

Automatic Web Page Reconstruction

Viliam Serečun

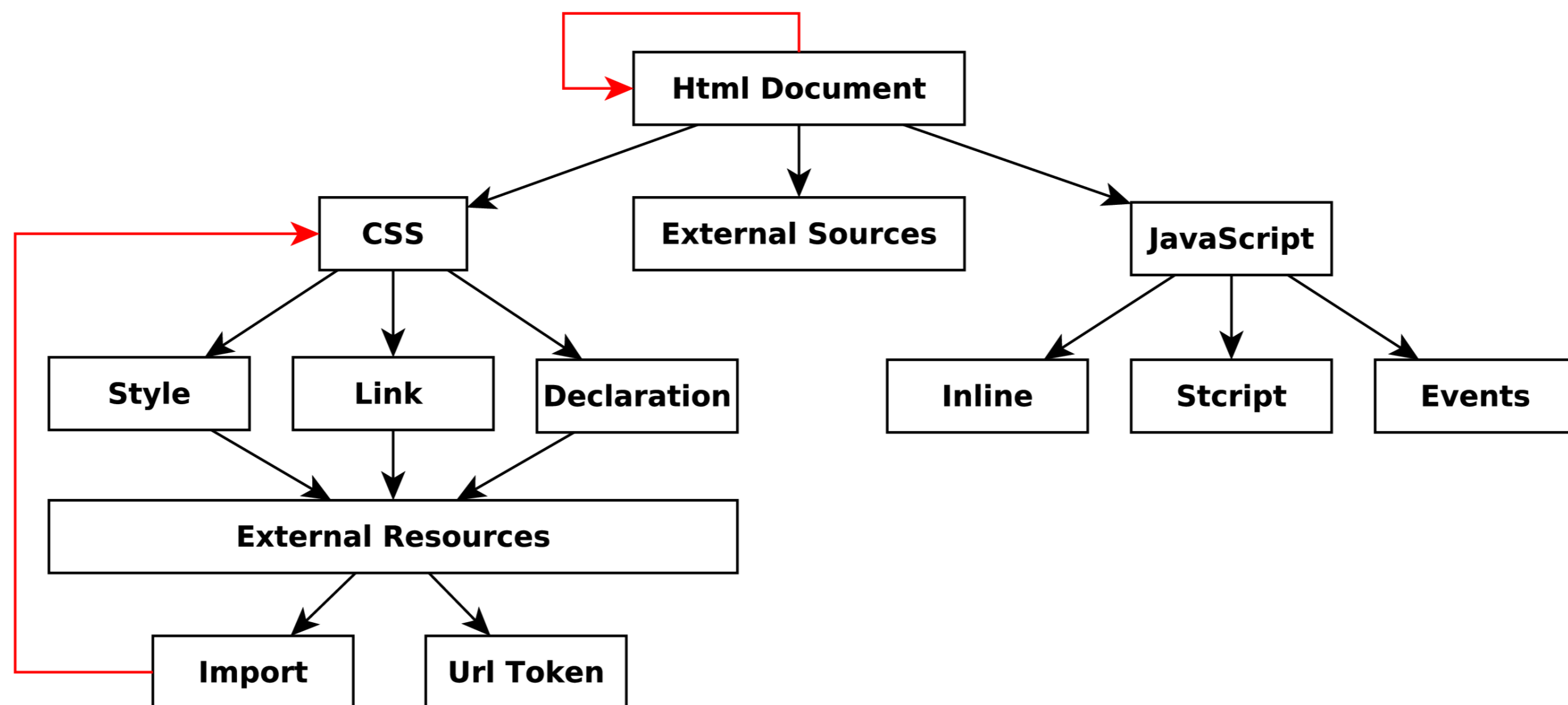


Lemmiwinks Framework

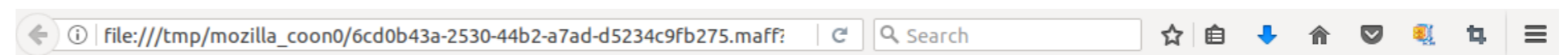
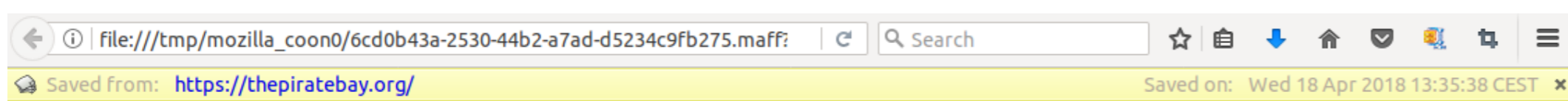
The Lemmiwinks framework provides set of libraries providing scraping and archiving functionality. The framework is possible to divide into four major parts.

HTTP Client	Parser	Archive	Extractor
<ul style="list-style-type: none"> • Selenium client • HTTP client • Selenium Pool 	<ul style="list-style-type: none"> • CSS parser • HTML parser 	<ul style="list-style-type: none"> • MAFF 	<ul style="list-style-type: none"> • Regular Expression • Generic Interface

Web Page Structure



Reconstructed Content



Type	Value
Bitcoin2	3HcEB6bi4TFPdvk31Pwz77DwAzIAZz2fMn
Litecoin2	LS78aoGtfuGCZ777x3Hmr6tcoW3WaYymx9
IP ADDRESS	104.27.217.28 443
	104.27.216.28 443
	2400:cb00:2048:1::681b:d81c 443
	2400:cb00:2048:1::681b:d91c 443
Timestamp	2018-04-18 13:35:38.203632