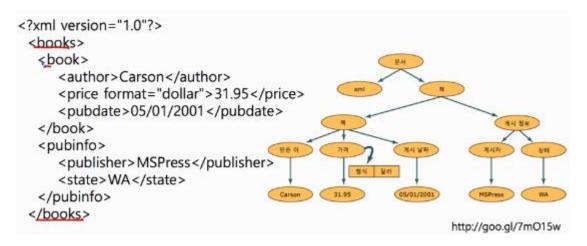
Chapter 15: XML과 JSON(eXtensible Markup Languages and JSON)

XML(eXtensible Markup Languages)

- -데이터의 구조와 의미를 설명하는 TAG(Markup)를 사용하여 표시하는 언어
- -TAG와 TAG사이에 값이 표시되고, 구조적인 정보를 표현할 수 있음



- -트리 형태로 구조를 표현
- -가장 많이 쓰이는 parser인 beautifulsoup으로 파싱
- -Scraping: 긁어오는 작업, Parsering: 해석하는 작업

beautifulsoup 모듈 사용 - 모듈 호출 from (bs4) import BeautifulSoup - 객체 생성 soup = BeautifulSoup(books xml, "lxml") - Tag 찾는 함수 find_all 생성 soup.find_all("author")

beautifulsoup 모듈 사용

- find all: 정규식과 마찬가지로 해당 패턴을 모두 반환
- find('invention-title')
 Tag 네임 = title
- get_text(): 반환된 패턴의 값 반환 (태그와 태그 사이)
 - <invention-title id="d2e43">
 Adjustable shoulder device for hard upper torso suit
 </invention-title>

JSON(JavaScript Object Notation)

- -Java Script의 데이터 객체 표현 방식
- -XML에 비해 데이터 용량이 적고, Code로의 전환이 쉽다.

-데이터 저장 및 읽기는 dict type과 상호 호환 가능