



# A Core Calculus for Documents

Or, Lambda: The Ultimate Document

WILL CRICHTON and SHRIRAM KRISHNAMURTHI, Brown University, USA

Passive documents and active programs now widely comingle. Document languages include Turing-complete programming elements, and programming languages include sophisticated document notations. However, there are no formal foundations that model these languages. This matters because the interaction between document and program can be subtle and error-prone. In this paper we describe several such problems, then taxonomize and formalize document languages as levels of a document calculus. We employ the calculus as a foundation for implementing complex features such as reactivity, as well as for proving theorems about the boundary of content and computation. We intend for the document calculus to provide a theoretical basis for new document languages, and to assist designers in cleaning up the unsavory corners of existing languages.

CCS Concepts: • **Applied computing** → **Document scripting languages**; • **Theory of computation** → **Type theory**.

Additional Key Words and Phrases: document languages, markup, templates

## ACM Reference Format:

Will Crichton and Shriram Krishnamurthi. 2024. A Core Calculus for Documents: Or, Lambda: The Ultimate Document. *Proc. ACM Program. Lang.* 8, POPL, Article 23 (January 2024), 28 pages. <https://doi.org/10.1145/3632865>

## 1 INTRODUCTION

We live in a golden age of document languages. We have time-tested methods for authoring stylized content, such as the widely-used Markdown language. Moreover, new commercially-backed languages like Typst [Mädje 2022], Markdoc ([markdoc.dev](https://markdoc.dev)), Quarto ([quarto.org](https://quarto.org)), and MDX ([mdxjs.com](https://mdxjs.com)) are providing ever more powerful ways of authoring documents. These languages are built within a rich tradition established by venerable systems like  $\text{\TeX}$  [Knuth and Bibby 1986] and Scribe [Reid 1980], and continuing with recent languages like Scribble [Flatt et al. 2009] and Pandoc ([pandoc.org](https://pandoc.org)).

Many of these systems are as much *programming* languages as document languages. Document authors want to systematically format data collections, abstract over similar text, create abbreviations, hide text for anonymous review, and so on. These tasks benefit from programmatic control over the document. Conversely, the lack of programmability in languages like HTML has generated a cottage industry of document *metalanguages*, often called *template* languages. For example, general-purpose languages with templates include PHP, Javascript (with JSX), and Lisp (via quasiquotes). Specialized template languages include Jinja for Python and Liquid for Ruby.

This proliferation of document languages raises foundational questions. What are the common characteristics of document languages? How do they relate? Are existing languages well-designed, or can we identify what appear to be flaws in their design? Given that documents are programs, can we reason about them? Our goal in this work is to shed light on these questions through the

---

Authors' address: Will Crichton; Shriram Krishnamurthi, Department of Computer Science, Brown University, Providence, Rhode Island, 02912, USA, [wcrichto@brown.edu](mailto:wcrichto@brown.edu).

---



This work is licensed under a Creative Commons Attribution 4.0 International License.

© 2024 Copyright held by the owner/author(s).

ACM 2475-1421/2024/1-ART23  
<https://doi.org/10.1145/3632865>

design of a *core calculus for documents* — that is, a formal model for the essential *computational* features of a document language. This paper describes the document calculus in four parts:

- (1) **We motivate the work with case studies about issues in the semantics of existing document languages** (Section 2). We show how document languages from both academia and industry can lead to unexpected behavior when composing content and computation. The case studies demonstrate the need to carefully study features like templates and interpolation.
- (2) **We incrementally describe the formal semantics of the document calculus** (Section 3). We construct the semantics in eight levels of the document language design space drawn from two key dimensions: document domain (strings or trees) and document constructors (literals, programs, template literals, template programs). We show how our choice of semantics corresponds to real-world document languages and also mitigates the issues in Section 2.
- (3) **We demonstrate how the document calculus can provide a foundation for modeling complex document features** (Section 4). We extend the calculus with three features: references, reforestation, and reactivity. Each feature is common to many document languages and stresses a different computational aspect of the calculus.
- (4) **We use the document calculus to formally reason about document programs** (Section 5). We formalize two useful theorems involving the document calculus. First, we show that our choice of template semantics produces well-typed terms. Second, we prove the correctness of a strategy for efficiently composing references and reactivity.

We conclude with related work (Section 6) and implications for language design (Section 7).

## 2 DOCUMENT LANGUAGES: THE BAD PARTS

The foundational concept of all document languages is the *template*. Templates are a kind of generalized data literal. Templates interleave computation (like expressions and variable bindings) into content (like strings and trees). The good part of a template is its brevity — with appropriate concrete syntax, a template can be more concise than a equivalent program without templates. This section is about the bad parts: *when templates go wrong*.

We present a series of case studies about how particular designs for template semantics cause problems at the boundary of content and computation. In each case study, we present a reasonable-looking program that works as expected. Then we show how a small change can quickly produce unreasonable results. These issues are not fatal flaws in the language; each has a workaround that would likely be known to seasoned users. Rather, we are simply drawing attention to common points of friction that could be both clarified and improved with a formal semantics.

### 2.1 PHP and the Global Mutable Buffer

PHP ([www.php.net](http://www.php.net)) is a popular programming language for web servers. A PHP program is itself a document, as any text placed outside a `<? tag ?>` is emitted to the client. However, templates in PHP are not pure. They do not construct values, but rather write to a global output buffer. This impure semantics requires that functions must be called in exactly the right place.

```

1 <? function mkelems($list) {
2   foreach ($list as $x) { ?>
3     <li><?= $x ?></li>
4   }
5 function mklist($list) { ?>
6   <ul><? mkelems($list) ?></ul>
7   <? }
8   mklist(["Hello", "World"]) ?>

```

For example, consider a PHP program that factors a bulleted-list generator into two functions. On the left, the function `mklist` wraps the result of `mkelems` in a `ul` tag. The `mkelems` function loops through the list and generates an `li` for each element. Text within a function but outside the question-mark-delimited ranges is emitted to the global buffer when the function is called.

Now, say the programmer wants to extend `mkelems` to return data about `$list` that is added as an attribute to the `ul` inside `mklist`. The programmer can easily modify `mkelems` to return e.g. the list length. But how should the programmer modify `mklist`? They must call `mkelems` *before* generating the `ul` in order to access the return value. But they also must call `mkelems` *after* generating the opening tag in order to position the list elements correctly.

```

1  <? function mkelems($list) {
2    foreach ($list as $x) { ?>
3      <li><?=$ $x ?></li>
4    <? }
5    return count($list);
6  }
7  function mklist($list) {
8    $n = mkelems($list); ?>
9    <ul data-n="<?=$ $n ?>" ?? </ul>
10  <? }
11  mklist(["Hello", "World"]) ?>

```

The programmer cannot easily compose the template mechanism with other concerns like returning auxiliary information. It is possible to work around this limitation through careful use of PHP’s output buffering facilities, but those APIs are also stateful and hence prone to errors. The key takeaway: an impure semantics for templates can restrict the ability of authors to design easily-composable document abstractions.

## 2.2 React and the Unresponsive Component

React ([reactjs.com](https://reactjs.com)) is a popular Javascript framework for writing reactive interfaces in the browser. A React program is a tree of components that encapsulate the state and view for a visual object. Ideally, React re-renders a component when its dependencies update. However, the kinds of components used in documents cannot always express their dependencies in terms understood by the framework.

```

1  function Toc() {
2    let [hdrs, setHdrs] = useState(()=>[]);
3    useEffect(() => {
4      let h1s = Array.from(
5        document.querySelectorAll('h1'));
6      setHdrs(h1s.map(node =>
7        node.textContent));
8    }, []);
9    return <ul>
10     {hdrs.map(text => <li>{text}</li>)}
11   </ul>;
12 }

```

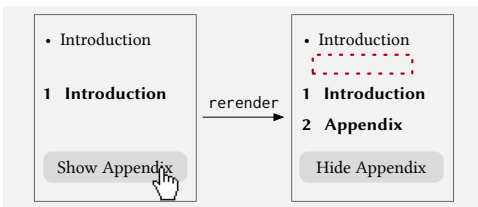
Say a programmer wants to implement a table of contents. A naïve implementation would be the program on the left. The headings array contains a list of the headings in the document, initially empty. The function `useEffect` indicates that the provided callback should execute after the component renders. The callback uses the browser’s DOM API to find all headers in the document. It extracts their text content and saves that array as local state. The returned template creates a bulleted list with a bullet per heading.

Here is an example application that uses the component. The `App` component creates a local boolean state `show` that is initialized to `false`. Then the template returns both a persistent header and a conditional header. A button is rendered that changes the condition on click, and then the table of contents is rendered.

```

1  function App() {
2    let [show, setShow] = useState(false);
3    return <<Toc />
4    <h1>Introduction</h1>
5    {show ? <h1>Appendix</h1> : null}
6    <button onClick={()=>setShow(!show)}>
7      {show ? "Hide" : "Show"} Appendix
8    </button></>;
9  }

```



This application will correctly render on the first pass, showing a table of contents with one bullet for “Introduction”. However, when the user clicks on the toggle button, the appendix header will appear, but the table of contents will not update, indicated by the dotted red rectangle.

The issue is that the header-query computation in the Toc component is not legible to the reactive runtime — React simply cannot express the concept that a component is dependent on the content of other components’ views. Therefore, React does not know to re-render the table of contents when App changes its heading structure. And this issue is not a chance mistake — as of September 2023, the top five results on Google for “react table of contents” are tutorials that all recommend this strategy. ToC implementations in other reactive frameworks like Svelte work similarly ([janosh/svelte-toc](#)). The key takeaway: computations over documents can have subtle dependency structures, which easily leads to bugs when combined with reactivity.

### 2.3 Scribble and the Improper Loop

Scribble [Flatt et al. 2009] is a document-oriented dialect of Racket. Scribble provides a template language that can interpolate computation via @-expressions, which desugar into standard Racket code. This desugaring can produce unexpected interactions with macros. For example, this program uses the `for/list` macro to map over a list of pairs:

```
1 @(define pairs
2   (list (list "A" "B") (list "C" "D")))
3 @itemlist{
4   @for/list[([p pairs])]{
5     @item{@(car p) @(cadr p)}
6   }
7 }
```

In this example, the `@itemlist` represents the list container, and `@item` represents a bullet in the list. The `for-loop` produces one bullet for each pair, creating the list:

- A B
- C D

Now say that the programmer wanted to change the code to flatten the list. A programmer might expect that factoring the `car` and `cadr` into separate `@items` should accomplish this task. However, Scribble instead drops the first bullet from each iteration, producing this list:

- B
- D

```
1 @(define pairs
2   (list (list "A" "B") (list "C" "D")))
3 @itemlist{
4   @for/list[([p pairs])]{
5     @item{@(car p)}
6     @item{@(cadr p)}
7   }
8 }
```

The cause of the bug is more apparent in this expression within the desugared Racket code:

```
1 (for/list ([p pairs])
2   (item (car p)) "\n" (item (cadr p))))
```

The issue is that `for/list` permits a “body” of `s`-expressions, where only the final `s`-expression becomes the value for each iteration. Scribble’s @-expression desugaring directly “pastes” the sequence of template elements into the `for/list` body, causing most of the template elements to be dropped. Notably, Racket’s web-server library uses Scribble’s @-expressions, and its documentation cites this bug as a common “gotcha” for users [McCarthy 2022, §7.3.2]. The key takeaway: the desugaring of templates to terms requires careful scrutiny to understand how it composes with other language features.

## 3 THE DOCUMENT CALCULUS

Section 2 shows that the complexity of a document language lies in more than its syntax: its semantics affect how well parts of the language compose together. However, a challenge in conceptualizing document language semantics is that there exists no formal foundations for describing

Table 1. Taxonomy of document languages based on the document calculus. A level is given as a particular pair document domain and document constructors. Each level corresponds to a specific family of existing document languages, with hyperlinks to the corresponding section of the paper. Example real-world languages in the family are provided along with a sample syntax from one of those languages (indicated with italics).

Domain	Ctors	Model	Example Languages	Example Syntax
String	Literal	$\mathcal{D}_{Lit}^{String}$	Text files, <i>quoted strings</i>	<code>"Hello World"</code>
	Program	$\mathcal{D}_{Prog}^{String}$	PLs with string APIs, such as <i>Javascript</i>	<code>"Hello" + " World"</code>
	Template Literal	$\mathcal{D}_{TLit}^{String}$	C <code>printf</code> , Python f-strings, <i>Javascript template literals</i> , Perl interpolated strings	<pre>let world = "World"; `Hello \${world}`</pre>
	Template Program	$\mathcal{D}_{TProg}^{String}$	C preprocessor, PHP, LaTeX, <i>jinja</i> (Python), Liquid (Ruby), Handlebars (Js)	<pre>{% set world = "World" %} Hello {{ world }}</pre>
Article	Literal	$\mathcal{D}_{Lit}^{Article}$	<i>CommonMark Markdown</i> , Pandoc Markdown, HTML, XML	<code>- Hello <b>World</b></code>
	Program	$\mathcal{D}_{Prog}^{Article}$	PLs with document APIs, such as <i>Javascript</i>	<pre>var ul =   document.createElement("ul"); // ...</pre>
	Template Literal	$\mathcal{D}_{TLit}^{Article}$	JSX Javascript, Scala 2, VB.NET, <i>Scribble Racket</i> , MDX Markdown, Lisp quasiquotes	<pre>@(define world "World") @itemlist{@item{   Hello @bold{@world}}}</pre>
	Template Program	$\mathcal{D}_{TProg}^{Article}$	<i>Typst</i> , Razor C#, Svelte Javascript, Markdoc Markdown	<pre>#let world = [World] - Hello *#world*</pre>

how a document language works. Our work aims to establish such a foundation by designing a *document calculus*, or a formal semantics for the core computational aspects of document languages.

First, we will establish a scope by asking: what is a document, and what is a document language? Within this paper, we consider a document to be “structured prose,” that is, plain text optionally augmented with styles (e.g., bold or italics) or hierarchy (e.g., paragraphs or sections) and interspersed with figures (e.g., images or tables). This definition includes objects like academic papers and news articles, and it excludes objects like source code, spreadsheets, and computational notebooks. It is useful to restrict the scope of documents because (a) many languages are often used to generate objects in the former set, and (b) those languages have commonalities which have not yet been carefully scrutinized via the lens of PL formalism, unlike e.g. spreadsheets [Bock et al. 2020]. We will give a formal definition of structured prose in the ensuing sections.

A document language, then, is a programming language that is commonly used to generate documents. Some document languages are specially designed for documents, such as Markdown, while others are general-purpose but commonly used to generate documents, such as PHP. In

reviewing the space of existing document languages, our key insight is that the design space can be decomposed along two dimensions:

- **Document domain:** the type of document generated by a document language. There are two main document domains: plain strings, and annotated trees of strings (which we call “articles”). Formally, we write this as:

$$\text{Domain } \mathcal{M} ::= \text{String} \mid \text{Article}$$

- **Document constructors:** the expressiveness of the operations for constructing a value in the document domain. There are four categories of expressiveness: literals (no computation), programs (computation over literals), templates literals (literals with interpolation), and template programs (literals with loops and variables). Formally, we write this as:


$$\text{Constructor } C ::= \text{Lit} \mid \text{Prog} \mid \text{TLit} \mid \text{TProg}$$

A language level is defined as an element of the cross product  $\text{Domain} \times \text{Constructor}$ . This taxonomy is useful because each level corresponds to multiple widely-used document languages, as shown in Table 1. This taxonomy also provides a natural progression for the development of the document calculus: starting with string literals, we can add successively more features until reaching article template programs.

The document calculus therefore consists of  $2 \times 4 = 8$  levels. Each level of the document calculus is written as  $\mathcal{D}_C^M$ , which consists of a document domain  $\mathcal{M}$ , document expressions  $\text{Expr}_C^M$  for the domain  $\mathcal{M}$  with constructors  $C$ , and a semantics that relates the two. This section will present each level by first giving examples of real-world document languages at that level, and then providing a formal definition for the level. We also provide an OCaml implementation of these semantics in the supplementary materials, using open functions [Löh and Hinze 2006] to match the incremental presentation of the semantics.

As with any model, the document calculus focuses on some aspects to the exclusion of others. For instance, concrete syntax is an essential aspect of any document language. However, our focus is on the computational aspects, so we postpone discussion of syntax to Section 7.2. As another example, we only model computation as interpolation, binding, conditionals, and loops. We believe these constructs capture the core commonalities amongst the languages in Table 1. But this decision inevitably omits various features we consider ancillary to the document language design space. For instance, template DSLs like Liquid and Handlebars provide a special facility for accessing the index of a loop iteration, but we do not include that feature in the document calculus.

### 3.1 The String Calculus

A string  $s \in \text{String}$  is a sequence of characters  $c \in \text{Char}$ , such as “x” or “T” or “”. We will present a sequence of document calculi  $\mathcal{D}_\bullet^{\text{String}}$  in the string domain.

**3.1.1 String Literals.** The first level is the string literal calculus  $\mathcal{D}_{\text{Lit}}^{\text{String}}$ . For example, text files (left) and string literals (right) are both examples of document languages at this level:

```
1 I'm suspicious of "strings".
```

```
1 "I'm suspicious of \"strings\"."
```

Formally,  $\mathcal{D}_{\text{Lit}}^{\text{String}}$  has no computation and therefore the simplest semantics:

$$\text{Expr}_{\text{Lit}}^{\text{String}} e ::= s$$

$$\text{Type}_{\text{Lit}}^{\text{String}} \tau ::= \text{String}$$

**3.1.2 String Programs.** The next level is the string program calculus  $\mathcal{D}_{\text{Prog}}^{\text{String}}$ , which supports both string-specific operations (like concatenation) and domain-general operations (like variable binding). This level models general-purpose programming languages with support for strings, such as this Javascript program (left) and Q program (right):

<pre>1 // Javascript 2 let x = "a"; 3 x + "b" + x</pre>	<pre>1 / Q 2 x: "a" 3 x, "b", x</pre>
---	---------------------------------------

Formally,  $\mathcal{D}_{\text{Prog}}^{\text{String}}$  is System F with a base type of strings and a few features relevant for later levels. Namely, string concatenation, fixpoints, sums, products, recursive types, and existential types:

$$\begin{array}{l}
 \text{Variable } x \quad \text{Type Variable } \alpha \quad \text{Label } \ell \\
 \text{Expr}_{\text{Prog}}^{\text{String}} e ::= e_{\text{Lit}}^{\text{String}} \mid e_1 + e_2 \mid \lambda(x : \tau). e \mid e_1 e_2 \mid x \mid \mathbf{fix}(x : \tau). e \mid \mathbf{let} x = e_1 \mathbf{in} e_2 \mid \\
 \quad \{(\ell : e_\ell)^*\} \mid e.\ell \mid \mathbf{inject} e \mathbf{at} \ell \mathbf{as} \tau \mid \mathbf{case} e \{(\ell(x) \Rightarrow e_\ell)^*\} \\
 \quad \mathbf{fold}_\tau e \mid \mathbf{unfold}_\tau e \mid \Lambda\alpha. e \mid e[\tau] \mid \mathbf{pack} e \mathbf{as} \exists\alpha.\tau \mid \mathbf{unpack} (x, \alpha) = e_1 \mathbf{in} e_2 \\
 \text{Type}_{\text{Prog}}^{\text{String}} \tau ::= \tau_{\text{Lit}}^{\text{String}} \mid \tau_1 \rightarrow \tau_2 \mid \{(\ell : \tau_\ell)^*\} \mid \langle(\ell : \tau_\ell)^*\rangle \mid \forall\alpha.\tau \mid \mu\alpha.\tau \mid \exists\alpha.\tau \mid \alpha
 \end{array}$$

The static and dynamic semantics of all the features are standard, so we provide them in Appendix A.1 for reference. But as a simple example, the “aba” program can be written as follows:

$$\mathbf{let} x = \text{“a”} \mathbf{in} x + \text{“b”} + x \mapsto^* \text{“aba”}$$

Note that not all of these features are essential for dealing with strings, e.g., recursive types will only be useful for representing tree documents. But rather than scattering this part of the language definition throughout the levels, we opted to introduce all the relevant System F features here. This enables the development of later levels to focus more on the purely document-related features.

In the remainder of the paper, we will refer to an assumed standard library of common types and operations containing the following:

$$\begin{array}{l}
 \tau \text{ list} \triangleq \mu\alpha. \langle \text{nil} : () \mid \text{cons} : \{\text{hd} : \tau, \text{tail} : \alpha\} \rangle \\
 \ell_{\mu\alpha.\tau} e \triangleq \mathbf{fold}_{\mu\alpha.\tau} \mathbf{inject} e \mathbf{at} \ell \mathbf{as} \tau[\alpha \rightarrow \mu\alpha.\tau] \\
 \text{map} : \forall\alpha, \beta. (\alpha \rightarrow \beta) \rightarrow \alpha \text{ list} \rightarrow \beta \text{ list} \\
 \text{flatten} : \forall\alpha. \alpha \text{ list list} \rightarrow \alpha \text{ list} \\
 \text{append} : \forall\alpha. \alpha \text{ list} \rightarrow \alpha \text{ list} \rightarrow \alpha \text{ list} \\
 \text{join} : \text{String list} \rightarrow \text{String}
 \end{array}$$

**3.1.3 String Template Literals.** The next level is the string template literal calculus  $\mathcal{D}_{\text{TLit}}^{\text{String}}$ . For example, the “aba” program can be written in Javascript (left) and Python (right) programs using string template literals:

<pre>1 // Javascript 2 let x = "a"; 3 `\${x}b\${x}`</pre>	<pre>1 # Python 2 x = "a" 3 f"{x}b{x}"</pre>
---	--

String template literals are variously called “template strings”, “format strings”, and “interpolated strings.” We specifically use the term “string template *literals*” to draw a distinction with “string template *programs*”. Template literals only support positional interpolation of expressions, while



template programs support additional template-level features such as variable-bindings and loops. This distinction is useful because both template literals and template programs can be found in real-world systems.

Formally, templates in  $\mathcal{D}_{\text{TLit}}^{\text{String}}$  are a list of template parts which can be either literals (strings) or interpolated expressions. Within an expression, a template is invoked with the `strtpl` operator:

$$\begin{aligned} \text{Template}_{\text{TLit}}^{\text{String}} \ t &::= [p^*] \\ \text{TPart}_{\text{TLit}}^{\text{String}} \ p &::= s \mid e \\ \text{Expr}_{\text{TLit}}^{\text{String}} \ e &::= e_{\text{Prog}}^{\text{String}} \mid \text{strtpl } t \end{aligned}$$

For example, the expression ``${x}b${x}`` would parse into the abstract syntax `strtpl [x, "b", x]`.

Templates are fundamentally about providing a concise representation of document programs, as opposed to increasing the expressiveness of the document language (in the sense used by Felleisen [1991]). Therefore, we do not provide an operational semantics directly for templates, but rather provide a translation from templates to the underlying calculus. (We provide a static semantics in Section 5.1). More precisely, the translation is specified as a family of functions over syntax kinds  $\alpha$  with the form  $\langle \cdot \rangle_{\alpha} : \alpha \rightarrow \text{Expr}$ .

Here, we arrive at a critical design question: *how should templates translate to terms?* As described in the Scribble case study in Section 2.3, the particular choice of desugaring will influence how well templates compose with other language features. For example, a “direct” desugaring for string template literals might look like this:

$$\begin{aligned} \langle [p_1, \dots, p_n] \rangle_{\text{Template}} &\stackrel{?}{=} \langle p_1 \rangle_{\text{TPart}} + \dots + \langle p_n \rangle_{\text{TPart}} \\ \langle \text{strtpl } t \rangle_{\text{Expr}} &\stackrel{?}{=} \langle t \rangle_{\text{Template}} \end{aligned}$$

In this desugaring, a template desugars to a term of type `String` produced by the concatenation of desugared template parts, and the `strtpl` operator is just the identity. This desugaring is, in fact, a perfectly valid semantics for languages only supporting string template literals. For instance, the ECMAScript 2024 specification [Guo et al. 2023, §13.2.8.6] describes a roughly comparable evaluation strategy for ECMAScript template literals.

However, this direct desugaring does not easily generalize to higher levels of the document calculus. For instance, if a template part can be a variable binding `set x = e`, it is not obvious how to desugar the binding under this general style of semantics. Or if we want to repurpose templates to generate trees of strings rather than just strings, then we want the desugaring to not collapse template parts into a single string too early.

Therefore, the  $\mathcal{D}_{\text{TLit}}^{\text{String}}$  semantics are carefully designed to support later levels. That semantics is given by the following desugaring:

$$\begin{aligned} \langle [] \rangle_{\text{Template}} &= [] \\ \langle p :: ps \rangle_{\text{Template}} &= \langle p \rangle_{\text{TPart}} :: \langle ps \rangle_{\text{Template}} \\ \langle s \rangle_{\text{TPart}} &= s \\ \langle e \rangle_{\text{TPart}} &= \langle e \rangle_{\text{Expr}} \\ \langle \text{strtpl } t \rangle_{\text{Expr}} &= \text{join } \langle t \rangle_{\text{Template}} \end{aligned}$$

This desugaring represents a few key design decisions *vis-à-vis* the direct desugaring. First, templates desugar to lists rather than strings. Second, the `strtpl` operator is now responsible for converting the list to a string by desugaring to a join. And third, the desugaring of a template is



defined inductively, providing the opportunity for later desugarings of template parts to access the tail of the template.

Another possible desugaring for string template literals could follow the example of PHP by desugaring to effectful commands on a global mutable buffer. However, a pure semantics for templates avoids the compositionality issue described in Section 2.1 because the output of a template can be, for instance, combined with auxiliary data in a single data structure. Therefore we consider the effectful desugaring an anti-pattern and focus only on pure desugarings.

**3.1.4 String Template Programs.** The final level of the document calculus in the string domain is the string template program calculus  $\mathcal{D}_{\text{TProg}}^{\text{String}}$ . String template literals reduce the notation required to concatenate strings and expressions. However, complex templates often involve interpolating expressions which contain nested templates, requiring additional content delimiters. For example, compare the string template literal in Javascript (left) with the string template program in Jinja (right):

```

1 // Javascript
2 var l = [1, 2, 3]
3 ...
4 Examples of addition include:
5 ${
6   l.map(n => `* ${n} + 1 = ${n + 1}`)
7   .join("\n")
8 }
9 ...

```

```

1 {# Jinja #}
2 {% set l = [1, 2, 3] -%}
3 Examples of addition include:
4 {% for n in l -%}
5 * {{ n }} + 1 = {{ n + 1 }}
6 {% endfor %}

```

String template programs (more often called “template languages”) offer concision by lifting computations such as binding and looping into the template. Intuitively, the difference is that in a normal program, content is delimited from computation, such as with quotes or backticks. In a template program, computation is delimited from content, such as with `{% percents %}` in Jinja.

Formally,  $\mathcal{D}_{\text{TProg}}^{\text{String}}$  models this concept by adding support for if-expressions, set-statements, and foreach-loops:

$$\text{TPart}_{\text{TProg}}^{\text{String}} p ::= p_{\text{TLit}}^{\text{String}} \mid \text{set } x = e \mid \text{if } e \text{ then } t_1 \text{ else } t_2 \mid \text{foreach } e \{x. t\}$$

Set-statements must be desugared in the context of the rest of the template, so their desugaring is defined as special case over the syntactic kind Template rather than TPart:

$$\langle (\text{set } x = e) :: ps \rangle_{\text{Template}} = \text{let } x = e \text{ in } \langle ps \rangle_{\text{Template}}$$

Observe here the importance of defining the desugaring of a template *inductively* so as to permit such special cases, as opposed to independently desugaring each template part.

The semantics of **if** and **foreach** are definable at the  $\mathcal{D}_{\text{TProg}}^{\text{String}}$  level; however, we will delay introducing them until reaching the article domain (Section 3.2.4). These template parts contain nested templates and therefore desugar to nested lists, which requires a flattening/splicing mechanism to un-nest. The explanation of these mechanisms will be more enlightening when contrasted against article template programs as well as string template programs.

## 3.2 The Article Calculus

String template programs are the highest level of the document calculus in the domain of strings. Therefore, we can now proceed by enriching the domain with additional structure. The most

common form of structured document is an *attributed tagged tree* like this:

$$\text{Node } n ::= \text{text } s \mid \text{node } (s, [(s, s)^*], n^*)$$

For example, the introduction to this paper would be modeled like this:

```
node ("section", [{"id", "intro"}], [
  node ("header", [], [text "Introduction"]),
  node ("para", [], [text "We live in a golden age of document languages.", ...])
])
```

Such trees can naturally represent many kinds of documents (e.g., HTML websites, XML data structures). We focus on the subset of tagged trees that represent *articles*: a tree that consists of *block* nodes (e.g., sections, paragraphs) and *inline* nodes (e.g., plain text, bold text). More precisely, we define an article as an attributed tagged tree that adheres to this schema:

$$\begin{aligned} \text{Article } a &::= b^* \\ \text{Block } b &::= \text{node } ("para", [], \kappa) \mid \text{node } ("section", [], a) \\ \text{Text } \kappa &::= \ell^* \\ \text{Inline } \ell &::= \text{text } s \mid \text{node } ("bold", [], \kappa) \end{aligned}$$

For simplicity, this definition provides a minimal set of elements that are sufficient to model interesting aspects of article languages, to the exclusion of some common elements like italicized text and bulleted lists. We will introduce additional elements and attributes as needed, such as when discussing references in Section 4.1.

We will present a sequence of document calculi  $\mathcal{D}_{\bullet}^{\text{Article}}$  in the article domain, examining how the mechanisms previously examined for computing with strings can be lifted onto trees.

**3.2.1 Article Literals.** The lowest level in the article domain is the article literal calculus  $\mathcal{D}_{\text{Lit}}^{\text{Article}}$ , analogous to the string literal calculus. This level models document languages like Markdown (left) and HTML (right):

<pre>1 &lt;!-- Markdown --&gt; 2 3 Hello [world]! 4 5 [world]: https://example.com</pre>	<pre>1 &lt;!-- HTML --&gt; 2 &lt;p&gt; 3   Hello 4   &lt;a href="https://example.com"&gt;world&lt;/a&gt;! 5 &lt;/p&gt;</pre>
--	--

Formally,  $\mathcal{D}_{\text{Lit}}^{\text{Article}}$  just consists of article literals  $a$ :

$$\text{Expr}_{\text{Lit}}^{\text{Article}} ::= a$$

Note that even languages like Markdown are not fully literal. As in the example above, the feature of named URLs requires interpretation to resolve named references to URL definitions. We will discuss how to model this particular aspect further in Section 4.1.

**3.2.2 Article Programs.** The next level is the article program calculus  $\mathcal{D}_{\text{Prog}}^{\text{Article}}$  of programs that construct articles through libraries in the base language. Article APIs usually look either like imperative widget trees as in Javascript (left), or functional combinators over lists as in Elm (right):

```

1 // Javascript
2 let p = document.createElement("p");
3 let hello =
4   document.createTextNode("Hello ");
5 let world =
6   document.createElement("strong");
7 world.textContent = "world";
8 p.appendChild(hello);
9 p.appendChild(world);

```

```

1 -- Elm
2 import Html exposing (text, p, strong)
3 main = p []
4   [ text "Hello "
5     , strong [] [ text "world" ]
6   ]

```

Formally,  $\mathcal{D}_{\text{Prog}}^{\text{Article}}$  is the combination of  $\mathcal{D}_{\text{Prog}}^{\text{String}}$  and  $\mathcal{D}_{\text{Lit}}^{\text{Article}}$ :

$$\text{Expr}_{\text{Prog}}^{\text{Article}} e ::= e_{\text{Prog}}^{\text{String}} \mid e_{\text{Lit}}^{\text{Article}}$$

We model articles in  $\mathcal{D}_{\text{Prog}}^{\text{Article}}$  as a sum of text nodes and structure nodes:

$$\begin{aligned} \tau \text{ struct-node} &\triangleq \{\text{name} : \text{String}, \text{attrs} : (\text{String} \times \text{String}) \text{ list}, \text{children} : \tau\} \\ \text{NodeTy} &\triangleq \mu\alpha. \langle \text{text} : \text{String} \mid \text{node} : \alpha \text{ list struct-node} \rangle \end{aligned}$$

Note that the type system of  $\mathcal{D}_{\text{Prog}}^{\text{Article}}$  is expressive enough to evaluate  $\text{if} \cdot \vdash e : \text{NodeTy list}$  for some document program  $e$ . However, the type system is not expressive enough to determine that  $e$  is actually an article, e.g., that there are no block nodes nested in an inline node. This limitation is present in all existing document languages, except article literal languages like Markdown where only the article subset of trees is syntactically expressible.

**3.2.3 Article Template Literals.** The next level is the article template literal calculus  $\mathcal{D}_{\text{TLit}}^{\text{Article}}$ . Article template literals are analogous to string template literals — they are a pithy form of document constructor with support for expression interpolation, but they evaluate to articles rather than strings. The most common form of article template literal is either HTML syntax as in JSX Javascript (left), or XML syntax as in Scala 2 (also left) and Visual Basic .NET (right):

```

1 // JSX Javascript and Scala 2
2 var items =
3   Array("Milk", "Eggs", "Cheese");
4 <article>
5   <p>Today I am going shopping for:</p>
6   <ul>
7     {items.map(item =>
8       <li><p>{item}</p></li>)}
9   </ul>
10 </article>

```

```

1 ' VB.NET
2 Dim items() =
3   {"Milk", "Eggs", "Cheese"}
4 <article>
5   <p>Today I am going shopping for:</p>
6   <ul>
7     <%= From item in items
8       Select <li>
9         <p><%= item %></p>
10    </li> %>
11   </ul>
12 </article>

```

Observe that these syntaxes represent templates in part because all text inside the tags is unlimited, in contrast to the examples in Section 3.2.2.

Formally, the semantics of  $\mathcal{D}_{\text{TLit}}^{\text{Article}}$  are quite simple given our careful setup from earlier. In Section 3.1.3 we defined Template and TPart to model string templates in  $\mathcal{D}_{\text{TLit}}^{\text{String}}$ . We will reuse those features, adding a new template part for tree nodes, and adding a new template expression

for articles:

$$\text{Expr}_{\text{TLit}}^{\text{Article}} e ::= e_{\text{Prog}}^{\text{Article}} \mid e_{\text{TLit}}^{\text{String}} \mid \mathbf{treetpl} \ t$$

$$\text{TPart}_{\text{TLit}}^{\text{Article}} p ::= p_{\text{TLit}}^{\text{String}} \mid \mathbf{node} \ (s, [(s, e)^*], t)$$

$$\begin{aligned} \llbracket \mathbf{treetpl} \ t \rrbracket_{\text{Expr}} &= \llbracket t \rrbracket_{\text{Template}}^{\text{NodeTy}} \\ \llbracket s \rrbracket_{\text{TPart}}^{\text{NodeTy}} &= \mathbf{text}_{\text{NodeTy}} \ s \\ \llbracket \mathbf{node} \ (s, at, t) \rrbracket_{\text{TPart}}^{\text{NodeTy}} &= \mathbf{node}_{\text{NodeTy}} \ \{ \mathbf{name} : s, \ \mathbf{attrs} : [(s, \llbracket e \rrbracket_{\text{Expr}})^*], \ \mathbf{children} : \llbracket t \rrbracket_{\text{Template}} \} \end{aligned}$$

For example, the following expression evaluates to the article containing a single paragraph with the text “Hello **World**”:

```
let x = “World” in treetpl [node (“para”, [], [“Hello”, node (“bold”, [], [x])])]
```

The key idea is that **strtpl**  $t$  should desugar to an expression of type `String`, which in turn relies on  $\llbracket t \rrbracket_{\text{Template}}^{\text{String}}$  to desugar to an expression of type `String` list. For tree templates, we want **treetpl**  $t$  to desugar to an expression of type `NodeTy` list, which in turn relies on  $\llbracket t \rrbracket_{\text{Template}}^{\text{NodeTy}}$  to desugar to an expression of type `NodeTy` list. Therefore, there are two key differences in the desugarings of **strtpl** and **treetpl**:

- The desugaring of **strtpl** wraps the template in a join, while the desugaring of **treetpl** does not.
- The desugaring of **strtpl** has string literal template parts desugar to terms of type `String`, while those same parts desugar to terms of type `NodeTy` inside a **treetpl**.

To implement the latter detail, we modify the desugaring function to be context-aware, notated with the superscript  $\llbracket \cdot \rrbracket^{\tau}$ , where a template should desugar to a list containing elements of type  $\tau$ . The **treetpl** desugaring enters the `NodeTy` context, which is assumed to be carried through where not explicitly written out. The **strtpl** desugaring similar enters the `String` context, where string literals are desugared according to the rule:

$$\llbracket s \rrbracket_{\text{TPart}}^{\text{String}} = s$$

**3.2.4 Article Template Programs.** The final level is the article template program calculus  $\mathcal{D}_{\text{TProg}}^{\text{Article}}$  of article templates with if-expressions, set-statements and foreach-loops. Examples include Typst (left) and Svelte Javascript (right):

```
1 // Typst
2 #let items = ("Milk", "Eggs", "Cheese")
3
4 Today I am going shopping for:
5
6 #for item in items [
7   - #item
8 ]
```

```
1 // Svelte Javascript
2 <script>
3   let items = ["Milk", "Eggs", "Cheese"];
4 </script>
5 <article>
6   <p>Today I am going shopping for:</p>
7   <ul>
8     {#each items as item}
9     <li><p>{item}</p></li>
10    {/each}
11  </ul>
12 </article>
```

Formally,  $\mathcal{D}_{\text{TProg}}^{\text{Article}}$  requires no additional features, and can be generated by composing the last level of the string calculus with the previous level of the article calculus:

$$\text{Expr}_{\text{TProg}}^{\text{Article}} e ::= e_{\text{TLit}}^{\text{Article}} \mid e_{\text{TProg}}^{\text{String}}$$

For example, the shopping list program can be expressed as the following template (imagining the article domain is enriched with bulleted lists and list items):

```
treetpl [
  set items = ["Milk", "Eggs", "Cheese"],
  node ("para", [], [text "Today I am going shopping for"]),
  node ("list", [], [foreach items {item. [node ("item", [], [item])]}])
]
```

Now we have sufficient points of reference to return to the question first raised in Section 3.1.4: what are the semantics of **foreach** and **if**? The crux of the problem is that these constructs contain nested templates, which affects the *dimensionality* of the term desugared from a template. To explain, consider a simple desugaring of **foreach** into a map:

$$(\text{foreach } e \{x. t\})_{\text{TPart}} \stackrel{?}{=} \text{map } (\lambda x. (t)_{\text{Template}}) (e)_{\text{Expr}}$$

Then consider the behavior of the example tree template under the simple desugaring. In particular, observe this part:

$$\begin{aligned} & ((\text{node } ("list", [], [\text{foreach items } \{item. [\text{node } ("item", [], [item])\}]))_{\text{TPart}} \\ &= \text{node } ("list", [], [\text{map } (\lambda item. [\text{node } ("item", [], [item])\}) \text{items}]) \\ &\xrightarrow{*} \text{node } ("list", [], [[[\text{node } ("item", [], ["Milk"]), \dots]]) \end{aligned}$$

Note that the child list of the “list” node is 3-dimensional! That certainly does not match the article schema provided at the top of Section 3.2. Any document language with a **foreach**-loop must somehow flatten the node list to one dimension; the key design question is where in the pipeline this should happen. A few different approaches can be found in existing languages:

*Avoid nested node lists with an imperative template semantics.* For example, Svelte will translate the shopping list program into 132 lines of Javascript. Part of that translation is a function that constructs the DOM in an imperative manner, like this:

```
1 insert(target, article, anchor);
2 append(article, p);
3 append(article, t1);
4 append(article, ul);
5
6 for (let i = 0; i < each_blocks.length; i += 1) {
7   each_blocks[i].m(ul, null);
8 }
```

While this translation avoids the issue of list dimensionality, the use of imperative template semantics can cause issues as described for PHP in Section 2.1. In the case of Svelte, one limitation is that templates cannot be nested inside expressions. For instance, this program is not valid Svelte:

```
1 <ul>
2   {items.map(item => <li><p>{item}</p></li>)}
3 </ul>
```

Avoid nested node lists with *unquote-splicing*. Quasiquotes are a kind of template language where the unquote-splicing operator can be used to reduce list dimensionality. For example, the shopping list could be written in Clojure with the Hiccup library ([weavejester.github.io](http://weavejester.github.io)) like this (noting the `~@` unquote-splice):

```
1 (def items ["Eggs", "Milk", "Cheese"])
2 (def item-lis (map #(h/html [:li [:p %]]) items))
3 (eval `(h/html [:ul ~@item-lis]))
```

In  $\mathcal{D}_{\text{TProg}}^{\text{Article}}$ , this strategy is modeled by introducing unquote-splicing via a **splice** template part:

$$\text{TPart}_{\text{TProg}}^{\text{Article}} p ::= p_{\text{TLit}}^{\text{Article}} \mid p_{\text{TProg}}^{\text{String}} \mid \mathbf{splice} e$$

$$\llbracket (\mathbf{splice} e) :: ps \rrbracket_{\text{Template}} = \text{append} (\llbracket e \rrbracket_{\text{Expr}}) (\llbracket ps \rrbracket_{\text{Template}})$$

Like a set-statement, a **splice** desugars in context as an append of the spliced head expression to the tail of the desugared template. Then **foreach** and **if** can be desugared into splices:

$$\begin{aligned} \llbracket (\mathbf{foreach} e \{x. t\}) :: ps \rrbracket_{\text{Template}}^{\text{NodeTy}} &= \llbracket (\mathbf{splice} \text{flatten} (\text{map} (\lambda x. \llbracket t \rrbracket_{\text{Template}}) (\llbracket e \rrbracket_{\text{Expr}})) :: ps) \rrbracket_{\text{Template}} \\ \llbracket (\mathbf{if} e \text{ then } t_1 \text{ else } t_2) :: ps \rrbracket_{\text{Template}}^{\text{NodeTy}} &= \llbracket (\mathbf{splice} \text{if} (\llbracket e \rrbracket_{\text{Expr}}) \text{ then } (\llbracket t_1 \rrbracket_{\text{Template}}) \text{ else } (\llbracket t_2 \rrbracket_{\text{Template}})) :: ps \rrbracket_{\text{Template}} \end{aligned}$$

Because the desugaring generates another template (without **foreach** or **if** but with **splice**), the desugaring function must be recursively invoked. We prove that this desugaring produces well-typed terms (with well-typed inputs) in Section 5.1.

Permit nested node lists as a document IR, and flatten the IR later. For example, document systems like Scribble, Typst, React, and ScalaTags ([com-lihaoyi.github.io/scalatags](http://com-lihaoyi.github.io/scalatags)) provide a document IR that permits arbitrary levels of nesting. After a document program is interpreted to a value in the IR, a visitor sweeps through each node and recursively flattens all node lists.

In  $\mathcal{D}_{\text{TProg}}^{\text{Article}}$ , this strategy can be modeled by introducing the concept of a *fragment* as an arbitrarily nested list of document content:

$$\begin{aligned} \tau \text{ fragment} &\triangleq \mu \alpha. \langle \text{base} : \tau \mid \text{children} : \alpha \text{ list} \rangle \\ \text{FNode} &\triangleq \mu \alpha. \langle \text{text} : \text{String} \mid \text{node} : \alpha \text{ fragment struct-node} \rangle \\ \text{NodeFrag} &\triangleq \text{FNode fragment} \end{aligned}$$

The desugaring for **foreach** and **if** can then follow the “simple” desugaring described earlier, with some additional constructors to build terms of the appropriate type:

$$\text{Expr}_{\text{TProg}}^{\text{Article}} e ::= \dots \mid \mathbf{fragtpl} t$$

$$\begin{aligned} \llbracket (\mathbf{fragtpl} t) \rrbracket_{\text{Expr}} &= \text{elim-frags} (\text{children}_{\text{NodeFrag}} (\llbracket t \rrbracket_{\text{Template}}^{\text{NodeFrag}})) \\ \llbracket (s) \rrbracket_{\text{TPart}}^{\text{NodeFrag}} &= \text{base}_{\text{NodeFrag}} (\text{text}_{\text{FNode}} s) \\ \llbracket (\mathbf{node} (s, at, t)) \rrbracket_{\text{TPart}}^{\text{NodeFrag}} &= \text{base}_{\text{NodeFrag}} (\text{node}_{\text{FNode}} (s, at, (\llbracket t \rrbracket_{\text{Template}}))) \\ \llbracket (\mathbf{foreach} e \{x. t\}) \rrbracket_{\text{TPart}}^{\text{NodeFrag}} &= \text{children}_{\text{NodeFrag}} (\text{map} (\lambda x. \text{children}_{\text{NodeFrag}} (\llbracket t \rrbracket_{\text{Template}}) e)) \\ \llbracket (\mathbf{if} e \text{ then } t_1 \text{ else } t_2) \rrbracket_{\text{TPart}}^{\text{NodeFrag}} &= \text{children}_{\text{NodeFrag}} (\text{if} (\llbracket e \rrbracket_{\text{Expr}}) \text{ then } (\llbracket t_1 \rrbracket_{\text{Template}}) \text{ else } (\llbracket t_2 \rrbracket_{\text{Template}})) \end{aligned}$$

It is worth noting that these semantics would be simpler in a dynamically-typed language, as a nested list could be expressed with the standard list type rather than a bespoke fragment type, and all the fragment constructors would be obviated.

Under this semantics, a template  $t$  should desugar to a term of type `NodeFrag`. To flatten the fragment, we introduce a function `elim-frags : NodeFrag → NodeTy list`:

$$\begin{aligned} \text{elim-frags } (\text{base}_{\text{NodeFrag}} (\text{text}_{\text{FNode}} s)) &= [\text{text}_{\text{NodeTy}} s] \\ \text{elim-frags } (\text{base}_{\text{NodeFrag}} (\text{node}_{\text{FNode}} (s, at, f))) &= [\text{node}_{\text{NodeTy}} (s, at, \text{elim-frags } f)] \\ \text{elim-frags } (\text{children}_{\text{NodeFrag}} l) &= \text{flatten } (\text{map } \text{elim-frags } l) \end{aligned}$$

Both the splicing and fragment strategies are reasonable ways to deal with the template dimensionality problem in a functional manner. It is ultimately a matter of taste as to which should be used in practice. The unquote-splicing strategy seems intuitively cleaner from the perspective of the language designer (no messy intermediate representation), although the fragment strategy seems nicer from the perspective of the document author (no worrying about getting just the right combination of splices).

Note as well that any of these strategies will avoid Scribble’s dropping-elements issue described in Section 2.3. The key idea is that templates desugar to lists, and in System F a list of strings or nodes cannot be mistaken for a sequence of expressions (unlike in Racket with its “implicit begin”). An equivalent  $\mathcal{D}_{\text{TProg}}^{\text{Article}}$  program to the one in Section 2.3 would produce the expected output.

## 4 EXTENDING THE DOCUMENT CALCULUS

Section 3 described the semantics of the document calculus, arguing that it models the features of popular document languages. The next two sections demonstrate how the document calculus can provide a foundation for describing higher-level document features and for reasoning about document programs. In this section, we extend the document calculus with three interesting document features: references, reforestation, and reactivity. Each feature requires a non-trivial change to the language’s semantics — references require staged computation (Section 4.1), reforestation requires a global analysis of document structure (Section 4.2), and reactivity requires a complex runtime (Section 4.3). We also provide an OCaml implementation of each feature in the supplemental materials (note that these implementations are shallowly embedded so as to avoid the verbosity of System F).

### 4.1 References

A common feature in document languages is to support identifiers on nodes that can be referred to elsewhere in a document. Specifically, we consider an extension to the article schema where sections can have string identifiers, and a `ref` element can refer to a section:

$$\begin{aligned} \text{Block } b ::= \dots \mid \text{node } (\text{"section"}, [(\text{"id"}, s)], a) \\ \text{Inline } \ell ::= \dots \mid \text{node } (\text{"ref"}, [(\text{"target"}, s)], []) \end{aligned}$$

The intended semantics are comparable to  $\text{T}_{\text{E}}\text{X}$ ’s, i.e., the displayed content of a section reference should be the number of the referenced section. This feature brings two challenges: checking for invalid references, and computing the content of a reference.

**4.1.1 Reference Validity.** As alluded to in Section 3.2.2, there are multiple conceptions of validity when thinking about document programs. For example, one form of validity is well-typedness of the input: a document expression  $e$  is valid if  $\cdot \vdash e : \text{NodeTy list}$ . Another form of validity is well-formedness of the output: a document expression  $e$  is valid if  $e \mapsto^* v$  and  $v \in \text{Article}$ . In the



wild, validity sometimes means parseability: the CommonMark specification [MacFarlane 2021] for Markdown states that “any sequence of characters is a valid CommonMark document”.

As the document domain is enriched with additional structure, well-formedness becomes an insufficient criterion for document validity. In the case of references, an article is not valid if it references an unknown identifier, analogous to a free variable. Therefore, we need to model validity via an auxiliary judgment that captures whether a document is valid beyond its syntactic structure.

Formally, we model reference validity first by constructing a identifier context:

$$\text{IdCtx} \triangleq (\text{String} * \text{int list}) \text{ list} \quad \Delta : \text{IdCtx}$$

This context maps identifiers to section numbers. We construct  $\Delta$  via the function:

$$\text{section-ids-at-depth}_\alpha : \text{int list} \rightarrow \alpha \rightarrow \text{IdCtx} * \text{int list}$$

The key case that deals with sections is as follows:

$$\begin{aligned} & \text{section-ids-at-depth}_{\text{NodeTy}} (k :: k^*) (\text{node}_{\text{NodeTy}} (\text{“section”}, [(\text{“id”}, id), \dots], \text{children})) = \\ & \quad \mathbf{let} (\Delta, \_) = \text{section-ids-at-depth}_{\text{NodeTy list}} (1 :: k :: ks) \text{ children in} \\ & \quad \mathbf{let} \Delta' = (id, k :: ks) :: \Delta \text{ in} \\ & \quad (\Delta', (k + 1) :: ks) \end{aligned}$$

In this case, given a current section numbering  $k :: ks$ , the section’s children are analyzed with a fresh subsection counter placed on the stack. The identifier context is updated with the current section’s ID, and the section number is incremented.

Let  $\text{section-ids}(n) = \text{section-ids-at-depth}(n, [1]).0$ . Then we can define a validity judgment  $\Delta \vdash \cdot \text{ valid}$ , where an article  $a$  is valid if  $\text{section-ids}(a) \vdash a \text{ valid}$ . Two representative inference rules are as follows:

$$\frac{\Delta \vdash b_1 \text{ valid} \quad \dots \quad \Delta \vdash b_n \text{ valid}}{\Delta \vdash [b_1, \dots, b_n] \text{ valid}} \quad \frac{s \in \text{dom}(\Delta)}{\Delta \vdash \text{node}_{\text{NodeTy}} (\text{“ref”}, [(\text{“target”}, s)], []) \text{ valid}}$$

The full validity judgment is provided in the OCaml implementation.

Representing references in documents has a similar flavor to representing binders in deeply embedded languages [Cave and Pientka 2012; Licata and Harper 2009], and could in theory be addressed with similar techniques. One important difference is that in documents, both identifiers and references can be placed anywhere in the document; referential structure is not strictly hierarchical as with lexically-scoped variables.

**4.1.2 Reference Content.** The validity judgment must notably be expressed in two stages – one to collect a context of identifiers ( $\text{section-ids}(a)$ ), and one to check for validity in that context ( $\Delta \vdash a \text{ valid}$ ). Similarly, the content of a reference must be generated in two stages. In  $\mathbb{L}\text{T}\text{E}\text{X}$ , for example, a reference to the next section in this document like `\ref{sec:reforesting}` will be replaced by the text “4.2”. This operation is non-local, because the document language cannot know the section number of a forward reference at the point of reference. Most document languages accomplish this task with a second pass over the document, such as the `.aux` file generated by LaTeX on a first-pass which is later rendered on a second-pass.

We model the generation of reference content in the document calculus as follows. Say that  $\cdot \vdash e : \text{NodeTy list}$  and  $e \mapsto^* v$ . Then we compute the final document  $v' = \text{render-refs}(v)$  where

$\text{replace-refs}_\alpha : \text{IdCtx} \rightarrow \alpha \rightarrow \alpha$  is a visitor over articles, with the key case as follows:

$$\begin{aligned} \text{replace-refs}_{\text{NodeTy}} (\Delta : \text{IdCtx}) (\text{node}_{\text{NodeTy}} (\text{"ref"}, [(\text{"target"}, id)], [], \Delta) : \text{NodeTy}) = \\ \text{text}_{\text{NodeTy}} (\text{section-number-to-string } \Delta[id]) \\ \text{render-refs } (d : \text{NodeTy}) = \text{replace-refs } (\text{section-ids } d) \ d \end{aligned}$$

More broadly, this extension reflects a key aspects of computing with documents: the validity and content of a document program’s output can be a global property of program, which requires commensurate features for non-local computation.

## 4.2 Reforestation

Another instance of non-local computation in documents is *reforestation*. In some article document languages, the document structure expressed by the programmer is often quite different from the final generated document structure. For example, languages like Typst, Scribble, and Markdown do not require paragraphs to be explicitly wrapped in a tag like `<p>`; rather, paragraphs are inferred based on line breaks. Another common operation is to permit the programmer to write sections linearly, and then to reconstruct the section hierarchy by grouping content between pairs of headers.

In a document language with reforestation, the user writes a template program which is initially evaluated into a “raw” document tree that is not a syntactically-valid article, but where all expressions have been reduced to a value. Then a second pass “reforests” the raw document into a syntactically-valid article by analyzing the global document structure of the input. For example, the decode function in Scribble [Flatt et al. 2009, p. 113] implements this functionality.

To model reforestation in the article calculus, we add a `flowtpl` primitive for reforested tree templates, which desugars into a tree template wrapped in a call to a reforest function:

$$\begin{aligned} \text{Expr}_{\text{TProg}}^{\text{Article}} e ::= \dots \mid \text{flowtpl } t \\ (\text{flowtpl } t)_{\text{Expr}} = \text{reforest } (\text{treetpl } t)_{\text{Expr}} \ [] \end{aligned}$$

The key detail is the implementation of  $\text{reforest} : \text{NodeTy list} \rightarrow \text{NodeTy list} \rightarrow \text{NodeTy list}$ . The specifics vary between languages, but a simple example that we can implement for  $\mathcal{D}_{\text{TProg}}^{\text{Article}}$  will collect inline elements into paragraphs. The function  $\text{reforest}$  iterates through a list of nodes  $n^*$  with an accumulator for the current paragraph  $par$ . It emits paragraphs upon encountering the end of a list, a double newline (as in Markdown), or a block node. For example, the document on the left would be reforested to the document on the right:

$$\begin{array}{ll} [\text{text}_{\text{NodeTy}} \text{"Hello"}, & [\text{node}_{\text{NodeTy}} (\text{"para"}, [], \\ \text{text}_{\text{NodeTy}} \text{"World"}, & \quad [\text{text}_{\text{NodeTy}} \text{"Hello"}, \text{text}_{\text{NodeTy}} \text{"World"}]), \\ \text{text}_{\text{NodeTy}} \text{"\n\n"}, & \text{node}_{\text{NodeTy}} (\text{"figure"}, [], [\dots]), \\ \text{node}_{\text{NodeTy}} (\text{"figure"}, [], [\dots]), & \text{node}_{\text{NodeTy}} (\text{"para"}, [], [\text{text}_{\text{NodeTy}} \text{"Post-figure"}])] \\ \text{text}_{\text{NodeTy}} \text{"Post-figure"}] & \end{array}$$

The precise definition of `reforest` is as follows:

$$\begin{aligned}
 \text{reforest } [] \text{ } par &= [\text{node}_{\text{NodeTy}} \text{ ("para", [], rev } par)] \\
 \text{reforest } (\text{text}_{\text{NodeTy}} \text{ "\n\n"} :: ns) \text{ } par &= (\text{node}_{\text{NodeTy}} \text{ ("para", [], rev } par)) :: \text{reforest } ns [] \\
 \text{reforest } ((\text{text}_{\text{NodeTy}} \text{ } s) :: ns) \text{ } par &= \text{reforest } ns (\text{text}_{\text{NodeTy}} \text{ } s :: par) \\
 \text{reforest } (\text{node}_{\text{NodeTy}} \text{ } (s, at, ns_c) :: ns) \text{ } par &= \begin{cases} \text{node}_{\text{NodeTy}} \text{ ("para", [], rev } par) :: \\ \text{node}_{\text{NodeTy}} \text{ } (s, at, \text{reforest } ns_c []) :: \\ \text{reforest } ns [] & \text{if is-block } s \\ \text{reforest } ns (\text{node}_{\text{NodeTy}} \text{ } (s, at, cs) :: par) & \text{otherwise} \end{cases}
 \end{aligned}$$

Reforestation again demonstrates how document computation requires analysis of the global structure of a document, such as by accumulating sequential elements into groups. The correctness condition for the `reforest` function is that it must generate a valid document that adheres to the Article schema. This condition generally assumes that the input `NodeTy` list also adheres to some intermediate schema as a precondition. For instance, the implementation above assumes that the input is already valid and does not, say, contain block nodes within inline nodes. A more aggressive implementation could attempt to repair an invalid document by, say, reordering invalid node nests. But in practice, document repair is most often performed during parsing rather than a later stage, as in HTML and Markdown.

### 4.3 Reactivity

Modern documents, especially those in the browser, can be *reactive* to signals such as a timer or user input. Such reactions include animations, interactive widgets, and explorable explanations. Many recently-developed document languages focus on reactivity, as we discuss in Section 6.4. Therefore, it would be valuable to model reactivity within the document calculus. This model enables us to reason about how reactivity interacts with features like references, as we will discuss in Section 5.2.

We model reactivity by blending ideas from two popular UI frameworks. First, we adopt functional reactive programming for UIs as in Elm [Czaplicki and Chong 2013], i.e., purely functional state management via signals. FRP is an appropriate paradigm for the document subset of UIs, and its purely functional nature fits well into our purely functional calculus. Second, we adopt UI components as in React, i.e., encapsulating model and view into a single object. Most existing reactive document languages use components (except Elm), so we reflect that fact in the model.

At the core of the reactive model are the types of components, instances, and nodes:

$$\begin{aligned}
 (\tau_{\text{props}}, \tau_{\text{state}}) \text{ component} &\triangleq \{ \text{init} : \tau_{\text{props}} \rightarrow \tau_{\text{state}}, \\ &\quad \text{update} : \text{Signal} \times \tau_{\text{state}} \rightarrow \tau_{\text{state}}, \\ &\quad \text{view} : \tau_{\text{state}} \rightarrow \text{ReactNode} \} \\
 \text{instance} &\triangleq \exists \tau_{\text{props}}, \tau_{\text{state}}. \\ &\quad \{ \text{id} : \text{InstId}, \\ &\quad \text{com} : (\tau_{\text{props}}, \tau_{\text{state}}) \text{ component} \\ &\quad \text{props} : \tau_{\text{props}}, \\ &\quad \text{state} : \tau_{\text{state}}, \\ &\quad \text{node} : \text{ReactNode} \}
 \end{aligned}$$

$$\text{ReactNode} \triangleq \mu \alpha. \langle \text{text} : \text{String} \mid \text{node} : \alpha \text{ list struct-node} \mid \text{inst} : \text{instance} \rangle$$

A component is described by an `init` function that converts properties into an initial state. The `update` function handles a signal (which we just assume `Signal = String` for simplicity) and returns an updated state. The `view` function returns a reactive node for a given state.

An instance is a particular reification of a component. It internally maintains a local state, and a child view, along with a unique ID that is generated by a function  $\text{gen-id} : () \rightarrow \text{InstId}$ . A component is instantiated with the following function:

```

instantiate  $\triangleq$   $\Lambda \tau_{\text{props}}, \tau_{\text{state}}. \lambda(\text{com} : (\tau_{\text{props}}, \tau_{\text{state}}) \text{ component}). \lambda(\text{props} : \tau_{\text{props}}).$ 
  let  $\text{state} = \text{com.init props}$  in
  let  $\text{node} = \text{com.view state}$  in
  pack  $\{\text{id} : \text{gen-id } (), \text{com}, \text{props}, \text{state}, \text{node}\}$  as instance

```

Finally, a reactive node `ReactNode` is a `NodeTy` with an additional case for instances. To integrate instances into templates, we add a `reacttpl` expression and a `component` template part:

$$\begin{aligned} \text{Expr}_{\text{TProg}}^{\text{Article}} e &:: \dots \mid \text{reacttpl } t \\ \text{TPart } p &:: \dots \mid \text{component } e_1 e_2 \\ \llbracket \text{reacttpl } t \rrbracket_{\text{Expr}} &= \llbracket t \rrbracket_{\text{Template}}^{\text{ReactNode}} \\ \llbracket \text{component } e_1 e_2 \rrbracket_{\text{TPart}}^{\text{ReactNode}} &= \text{inst}_{\text{ReactNode}} (\text{instantiate } e_1 e_2) \end{aligned}$$

For example, the following document uses a counter component that appends to a string every time the component is clicked:

```

let  $\text{counter} : (\text{String}, \text{String} \times \text{String}) \text{ component} = \{$ 
   $\text{init} : \lambda p. (p, \text{""}),$ 
   $\text{update} : \lambda e. \lambda(p, s). \text{if } e = \text{"click"} \text{ then } (p, p + s) \text{ else } (p, s),$ 
   $\text{view} : \lambda(\_, s). \text{text}_{\text{NodeTy}} s$ 
 $\}$  in
reacttpl  $[\text{"The number of clicks is"}, \text{component counter } \text{"}"]$ 

```

To make the document reactive, we must provide it a runtime. The runtime consists of two functions:

- `doc-step` :  $(\text{InstId}, \text{Signal}) \text{ map} \rightarrow \text{ReactNode} \rightarrow \text{ReactNode}$  takes a reactive document and a set of signals for each instance, and updates the state of each component with the signal.
- `doc-view` :  $\text{ReactNode} \rightarrow \text{NodeTy}$  replaces instance nodes with their children, creating the final article to display.

Starting with an initial reactive document program  $d_0$ , the runtime iteratively generates views and steps in the following pattern:

$$\begin{array}{ccccc} d_0 & \xrightarrow{\text{doc-step}} & d_1 & \xrightarrow{\text{doc-step}} & d_1 & \xrightarrow{\text{doc-step}} & \dots \\ \downarrow \text{doc-view} & & \downarrow \text{doc-view} & & \downarrow \text{doc-view} & & \\ a_0 & & a_1 & & a_2 & & \end{array}$$

The key case of the step function for instances is as follows:

```

doc-step (signals : (InstId, Signal) map) (instReactNode inst : ReactNode) =
  if inst.id ∉ signals then instReactNode {inst with node = doc-step signals inst.node} else
  let state' = inst.com.update (signals[inst.id]) inst.state in
  let node' = inst.com.view state' in
  let node'' = reconcile signals node node' in
  instReactNode {inst with state'; children''}

```

When an instance receives a signal, then the update function generates the new state, and the view function generates the new view. However, simply returning the new view would erase all the state contained in child instances. Therefore, we must *reconcile* the old and new views, expressed with the reconcile function. The key case is as follows:

$$\text{reconcile } \text{signals } (\text{inst}_{\text{ReactNode}} i) (\text{inst}_{\text{ReactNode}} i') = \begin{cases} \text{doc-step } \text{signals } (\text{inst}_{\text{ReactNode}} i) & \text{if } i.\text{com} = i'.\text{com} \wedge i.\text{props} = i'.\text{props} \\ \text{inst}_{\text{ReactNode}} i' & \text{otherwise} \end{cases}$$

If an instance has the same component and properties as before<sup>1</sup>, then reconciliation persists its state and recursively steps the instance. Otherwise, the new instance is returned.

Finally, the view function eliminates all instances from the node tree, with the key case as follows:

$$\text{doc-view } (\text{inst}_{\text{ReactNode}} i) = \text{doc-view } i.\text{node}$$

This runtime system is sufficient to model an Elm/React-like reactive document language, including per-component state and reconciliation on state updates. We provide an example of formal reasoning about this system in Section 5.2.

## 5 REASONING WITH THE DOCUMENT CALCULUS

Finally, we demonstrate the value of the document calculus as a *formal* foundation (in addition to being a conceptual foundation) by reasoning about the semantics of document programs. Specifically, we prove two theorems: first, we prove that the template desugaring always produces terms of the correct type (Section 5.1). Second, we show how to design a provably correct implementation strategy for efficiently composing references and reactivity (Section 5.2).

### 5.1 Templates Desugar to Well-Typed Terms

We would like to be able to say that our particular desugaring of templates is “correct” by some metric. For example, the **foreach** desugaring in Section 3.2.4 involves both a **splice** and a **flatten** — we should be unable to prove some correctness theorem if the desugaring omitted either construct.

One such theorem is the statement that *templates desugar to well-typed terms*. Specifically, a sugared expression **trreetpl**  $t$  should desugar to an expression with type `NodeTy list`. A sugared expression **strtpl**  $t$  should desugar to an expression with type `String`. Of course, desugared template terms are only well-typed if used properly. For instance, a program cannot interpolate an expression of the wrong type, or use a unbound variable. That is to say: well-typed inputs lead templates to desugar to well-typed terms.

To capture these ideas, we extend the type system to describe the types of templates. The typing judgments for templates are systematically constructed from their desugaring, roughly following

<sup>1</sup>Note that because  $i$  and  $i'$  are existentially-typed, it is not inherently type-safe to compare their fields. The runtime must provision some way of first determining whether the two instances have the same type.

the “type resugaring” method developed by [Pombrio and Krishnamurthi \[2018\]](#). To deal with the fact that template desugaring is dependent on context, we add a new kind of fact to the typing context that indicates the current template context:

$$\Gamma ::= \dots \mid \text{tpl } \tau$$

Then the typing rules for templates are as follows:

$$\begin{array}{c}
\text{T-STRTPL} \\
\frac{\Gamma, \text{tpl } \text{String} \vdash t : \text{String list}}{\Gamma \vdash \text{strtpl } t : \text{String}}
\end{array}
\quad
\begin{array}{c}
\text{T-TREETPL} \\
\frac{\Gamma, \text{tpl } \text{NodeTy} \vdash t : \text{NodeTy list}}{\Gamma \vdash \text{treetpl } t : \text{NodeTy list}}
\end{array}
\quad
\begin{array}{c}
\text{T-NILTPL} \\
\frac{}{\Gamma, \text{tpl } \tau \vdash [] : \tau \text{ list}}
\end{array}$$
  

$$\begin{array}{c}
\text{T-CONSTPL} \\
\frac{\Gamma, \text{tpl } \tau \vdash p : \tau \quad \Gamma, \text{tpl } \tau \vdash ps : \tau \text{ list}}{\Gamma, \text{tpl } \tau \vdash (p :: ps) : \tau \text{ list}}
\end{array}
\quad
\begin{array}{c}
\text{TP-STR} \\
\frac{}{\Gamma, \text{tpl } \tau \vdash s : \tau}
\end{array}
\quad
\begin{array}{c}
\text{TP-NODE} \\
\frac{\Gamma, \text{tpl } \tau \vdash t : \tau \text{ list} \quad \forall i. \Gamma \vdash e_i : \text{String}}{\Gamma, \text{tpl } \tau \vdash \text{node } (s, [(s_i, e_i)*], t) : \tau}
\end{array}$$
  

$$\begin{array}{c}
\text{TP-SET} \\
\frac{\Gamma \vdash e : \tau_e \quad \Gamma, \text{tpl } \tau, x : \tau_e \vdash ps : \tau \text{ list}}{\Gamma, \text{tpl } \tau \vdash (\text{set } x = e :: ps) : \tau \text{ list}}
\end{array}
\quad
\begin{array}{c}
\text{TP-SPLICE} \\
\frac{\Gamma \vdash e : \tau \text{ list} \quad \Gamma, \text{tpl } \tau \vdash ps : \tau \text{ list}}{\Gamma, \text{tpl } \tau \vdash (\text{splice } e :: ps) : \tau \text{ list}}
\end{array}$$
  

$$\begin{array}{c}
\text{TP-FOREACH} \\
\frac{\Gamma \vdash e : \tau_e \text{ list} \quad \Gamma, \text{tpl } \tau, x : \tau_e \vdash t : \tau \text{ list} \quad \Gamma, \text{tpl } \tau \vdash ps : \tau \text{ list}}{\Gamma, \text{tpl } \tau \vdash (\text{foreach } e \{x. t\} :: ps) : \tau \text{ list}}
\end{array}
\quad
\begin{array}{c}
\text{TP-IF} \\
\frac{\Gamma \vdash e : \text{bool} \quad \Gamma, \text{tpl } \tau \vdash t_1 : \tau \text{ list} \quad \Gamma, \text{tpl } \tau \vdash t_2 : \tau \text{ list} \quad \Gamma, \text{tpl } \tau \vdash ps : \tau \text{ list}}{\Gamma, \text{tpl } \tau \vdash (\text{if } e \text{ then } t_1 \text{ else } t_2 :: ps) : \tau \text{ list}}
\end{array}$$

In general, all templates desugar to terms of type  $\tau$  list for some element type  $\tau$ . The rules for each template part lay out the conditions under which the overall template is well-typed. For example, a **foreach** is well-typed if the input  $e$  is a list, if the nested template  $t$  is well-typed under the binding  $x$ , and if the tail  $ps$  is well-typed. If the rules are formulated correctly, then the following theorem should hold:

**THEOREM 5.1 (DESUGARING PRESERVES TYPES).** *Let  $e \in \text{Expr}_{\text{TPProg}}^{\text{Article}}$ . If  $\Gamma \vdash e : \tau$  then  $\Gamma \vdash \langle e \rangle_{\text{Expr}} : \tau$ .*

We give the full proof by induction over the derivation of  $\Gamma \vdash e : \tau$  in Appendix A.2, but here we can provide the intuition for one case, again focusing on **foreach**. Recall the desugaring of **foreach** in **treetpl**:

$$\begin{aligned}
\langle (\text{foreach } e \{x. t\} :: ps) \rangle_{\text{Template}}^{\text{NodeTy}} &= \langle (\text{splice flatten (map } (\lambda x. \langle t \rangle_{\text{Template}}) \langle e \rangle_{\text{Expr}})) :: ps \rangle_{\text{Template}} \\
&= \text{append (flatten (map } (\lambda x. \langle t \rangle_{\text{Template}}) \langle e \rangle_{\text{Expr}})) \langle ps \rangle_{\text{Template}}
\end{aligned}$$

By the inductive hypothesis, we can assume that:

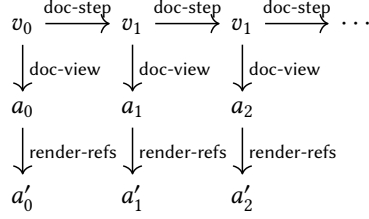
$$\Gamma \vdash \langle e \rangle_{\text{Expr}} : \tau_e \text{ list} \quad \Gamma, x : \tau_e \vdash \langle t \rangle_{\text{Template}} : \tau \text{ list} \quad \Gamma \vdash \langle ps \rangle_{\text{Template}} : \tau \text{ list}$$

Then the type of the desugared term can be systematically derived in standard fashion. The map term has type  $\tau$  list list. The flatten term therefore has type  $\tau$  list. The append term therefore has type  $\tau$  list. We conclude that the sugared and desugared terms have the same type.

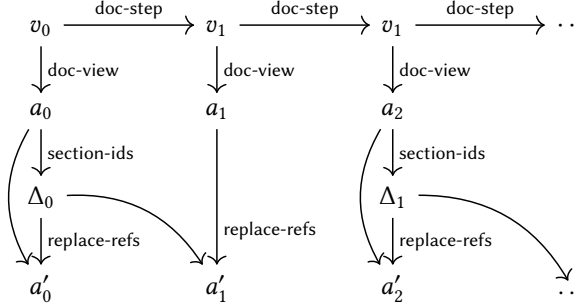
## 5.2 Correctly Composing References and Reactivity

As shown in the React case study in Section 2.2, it requires careful thought to correctly compose efficient reactivity with document features like section references. The content of a reference is a global property of a document based on the number of sections and the location of each section label. This global computation is conceptually at odds with reactivity, which is oriented towards localizing computation to components that aren't aware of their sibling or parent components.

A simple approach to composing these extensions is to postprocess every reactively-generated document. The simple reactive runtime looks like this:



However, this strategy is needlessly inefficient. For example, the counter component described earlier in Section 4.3 would never affect the section ordering, and therefore never affect the content of a reference. The  $\text{IdCtx}$   $\Delta$  could be computed once on  $a_0$  and then reused for all subsequent computations, or at least until  $\Delta$  is invalidated. For example, if the context was persisted after the first step and invalidated on the second step, then such a strategy would look like this:



These two strategies can be formalized as functions `simple` and `incr` that take a given article  $a_i$  and produce a final article  $a'_i$ . Their semantics are as follows:

$$\begin{aligned}
 \text{simple}(a_i) &= \text{render-refs}(a_i) \\
 \Delta_i^{\text{incr}} &= \begin{cases} \text{section-ids}(a_i) & \text{if } i = 0 \vee \text{dirty}(v_{i-1}, v_i) \\ \Delta_{i-1}^{\text{incr}} & \text{otherwise} \end{cases} \\
 \text{incr}(a_i) &= \text{replace-refs}(a_i, \Delta_i^{\text{incr}})
 \end{aligned}$$

The `dirty` function is the key logic that determines whether  $\Delta$  should be recomputed on a given step. Before considering a specific implementation of `dirty`, we can first articulate a correctness condition for this optimization: the incremental strategy should produce an equivalent document as the naive strategy for all inputs. Formally:

**THEOREM 5.2 (CORRECTNESS OF INCREMENTAL STRATEGY FOR SECTION REFERENCES).** *Let  $e \in \text{Expr}_{\text{TProg}}^{\text{Article}}$  where  $\cdot \vdash e : \text{ReactNode}$ . Let  $e \mapsto^* v_0$ . Let  $i \in \mathbb{N}$  and  $v_i = \text{doc-step}^i(v_0)$ . Let  $a_i = \text{doc-view}(v_i)$ . Then  $\text{simple}(a_i) = \text{incr}(a_i)$ .*



This theorem reduces to the lemma that `dirty` will always be true if the section IDs have changed from one document to the next, or formally:

LEMMA 5.3 (DIRTY FUNCTION ALWAYS CATCHES A CHANGE TO SECTION NUMBERING).

$$\text{section-ids}(a_{i-1}) \neq \text{section-ids}(a_i) \implies \text{dirty}(v_{i-1}, v_i)$$

For example, a simple implementation of `dirty` can recognize that the document structure can only change as a result of components. In this implementation, `dirty` only returns true if any component’s children includes a section either before or after the step. Say we have a function `descendents : ReactNode → String set` that returns the types of nodes descendent from the input. Then `dirty` is as follows:

$$\begin{aligned} \text{dirty}(\text{inst}_{\text{ReactNode}} \text{inst}_{i-1}, \text{inst}_{\text{ReactNode}} \text{inst}_i) = \\ \text{“section”} \in (\text{descendents } \text{inst}_{i-1}.\text{node} \cup \text{descendents } \text{inst}_i.\text{node}) \\ \vee \text{dirty}(\text{inst}_{i-1}.\text{node}, \text{inst}_i.\text{node}) \end{aligned}$$

In Appendix A.2 we give a proof of how the theorem reduces to the lemma, as well as the proof of the lemma for this particular definition of `dirty`. More broadly, the point is that this theorem demonstrates how the document calculus provides a foundation for reasoning about aspects such as how global document dependencies compose with local reactive computations.

## 6 RELATED WORK

The impetus for this work is that, in fact, very little work has attempted to provide a formal foundation for document languages from a computational perspective. An informal description of the Scribe language [Reid 1980] was published in POPL 43 years ago. *The TeXbook* [Knuth and Bibby 1986] gives a fairly precise specification for TeX, but its principal concerns are parsing and rendering—less so the computation in the middle. The @-syntax of Scribble [Flatt et al. 2009] is well-defined but its metatheory is not, leading in part to issues as in Section 2.3. Within efforts to formalize languages with templates like PHP [Filaretti and Maffei 2014], templates are usually a small footnote within the broader project, rather than a central focus of investigation.

In the rest of this section, we focus on understanding the practical systems that we model in this paper. Document languages have come a long way since the 1960s when “only a few dozen people in the world knew how to typeset mathematical formulas” [Knuth 1996]. Most academic research on document languages in the 20th century focused on vocabulary and abstractions for the graphical aspects of documents (Section 6.1). Such work largely continues today in the form of the ever-growing complexity of web browser rendering engines. The 1990s saw an explosion of languages for generating strings with templates (Section 6.2). In the new millennium, document language research has shifted focus to the aspects more at the heart of our paper, namely using computation to generate articles (Section 6.3). Today, the most complex interactions between content and computation can be found in reactive document systems (Section 6.4).

### 6.1 Markup Systems

“Markup languages” have a long history as programming languages for marking up documents that are presented on paper (literally via printing, or metaphorically in a PDF), or presented in the browser. Coombs et al. [1987] developed an early markup theory that distinguished “procedural markup”, or low-level graphical commands, from “descriptive markup”, or high-level structuring of a document. Descriptive markup, later called a “ordered hierarchy of content objects” [DeRose et al. 1997], formed the basis of systems such as SGML [Goldfarb 1990] and HyTime [Newcomb et al. 1991] that would go on to inspire HTML and XML. The article domain of the document calculus

defined in Section 3.2 is a model for descriptive markup, the predominant model for document languages used today. (The notable exception to this is  $\text{\TeX}$ , which evaluates into procedural markup but uses “environments” to attempt to simulate the experience of descriptive markup.)

Early markup systems had relatively primitive support for computation. The Scribe system [Reid 1980] only supported custom environments that were composed out of a fixed set of formatting attributes. This tradition has continued with modern markup languages. Languages like Markdown [MacFarlane 2021], AsciiDoc ([asciidoc.org](https://asciidoc.org)), reStructuredText ([docutils.sourceforge.io](https://docutils.sourceforge.io)), Markdoc ([markdoc.dev](https://markdoc.dev)), and Pandoc ([pandoc.org](https://pandoc.org)) all have little native support for anything resembling computation. At most, these languages have capabilities for resolving references or performing textual substitution for global variables defined in an external config file.

One exception is  $\text{\TeX}$ , which has a powerful macro system. It is notable that only 1 out of 27 chapters of *The  $\text{\TeX}$ book* [Knuth and Bibby 1986] concerns macros, a reflection of how many tasks  $\text{\TeX}$  had to juggle at the time of its inception. Later computational markup systems like Typst [Mädje 2022] reflect a significantly greater separation of concerns between computation and rendering.

## 6.2 String Template Systems

Unhygienic macro systems like the C preprocessor and the M4 processor [Kernighan and Ritchie 1977] can be viewed as the ur-template-systems (what we call a *string template program* in the document calculus). Format strings (i.e., *string template literals*) also date back to the earliest programming languages such as the PICTURE clause in COBOL, WRITE command in Fortran, and printf function in Algol 68. Variable interpolation in strings can be found in several early shell languages, and was later adopted by Perl and Tcl. If-statements and for-loops inside templates were popularized by PHP, which used these facilities primarily to generate strings of HTML. However, PHP is today the only widely-used general-purpose programming language (to our knowledge) with built-in support for string template programs — such features in other languages are expressed usually through domain-specific languages such as Jinja ([jinja.palletsprojects.com](https://jinja.palletsprojects.com)) for Python, Handlebars ([handlebarsjs.com](https://handlebarsjs.com)) for Javascript, or Liquid ([shopify.github.io/liquid](https://shopify.github.io/liquid)) for Ruby.

Using string template programs to generate article literals can be prone to error, as discussed in the PHP case study in Section 2.1. This observation motivated Parr [2004], who gave one of the first formal models of string template literals. Parr’s model is roughly equivalent to the  $\mathcal{D}_{\text{TLit}}^{\text{String}}$  level of the document calculus. Parr’s goal was to reason about the computability of template languages (so as to demonstrate that a restricted template language was not Turing-complete), while our goal is to provide a concise model for a wide variety of template features.

## 6.3 Article Template Systems

As XML and HTML gained popularity, many languages and libraries were developed to provide better ways of creating and analyzing tree-shaped documents. XDuce [Hosoya and Pierce 2003], XML Schema [Siméon and Wadler 2003], and CDuce [Benzaken et al. 2003] provided for typed processing of XML documents, with a focus on encoding domain-specific XML schemata into the type system. The JTwig extension [Christensen et al. 2003] to Java supported XML templates with holes to address the problems of generating structured documents via strings. “Syntax-safe” string template engines were designed to ensure that generated strings matched a schema [Arnoldus et al. 2007; Heidenreich et al. 2009]. Scala even supported XML template literals upon its release, although XML support has since been deprecated.

Metaprogramming systems share many similarities with article template systems; programs are trees, and articles are trees. As described in Section 3.2.4, HTML libraries in modern Lisps permit the use of quasiquotes to generate HTML. To reduce the verbosity of writing string content within

Lisp-embedded articles, Scribe’s Sk-expressions [Gallesio and Serrano 2005] and Scribble’s @-expressions [Flatt et al. 2009] provide for concise article templates. Our goal of providing statically-typed templates also overlaps with typed metaprogramming systems like MetaML [Taha and Sheard 1997], Template Haskell [Sheard and Jones 2002], and Scala macros [Burmako 2013], just to name a few notable systems among many.

A goal of some article template systems is to statically check for the validity of documents, i.e., that a node tree matches a given document schema. XDuce and CDuce are specially designed for this purpose by supporting the language of regular trees as types. Haskell libraries like `type-of-html` (knupfer/type-of-html) show how specific schema (like HTML) can be encoded into a sufficiently expressive type system of a general-purpose language. Our goal is to integrate templates into System F in as simple a manner as possible, so the document calculus only checks for the weaker property of well-formedness.

## 6.4 Reactive Article Template Systems

While interest in XML templates has since waned, interest in HTML templates has grown dramatically, especially focusing on templates that describe reactive HTML documents. Recent years have seen many new languages for authoring reactive articles: Idyll [Conlen and Heer 2018], MDX (mdxjs.com), Observable (observablehq.com), and Living Papers (uwdata/living-papers). In particular, the JSX extension to Javascript, originally created by the developers of React (reactjs.com), is now widely adopted within the Javascript ecosystem. Frameworks like Vue (vuejs.org) and SolidJS (solidjs.com) use JSX, and Svelte (svelte.dev) uses a JSX-like syntax.

Notably, these reactive JS frameworks all provide substantively different desugarings for JSX into vanilla Javascript. The desugaring provided in Section 4.3 is most similar to React’s, where the desugaring is straightforward and the runtime does most of the work. However, more recent frameworks like Svelte have adopted much more complex desugarings to improve efficiency. Svelte statically analyzes its templates for data dependencies to determine when components should react to state changes, avoiding the cost of dynamic dependency analysis. This trend provides a fertile ground for future PL research that can build on the foundations of the document calculus. Just as one example, Svelte’s dependency analysis is deeply unsound, as it is not sensitive to fields or aliases (see: svelte.dev). Future document languages will need firm theoretical foundations to correctly analyze and desugar complex templates.

## 7 DISCUSSION

This paper has presented the document calculus, a formal model for how templates interleave content and computation to produce strings and articles. Our immediate goal with this work is to provide a formal model that can undergird any theoretical investigation into document languages. Document languages have long been a subject with a plethora of practice but only tacit theory, especially with regards to the computation/content boundary.

Our long-term goal with this work is to provide conceptual clarity to designers of document languages. We hope that the vocabulary and semantics of the document calculus can guide future designs (this paper came out of the authors’ own work in designing a document language). We conclude by discussing actionable takeaways for document language designers (Section 7.1), and then discuss one of the major challenges unaddressed in this paper: concrete syntax (Section 7.2).

### 7.1 Implications for Language Designers

The taxonomy in Table 1 provides a high-level vocabulary for talking about the design space of document language. A document language designer should ask: which domain and constructors are most appropriate for their context of use? With regards to constructors, template literals can be

quite powerful with an expressive language of expressions. But an imperative language with limited expressions would probably need template programs or else the template DSL is too limiting.

With regards to domain, a language designer should be aware that designing templates for both strings and articles does not require wholly different features. A shared template language can be used across both domains; the language just needs different syntaxes for invoking a template in each domain context (i.e., a string template `strtpl` versus a tree template `treetpl`).

The semantics in Sections 3.1 and 3.2 provide one possible implementation strategy for template desugaring. We strongly recommend a pure strategy over an impure strategy for the reasons discussed in the PHP case study (Section 2.1). We recommend a variable-binding desugaring that permits lexical scope and does not require a single global context (Section 3.1.3). We also recommend carefully considering the dimensionality of lists produced by each template feature to avoid dimension mismatch, such as by providing a splice/quasiquote feature (Section 3.2.4).

Designers should be aware that reducing expressions to values is not likely to be the last step in the document generation pipeline (Section 4). Global passes such as section numbering (Section 4.1) and reforestation (Section 4.2) should execute on the reduced “raw” document. These passes have a subtle interaction with reactivity (Section 4.3). We provide an example for how these separate concerns can be composed in Section 5.2.

## 7.2 Concrete Syntax for Language Users

Concrete syntax is not a common concern in programming languages research, where it is assumed to be handled through standard parsing techniques. But concrete syntax is *essential* to document languages, moreso than most other kinds of programming languages (on par with languages for novices or compact DSLs). The fundamental utility of a document language is predicated on its syntactic convenience. Most authors would not want to write at only the  $\mathcal{D}_{\text{Prog}}^{\text{Article}}$  level, like this:

```
1 [text("We live in a "), bold(text("golden age")), text(" of documents.")]
```

In that sense, this paper’s subtitle is deliberately inaccurate: lambda is not the ultimate document. As Olin Shivers wrote, “lambda is not a universally sufficient value constructor,” and that holds true for constructing documents as well. To that end, future work on document languages should investigate the design of syntaxes that trade-off intuitiveness, error-tolerance, and systematicity.

For instance, Markdown’s syntax is designed to be reasonably intuitive and maximally error-tolerant, at the expense of systematicity. Taking one example from MacFarlane [2018], Markdown does not have a consistent strategy for parsing lists adjacent to a paragraph. A different behavior occurs depending on whether the list number is equal to 1 or not, as shown in this Markdown program (left) with its HTML output according to CommonMark (right):

```
1 A paragraph
2 1. A list
3
4 Another paragraph
5 2. Another list
```

```
1 <p>A paragraph</p>
2 <ol><li>A list</li></ol>
3 <p>Another paragraph
4 2. Another list</p>
```

More generally, the widespread adoption of Markdown has demonstrated the strong desire for a concise document syntax. Yet, authors also want computation to simplify authoring of complex documents, as evinced by both the enduring usage of  $\text{\LaTeX}$  and the proliferation of “Markdown++” successor languages. It is an open question how to get the best of both worlds — a human-friendly, concise syntax with a principled, powerful semantics. Now is clearly the time to revisit the accumulated design decisions of past languages to build the foundations for a better-documented future.

## DATA AVAILABILITY STATEMENT

The archival version of the artifact for this document is hosted on Zenodo [Crichton and Krishnamurthi 2023]. The latest version of the artifact is hosted on GitHub (cognitive-engineering-lab/document-calculus).

## ACKNOWLEDGMENTS

This work was partially supported by a gift from Amazon and by the US NSF under Grant No. 2319014. We are grateful to the numerous individuals who have worked on the document languages that influenced our work.

## REFERENCES

- Jeroen Arnoldus, Jeanot Bijpost, and Mark van den Brand. 2007. Repleo: A Syntax-Safe Template Engine. In *Proceedings of the 6th International Conference on Generative Programming and Component Engineering* (Salzburg, Austria) (GPCE '07). Association for Computing Machinery, New York, NY, USA, 25–32. <https://doi.org/10.1145/1289971.1289977>
- Véronique Benzaken, Giuseppe Castagna, and Alain Frisch. 2003. CDuce: An XML-Centric General-Purpose Language. In *Proceedings of the Eighth ACM SIGPLAN International Conference on Functional Programming* (Uppsala, Sweden) (ICFP '03). Association for Computing Machinery, New York, NY, USA, 51–63. <https://doi.org/10.1145/944705.944711>
- Alexander Asp Bock, Thomas Bøgholm, Peter Sestoft, Bent Thomsen, and Lone Leth Thomsen. 2020. On the semantics for spreadsheets with sheet-defined functions. *J. Comput. Lang.* 57 (2020), 100960.
- Eugene Burmako. 2013. Scala Macros: Let Our Powers Combine! On How Rich Syntax and Static Types Work with Metaprogramming. In *Proceedings of the 4th Workshop on Scala* (Montpellier, France) (SCALA '13). Association for Computing Machinery, New York, NY, USA, Article 3, 10 pages. <https://doi.org/10.1145/2489837.2489840>
- Andrew Cave and Brigitte Pientka. 2012. Programming with Binders and Indexed Data-Types. In *Proceedings of the 39th Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages* (Philadelphia, PA, USA) (POPL '12). Association for Computing Machinery, New York, NY, USA, 413–424. <https://doi.org/10.1145/2103656.2103705>
- Aske Simon Christensen, Anders Møller, and Michael I. Schwartzbach. 2003. Extending Java for High-Level Web Service Construction. *ACM Trans. Program. Lang. Syst.* 25, 6 (nov 2003), 814–875. <https://doi.org/10.1145/945885.945890>
- Matthew Conlen and Jeffrey Heer. 2018. Idyll: A Markup Language for Authoring and Publishing Interactive Articles on the Web. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology* (Berlin, Germany) (UIST '18). Association for Computing Machinery, New York, NY, USA, 977–989. <https://doi.org/10.1145/3242587.3242600>
- James H. Coombs, Allen H. Renear, and Steven J. DeRose. 1987. Markup Systems and the Future of Scholarly Text Processing. *Commun. ACM* 30, 11 (nov 1987), 933–947. <https://doi.org/10.1145/32206.32209>
- Will Crichton and Shriram Krishnamurthi. 2023. Artifact for "A Core Calculus for Documents". <https://doi.org/10.5281/zenodo.8409115>
- Evan Czaplicki and Stephen Chong. 2013. Asynchronous Functional Reactive Programming for GUIs. In *Proceedings of the 34th ACM SIGPLAN Conference on Programming Language Design and Implementation* (Seattle, Washington, USA) (PLDI '13). Association for Computing Machinery, New York, NY, USA, 411–422. <https://doi.org/10.1145/2491956.2462161>
- Steven J. DeRose, David G. Durand, Elli Mylonas, and Allen H. Renear. 1997. What is Text, Really? *SIGDOC Asterisk J. Comput. Doc.* 21, 3 (aug 1997), 1–24. <https://doi.org/10.1145/264842.264843>
- Matthias Felleisen. 1991. On the expressive power of programming languages. *Science of Computer Programming* 17, 1 (1991), 35–75. [https://doi.org/10.1016/0167-6423\(91\)90036-W](https://doi.org/10.1016/0167-6423(91)90036-W)
- Daniele Filaretti and Sergio Maffei. 2014. An Executable Formal Semantics of PHP. In *ECOOP 2014 – Object-Oriented Programming*, Richard Jones (Ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, 567–592.
- Matthew Flatt, Eli Barzilay, and Robert Bruce Findler. 2009. Scribble: Closing the Book on Ad Hoc Documentation Tools. In *Proceedings of the 14th ACM SIGPLAN International Conference on Functional Programming* (Edinburgh, Scotland) (ICFP '09). Association for Computing Machinery, New York, NY, USA, 109–120. <https://doi.org/10.1145/1596550.1596569>
- Erick Gallezio and Manuel Serrano. 2005. Skribe: a functional authoring language. *Journal of Functional Programming* 15, 5 (2005), 751–770. <https://doi.org/10.1017/S0956796805005575>
- Charles F Goldfarb. 1990. *The SGML handbook*. Vol. 10. Oxford University Press, Oxford, UK.
- Shu-yu Guo, Michael Ficarra, and Kevin Gibbons. 2023. ECMAScript 2024 Language Specification. <https://tc39.es/ecma262/>.
- Florian Heidenreich, Jendrik Johannes, Mirko Seifert, Christian Wende, and Marcel Böhme. 2009. Generating Safe Template Languages. In *Proceedings of the Eighth International Conference on Generative Programming and Component Engineering* (Denver, Colorado, USA) (GPCE '09). Association for Computing Machinery, New York, NY, USA, 99–108. <https://doi.org/10.1145/1621607.1621624>



- Haruo Hosoya and Benjamin C. Pierce. 2003. XDuce: A Statically Typed XML Processing Language. *ACM Trans. Internet Technol.* 3, 2 (may 2003), 117–148. <https://doi.org/10.1145/767193.767195>
- Brian W. Kernighan and Dennis M. Ritchie. 1977. The M4 Macro Processor.
- Donald E. Knuth. 1996. Digital Typography. (1996). Commemorative lecture of the Kyoto Prize.
- Donald E. Knuth and Duane Bibby. 1986. *The TeXbook*. Addison-Wesley, Boston, Massachusetts, USA.
- Daniel R. Licata and Robert Harper. 2009. A Universe of Binding and Computation. In *Proceedings of the 14th ACM SIGPLAN International Conference on Functional Programming (Edinburgh, Scotland) (ICFP '09)*. Association for Computing Machinery, New York, NY, USA, 123–134. <https://doi.org/10.1145/1596550.1596571>
- Andres Löh and Ralf Hinze. 2006. Open Data Types and Open Functions. In *Proceedings of the 8th ACM SIGPLAN International Conference on Principles and Practice of Declarative Programming (Venice, Italy) (PPDP '06)*. Association for Computing Machinery, New York, NY, USA, 133–144. <https://doi.org/10.1145/1140335.1140352>
- John MacFarlane. 2018. Beyond Markdown. <https://johnmacfarlane.net/beyond-markdown.html>.
- John MacFarlane. 2021. CommonMark Spec Version 0.30. <https://spec.commonmark.org/0.30/>.
- Laurenz Mädje. 2022. *Typst: A Programmable Markup Language for Typesetting*. Master's thesis. Technical University of Berlin, Berlin, Germany.
- Jay McCarthy. 2022. Web Applications in Racket. <https://docs.racket-lang.org/web-server/>.
- Steven R. Newcomb, Neill A. Kipp, and Victoria T. Newcomb. 1991. The “HyTime”: Hypermedia/Time-Based Document Structuring Language. *Commun. ACM* 34, 11 (nov 1991), 67–83. <https://doi.org/10.1145/125490.125495>
- Terence John Parr. 2004. Enforcing Strict Model-View Separation in Template Engines. In *Proceedings of the 13th International Conference on World Wide Web (New York, NY, USA) (WWW '04)*. Association for Computing Machinery, New York, NY, USA, 224–233. <https://doi.org/10.1145/988672.988703>
- Justin Pombrio and Shriram Krishnamurthi. 2018. Inferring Type Rules for Syntactic Sugar. In *Proceedings of the 39th ACM SIGPLAN Conference on Programming Language Design and Implementation (Philadelphia, PA, USA) (PLDI 2018)*. Association for Computing Machinery, New York, NY, USA, 812–825. <https://doi.org/10.1145/3192366.3192398>
- Brian K. Reid. 1980. A High-Level Approach to Computer Document Formatting. In *Proceedings of the 7th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages (Las Vegas, Nevada) (POPL '80)*. Association for Computing Machinery, New York, NY, USA, 24–31. <https://doi.org/10.1145/567446.567449>
- Tim Sheard and Simon Peyton Jones. 2002. Template Meta-Programming for Haskell. In *Proceedings of the 2002 ACM SIGPLAN Workshop on Haskell (Pittsburgh, Pennsylvania) (Haskell '02)*. Association for Computing Machinery, New York, NY, USA, 1–16. <https://doi.org/10.1145/581690.581691>
- Jérôme Siméon and Philip Wadler. 2003. The Essence of XML. In *Proceedings of the 30th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages (New Orleans, Louisiana, USA) (POPL '03)*. Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/604131.604132>
- Walid Taha and Tim Sheard. 1997. Multi-Stage Programming with Explicit Annotations. In *Proceedings of the 1997 ACM SIGPLAN Symposium on Partial Evaluation and Semantics-Based Program Manipulation (Amsterdam, The Netherlands) (PEPM '97)*. Association for Computing Machinery, New York, NY, USA, 203–217. <https://doi.org/10.1145/258993.259019>

Received 2023-07-11; accepted 2023-11-07