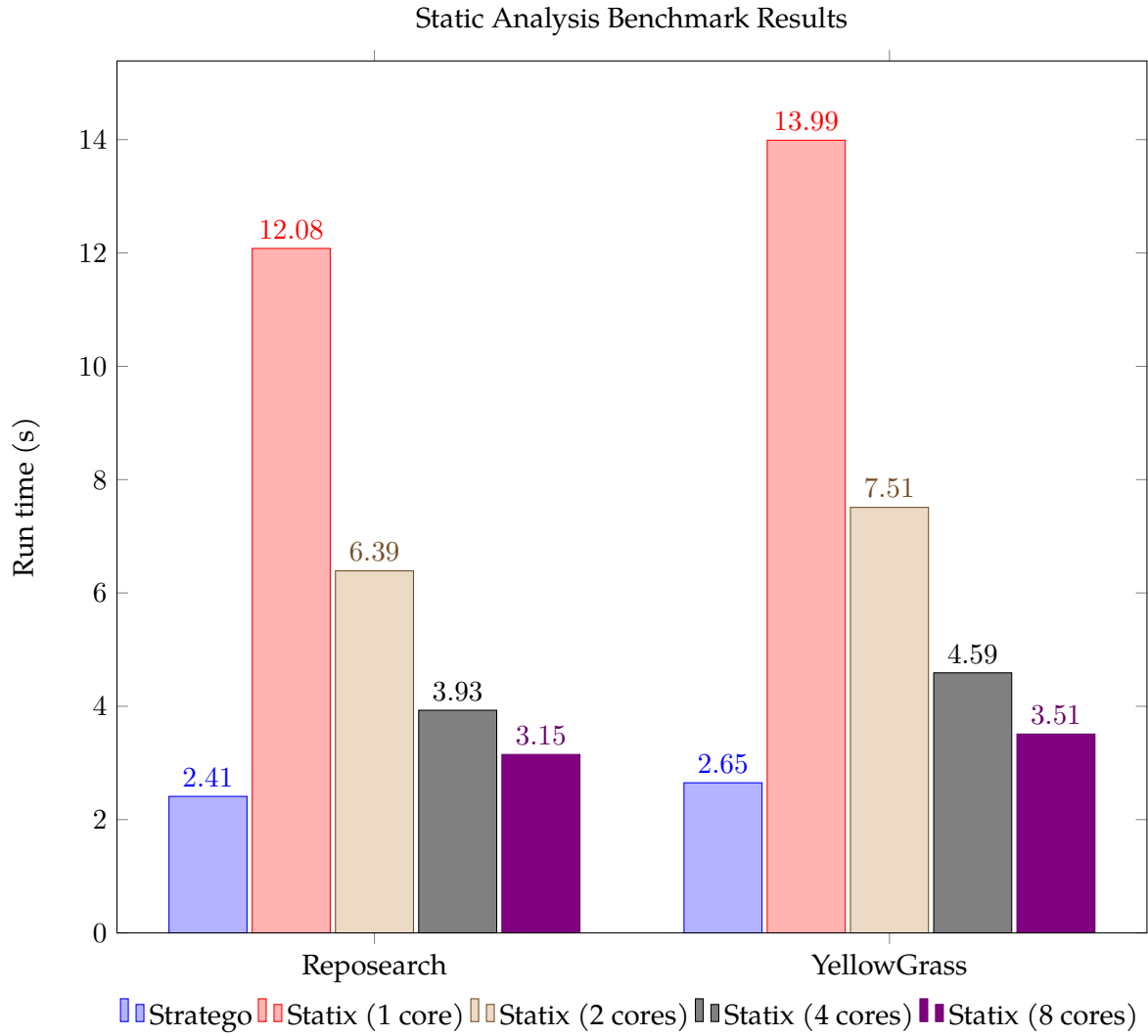


Figure 5.8: Run time of the WebDSL static analysis in Stratego vs. Statix

sages. A lack of provided list functions such as a *zip*, *map* or *filter* function requires multiple definitions of these, essentially duplicates but for different types.

Next, we argue that support for string manipulation would increase the quality of error messages, and reduce code duplication by being able to pass strings to other rules and adjusting them for a specific error message. Next to better error messages, this would also increase the range of language features Statix can support, as discussed in Section 4.9.

Lastly, the extensive amount of time required to analyze the Statix specification itself is a thorn in the side from a language engineer perspective, as it requires more than a minute to run the analysis on the WebDSL Statix specification.

Chapter 6

Related work

In this chapter we discuss work related to compiler front-ends of domain-specific languages. First, we examine published research related to WebDSL that address its grammar and static analysis. Second, we examine work concerning SDF3 and Statix and recent case studies using these meta languages. To conclude, we discuss related work on alternative approaches for implementing the front-end of domain-specific languages.

6.1 WebDSL

The paper of Visser (2007) introduced WebDSL as a case study of Stratego. The work lists most language constructs of the first WebDSL version, combined with SDF2 snippets of their syntax and Stratego code that demonstrates how code is generated for the WebDSL construct. Despite type checking not being mentioned for all language constructs, an extensive example plus implementation in Stratego is given for type checking a for-loop and the declaration and resolving of a variable in WebDSL. Other work by Groenewegen and Visser present linguistically integrated extensions of WebDSL regarding access control (2008) and data validation (2013).

The work of Hemel et al. (2011) identifies static consistency checking as lacking for web frameworks that were modern at the time, and they provide an analysis of how the web frameworks deal with certain consistency checks. Hemel et al. argue that domain-specific languages should be designed for consistency checking, providing a deep-dive on WebDSL as example. Additionally, they present how to perform such consistency checks with Stratego, essentially explaining an early version of the current WebDSL static analysis in Stratego.

The latest publication on WebDSL by Groenewegen et al. (2020) reflects on the WebDSL language as a whole, and provides an experience report of using WebDSL for over 10 years for increasingly ambitious applications.

6.2 Statix

Scope graphs were introduced by Neron et al. (2015) as a language-independent framework for describing name binding in programming languages. Van Antwerpen et al. (2016) build upon this work and extend the scope graph framework with generalized edge labels and introduce a constraint language with a solver that is able to express name binding and typing constraints. In a subsequent publication, Van Antwerpen et al. (2018) further extend the scope graph framework to increase the range of language constructs that can be modelled, such as parameterized types. Additionally, this work introduces Statix as a declarative language to specify type systems. In later work, the performance of Statix was boosted by Van Antwerpen and Visser (2021) through the introduction of the concurrent Statix solver, and by Zwaan et al. (2022) who introduced an incremental Statix solver.

6.2.1 Case Studies using Statix

The paper that introduced scope graphs by Neron et al. (2015) contains several illustrations on how scope graphs can model the name binding structure of programs. In the paper, the authors illustrate how to model various concepts of the Language with Modules and Records (LMR) such as let-bindings and, unsurprisingly, modules and records. The extended version (2015) contains examples on definition-before-use, Java packages and imports and C# partial classes. Van Antwerpen et al. (2016) use the LMR example again, but in addition to showing the scope graph, they list the typing constraints for certain snippets, and show an algorithm with rules that dictate how to traverse any LMR program and generate constraints.

The work that introduced Statix (Antwerpen, Bach Poulsen, et al. 2018) contained case studies on simply-typed lambda calculus that shows records and structural subtyping, Featherweight Java that presents classes and nominal subtyping. Lastly, the paper contains a case study on System F that illustrates Statix' ability to deal with parametric polymorphism.

The research of Rouvoet et al. (2020) on sound scheduling of name resolution queries also contains various case studies of languages used in practice, modelled in MiniStatix: the core constraint language of Statix, with some extras implemented in Haskell. The case studies are on a subset of name resolution for Java and Scala, and the whole of LMR. In their evaluation they used a combination of valid and invalid programs, similar to this thesis. In contrast to this thesis, Rouvoet et al. do not use error message expectations, simply a fail or succeed expectation but the aim of their evaluation was different from ours.

Van Antwerpen et al. (2021) implement a subset of Java in Statix to evaluate their work on real-world codebases, similar to what we show in this thesis. However, their work is not evaluated on erroneous programs, as their goal is to benchmark the parallel Statix solver in terms of run time.

In addition to published research, there are multiple Master's theses that contain Statix case studies:

Aerts (2019) In his Master's thesis, Aerts described, implemented and evaluated an approach for incrementalizing Statix based on separate compilation and dynamic dependency detection. In the evaluation, Aerts implemented a simplified version of Java in Statix that focuses on type-dependent name resolution.

Zwaan (2021) Zwaan provided a runtime that is able to handle composed Statix specifications. He validated his work by integrating Statix specifications of a small subset of SDF3 (Mini-SDF) and Stratego (Mini-STR). Additionally, his Master's thesis contains another case study where the specifications of a small toy-language with expressions, record types, functions and modules (Mod) and a subset of SQL (Mini-SQL).

Wilms (2022) In Wilms' Master's thesis, he introduced PIE DSL 2, a successor of domain specific language accompanying incremental build system PIE (Konat et al. 2019). As part of the PIE DSL 2 implementation, Wilms implemented the static semantics in Statix which modelled many interesting features such as a module system similar to Java, class inheritance and parameterized types.

6.3 Alternative Approaches

The Spoofox language workbench provides meta-DSLs such as SDF3, Statix and Stratego for implementing all aspects of a domain-specific language. However, many alternative approaches exist.

One such approach to defining a static analysis is by using Datalog. Datalog is a subset of Prolog and similar to Statix, it is a declarative logic programming language. Where Statix builds and queries a scope graph, Datalog builds and queries a deductive database. Recent work by Pacak et al. (2020) utilize the performance of Datalog’s incremental solvers by expressing algorithmic typing rules in Datalog.

Xtext (Efftinge and Völter 2006) is a language workbench that generates with a heavy focus on generated tooling, such as a compiler infrastructure and IDE support, based on the Xtext grammar language. Xtext utilizes the Eclipse Modeling Framework (Steinberg et al. 2009) and provides an API to customize behavior of the generated tooling using dependency injection (Eysholdt and Behrens 2010). For example, this customization allows for changing the scoping and name binding rules, or adjusting the code generator.

Rascal is a metaprogramming language, developed by the CWI SWAT group (2009). A commonality of Spoofax and Rascal is that both are in a sense based on the ASF+SDF formalism (Klint 1993). A complete language implementation can be defined using Rascal, including a generated parser, static analysis and code generation. Rascal supports the generation of constraints for program analysis, but does not provide a built-in constraint solver, as they believe hand-crafted specializations are important for making program analysis scale. In the paper titled *Rascal, 10 years later* (2019), Klint et al. mentions constraint-based type checking as an ongoing project.

The MPS platform by JetBrains (Dmitriev 2004) is a commercial language workbench with a focus on projectional editing, using representations such as tables and graphs. The first language developed in JetBrains MPS was BaseLanguage, a dialect of Java. The other language definition DSLs of MPS build on BaseLanguage (Pech, Shatalin, and Völter 2013). Similar to Xtext, MPS provides many IDE functionalities such as type checking and even code completion out-of-the-box. A recent experimental feature of JetBrains MPS named CodeRules¹ allows for type inference and type checking using logical programming with constraints, similar to Statix.

¹<https://jetbrains.github.io/mps-coderules/about>

Chapter 7

Conclusion

In this thesis, we have presented a new front-end for WebDSL. WebDSL is a domain-specific language for web programming, inspired by multiple programming language paradigms. WebDSL is used to create applications such as WebLab and conf.researchr.org, which have thousands of daily users.

We have shown the conversion of the WebDSL grammar from an SDF2 specification to an SDF3 that is disambiguated without post-parse filters, and where the definition of sorts and constructors can be reused for Statix. The grammar formalism SDF3 generates a parse table which can be executed to efficiently transform textual programs into abstract syntax trees that are used in subsequent components of the compilation chain, such as static analysis.

Next, we presented the static semantics of WebDSL modelled in Statix. Statix is a declarative constraint-based programming language using the concept of scope graphs to model program structures and types. Statix comes with a built-in constraint solver that schedules the constraints in a sound way, builds and queries the scope graph and is able to show error messages for failing constraints.

The challenges of implementing a modernized WebDSL front-end are documented in this thesis, and we provided qualitative feedback on how to further improve the meta-DSLs SDF3 and Statix. The most notable challenges include but are not limited to:

- Disambiguating the WebDSL grammar without the use of post-parse filters `{prefer}` and `{avoid}`;
- Modifying the grammar such that it adheres to the requirements of the Statix signature generator;
- Expressing language constructs such as type extension, overloading, generated definitions and static code template expansion in Statix;
- Expressing the WebDSL module system in Statix and defining a revised module system;
- Reducing the run time of the WebDSL static analysis in Statix from over 8 hours to 4 seconds on real world applications.

Lastly, the resulting modernized front-end of WebDSL was evaluated in terms of correctness and run time performance using large test suites and WebDSL applications that are used in practice.

7.1 Research Questions

The aim of this thesis was to answer the following research questions.

- **RQ1** Is it possible to define the WebDSL syntax in SDF3, without the use of post-parse filters?
- **RQ2** How does the run time efficiency of the parser generated from WebDSL in SDF3 compare to the current WebDSL parser generated from SDF2?
- **RQ3** How does the maintainability of WebDSL in SDF3 compare to the WebDSL syntax definition in SDF2?
- **RQ4** Is a Statix implementation of the WebDSL static semantics able to catch the same errors and warnings as the current implementation in Stratego?
- **RQ5** How does the run time efficiency of the WebDSL static semantics in Statix compare to the current WebDSL implementation?
- **RQ6** How does the maintainability of the WebDSL static semantics in Statix compare to the current implementation?

RQ1 In Chapter 3 we showed that it is possible to define the WebDSL syntax in SDF3 without the `{prefer}` and `{avoid}` annotations. Doing so required the addition of sorts for disambiguation purposes, which is not idiomatic. Additionally, indexed non-transitive priority rules were necessary to get rid of all observed ambiguities. In Section 5.1.1 we showed that the WebDSL syntax test suite, which contains regression tests from production bugs can be parsed by the parser generated from the WebDSL SDF3 definition.

RQ2 In Section 5.1.2 we showed that the parse table generated from the WebDSL SDF3 definition is larger than the parse table generated from the current WebDSL syntax definition in SDF2. Even though the described grammar did not change, it is implemented differently using more sorts and constructors, leading to an increase in parse table size which resulted in a slightly slower parser.

RQ3 While the maintainability of the WebDSL SDF3 definition is subjective, we argue that it is more readable due to the increased similarity to other grammar formalisms in SDF3, such as EBNF. Additionally, the implementation without post-parse filters does increase the directness of finding an ambiguity in the WebDSL SDF3 definition. However, the compliance with Statix Signature Generator and the disambiguation without post-parse filters introduce additional sorts which unnecessarily complicate the grammar. For an objective answer to this question, user studies are required.

RQ4 The Statix definition the WebDSL static semantics is not able to report the same errors and warnings as the current implementation in Stratego. For example, Statix is missing String manipulation which is necessary to encode certain WebDSL features such as static code template expansion. In Section 5.2.1 we use the analysis test suite of the current implementation, which contains concrete test expectations, namely if an error should be shown, what the error should be and how many errors should be shown. From this test suite, 302 out of 512 tests succeed and a discussion of the failed tests can be found in Section 5.2.1. When the test suite is simplified to test whether correct programs pass and incorrect programs show an error, 416 out of the 512 tests succeed.

RQ5 Section 5.2.2 that the static analysis generated from the WebDSL Statix specification is slightly slower than the current implementation in Stratego. While earlier iterations took over 8 hours to complete, Section 4.7.1 shows how we reduced its run time to several seconds.

RQ6 The meta-language Statix is developed specifically for implementing the static semantics of programming languages, in contrast to Stratego which is broadly applicable. The Statix syntax resembles that of formal inference rules and makes the individual rules more readable than their Stratego implementation. However, encoding the WebDSL static semantics in Statix requires many boilerplate rules and code duplication due to a lack of a standard library or string manipulation. For an objective answer to this question, user studies are required.

7.2 Future work

While the modernized WebDSL front-end using SDF3 and Statix is promising, there are many possibilities for improving and extending the work shown in this thesis.

Increased Engineering Effort We argue that the correctness of the modernized WebDSL static analysis scales with the engineering effort put in to the Statix specification (see Section 5.2.1). For the development of web applications with WebDSL, catching more erroneous programs before compilation and reporting telling error messages is essential, especially since the current implementation in Stratego is capable of doing so.

Implement Revised Module System in Compiler Back-End In this thesis we described and implemented a revised module system for WebDSL in Statix (see Section 4.7) that can be described as a traditional module system where (with some exceptions) referencing a declaration made in another module requires importing that module. Additionally, it supports wildcard imports for pragmatic purposes during WebDSL development. Currently, the WebDSL compiler adheres to the module system as listed in the original WebDSL paper (Visser 2007), where it is described as follows: *“a very simple module system has been chosen that supports distributing functionality over files, without separate compilation”*.

Connect Modernized Front-End to Existing Compiler Back-End The WebDSL static analysis implemented in Stratego is not solely used for error reporting in the IDE. It generates signatures for definitions as discussed in Section 4.9 for which code should be generated, and it creates dynamic rules with name binding and type checking information on which the code generator depends. Re-implementing the current WebDSL back-end to use the Statix analysis results from the Statix Stratego API¹ is a possibility, but this may require a significant amount of time. An alternative is to write a connecting piece of software in Stratego that takes the Statix analysis result as input and generates the correct dynamic rules and AST terms such that the current WebDSL back-end can largely be used as is.

In the current Stratego implementation, the code generation phase and type-checking phase are intertwined; additional WebDSL snippets are generated based on typing information, which could have non-local consequences and requires a type check of the program again (Hemel, Kats, et al. 2010). This process is currently not implemented in the modernized static analysis, since it is not necessary for presenting the developer with early feedback. When the front-end is being connected to the back-end, analyzing the program multiple times with added generated code would be necessary. We suspect this would increase the run time of the whole compilation chain if the incremental Statix solver by Zwaan et al. (2022) is utilized.

Evaluate the Incremental Statix Solver The recently published incremental Statix solver by Zwaan et al. (2022) is promising in terms of speeding up the run time of executing Statix specifications. Their work already uses the WebDSL specification that we presented in this

¹<https://www.spoofax.dev/references/statix/stratego-api/>

thesis, but it would be interesting to further evaluate the run time on other (closed-source) applications such as WebLab and conf.researchr.org, and the impact of specific WebDSL language constructs on the incrementality.

Evaluate the Impact of Post-Parse Filters `{prefer}` and `{avoid}` on Parsing Run Time In this thesis we have shown how we disambiguated the WebDSL grammar in SDF3 without the use of post-parse filters (see Section 3.5). However, we do not provide insight into the impact of `{prefer}` and `{avoid}` on the run time of the generated parser, because we developed the SDF3 grammar from the start with the intention of leaving out the post-parse filters.

Bibliography

- Aerts, Taico (2019). “Incrementalizing Statix: A Modular and Incremental Approach for Type Checking and Name Binding using Scope Graphs”. MA thesis. Delft University of Technology. URL: <http://resolver.tudelft.nl/uuid:3e0ea516-3058-4b8c-bfb6-5e846c4bd982>.
- Antwerpen, Hendrik van, Casper Bach Poulsen, et al. (Oct. 2018). “Scopes as Types”. In: *Proc. ACM Program. Lang.* 2.OOPSLA. DOI: 10.1145/3276484. URL: <https://doi.org/10.1145/3276484>.
- Antwerpen, Hendrik van, Pierre Néron, et al. (2016). “A constraint language for static semantic analysis based on scope graphs”. In: *Proceedings of the 2016 ACM SIGPLAN Workshop on Partial Evaluation and Program Manipulation, PEPM 2016, St. Petersburg, FL, USA, January 20 - 22, 2016*. Ed. by Martin Erwig and Tiark Ropf. ACM, pp. 49–60. ISBN: 978-1-4503-4097-7. DOI: 10.1145/2847538.2847543. URL: <http://doi.acm.org/10.1145/2847538.2847543>.
- Antwerpen, Hendrik van and Eelco Visser (2021). “Scope States: Guarding Safety of Name Resolution in Parallel Type Checkers”. In: *35th European Conference on Object-Oriented Programming, ECOOP 2021, July 11-17, 2021, Aarhus, Denmark (Virtual Conference)*. Ed. by Anders Møller and Manu Sridharan. Vol. 194. LIPIcs. Schloss Dagstuhl - Leibniz-Zentrum für Informatik. ISBN: 978-3-95977-190-0. DOI: 10.4230/LIPIcs.ECOOP.2021.1. URL: <https://doi.org/10.4230/LIPIcs.ECOOP.2021.1>.
- Backus, John W. et al. (1963). “Revised report on the algorithmic language ALGOL 60”. In: *Comput. J.* 5.4, pp. 349–367. DOI: 10.1093/comjnl/5.4.349. URL: <https://doi.org/10.1093/comjnl/5.4.349>.
- Becker, Brett A. et al. (2019). “Compiler Error Messages Considered Unhelpful: The Landscape of Text-Based Programming Error Message Research”. In: *Proceedings of the Working Group Reports on Innovation and Technology in Computer Science Education, ITiCSE 2019, Aberdeen, Scotland Uk, July 15-17, 2019*. Ed. by Bruce Scharlau et al. ACM, pp. 177–210. ISBN: 978-1-4503-6895-7. DOI: 10.1145/3344429.3372508. URL: <https://doi.org/10.1145/3344429.3372508>.
- Brand, Mark G. J. van den et al. (2000). “Efficient annotated terms”. In: *Software: Practice and Experience* 30.3, pp. 259–291. DOI: 10.1002/(SICI)1097-024X(200003)30:3%3C259::AID-SPE298%3E3.0.CO;2-Y. URL: [https://doi.org/10.1002/\(SICI\)1097-024X\(200003\)30:3%3C259::AID-SPE298%3E3.0.CO;2-Y](https://doi.org/10.1002/(SICI)1097-024X(200003)30:3%3C259::AID-SPE298%3E3.0.CO;2-Y).
- Bravenboer, Martin, Arthur van Dam, et al. (2006). “Program Transformation with Scoped Dynamic Rewrite Rules”. In: *Fundamenta Informaticae* 69.1-2, pp. 123–178. URL: <https://content.iospress.com/articles/fundamenta-informaticae/fi69-1-2-06>.
- Bravenboer, Martin, Karl Trygve Kalleberg, et al. (2008). “Stratego/XT 0.17. A language and toolset for program transformation”. In: *Science of Computer Programming* 72.1-2, pp. 52–70. DOI: 10.1016/j.scico.2007.11.003. URL: <http://dx.doi.org/10.1016/j.scico.2007.11.003>.

- Chastelet, Elmer van (2013). "A Domain-Specific Language for Internal Site Search". MA thesis.
- Denkers, Jasper (2018). "A Modular SGLR Parsing Architecture for Systematic Performance Optimization". MA thesis. Delft University of Technology. URL: <http://resolver.tudelft.nl/uuid:7d9f9bcc-117c-4617-860a-4e3e0bbc8988>.
- Dmitriev, Serguei (2004). "Language Oriented Programming: The Next Programming Paradigm". In.
- Efftinge, Sven and Markus Völter (2006). "oAW xText: A framework for textual DSLs". In: *Workshop on Modeling Symposium at Eclipse Summit*.
- Eysholdt, M. and H. Behrens (2010). "Xtext: implement your language faster than the quick and dirty way". In: *Proceedings of the ACM international conference companion on Object oriented programming systems languages and applications companion*. ACM, pp. 307–309.
- Groenewegen, Danny M., Elmer van Chastelet, and Eelco Visser (2020). "Evolution of the WebDSL Runtime: Reliability Engineering of the WebDSL Web Programming Language". In: *Conference Companion of the 4th International Conference on Art, Science, and Engineering of Programming*. '20. Porto, Portugal: Association for Computing Machinery, pp. 77–83. ISBN: 9781450375078. DOI: 10.1145/3397537.3397553. URL: <https://doi.org/10.1145/3397537.3397553>.
- Groenewegen, Danny M. and Eelco Visser (2008). "Declarative Access Control for WebDSL: Combining Language Integration and Separation of Concerns". In: *Proceedings of the Eighth International Conference on Web Engineering, ICWE 2008, 14-18 July 2008, Yorktown Heights, New York, USA*. Ed. by Daniel Schwabe, Francisco Curbera, and Paul Dantzig. IEEE, pp. 175–188. ISBN: 978-0-7695-3261-5. DOI: 10.1109/ICWE.2008.15. URL: <http://dx.doi.org/10.1109/ICWE.2008.15>.
- (2013). "Integration of data validation and user interface concerns in a DSL for web applications". In: *Software and Systems Modeling* 12.1, pp. 35–52. DOI: 10.1007/s10270-010-0173-9. URL: <http://dx.doi.org/10.1007/s10270-010-0173-9>.
- Heering, Jan et al. (1989). "The syntax definition formalism SDF - reference manual". In: *SIGPLAN Notices* 24.11, pp. 43–75. DOI: 10.1145/71605.71607. URL: <http://doi.acm.org/10.1145/71605.71607>.
- Hemel, Zef, Danny M. Groenewegen, et al. (2011). "Static consistency checking of web applications with WebDSL". In: *Journal of Symbolic Computation* 46.2. Automated Specification and Verification of Web Systems, pp. 150–182. ISSN: 0747-7171. DOI: <https://doi.org/10.1016/j.jsc.2010.08.006>. URL: <http://www.sciencedirect.com/science/article/pii/S0747717110001367>.
- Hemel, Zef, Lennart C. L. Kats, et al. (2010). "Code generation by model transformation: a case study in transformation modularity". In: *Software and Systems Modeling* 9.3, pp. 375–402. DOI: 10.1007/s10270-009-0136-1.
- Kats, Lennart C. L. and Eelco Visser (2010). "The Spoofox language workbench: rules for declarative specification of languages and IDEs". In: *Proceedings of the 25th Annual ACM SIGPLAN Conference on Object-Oriented Programming, Systems, Languages, and Applications, OOPSLA 2010*. Ed. by William R. Cook, Siobhán Clarke, and Martin C. Rinard. Reno/Tahoe, Nevada: ACM, pp. 444–463. ISBN: 978-1-4503-0203-6. DOI: 10.1145/1869459.1869497. URL: <https://doi.org/10.1145/1869459.1869497>.
- Klint, Paul (1993). "A Meta-Environment for Generating Programming Environments". In: *ACM Transactions on Software Engineering Methodology* 2.2, pp. 176–201. DOI: 10.1145/151257.151260. URL: <http://doi.acm.org/10.1145/151257.151260>.
- Klint, Paul, Tijs van der Storm, and Jurgen J. Vinju (2009). "RASCAL: A Domain Specific Language for Source Code Analysis and Manipulation". In: *Ninth IEEE International Working Conference on Source Code Analysis and Manipulation, SCAM 2009, Edmonton, Alberta, Canada, September 20-21, 2009*. IEEE Computer Society, pp. 168–177. ISBN: 978-0-7695-3793-

1. DOI: 10.1109/SCAM.2009.28. URL: <http://doi.ieeecomputersociety.org/10.1109/SCAM.2009.28>.
- (2019). “Rascal, 10 Years Later”. In: *19th International Working Conference on Source Code Analysis and Manipulation, SCAM 2019, Cleveland, OH, USA, September 30 - October 1, 2019*. IEEE, p. 139. ISBN: 978-1-7281-4937-0. DOI: 10.1109/SCAM.2019.00023. URL: <https://doi.org/10.1109/SCAM.2019.00023>.
- Knuth, Donald E. (1965). “On the translation of languages from left to right”. In: *Information and control* 8.6.
- Konat, Gabriël et al. (2019). “Precise, Efficient, and Expressive Incremental Build Scripts with PIE”. In: *Second Workshop on Incremental Computing (IC 2019)*.
- Neron, Pierre, Andrew Tolmach, et al. (2015). “A Theory of Name Resolution”. In: *Programming Languages and Systems*. Ed. by Jan Vitek. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 205–231. ISBN: 978-3-662-46669-8.
- Neron, Pierre, Andrew P. Tolmach, et al. (Jan. 2015). *A Theory of Name Resolution with extended Coverage and Proofs*. Technical Report TUD-SERG-2015-001. Extended version of ESOP 2015 paper “A Theory of Name Resolution”. Software Engineering Research Group. Delft University of Technology.
- Pacak, Andre, Sebastian Erdweg, and Tamas Szabo (2020). “A systematic approach to deriving incremental type checkers”. In: *Proceedings of the ACM on Programming Languages* 4.OOPSLA. DOI: 10.1145/3428195. URL: <https://doi.org/10.1145/3428195>.
- Pech, Vaclav, Alex Shatalin, and Markus Völter (2013). “JetBrains MPS as a tool for extending Java”. In: *Proceedings of the 2013 International Conference on Principles and Practices of Programming on the Java Platform: Virtual Machines, Languages, and Tools, Stuttgart, Germany, September 11-13, 2013*. Ed. by Martin Plümiche and Walter Binder. ACM, pp. 165–168. ISBN: 978-1-4503-2111-2. DOI: 10.1145/2500828.2500846. URL: <http://doi.acm.org/10.1145/2500828.2500846>.
- Pelsmaeker, Daniël A. A., Hendrik van Antwerpen, and Eelco Visser (2019). “Towards Language-Parametric Semantic Editor Services Based on Declarative Type System Specifications (Brave New Idea Paper)”. In: *33rd European Conference on Object-Oriented Programming, ECOOP 2019, July 15-19, 2019, London, United Kingdom*. Ed. by Alastair F. Donaldson. Vol. 134. LIPIcs. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik. ISBN: 978-3-95977-111-5. DOI: 10.4230/LIPIcs.ECOOP.2019.26. URL: <https://doi.org/10.4230/LIPIcs.ECOOP.2019.26>.
- Rafalski, Timothy et al. (2019). “A Randomized Controlled Trial on the Wild Wild West of Scientific Computing with Student Learners”. In: *Proceedings of the 2019 ACM Conference on International Computing Education Research, ICER 2019, Toronto, ON, Canada, August 12-14, 2019*. Ed. by Robert McCartney et al. ACM, pp. 239–247. ISBN: 978-1-4503-6185-9. DOI: 10.1145/3291279.3339421. URL: <https://doi.org/10.1145/3291279.3339421>.
- Rekers, Jan (Jan. 1992). “Parser Generation for Interactive Environments”. PhD thesis. Amsterdam, The Netherlands: University of Amsterdam. URL: <https://homepages.cwi.nl/~paulk/dissertations/Rekers.pdf>.
- Rouvoet, Arjen et al. (2020). “Knowing when to ask: sound scheduling of name resolution in type checkers derived from declarative specifications”. In: *Proceedings of the ACM on Programming Languages* 4.OOPSLA. DOI: 10.1145/3428248. URL: <https://doi.org/10.1145/3428248>.
- Runeson, Per and Martin Höst (2009). “Guidelines for conducting and reporting case study research in software engineering”. In: *Empirical Software Engineering* 14.2, pp. 131–164. DOI: 10.1007/s10664-008-9102-8. URL: <http://dx.doi.org/10.1007/s10664-008-9102-8>.
- Souza Amorim, Luis Eduardo de and Eelco Visser (2020). “Multi-purpose Syntax Definition with SDF3”. In: *Software Engineering and Formal Methods - 18th International Conference, SEFM 2020, Amsterdam, The Netherlands, September 14-18, 2020, Proceedings*. Ed. by Frank S. de Boer and Antonio Cerone. Vol. 12310. Lecture Notes in Computer Science. Springer,

- pp. 1–23. ISBN: 978-3-030-58768-0. DOI: 10.1007/978-3-030-58768-0_1. URL: https://doi.org/10.1007/978-3-030-58768-0_1.
- Steinberg, Dave et al. (2009). *Eclipse Modeling Framework*. 2nd ed. Addison-Wesley.
- Visser, Eelco (July 1997a). *Scannerless Generalized-LR Parsing*. Tech. rep. P9707. Programming Research Group, University of Amsterdam.
- (Sept. 1997b). “Syntax Definition for Language Prototyping”. PhD thesis. University of Amsterdam.
- (2007). “WebDSL: A Case Study in Domain-Specific Language Engineering”. In: *Generative and Transformational Techniques in Software Engineering II, International Summer School, GTTSE 2007*. Ed. by Ralf Lämmel, Joost Visser, and João Saraiva. Vol. 5235. Lecture Notes in Computer Science. Braga, Portugal: Springer, pp. 291–373. ISBN: 978-3-540-88642-6. DOI: 10.1007/978-3-540-88643-3_7. URL: http://dx.doi.org/10.1007/978-3-540-88643-3_7.
- Visser, Eelco, Zine-El-Abidine Benaissa, and Andrew P. Tolmach (1998). “Building Program Optimizers with Rewriting Strategies”. In: *Proceedings of the third ACM SIGPLAN international conference on Functional programming*. Ed. by Matthias Felleisen, Paul Hudak, and Christian Queinnec. Baltimore, Maryland, United States: ACM, pp. 13–26. DOI: 10.1145/289423.289425. URL: <http://doi.acm.org/10.1145/289423.289425>.
- Vollebregt, Tobi, Lennart C. L. Kats, and Eelco Visser (2012). “Declarative specification of template-based textual editors”. In: *International Workshop on Language Descriptions, Tools, and Applications, LDTA '12, Tallinn, Estonia, March 31 - April 1, 2012*. Ed. by Anthony Sloane and Suzana Andova. ACM, pp. 1–7. ISBN: 978-1-4503-1536-4. DOI: 10.1145/2427048.2427056. URL: <http://doi.acm.org/10.1145/2427048.2427056>.
- Wilms, Ivo (2022). “Extending the Domain Specific Language for the Pipelines for Interactive Environments build system”. MA thesis. Delft University of Technology. URL: <http://resolver.tudelft.nl/uuid:567a7faf-1460-4348-8344-4746a18fb0b1>.
- Zwaan, Aron (2021). “Composable Type System Specification using Heterogeneous Scope Graphs”. MA thesis. Delft University of Technology. URL: <http://resolver.tudelft.nl/uuid:68b7291c-0f81-4a70-89bb-37624f8615bd>.
- Zwaan, Aron, Hendrik van Antwerpen, and Eelco Visser (2022). “Incremental Type-Checking for Free: Using Scope Graphs to Derive Incremental Type-Checkers”. In: *Proceedings of the ACM on Programming Languages* 6.OOPSLA2. DOI: 10.1145/3276484. URL: <https://doi.org/10.1145/3563303>.

Acronyms

AJAX Asynchronous JavaScript and XML

AST Abstract Syntax Tree

DSL Domain-specific Language

HTML HyperText Markup Language

IDE Integrated Development Environment

SQL Structured Query Language

URL Uniform Resource Locator