

XML 데이터베이스의 설계 및 구현

김 현

한국학중앙연구원 인문정보학교실

hyeon@aks.ac.kr



이 저작물(PPT)의 인용 표시 방법:

김현, 「XML 데이터베이스 설계 및 구현」, ‘전자문서와 하이퍼텍스트’ 수업 자료 (2022)



1. XML 데이터베이스 설계

2. XML 데이터베이스 구현

```
<?xml version="1.0" encoding="utf-8"?>
<?xml-stylesheet type="text/xsl" href="dataset.xsl"?>
<!DOCTYPE dataset SYSTEM "dataset.dtd">
```

```
<dataset>
```

```
<항목 레벨="1" 노드번호="1">
<제목><지명>서울</지명>의 문화</제목>
</항목>
```

```
<항목 레벨="2" 노드번호="6">
<제목>인물</제목>
</항목>
```

```
<항목 레벨="3" 노드번호="7">
<제목><인명 유형="여성" 식별자="강정일당|姜靜一堂">정일당 강씨
(靜一堂姜氏, 1772~1832)</인명></제목>
```

```
<메타데이터>
```

```
<분야>삶의 주체</분야>
<유형>인물/여성</유형>
<시대>조선시대</시대>
```

```
</메타데이터>
```

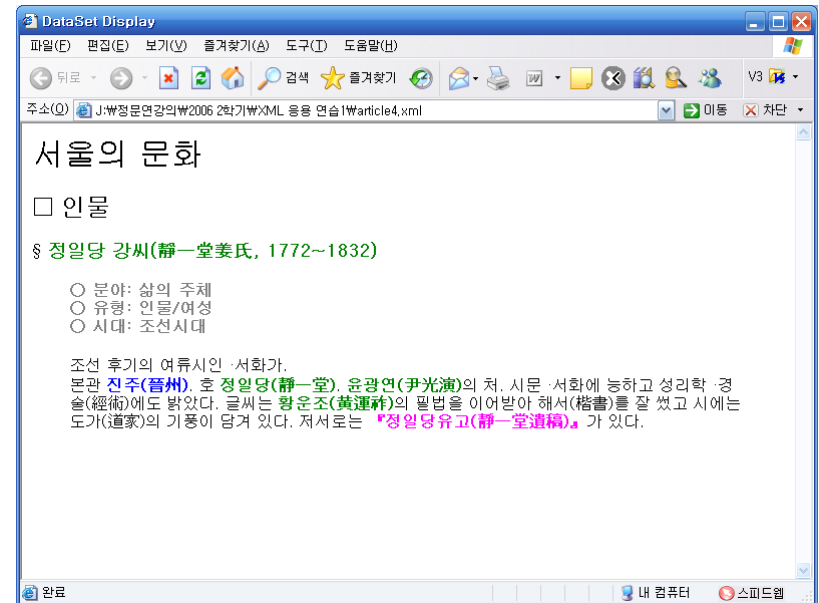
```
<본문>
```

```
<문단>조선 후기의 여류시인 ·서화가.</문단>
<문단>본관 <지명>진주(晉州)</지명>. 호 <인명 유형="여성"
식별자="강정일당|姜靜一堂">정일당(靜一堂)</인명>. <인명
유형="성명" 식별자="윤광연|尹光演">윤광연(尹光演)</인명>
의 처. 시문 ·서화에 능하고 성리학 ·경술(經術)에도 밝았다. 글
씨는 <인명 유형="외국" 식별자="황운조|黃運祚">황운조(黃運
祚)</인명>의 필법을 이어받아 해서(楷書)를 잘 썼고 시에는 도
가(道家)의 기풍이 담겨 있다. 저서로는 <서명>『정일당유고
(靜一堂遺稿)』</서명>가 있다.</문단>
```

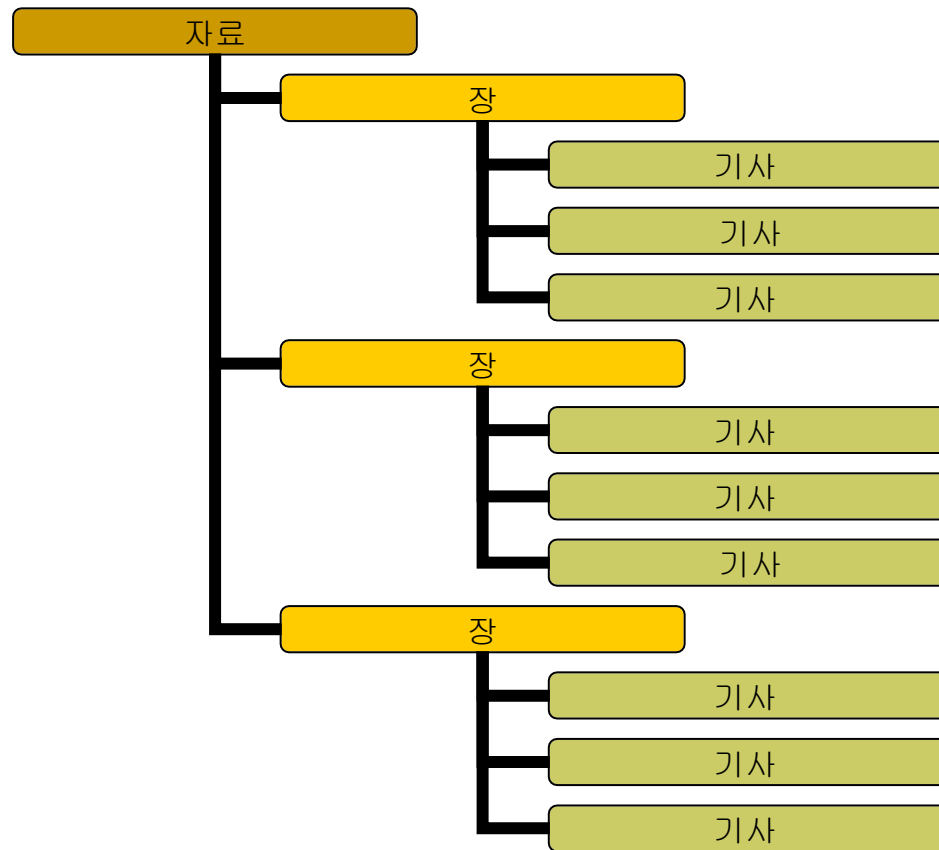
```
</본문>
```

```
</항목>
```

```
</dataset>
```



디자인(1): 노드 구성



❖ 제1수준: 자료

<항목>

<제목></제목>

</항목>

❖ 제2수준: 장

<항목>

<제목></제목>

</항목>

❖ 제3수준: 기사

<항목>

<제목></제목>

<메타데이터>

<분야></분야>

<유형></유형>

<지역></지역>

<시대></시대>

</메타데이터>

<본문>

<문단></문단>

</본문>

</항목>

❖ 텍스트 내의 정보 요소

<인명>, <지명>, <서명> ⇒ <제목>, <문단> 요소의 하위 요소가 될 수 있음

❖ 콘텐츠 목차 서비스

‘자료 > 장 > 기사’의 자료 구조를 계층적 목차 형태로 보여 주고, 이를 통해 단위 기사에 접근할 수 있도록 함.

❖ 디렉토리 서비스

‘분야’, ‘유형’, ‘지역’, ‘시대’별로 정리된 기사 목록을 보여 주고, 이를 통해 단위 기사에 접근할 수 있도록 함.

❖ 정보 검색 서비스

텍스트 안에 존재하는 ‘인명’, ‘지명’, ‘서명’을 검색하거나, 임의의 문자열을 검색할 수 있도록 함.

1. 단위 노드의 XML 문서는 더 이상 분해하지 않고, 하나의 XML 컬럼에 적재한다.
2. 주요 메타데이터 요소는 별도 테이블의 독립 컬럼 데이터로 관리한다.
3. XML 본문 컬럼과 메타데이터 컬럼의 데이터는 항상 일치되도록 한다.

- ❖ ‘단위 노드’로 취급되는 원시 XML 문서는 하나의 ‘XML 컬럼’에 담기도록 한다.

기사번호	국문기사
10001976	<항목 ID="10001976"> <항목명> <대표항목명>장서각</대표항목명> <삽도><시청각 ID="P0148">장서각 - 전경</시청각></삽도> </항목>
10000312	<항목 ID="10000312"> <항목명> <대표항목명>백헌집</대표항목명> <삽도><시청각 ID="P0149">백헌집 표지</시청각></삽도> </항목>
10002014	<항목 ID="10002014"> <항목명> <대표항목명>정일당 강씨 사당</대표항목명> <삽도><시청각 ID="P0336">정일당 강씨 사당 - 전경</시청각></삽도> </항목>

- ❖ 주요 메타데이터 요소는 별도의 정형 데이터로 관리한다.

기사번호	국문기사						
10001965	<항목 ID="10001965"> <항목명> <대표항목명>『입학도설』</대표항목명> <글항목명>입학도설</글항목명> <한자항목명>入學圖說</한자항목명> </항목명> <메타데이터> <분야>문화유산/기록유산</분야> <유형>문헌/단행본</유형> <지역>경기도 성남시 분당구 운중동</지역> <시대>조선/조선 전기</시대> <위치 연관="2">공간 식별자="성남:한국학중앙연구원 장서각" /></위치> </메타데이터> <삼도><시청각 ID="1219">입학도설</시청각></삼도> </항목>						
기사번호	한글제목	한자제목	분야	유형	지역	시대	위치
10001965	입학도설	入學圖說	문화유산/기록유산	문헌/단행본	경기도 성남시 분당구 운중동	조선/조선 전기	성남:한국학중앙연구원 장서각

- ❖ XML 본문과 여기에서 추출된 정형 데이터는 항상 일치되도록 한다
 - XML 본문과 추출된 정형 데이터의 동기화: '가상 뷰' 또는 '트리거' 사용

트리거(trigger):

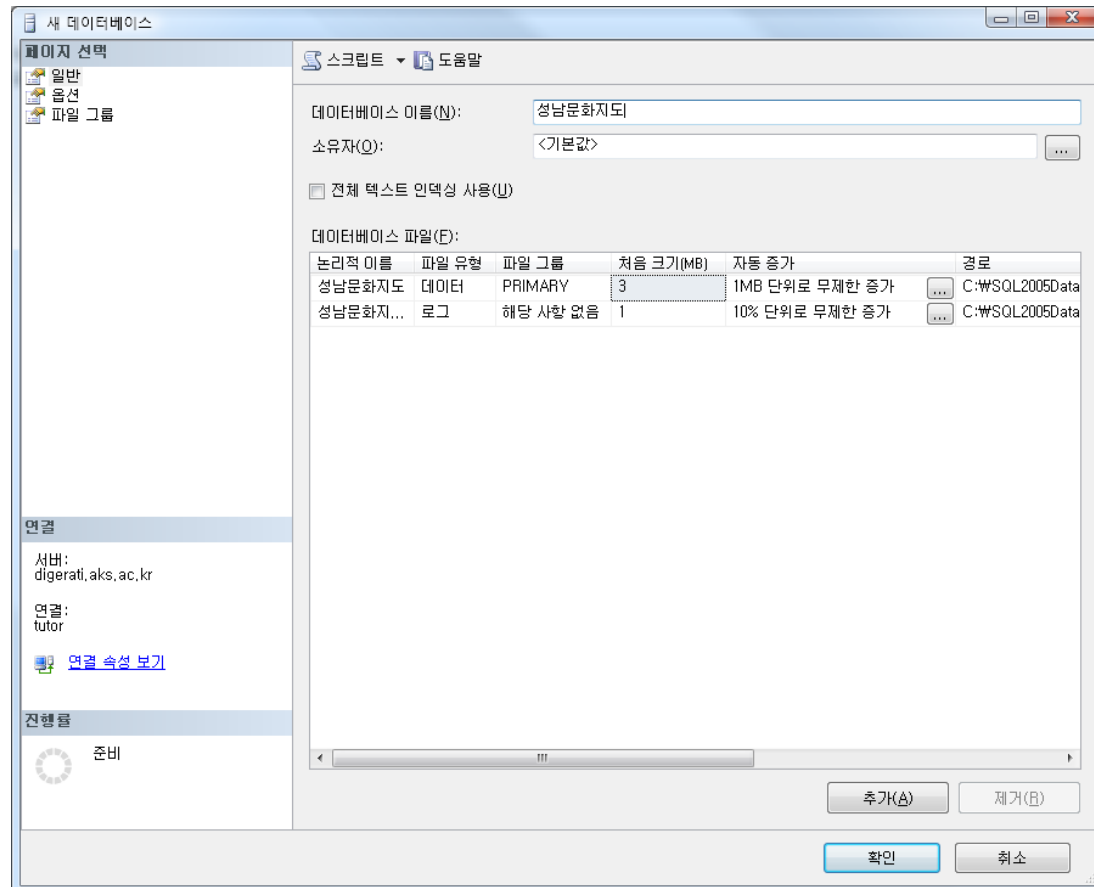
- 데이터의 변경시 자동적으로 동작하도록 예정된 프로그램
- a SQL procedure that initiates an action when an event (INSERT, DELETE or UPDATE) occurs.

The slide features a minimalist design with a vertical green line on the left and a horizontal green line intersecting it. Two green rectangular boxes are positioned at the intersection: a light green box above a darker green box. The text '1. XML 데이터베이스 설계' is inside the light green box, and '2. XML 데이터베이스 구현' is inside the darker green box.

1. XML 데이터베이스 설계

2. XML 데이터베이스 구현

- Microsoft SQL Server Management Studio
- 데이터베이스 > 새 데이터베이스



❖ 원시 XML 문서의 저장 공간: [텍스트]

열 이름	데이터 타입	데이터 소스	특성
기사번호	nvarchar(20)	'/항목' 요소의 'ID' 속성	Not null, primary key
국문기사	xml	입력 데이터	Not null
영문기사	xml	입력 데이터	Null

USE 성남문화지도

```
CREATE TABLE 텍스트(  
    기사번호 nvarchar(20) NOT NULL,  
    국문기사 xml NOT NULL,  
    영문기사 xml NULL,  
    CONSTRAINT [PK_텍스트] PRIMARY KEY CLUSTERED(기사번호 ASC)  
)
```

❖ 자료번호 자동 추출을 위한 트리거 설정: [텍스트].[기사번호]

```
CREATE TRIGGER 번호처리 ON 텍스트 INSTEAD OF INSERT
AS
    INSERT INTO 텍스트(기사번호, 국문기사)
    SELECT
        t.국문기사.value('(/항목/@ID)[1]', 'nvarchar(20)')
            AS 기사번호,
        t.국문기사 AS 국문기사
    FROM inserted t
```

❖ 메타데이터 관리를 위한 가상 뷰: [안내]

열이름	데이터 형식	컬럼	데이터 경로
기사번호	nvarchar(20)	기사번호	
한글제목	nvarchar(max)	국문기사	/항목/항목명/대표항목명
한자제목	nvarchar(max)	영문기사	/항목/항목명/한자항목명
영문제목	nvarchar(max)	영문기사	/항목/항목명/대표항목명
국문정의	nvarchar(max)	국문기사	/항목/본문/문단[1]
영문정의	nvarchar(max)	영문기사	/항목/본문/문단[1]
분야	nvarchar(40)	국문기사	/항목/메타데이터/분야
유형	nvarchar(40)	국문기사	/항목/메타데이터/유형
지역	nvarchar(40)	국문기사	/항목/메타데이터/지역
시대	nvarchar(40)	국문기사	/항목/메타데이터/시대
위치	nvarchar(40)	국문기사	/항목/메타데이터/위치/공간/@식별자
위치연관	int	국문기사	/항목/메타데이터/위치/@연관

❖ 가상 뷰 생성을 위한 SQL 문

```
CREATE VIEW 안내 AS
SELECT
    기사번호,
    국문기사.value('/항목/항목명/대표항목명')[1], 'nvarchar(max)') AS 한글제목,
    국문기사.value('/항목/항목명/한자항목명')[1], 'nvarchar(max)') AS 한자제목,
    영문기사.value('/항목/항목명/대표항목명')[1], 'nvarchar(max)') AS 영문제목,
    국문기사.value('/항목/본문/문단')[1], 'nvarchar(max)') AS 한글정의,
    영문기사.value('/항목/본문/문단')[1], 'nvarchar(max)') AS 영문정의,
    국문기사.value('/항목/메타데이터/분야')[1], 'nvarchar(40)') AS 분야,
    국문기사.value('/항목/메타데이터/유형')[1], 'nvarchar(40)') AS 유형,
    국문기사.value('/항목/메타데이터/지역')[1], 'nvarchar(40)') AS 지역,
    국문기사.value('/항목/메타데이터/시대')[1], 'nvarchar(40)') AS 시대,
    국문기사.value('/항목/메타데이터/위치/공간/@식별자')[1], 'nvarchar(40)') AS 위치,
    국문기사.value('/항목/메타데이터/위치/@연관')[1], 'int') AS 위치연관
FROM 텍스트
```


❖ XML 문서를 [국문기사] 컬럼에 적재

```
INSERT INTO 텍스트(국문기사) VALUES(
  CONVERT(xml,
    N'<항목 ID="10001642">
      <항목명>
        <대표항목명><인명>이경석</인명> 선생 묘</대표항목명>
        <한글항목명>이경석 선생 묘</한글항목명>
        <한자항목명>李景奭 先生 墓</한자항목명>

      </항목명>
      <메타데이터>
        <이칭별칭></이칭별칭>
        <키워드>묘소|무덤|신도비</키워드>
        <분야>역사/전통시대|문화유산/유형유산</분야>
        <유형>유적/능묘</유형>
        <지역>경기도 성남시 분당구 석운동</지역>
        <시대>조선/조선 후기</시대>

      .....

    </항목>', 1)
)
```

❖ XML 문서를 [영문기사] 컬럼에 적재

```
UPDATE 텍스트
SET 영문기사 = CONVERT(xml,
N'<항목 ID="10001642">
    <항목명>
        <대표항목명>Tomb of <인명>Yi Gyeongseok</인명></대표항목명>
    </항목명>
    <메타데이터>
        <분야>History / Traditional Period | Cultural Heritage / Tangible Cultural Heritage</분야>
        <유형>Archaeological Sites / Tombs</유형>
        <지역>San 16-18 Seogun-dong, Bundang-gu, Seongnam-si, Gyeonggi-do</지역>
        <시대>Joseon</시대>
    .....
</항목>', 1)
WHERE 기사번호 = '10001642'
```

❖ 메타데이터 및 전자문서 연계 조회

```
SELECT 안내.*, 텍스트.국문기사 FROM 안내, 텍스트 WHERE 안내.기사번호=텍스트.기사번호
```

기사번호	한글제목	한자제목		분야	유형	지역	시대	위치	전자문서
10001642	이경석 선생 묘	李景奭 先生 墓	...	문화유산 /유형유산	유적/능묘	경기도 성남 시 분당구 석운동	조선/조선 후기	성남:이경석 묘 및 신도비	<항목 ID="10001642"> <항목명> </항목명> </항목>
10001657	이곤 사당	李坤 祠堂	...	종교/유 교	유적/건물	경기도 성남 시 분당구 판교동	현대/현대	성남:이곤사 당	<항목 ID="10001657"> <항목명> </항목명> </항목>
10001976	장서각	藏書閣	...	문화·교 육/교육	지명/시설	성남시 분당 구 운중동	근대/일제 강점기 현 대/현대	성남:한국학 중앙연구원 장서각	<항목 ID="10001976"> <항목명> </항목명> </항목>

❖ XML 전자문서의 갱신에 따라 메타데이터 뷰 자동 갱신

❖ XML 데이터의 갱신

```
update 텍스트 set 국문기사=N'.....' where 기사번호='.....'
```

❖ 데이터 삭제

```
Delete 텍스트 where 기사번호='.....'
```