# Problem Statement

The literature loader obtains articles from pubmed. Periodically the XML format changes and a new DTD is published. In order to accommodate the latest changes and correctly parse the XML. The updated java classes corresponding to the new DTD need to be prepared. This is accomplished with jaxb, a command line utility.

The following message is given when the literature loader encountered a new DTD version while parsing the XML document obtained from pubmed.

Enter: validateAgainstDTD

https://dtd.nlm.nih.gov/ncbi/pubmed/out/pubmed\_240101.dtd

Found new DTD version

DTD change in loadArticles function validateAgainstDTD - new dtd: https://dtd.nlm.nih.gov/ncbi/pubmed/out/pubmed\_240101.dtd

# Task

### Prepare the java POJO by running jaxb with the latest DTD obtained from pubmed.

curl -LOk https://dtd.nlm.nih.gov/ncbi/pubmed/out/pubmed\_240101.dtd

curl -LOk https://dtd.nlm.nih.gov/ncbi/pubmed/out/pubmed\_230101.dtd

curl -LOk <https://dtd.nlm.nih.gov/ncbi/pubmed/out/pubmed_190101.dtd>

The previous jaxb generated POJO from the DTD pubmed\_190101.dtd is available here:

https://github.com/pellst/xbdevops\_testcase/tree/main/lit\_loader\_dtd\_upd/src/generated\_ncbi\_pubmed190101/org/xenbase/literature/jaxb/beans/pubmed

The folder has been prepared for submitting the pull request for the changes to be applied to the literature loader based on the DTD pubmed\_240101.dtd

<https://github.com/pellst/xbdevops_testcase/tree/main/lit_loader_dtd_upd/src/todo_generated_ncbi_pubmed240101/org/xenbase/literature/jaxb/beans/pubmed>

Please review the instructions per the section titled JAXB – compilation of DTD given below. Using these instructions run jaxb with the DTD pubmed\_240101.dtd and output by giving the required path “-p org.xenbase.literature.jaxb.beans.pubmed”.

### Prepare a pull request to submit your changes to the github project so that they can be deployed to the literature loader code base.

## JAXB - compilation of DTD

Literature Loader uses *jaxB* to generate Java classes from XML schema or, in our case, DTD. The whole *jaxB* is deprecated starting from JDK 11. The good news is that we continue to use jaxB thanks to the resources here: <https://blog.payara.fish/how-to-update-an-application-using-jaxb-from-jdk-8-to-jdk-11> <https://projects.eclipse.org/projects/ee4j.jaxb-impl> <https://eclipse-ee4j.github.io/jaxb-ri/>

[jaxB and xjc in Java 11](https://dzone.com/articles/running-jaxb-xjc-compiler-with-openjdk-11)

Moreover, DTD as a source never was fully implemented even in Java 8. If you call **xjc** in command line, you’ll see the list of possible parameters

$ xjc

...

-dtd  : treat input as XML DTD (experimental,unsupported)

...

jaxB binding generation:

C:\java\jdk1.8.0\_101\bin\xjc.exe -dtd "<https://eutils.ncbi.nlm.nih.gov/eutils/dtd/20060628/esearch.dtd>" -d C:\Users\xbsrc\active\_wip\issues\year2017\jira633\_literature\_loader\jaxb\generated\_ncbi\_esearch -p org.xenbase.literature.jaxb.beans.esearch

C:\java\jdk1.8.0\_101\bin\xjc.exe -dtd "<https://dtd.nlm.nih.gov/ncbi/pubmed/out/pubmed_180101.dtd>" -d C:\Users\xbsrc\active\_wip\issues\year2017\jira633\_literature\_loader\jaxb\generated\_ncbi\_pubmed -p org.xenbase.literature.jaxb.beans.pubmed

rem old version C:\java\jdk1.8.0\_101\bin\xjc.exe -dtd "<https://www.nlm.nih.gov/databases/dtd/nlmcatalogrecordset_170101.dtd>" -d C:\Users\xbsrc\active\_wip\issues\year2017\jira633\_literature\_loader\jaxb\generated\_ncbi\_nlmcat -p org.xenbase.literature.jaxb.beans.nlmcat

C:\java\jdk1.8.0\_101\bin\xjc.exe -dtd "<https://www.nlm.nih.gov/databases/dtd/nlmcatalogrecordset_170601.dtd>" -d C:\Users\xbsrc\active\_wip\issues\year2017\jira633\_literature\_loader\jaxb\generated\_ncbi\_nlmcat -p org.xenbase.literature.jaxb.beans.nlmcat

rem failed elink gen with DTD

C:\java\jdk1.8.0\_101\bin\xjc.exe -dtd "<https://eutils.ncbi.nlm.nih.gov/eutils/dtd/20101123/elink.dtd>" -d C:\Users\xbsrc\active\_wip\issues\year2017\jira633\_literature\_loader\jaxb\generated\_ncbi\_elink -p org.xenbase.literature.jaxb.beans.elink

It could be necessary to edit DTD file locally, so it can be then processed by **xjc**.

$/opt/IBM/WebSphere/AppServer/java/8.0/bin/xjc -d generated\_ncbi\_pubmed -p org.xenbase.literature.jaxb.beans.pubmed -dtd ./pubmed\_230101.dtd

When you try to parse DTD as a local file, **xjc** generates errors about missed libraries related to **mathml3**. These libraries are automatically loaded when you parse remote DTD, but for a local file they are expected to exist in some particular directories relatively to the main DTD file. Different DTDs may require different missed libraries.

The missed libraries can be downloaded from

<https://github.com/biopython/biopython/tree/master/Bio/Entrez/DTDs>

One discovered issue is that **xjc** doesn't like any spaces before closing **>** for XML elements. When you try to parse remote DTD and see an error like that

[ERROR] Either an attribute declaration or ">" is expected, not ">"

line 54 of <https://dtd.nlm.nih.gov/ncbi/pubmed/out/pubmed_230101.dtd>

then you need to load DTD locally and edit it. It can’t be the original

<!ATTLIST PubmedArticleSet

>

It can’t be the

<!ATTLIST PubmedArticleSet >

It has to be

<!ATTLIST PubmedArticleSet>

As a safe bet it's recommended to have all attributes and elements in one long line starting with <! and ended with >

jaxB dependency on mathml3:

<https://www.nlm.nih.gov/pubs/techbull/ma18/ma18_pm_display_formulas.html>

<ftp://ftp.ncbi.nlm.nih.gov/pubmed/sample-2018-06-01/>

sustainment jira ticket:

<https://zft3.xenbase.org/browse/SUS-3>

## Reference:

pmid 34712818

https://eutils.ncbi.nlm.nih.gov/entrez/eutils/efetch.fcgi?db=pubmed&retmode=xml&id=34712818

reference PMID 22975720

<https://eutils.ncbi.nlm.nih.gov/entrez/eutils/efetch.fcgi?db=pubmed&retmode=xml&id=22975720>

<https://github.com/pellst/xbdevops_testcase/tree/main/lit_loader_dtd_upd>

<https://github.com/pellst/xbdevops_testcase/blob/main/lit_loader_dtd_upd/JAXB_literature_loader_dtd_prep.docx>

https://github.com/pellst/xbdevops\_testcase/blob/main/schemaspy\_mod\_xb/schemaspy\_devel31/index.html

https://xenbase-bio1.ucalgary.ca/schemaspy/prod/schemaspy\_xbprod27/tables/PAPER.html

paper

https://xenbase-bio1.ucalgary.ca/schemaspy/prod/schemaspy\_xbprod27/tables/PAPER\_CITATION.html

paper\_citation