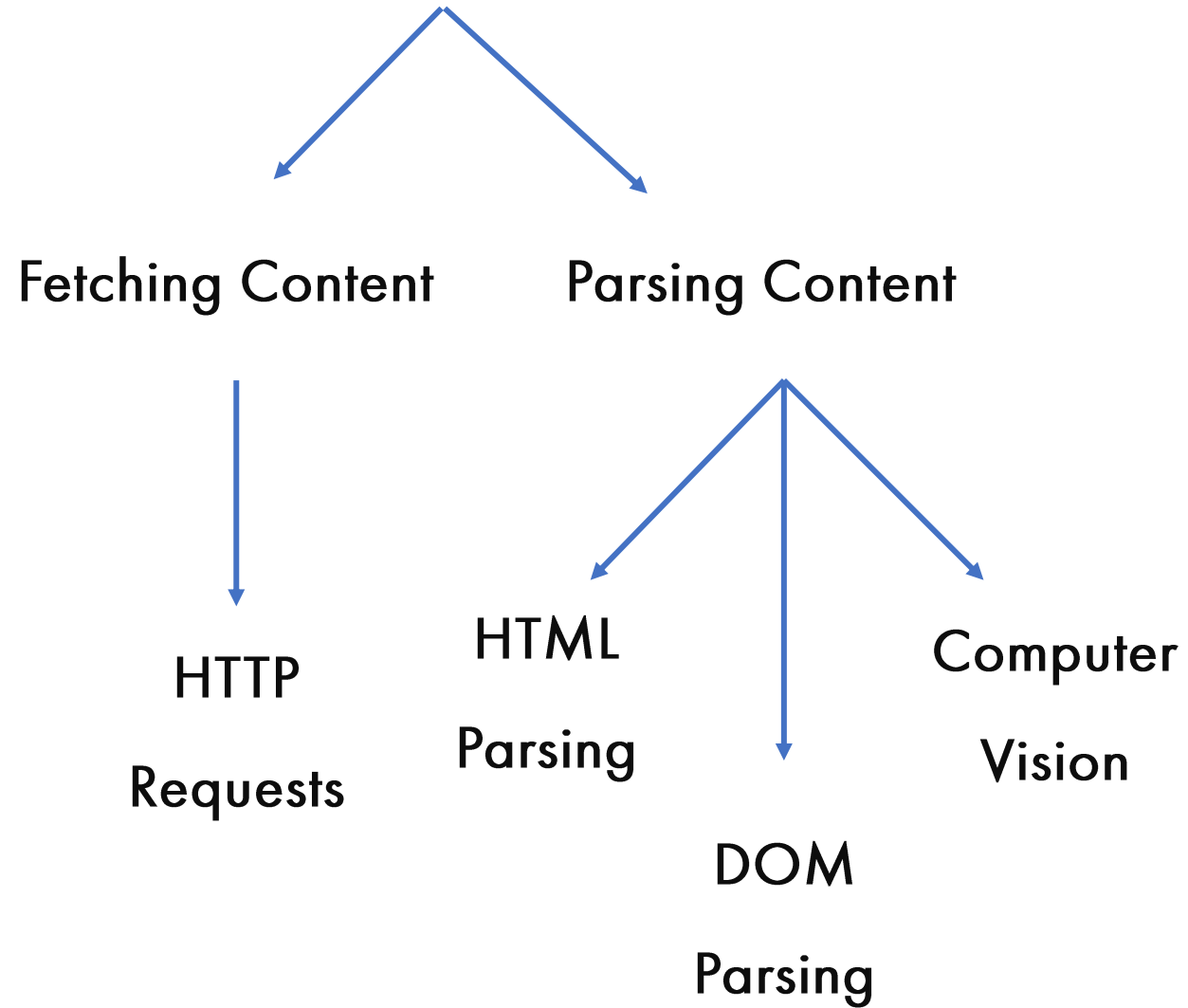
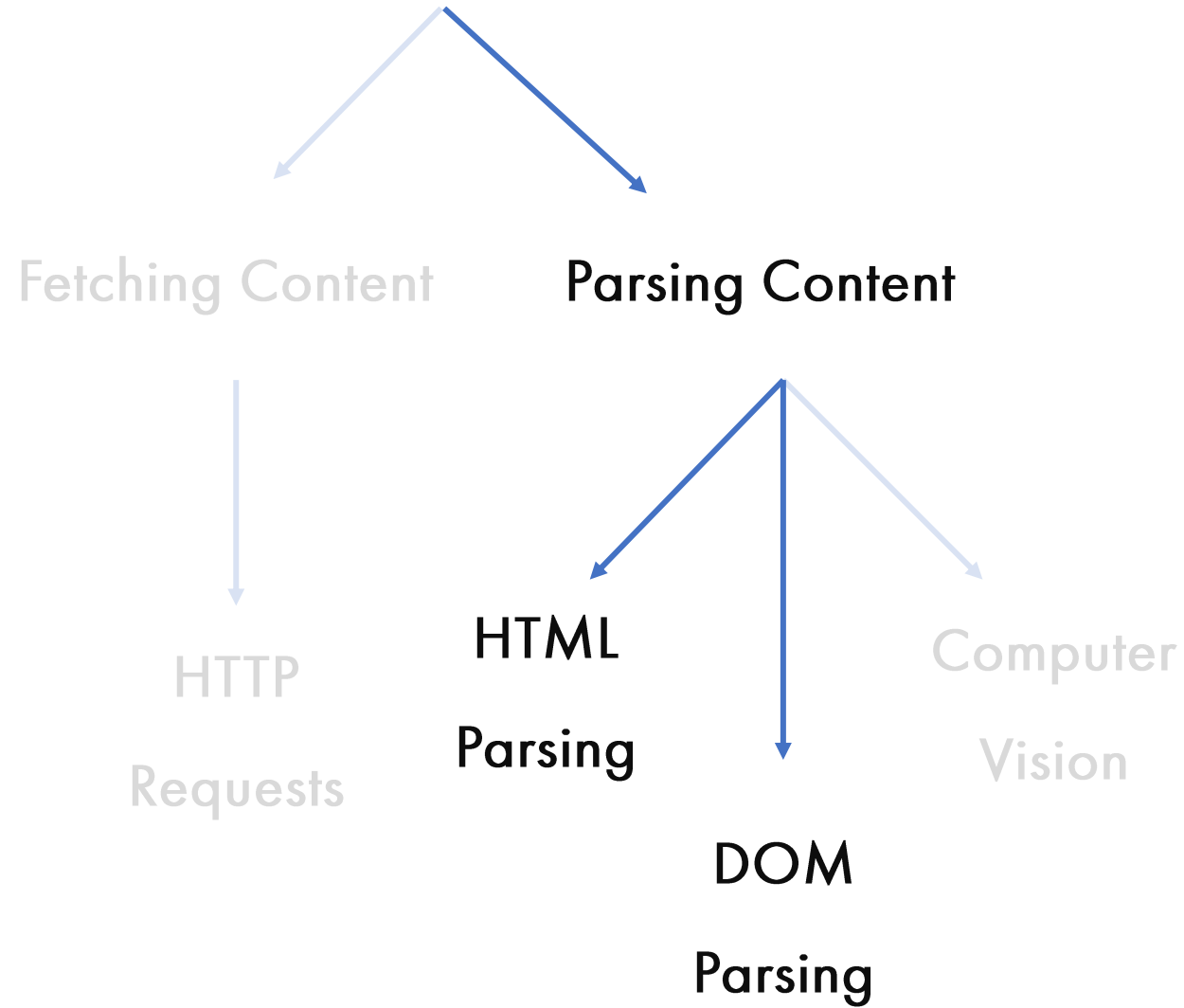


Parsing HTML Content

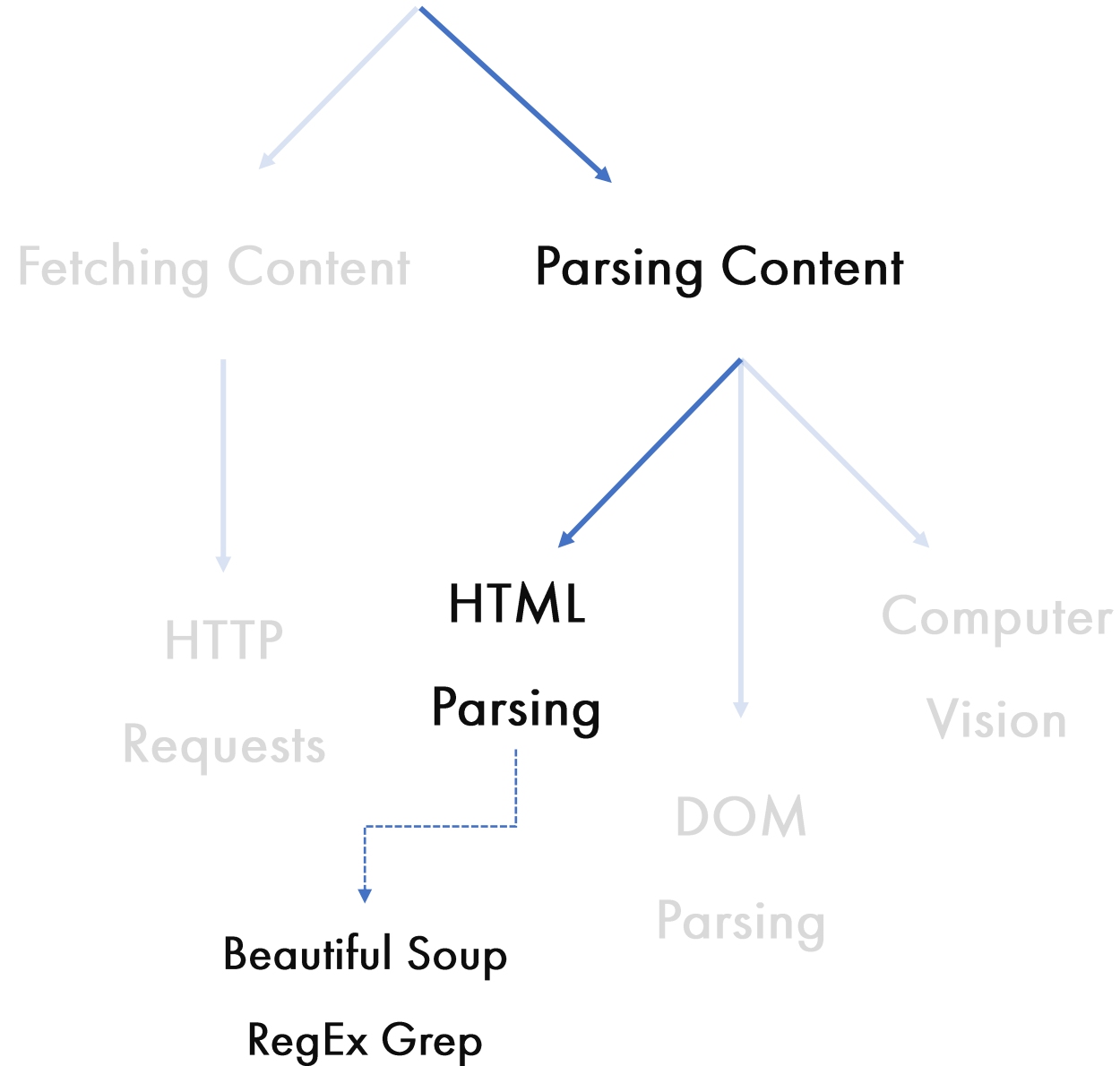
Web Scrapping



Web Scrapping



Web Scraping



Beautiful Soup

Python package for parsing HTML and XML, including those with malformed markup such as missing tags.

Beautiful Soup



Mitigates weaknesses of regular expressions

- Global: Forms parse tree of entire HTML
- Relatively simple to use
- Robust to problems in markup being parsed