

XPath의 이해

김 현

한국학중앙연구원 인문정보학교실

hyeon@aks.ac.kr



이 저작물(PPT)의 인용 표시 방법:
김현, 「XPath의 이해」, '전자문서와 하이퍼텍스트' 수업 자료 (2022)

1. XPath란?

2. 필터

3. Xpath 함수

❖ XPath란?

- XML 문서 내의 특정 부분(element, attribute)을 찾아가기 위해 사용되는 경로 표기 언어
- 1999년 11월 W3C 권고안으로 XPath 1.0 발표.

❖ XPath의 용도

- 스타일 쉬트
 - 스타일을 적용할 요소를 찾아가는 경로 표시자로 사용
 - xsl:template 요소의 match 속성 혹은 xsl:apply-templates, xsl:value-of 요소의 select 속성에 사용.
- 데이터베이스
 - 데이터 상에서 검색 대상 요소를 찾아내는 경로 표시자로 사용
 - XML Data Type Method에서 사용

❖ 노드(node)

- XML 문서를, 루트(root)라는 하나의 개체에서 시작하여 여러 개의 연결된 개체로 벌어나간 나무(tree) 형태의 자료로 이해하고, 서로 연결된 각각의 구성 요소를 노드(node)라고 명명
- XML 문서 상에서 명시적으로 확인되는 노드는 요소(element)와 속성(attribute)이지만, XPath에서는 정교하고 유연한 조작을 위해 다음을 모두 노드로 간주
 - 루트(root)
 - 요소(element)
 - 속성(attribute)
 - 텍스트(text)
 - 이름공간(namespace)
 - 프로세싱 명령문(processing instruction)
 - 주석(comment)
- 노드 집합(node collection) : XML 문서상의 모든 노드 중에서 특정 조건에 부합하는 노드들의 집합

루트 노드와 문맥 노드

❖ 루트 노드(root node)

- XML 문서 자체를 지시하는 가상적인 최상위 노드
- 문서 루트(document root)라고도 하며, ("/")로 표기

❖ 문맥 노드(context node)

- 템플릿(template)의 적용의 시작점이 되는 노드

❖ 위치 경로: 문맥 노드의 위치를 지정하는 방법

- /
 - 루트 노드를 선택
- /tag
 - 루트 노드의 자식인 모든 <tag> 요소를 선택
 - 문맥 노드가 루트 노드에서 <tag> 요소로 이동
- tag
 - 문맥 노드의 자식인 모든 <tag> 요소를 선택
 - 문맥 노드가 상위 노드에서 <tag> 요소로 이동
- /tag/stag/sstag
 - 루트 노드의 자식 <tag> 요소를 찾은 후
 - <tag> 요소의 자식 <stag> 요소를 찾은 후
 - <stag> 요소의 자식 <sstag> 요소를 선택
 - 문맥 노드는 최종적으로 선택된 <sstag> 요소로 이동
- @attr
 - 문맥 노드의 attr 속성을 선택
- .
 - 문맥 노드 자체를 선택
- ..
 - 문맥 노드의 상위 노드를 선택

❖ 재귀적 내림 연산자(Recursive Descent Operator) "//"

- 찾고자 하는 노드의 위치와 상관없이 노드 이름을 찾아 경로 이동
- **//인명** → 문서 내의 모든 <인명> 요소를 선택.
- **//@id** → 문서 내의 모든 id 속성을 선택

❖ OR 연산자 "|"

- 어떤 노드가 하나 이상의 패턴과 일치할 경우 템플릿을 활성화
- **인명 | 지명** → 노드 이름이 <인명> 또는 <지명>인 경우의 처리 지시

1. XPath란?

2. 필터

3. Xpath 함수

❖ 노드 선택 필터 "[]"

- 특정 조건을 만족하는 노드의 선택
- 필터는 문맥 노드의 위치에 영향을 주지 않는다.
- 분야[소분야]
 - 문맥 노드의 자식 중 <소분야> 자식 요소를 가진 모든 <분야> 요소 선택
 - 문맥 노드가 <소분야> 요소가 아니라 <분야> 요소까지만 이동
- 인명[@id]
 - 문맥 노드의 자식 중 id 속성을 가진 모든 <인명> 요소만 선택
- 분야[소분야 = 'A']
 - 문맥 노드의 자식 중 <소분야> 요소의 값이 "A"인 자식 노드를 가진 모든 <분야> 요소 선택
- 유형[. = 'B']
 - 문맥 노드의 자식 중 값이 'B'인 모든 <유형> 요소
- //명단[인명/@id = 'C']
 - id 속성 값이 'C' 인 <인물> 요소를 자식으로 가진 모든 <명단> 요소

※응용

- /프로젝트[@code = 'D']/명단/인명[@id = 'C']

❖ 필터 조건문에 사용하는 연산자

❖ 비교 연산자

연산자	기능
=	같다
!=	같지 않다
<	작다
<=	작거나 같다
>	크다
>=	크거나 같다

❖ 관계 연산자

연산자	기능
and	두 조건을 모두 만족
or	적어도 하나를 만족

1. XPath란?

2. 필터

3. Xpath 함수

XPath 함수(1)

❖ 노드 함수

- **node()**
 - 노드 그 자체를 반환
- **text()**
 - 노드의 PCDATA 내용을 반환

❖ 위치 함수

- **count()**
 - 노드 집합에서 노드의 갯수를 반환
- **position()**
 - 노드 집합 내의 노드의 위치를 문서 순서대로 반환
 - "tag[position() = 2]"는 "tag[2]"와 동일
- **last()**
 - 노드 집합에서 마지막 노드의 위치를 반환

❖ 수치 함수

- **number()**
 - PCDATA를 수치값으로 변환하여 반환
- **sum()**
 - 노드 집합 내의 모든 수치값을 합하여 반환

❖ 부울 함수

- **boolean()**
 - XPath expression을 진. 위를 판단하여 true 혹은 false를 반환
 - 수치값: 0인 경우-> false, 0이 아닌 경우 -> true
 - 문자열: 1개 이상의 문자로 구성된 경우 -> true
 - 노드 집합: empty인 경우 -> false

❖ 문자열 함수

- **string()**
 - 임의의 값을 문자열로 변환하여 반환
- **string-length()**
 - 공백을 포함한 문자열의 길이를 반환
- **concat()**
 - 2개 이상의 문자열을 합하여 하나의 문자열로 반환
- **contains()**
 - 두 번째 문자열의 내용이 첫 번째 문자열에 포함되었는지 진위를 반환
- **substring(), substring-after(), substring-before()**
 - 문자열의 일부를 반환
- **translate()**
 - 두 번째 문자열에 속하는 문자를 세 번째 문자열에 속하는 문자로 변환