

KoreaMed XML 만들기 (2)

이 윤 혜
KoreaMed 정보관리팀

1. XML KoreaMed 파일 점검하기

a. Explorer에서 띄어보기 (Explorer 5.0 이상에서 가능)

- * XML 파일이 있는 디렉토리에 KoreaMed DTD (KoreaMed.dtd)가 꼭 있어야만 한다.
- * XML 파일의 이름은 학술지약어명, 연도, 권, 호 순으로 생성하는 것이 좋다.
e.g., jkms2001v16n4.xml (J Korean Med Sci 2001년 8월호, 16권 4호)

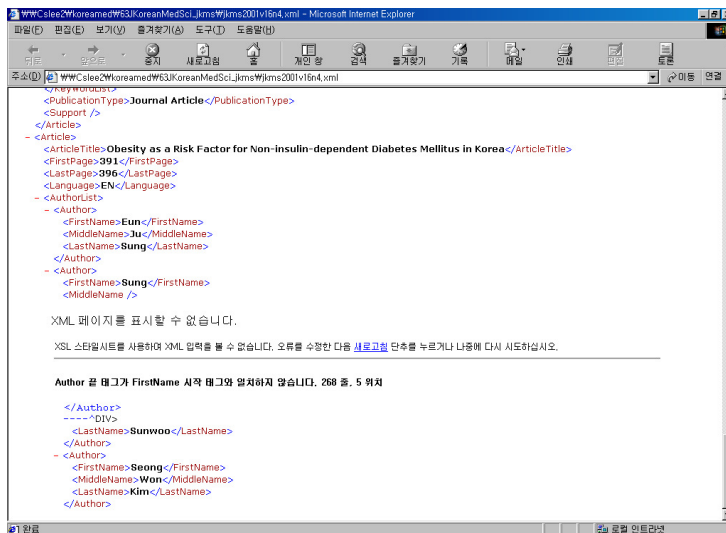


그림 1. Explorer에서 파일 오류 점검

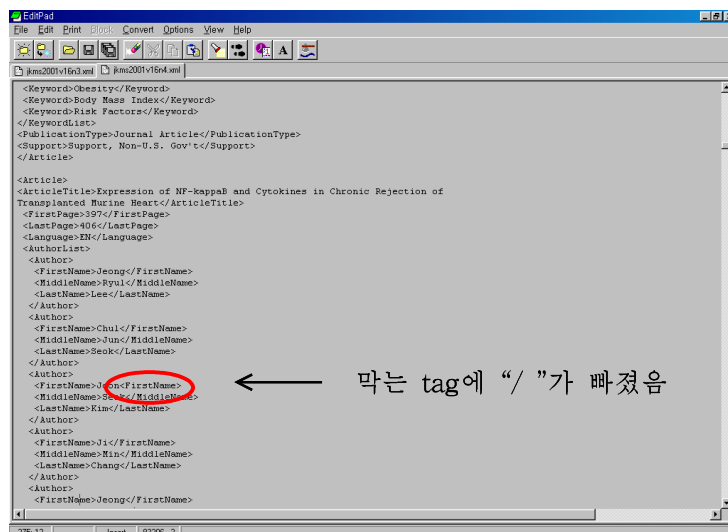


그림 2. Editpad에서 오류 확인

b. 오류 수정하기

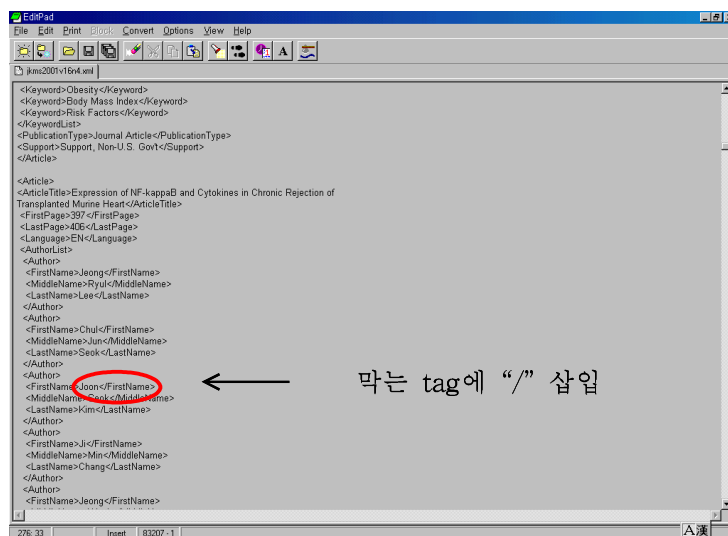


그림 3. Editpad에서 오류 수정

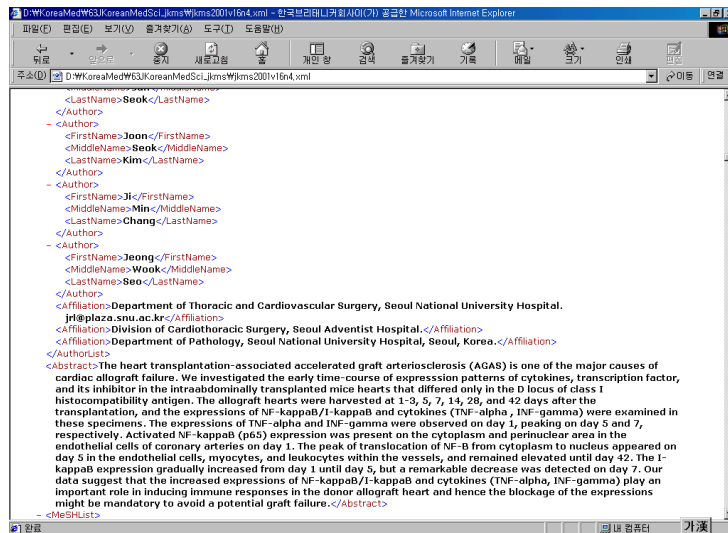


그림 4. 오류 없는 파일

2. 학술지 전자출판 파일을 text 파일로 변환할 때 (cut & paste)의 문제점

a. 문자가 사라지거나 깨지는 경우

- * Mac 파일을 텍스트 파일로 변환할 때, 변하거나 사라지는 문자.

e.g., $\alpha \rightarrow a$, $\beta \rightarrow b$, $\gamma \rightarrow r$, $\delta \rightarrow d$, $\mu \rightarrow m$

e.g., μ 가 사라지며 빈칸으로 남는다.

- * 한글 파일이나 pdf 파일에서 텍스트 파일로 변환할 때, 깨지거나 사라지는 문자.

e.g., “ (여는 따옴표), ” (닫는 따옴표), ‘ (여는 작은따옴표), ’ (닫는 작은따옴표),

<, >, α , β , γ , δ , μ

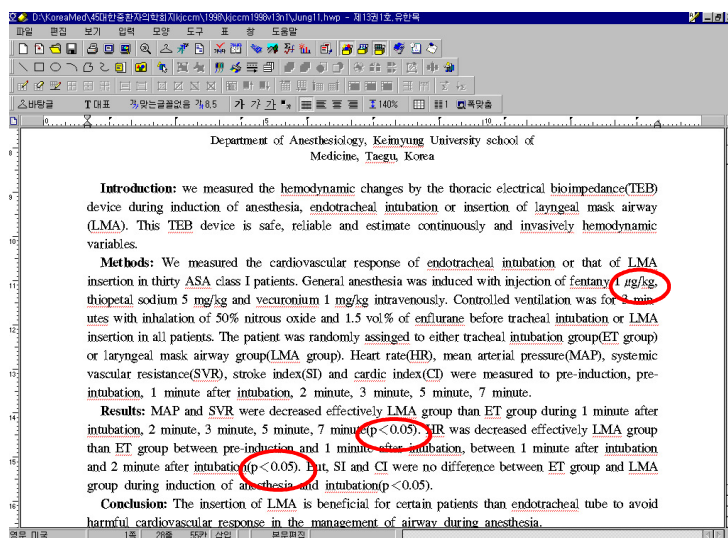


그림 5. 학술지 출판 파일 (한글 hwp 파일)

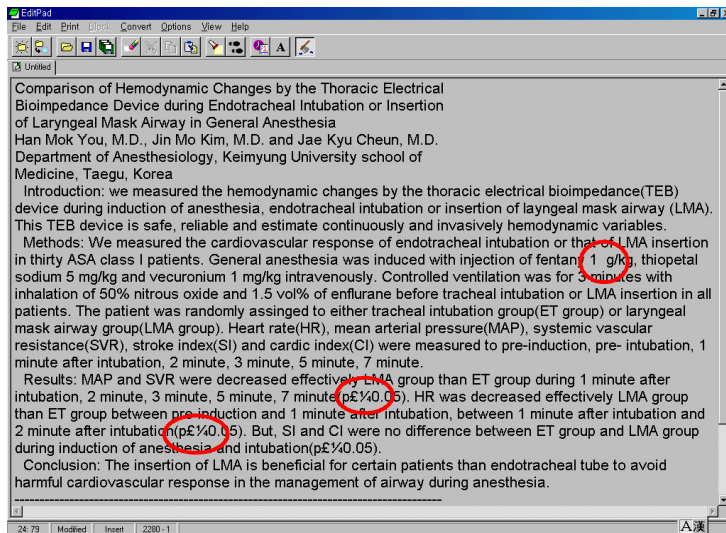


그림 6. 텍스트 파일로 변환했을 때 문자가 사라지거나 깨진 경우

b. Hyphen이 있는 경우

- * 한글 문서에 자동하이픈이 설정되어 있는 경우에는, 인쇄파일에 hyphen이 있더라도 텍스트 파일로 변환했을 때 hyphen은 자동으로 사라진다.
- * 그러나 일부러 hyphen을 넣어서 작업한 한글 파일이나 pdf 파일의 경우에는, 그 부분을 텍스트 파일로 변환하면 hyphen은 그대로 따라온다. 따라서 이런 파일의 경우에는 hyphen을 일일이 삭제해야 한다.

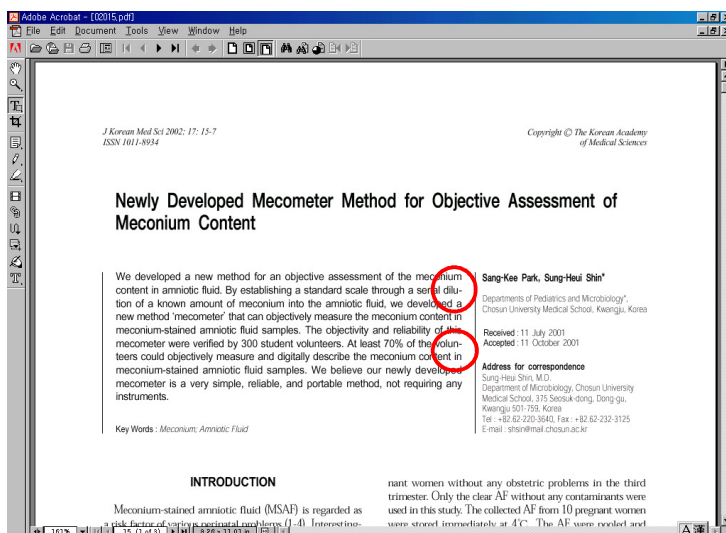


그림 7. hyphen이 있는 출판 파일

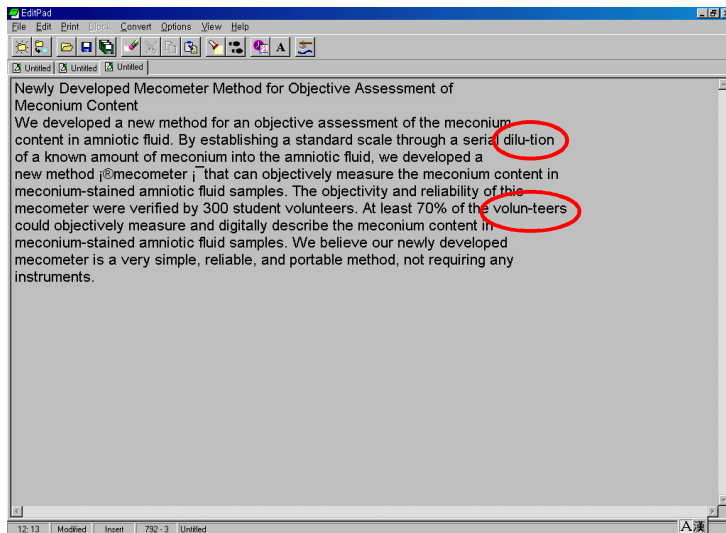


그림 8. hyphen이 그대로 따라오는 경우

c. 빈칸 (extra space)이 많이 있는 경우

- * 초록에 필요 이상의 빈칸이 있는 경우, 빈칸을 지우는 것이 좋다. KoreaMed 검색 후 뜨는 abstract 화면이나 full record 화면에 빈칸이 그대로 나타나기 때문이다.

3. 특수문자

a. XML 예약문자

- * <, >, &는 XML tag와 entity를 표시하는 XML 예약문자
- * 따라서 XML 문서에 그대로 사용할 수 없다.

표 1. XML 예약문자

특수문자	KoreaMed		PubMed	
	입력	display	입력	display
< (less-than sign)	<	<	<	<
> (greater-than sign)	>	>	>	>
& (ampersand)	&	&	&	&

b. 특수문자 처리방법

- * α, β, γ 등의 그리스 문자
alpha, beta, gamma 등으로 풀어 쓴다.
- * 위첨자, 아래첨자 등
화학기호일 경우는 무시

표 2. KoreaMed에서 많이 사용되는 특수문자 처리예

특수문자	KoreaMed		PubMed	
	입력	display	입력	display
α	alpha	alpha	&agr;	alpha
β	beta	beta	&Bgr;	beta
γ	gamma	gamma	&ggr;	gamma
ϵ	Epsilon	Epsilon	&egr;	Epsilon
δ	delta	delta	&dgr;	delta
<	<	<	<	<
>	>	>	>	>
&	&	&	&	&
\pm	+/-	+/-	±	+/-
m^2	m2	m2	m²	m2
10^{-6}	10(-6)	10(-6)	10(-6)	10(-6)
10_6	&Bgr;10(6)	10(6)	10<inf>(6)</inf>	10(6)
$height^{2.239}$	height(2.239)	height(2.239)	height^{2.239}	height(2.239)
H ₂ O	H2O	H2O	H2O	H2O

c. Scientific notation

- * Scientific notation 중 일부 (μ g, μ L 등의 단위)는 full spelling을 표기하지 않으면 의미가 상실된다.

표 3. Full spelling을 표기하지 않으면 scientific notation의 의미 상실하는 특수문자

특수문자	풀어쓰기	KoreaMed 입력 방법
μ g	mug	microgram
μ l	mul	microliter
°		degree
°C		degrees C

KoreaMed에서 많이 사용되는 특수문자의 처리 예 (표 1, 표 2, 표 3)는 대한의학학술지편집인협의회 web site (<http://kamje.kams.or.kr>)에서 내려받을 수 있습니다.

4. KoreaMed XML 파일 올리기와 색인하기

a. XML 파일 올리기 (uploading)

- * 완성된 XML 파일을 KoreaMed 데이터베이스에 추가하는 과정
- * 파일이 완성 되는대로 즉시 처리

* 파일에 오류가 있는지 검사하고 message 제공

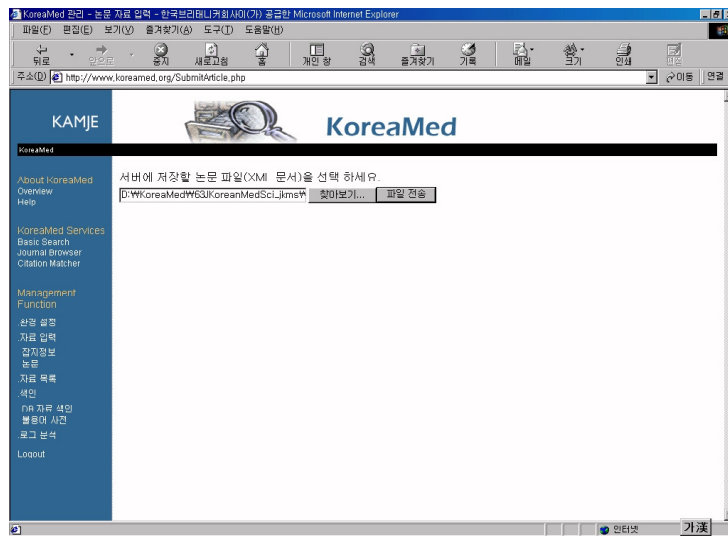


그림 9. XML 파일 uploading 화면



그림 10. KoreaMed XML data parsing 오류 검사 결과

b. 색인하기 (indexing)

* 색인작업 중에는 서비스가 중지되며 다음 message가 뜬다.

“Service is temporarily closed due to the indexing and/or halting by administrator. Please try again. It will take maximum 20 minutes. Thank you for your patience.”

* 약 10분 정도 소요.

* 현재는 1일 1회 색인작업 및 파일 갱신

* XML 파일을 KoreaMed 시스템에 uploading한 후 색인하기를 시행하기 이전까지는, Journal Browser를 통하여 파일이 데이터베이스에 추가된 것을 확인할 수 있

으나 검색을 시행하면 "No items found"로 처리된다.

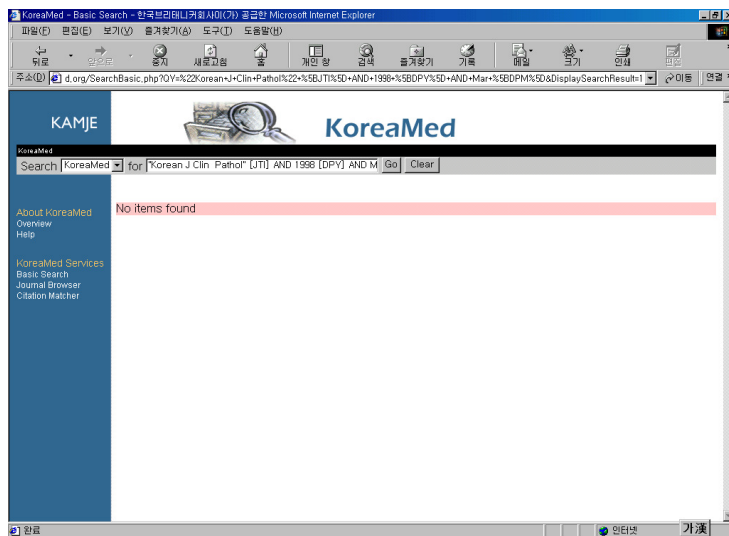


그림 11. 색인작업이 아직 시행되지 않아서 검색결과가 "No items found"