

Crawling

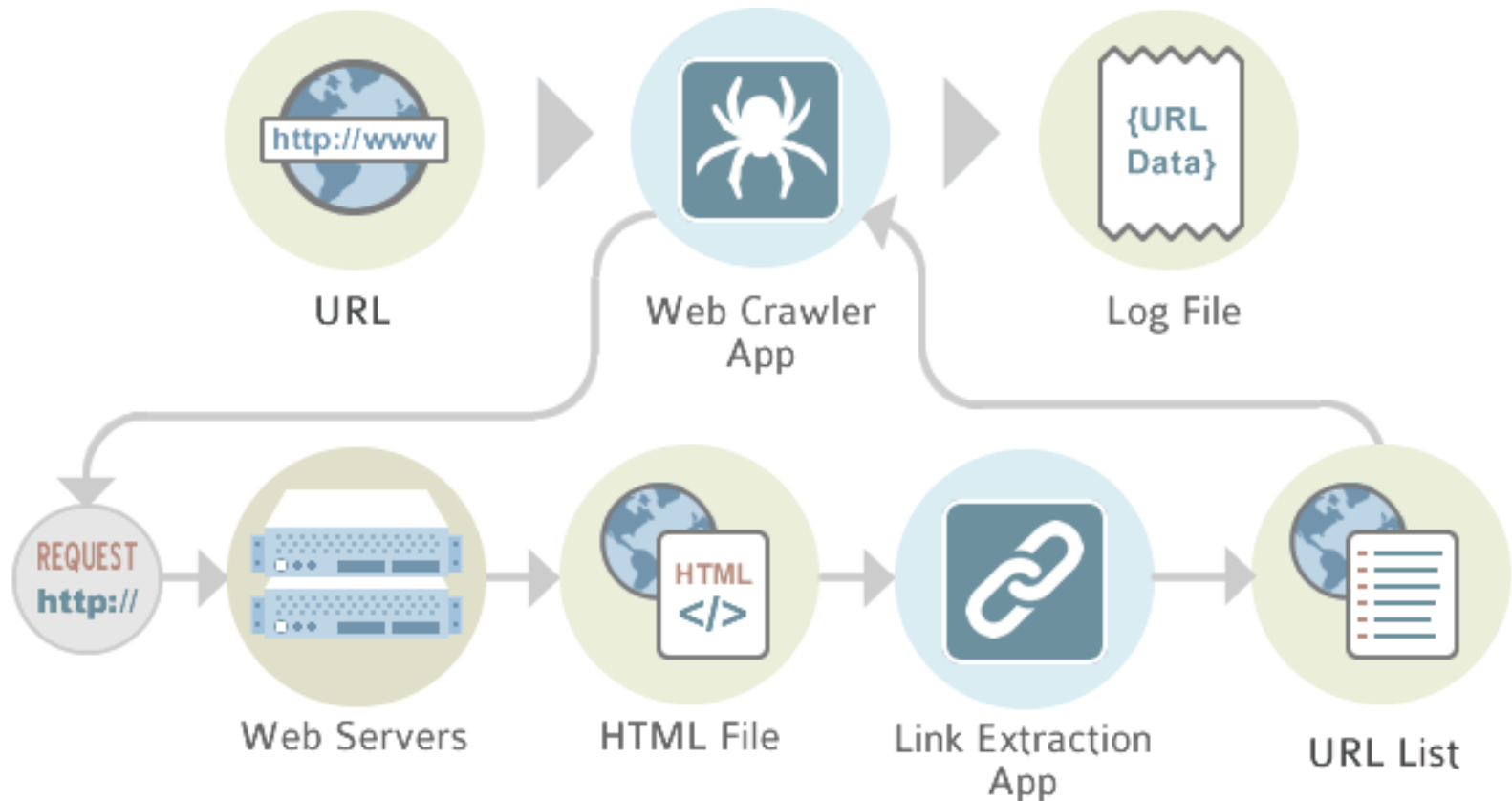
Crawling

WHAT Web crawlers are known by a variety of names *spiders, bots or web crawlers*
Systematically browses the World Wide Web

WHY Typically for the purpose of **Web indexing**
Specifically, they index what **words** are used on a website and in what context

Architecture

Process of **repetitively** finding and fetching **hyperlinks** starting from a list of starting URLs



Scrapy

Scrapy	Web-spider or web scraper framework
BeautifulSoup	Parsing library

Scrapy vs. BeautifulSoup

ScrapingAuthority.com

WHAT IS IT?

Scrapy



BeautifulSoup

