


DIGITAL HUMANITIES

# XML의 이해

---

 AI 고전번역학 - 동양 서화 제발(題跋) AI 번역 및 큐레이션

 강혜원

 2026.04.03.



# 왜 XML이 필요한가

고전 한문 텍스트의 복합적 정보 구조와 디지털화의 필요성

[https://sillok.history.go.kr/id/kaa\\_10107017\\_001](https://sillok.history.go.kr/id/kaa_10107017_001)

## 기존 방식의 한계

고전 한문 텍스트는 원문, 번역, 주석, 인물, 지명 등 여러 층의 정보가 동시에 존재함.

기존 방식인 Word/PDF는 사람이 읽기에는 좋지만, 컴퓨터는 '무엇이 무엇인지' 의미를 알지 못함.

컴퓨터는 이 텍스트를 단순한 문자열로만 인식함. "이게 원문인지, 번역인지, 주석인지" 알 수 없음.

문자열 → 의미 없음

子曰學而時習之不亦說乎

→ 컴퓨터 입장: 이 텍스트를 단순한 문자열로만 인식. 원문인지, 번역인지, 주석인지 알 수 없음.



태조실록1권, 태조 1년 7월 17일 병신 1/2 기사 / 1392년 명 홍무(洪武) 25년 태조가 백관의 추대를 받아 수창궁에서 왕위에 오르다

丙申/十七日丙申, 太祖卽位于壽昌宮。先是, 是月十二日辛卯, 恭讓將幸太祖第, 置酒與之同盟, 儀仗已列。侍中裴克廉等白王大妃曰: "今王昏暗, 君道已失, 人心已去, 不可爲社稷生靈主, 請廢之。"

## XML로의 전환

연구 데이터를 컴퓨터가 이해할 수 있는 구조로 만들 필요.

검색·연결·분석·공유가 가능한 데이터로 전환이 필요함.

구조화 → 의미 있음

<original>子曰學而時習之不亦說乎 </original>  
<translation> 배우고 때때로 익히면 기쁘지 아니한가 </translation>

→ 컴퓨터 입장: "子曰學而時習之不亦說乎"는 원문이고, 배우고 때때로 익히면 기쁘지 아니한가"는 번역이다"

```
<?xml version="1.0" encoding="UTF-8"?>
<text id="waa_10107017_001">
  <title>
    <mainTitle>태조가 백관의 추대를 받아 수창궁에서 왕위에 오르다</mainTitle>
  </title>
  <date>
    <dateOccurred date="1392-07-17L0" type="서기"/>
  </date>
  <content>
    <paragraph>○丙申/十七日丙申, <index num="0002237_0" ref="M_0000889" sort="K" type="이름">太祖</index>卽位于<index num="0002238_0" sort="K" type="지명">壽昌宮</index>。先是, 是月十二日辛卯, <index num="0002239_0" ref="M_0001101" sort="K" type="이름">恭讓</index>將幸<index num="0002240_0" ref="M_0000889" sort="K" type="이름">太祖</index>第, 置酒與之同盟, 儀仗已列。侍中<index num="0002241_0" ref="M_0002701" sort="K" type="이름">裴克廉</index>等白王大妃曰: "今王昏暗, 君道已失, 人心已去, 不可爲社稷生靈主, 請廢之。" </paragraph>
    <translation>태조가 수창궁(壽昌宮)에서 왕위에 올랐다. 이보다 먼저 이달 12일에 공양왕(恭讓王)이 장차 태조의 사제(私第)로 거둥하여 술자리를 베풀고 태조와 더불어 동맹(同盟)하려고 하여 의장(儀仗)이 이미 늘어섰는데, 시중(侍中) 배극령(裴克廉) 등이 왕대비(王大妃)에게 아뢰었다. "지금 왕이 혼암(昏暗)하여 임금의 도리를 이미 잃고 인심도 이미 떠나갔으므로, 사직(社稷)과 백성의 주재자(主宰者)가 될 수 없으니 이를 폐하기를 청합니다."</translation>
    <newTranslation>태조가 수창궁(壽昌宮)에서 즉위하였다. 이보다 앞선 7월 12일 신묘일에, 공양왕이 태조의 사저(私邸)에 거둥하여 술자리를 베풀고 태조와 동맹(同盟)하려고 의장(儀仗)이 이미 도열하여 있었다. 이때 시중(侍中) 배극령(裴克廉) 등이 왕대비에게 아뢰기를 "지금 왕이 어둔해서 임금의 도리를 이미 잃고 인심도 떠나갔으므로, 사직(社稷)과 백성들의 주인이 될 수 없으니, 그를 폐하소서." 하였다.</newTranslation>
  </content>
</text>
```



# XML이란 무엇인가

데이터를 구조적으로 표현하는 마크업 언어

## XML의 정의

XML = eXtensible Markup Language

데이터를 구조적으로 표현하는 마크업 언어임.

### 핵심 정의

#### "텍스트에 의미 구조를 붙이는 방식"

XML은 단순한 텍스트가 아니라, 태그를 통해 데이터의 의미와 구조를 명확히 표현하는 방식  
이를 통해 컴퓨터가 텍스트의 각 부분이 무엇을 의미하는지 이해할 수 있게 됨

## XML의 특징

XML은 사람이 읽을 수 있고, 기계도 처리 가능한 텍스트 기반 형식임.

**가독성**  
사람이 읽을 수 있는 텍스트 형식

**기계 처리**  
컴퓨터가 이해하고 처리 가능

**구조화**  
데이터의 의미와 구조를 명확히 표현

**교환성**  
다양한 시스템 간 데이터 교환 용이

**i** XML은 텍스트에 의미 구조를 붙이는 방식으로, 고전 한문 텍스트를 구조적으로 관리할 수 있게 함.

# </> XML의 핵심 개념 3가지

— 데이터 구조화의 기본 원리 —

## 1. 태그(Tag)

의미를 붙이는 이름표

- 태그는 텍스트에 의미를 부여하는 **이름표**
- 시작 태그와 끝 태그로 정보를 감싸서 그 역할을 명확히 함.

```
<original>子曰學而時習之</original>
```

시작태그 → 내용 → 끝태그

## 2. 계층 구조

부모-자식 관계

- XML 문서는 **트리 구조**를 가짐
- 큰 상자(부모) 안에 작은 상자(자식)가 들어가는 방식으로 정보를 조직

```
<text>  
<original>...</original>  
<translation>...</translation>  
</text>
```

루트 요소 → 하위 요소

## 3. 자기설명성

의미가 명확한 구조

- 태그 이름 자체가 **정보의 역할**을 설명
- 예: <title>은 제목, <author>는 작가를 의미

```
<title>제목</title>  
<author>작가</author>
```

태그명 = 의미

**XML은 형태보다 의미 중심으로 텍스트를 조직하는 방식.**  
태그를 통해 정보의 역할을 명시하고, 계층 구조로 데이터를 체계적으로 관리하며, 자기설명성을 통해 누구나 쉽게 이해 가능.



# HTML과 XML의 차이

화면 표시 vs 의미 구조, 보기 vs 이해

## HTML vs XML 비교표

구분	HTML	XML
목적	<b>화면 표시</b> 웹페이지를 보기 좋게 표시하는 것이 목적 <b>보기</b>	<b>의미 구조</b> 데이터의 의미와 구조를 명확히 표현 <b>이해</b>
태그	<b>미리 정해져 있음</b> <h1>, <p>, <div> 등 미리 정의된 태그	<b>직접 정의</b> 연구 목적에 맞춰 직접 태그를 설계
역할	<b>보기</b> 어떻게 보일지에 초점	<b>이해</b> 무엇을 의미하는지에 초점

### ✓ 핵심 메시지



**HTML = 어떻게 보일까** (화면 표시, 보기 좋게)

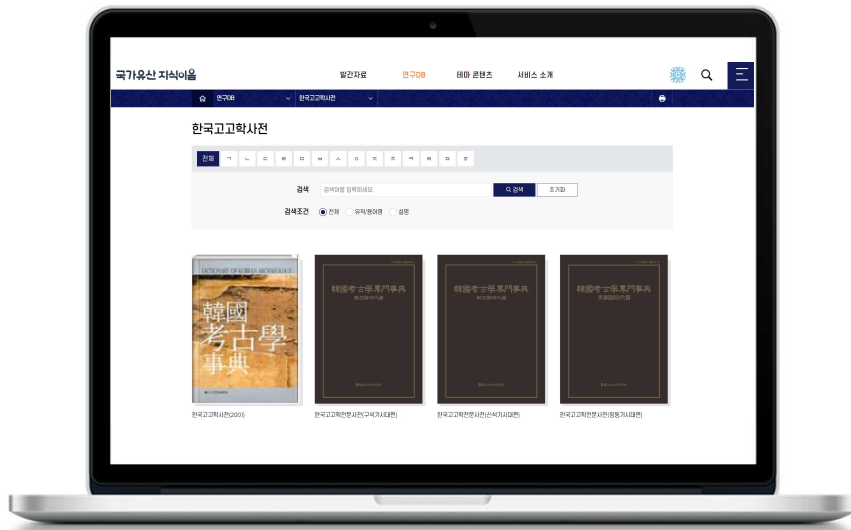
**XML = 무엇을 의미할까** (의미 구조, 이해 가능하게)



**XML은 단순 텍스트를 지식 데이터로 변환하는 기반임.**

고전 텍스트의 복합적 정보를 구조화하여, 검색·분석·공유·재사용이 가능한 디지털 자료로 만드는 핵심 기술임.

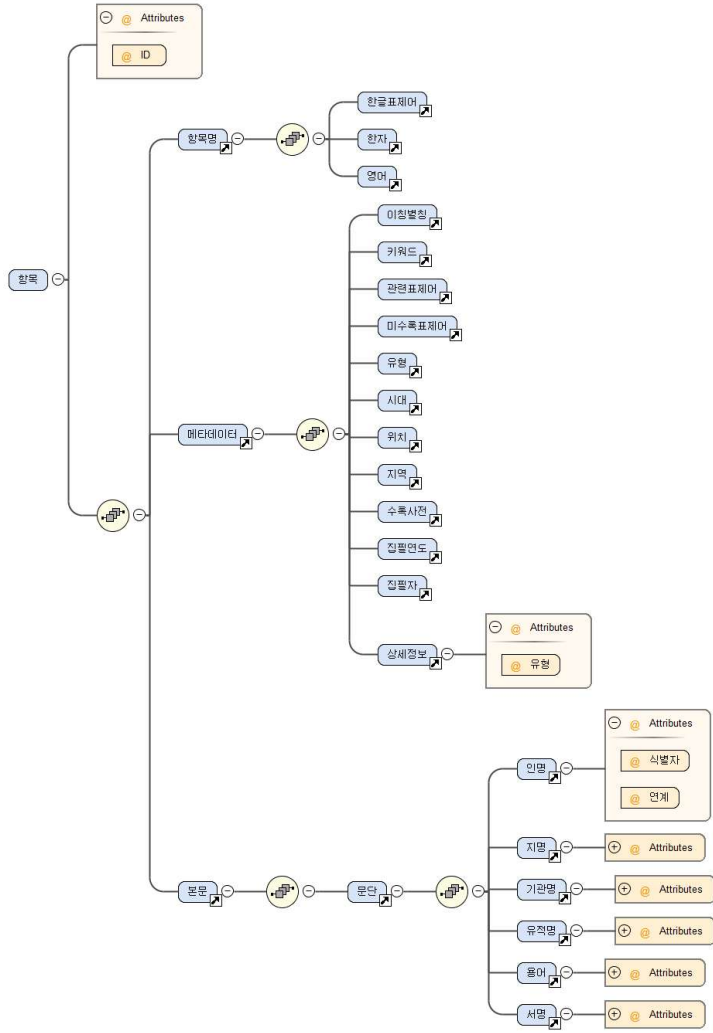
## </> XML 문서 예시 - 한국고고학사전



- **한국고고학사전**: 국립문화유산연구원에서 2001년부터 발간하고 있는 국내 유일의 고고학 전문사전
- 한국 고고학을 이해하는 데 필수적인 용어 정의와 함께 그간 우리나라에서 발굴조사한 중요 유적 뿐만 아니라, 세계 각지의 중요 고고학 이론 및 관련 유적에 관한 설명 수록
- 총론에 해당하는 『한국고고학사전(2001)』  
+ 주제별(고분, 성곽·봉수, 고분 유물, 생산유적) + 시대별(구석기, 신석기, 청동기) + 영문판 2종  
👉 **총 11종** 발간 (현재 증보 편찬 진행 중)
- 국내의 거의 모든 고고학 전문가들이 참여하여 한국 고고학계의 연구성과를 집대성한다는 점에서 한국 고고학 집단지성의 역량을 보여주는 유례없는 성과로 평가
- 2021년도 국립문화재연구원 문화유산 연구지식포털 이용 건수 총 2,189,311건 중 한국고고학사전 이용 건수는 총 407,518건



# </> XML 문서 예시 - 한국고고학사전



```

<?xml version="1.0" encoding="UTF-8"?>
<항목 ID="800-0003">
  <항목명>
    <한글표제어>가락바퀴</한글표제어>
    <한자>紡錘車</한자>
    <영어>Spindle whorl</영어>
  </항목명>
  <메타데이터>
    <이칭별칭>방추차</이칭별칭>
    <키워드>그물무늬 토기 | 점열무늬 | 새김 줄무늬</키워드>
    <관련표제어>그물무늬 토기 | 점열무늬</관련표제어>
    <미수록표제어>새김 줄무늬</미수록표제어>
    <유형>개념</유형>
    <시대>신석기</시대>
    <위치>대한민국</위치>
    <지역>진안 갈매리 유적 | 부산 동삼동 조도 조개더미 | 김해 수가리 조개더미 | 강릉 초당동 유적</지역>
    <수록사전>한국고고학전문사전(신석기시대편)</수록사전>
    <집필연도>2024</집필연도>
    <집필자>이상규</집필자>
    <상세정보 type="시간" 유형="관련 시기">신석기 시대</상세정보>
    <상세정보 type="공간" 유형="관련 지역">한반도 전 지역</상세정보>
    <상세정보 type="공간" 유형="관련 유적" 식별자="강릉초당동유적">강릉 초당동 유적</상세정보>
    <상세정보 type="공간" 유형="관련 유적" 식별자="김해수가리조개더미">김해 수가리 조개더미</상세정보>
    <상세정보 type="공간" 유형="관련 유적" 식별자="진안갈매리유적">진안 갈매리 유적</상세정보>
    <상세정보 type="공간" 유형="관련 유적" 식별자="부산동삼동조도조개더미">부산 동삼동 조도 조개더미</상세정보>
    <상세정보 유형="상격">출토 유물</상세정보>
    <상세정보 유형="관련 개념">방적구</상세정보>
  </메타데이터>
  <본문>
    <문단><용어 연계="1" 식별자="가락바퀴">가락바퀴</용어>는 실을 뽑는 방적구이며, ‘<용어 연계="1" 식별자="가락바퀴">방추차</용어>’ 라고도 한다. 가락바퀴 중앙에 둘러 있는 구멍에 가락(막대)을 끼워 넣고 동식물에서 얻은 섬유를 회전시켜 꼬아 실을 만든다. 한반도 <용어 연계="1" 식별자="신석기시대">신석기 시대</용어>의 가락바퀴는 대부분 흙을 구워 만들었으며, 지름은 4-5cm 정도이다. 평면은 둥글고, 단면은 주판알형이나 원판형(圓板形), 구형(球形), 원추형(圓錐形)을 이룬다. 가락바퀴의 표면에는 <용어 연계="2" 식별자="새김줄무늬">새김 줄무늬[沈縵文]</용어>, <용어 연계="1" 식별자="점열무늬">점열무늬</용어> 등의 무늬가 시문되기도 한다. <용어 연계="1" 식별자="새김줄무늬토기">새김 줄무늬 토기</용어> 조각을 가락바퀴로 재가공하여 새김 줄무늬가 시문된 것도 있다.</문단>
    <문단>가락바퀴는 신석기 시대 중기 이후 유적에서 많이 출토되나 양은 적은 편이다. 중기 이전의 가락바퀴 양상은 분명하지 않으나 <유적명 연계="1" 식별자="부산동삼동조도조개더미">부산 동삼동 조개더미</유적명>에서 출토된 토기 표면에 찍힌 그물 흔적으로 보아 낚시줄이나 그물 등을 제작하는 데 사용되었으리라고 추정된다. 가락바퀴는 한반도 신석기 시대 사람들이 동식물성 섬유로 실을 뽑아 직물을 제작했음을 보여주는 중요한 자료이다.</문단>
    <문단>일본에서는 <용어 연계="1" 식별자="야요이시대">야요이(弥生) 시대</용어> <유적명 연계="2" 식별자="키토라가와유적">키토라가와(鬼虎川) 유적</유적명>에서 가락바퀴의 중앙 구멍에 나무로 만든 목재 가락이 낀 상태로 출토된 것이 있으며, 중국 윈난성(雲南省) 일대 소수 민족들이 가락바퀴를 이용하여 실을 짜는 민족지 사례도 알려져 있다.</문단>
  </본문>
</항목>
  
```

# </> XML 문서 예시 - 한국고고학사전

## 신석기 시대 가락바퀴[新石器時代紡錘車]

문서 토론

읽기 편집 원본 편집 역사 보기 ★

### 설명 [ 편집 | 원본 편집 ]

가락바퀴는 실을 뽑는 방직구이며, '방추차'라고도 한다. 가락바퀴 중앙에 뚫려 있는 구멍에 가락(막대)을 끼워 넣고 동식물에서 얻은 섬유를 회전시켜 꼬아 실을 만든다. 한반도 신석기 시대의 가락바퀴는 대부분 흙을 구워 만들었으며, 지름은 4-5cm 정도이다. 평면은 등글고, 단면은 주판알형이나 원판형(圓板形), 구형(球形), 원추형(圓錐形)을 이룬다. 가락바퀴의 표면에는 새김 줄무늬[沈線文], 점열무늬 등의 무늬가 시문되기도 한다. 새김 줄무늬 토기 조각을 가락바퀴로 재가공하여 새김 줄무늬가 시문된 것도 있다.

가락바퀴는 신석기 시대 중기 이후 유적에서 많이 출토되나 양은 적은 편이다. 중기 이전의 가락바퀴 양상은 분명하지 않으나 부산 동삼동 조개더미에서 출토된 토기 표면에 찍힌 그물 흔적으로 보아 낚시줄이나 그물 등을 제작하는 데 사용되었으리라고 추정된다. 가락바퀴는 한반도 신석기 시대 사람들이 동식물성 섬유로 실을 뽑아 직물을 제작했음을 보여 주는 중요한 자료이다.

일본에서는 아요이(弥生) 시대 키토라가와(鬼虎川) 유적에서 가락바퀴의 중앙 구멍에 나무로 만든 목재 가락이 낀 상태로 출토된 것이 있으며, 중국 윈난성(雲南省) 일대 소수 민족들이 가락바퀴를 이용하여 실을 짜는 민족지 사례도 알려져 있다.

### 이미지 [ 편집 | 원본 편집 ]



신석기 시대 가락바퀴	
기본 정보	
동의어	방추차
시대	신석기 시대
지역	한반도 전 지역
관련 정보	
유적	강릉 조당동 유적, 김해 수가리 조개더미, 진안 갈머리 유적
사전 정보	
수록 사전	한국고고학전문사전(신석기 시대 편)
집필 연도	2024
집필자	이상규

id	항목명	식별자	태깅어휘	종류	연계
600-1586	송국리식 집자리	한강	한강	지명	2
600-1586	송국리식 집자리	낙동강	낙동강	지명	2
600-1586	송국리식 집자리	금강	금강	지명	2
600-1586	송국리식 집자리	영산강	영산강	지명	2
600-1586	송국리식 집자리	보성강	보성강	지명	2
600-1586	송국리식 집자리	부여송국리유적	부여 송국리 유적	유적명	1
600-1586	송국리식 집자리	광주송암동유적	광주 송암동	유적명	2
600-1586	송국리식 집자리	서산휴암리유적	서산 휴암리	유적명	1
600-1586	송국리식 집자리	영암장천리유적	영암 장천리	유적명	1
600-1586	송국리식 집자리	거창대야리유적	거창 대야리 유적	유적명	1
600-1586	송국리식 집자리	집자리	집자리	용어	1
600-1586	송국리식 집자리	타원형구덩이	타원형 구덩이	용어	2
600-1586	송국리식 집자리	기둥구멍	기둥 구멍	용어	2
600-1586	송국리식 집자리	저장구덩이	저장 구덩이(貯藏孔)	용어	1
600-1586	송국리식 집자리	화덕자리	화덕 자리(爐址)	용어	1
600-1586	송국리식 집자리	방사성탄소연대측정법	방사성 탄소 연대	용어	1
600-1586	송국리식 집자리	송국리유형	송국리 유형	용어	1
600-1586	송국리식 집자리	송국리식토기	송국리식 토기	용어	1
600-1586	송국리식 집자리	흙자귀	흙자귀(有溝石斧)	용어	1
600-1586	송국리식 집자리	삼각형돌칼	삼각 돌칼(三角形石刀)	용어	1
600-1586	송국리식 집자리	플라스크모양붉은간토기	플라스크 모양의 붉은 간 토기(赤色磨研土器)	용어	1
600-1586	송국리식 집자리	간돌검	간 돌검(磨製石劍)	용어	1
600-1586	송국리식 집자리	돌살촉	돌살촉(石鏃)	용어	1
600-1586	송국리식 집자리	돌도끼	돌도끼(石斧)	용어	1
600-1586	송국리식 집자리	돌널무덤	돌널무덤(石棺墓)	용어	1
600-1586	송국리식 집자리	돌뚜껑용무덤	돌뚜껑용무덤(石蓋土壙墓)	용어	1
600-1586	송국리식 집자리	독무덤	독무덤(甕棺墓)	용어	1
600-1586	송국리식 집자리	이음독널	이음독널(合口式甕棺)	용어	2
600-1586	송국리식 집자리	삼각덧띠토기	삼각 덧띠 토기(粘土帶土器)	용어	2

### 참고문헌 [ 편집 | 원본 편집 ]

- 최득준. (2011). 한반도 신석기시대 방추차에 대한 소고. *한국신석기연구*, 21, 103-127. <https://www.riss.kr/link?id=A102089386>

분류: 한국고고학사전 | 한국고고학전문사전(신석기 시대 편)