NCafe Archiver Documentation

Julius Hamilton

July 20, 2024

1 Introduction

This is the documentation for the NCafe Archiver project.

2 Crawling Strategy

We are going to build a categorical model of web scraping in Coq. This model will help us define and execute our crawling strategy more precisely and formally.

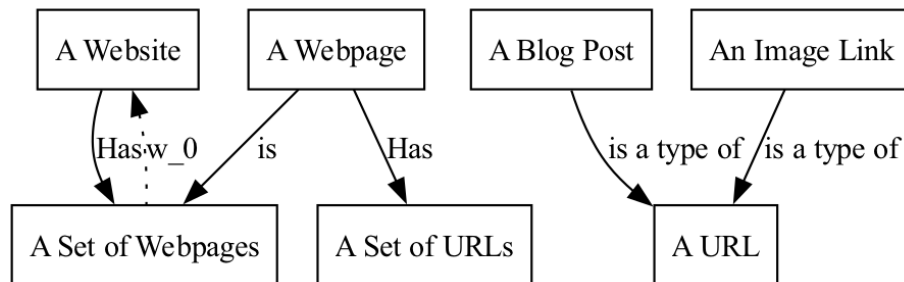## 3 Website Structure and Types of Links

Figure 1: Graph generated by Graphviz

4 Coq Code for Link Types

```
Inductive LinkType :=
| BlogArticle
| Other.
```

5   Home Page Source Code (Abridged)

```
<html>
    <head>
        <title>Home Page</title>
    </head>
    <body>
        <a href="/blog-post-1">Blog Post 1</a>
        <a href="/about-us">About Us</a>
        <!-- Other links -->
    </body>
</html>
```