

# Instructions to Run bioPDFX

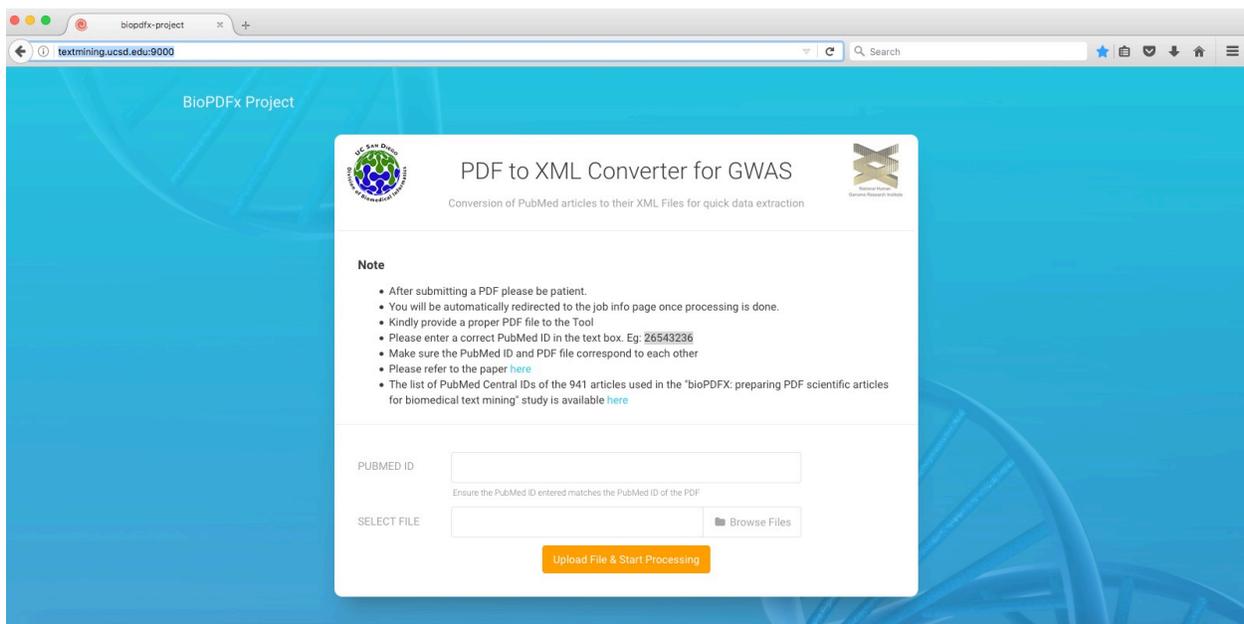
**Step 1.** Obtain a PubMed ID. For example, “26543236”.

**Step 2.** Obtain the PDF file from PubMed Central or the publisher of the paper. In our example, the PDF file of the paper can be found in the following PubMed Central link:

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4741908/>

**Step 3.** Open a Web browser (Firefox recommended) and connect to the bioPDFX service:

<http://textmining.ucsd.edu:9000/>



14 **Step 4.** Type or paste the PubMed ID in the “PUBMED ID” field:

15

**UC SAN DIEGO**  
Division of Biomedical Informatics

# PDF to XML Converter for GWAS

Conversion of PubMed articles to their XML Files for quick data extraction

**Note**

- After submitting a PDF please be patient.
- You will be automatically redirected to the job info page once processing is done.
- Kindly provide a proper PDF file to the Tool
- Please enter a correct PubMed ID in the text box. Eg: 26543236
- Make sure the PubMed ID and PDF file correspond to each other
- Please refer to the paper [here](#)
- The list of PubMed Central IDs of the 941 articles used in the "bioPDFX: preparing PDF scientific articles for biomedical text mining" study is available [here](#)

PUBMED ID

Ensure the PubMed ID entered matches the PubMed ID of the PDF

SELECT FILE

16

17

18

19 **Step 5.** Upload the PDF file by clicking the “Browse Files” button:

20

The screenshot shows the 'PDF to XML Converter for GWAS' web interface. At the top left is the UC San Diego Division of Biomedical Informatics logo, and at the top right is the National Human Genome Research Institute logo. The main heading is 'PDF to XML Converter for GWAS' with the subtitle 'Conversion of PubMed articles to their XML Files for quick data extraction'. Below this is a 'Note' section with a bulleted list of instructions. The instructions include: 'After submitting a PDF please be patient.', 'You will be automatically redirected to the job info page once processing is done.', 'Kindly provide a proper PDF file to the Tool', 'Please enter a correct PubMed ID in the text box. Eg: 26543236', 'Make sure the PubMed ID and PDF file correspond to each other', 'Please refer to the paper [here](#)', and 'The list of PubMed Central IDs of the 941 articles used in the "bioPDFX: preparing PDF scientific articles for biomedical text mining" study is available [here](#)'. Below the note is a form with a 'PUBMED ID' label and a text box containing '26543236'. A note below the text box says 'Ensure the PubMed ID entered matches the PubMed ID of the PDF'. Below that is a file selection area with a 'SELECT FILE' label, a text box containing '26543236.pdf', and a 'Browse Files' button. At the bottom of the form is an orange button labeled 'Upload File & Start Processing'.

**Note**

- After submitting a PDF please be patient.
- You will be automatically redirected to the job info page once processing is done.
- Kindly provide a proper PDF file to the Tool
- Please enter a correct PubMed ID in the text box. Eg: 26543236
- Make sure the PubMed ID and PDF file correspond to each other
- Please refer to the paper [here](#)
- The list of PubMed Central IDs of the 941 articles used in the "bioPDFX: preparing PDF scientific articles for biomedical text mining" study is available [here](#)

PUBMED ID

Ensure the PubMed ID entered matches the PubMed ID of the PDF

SELECT FILE

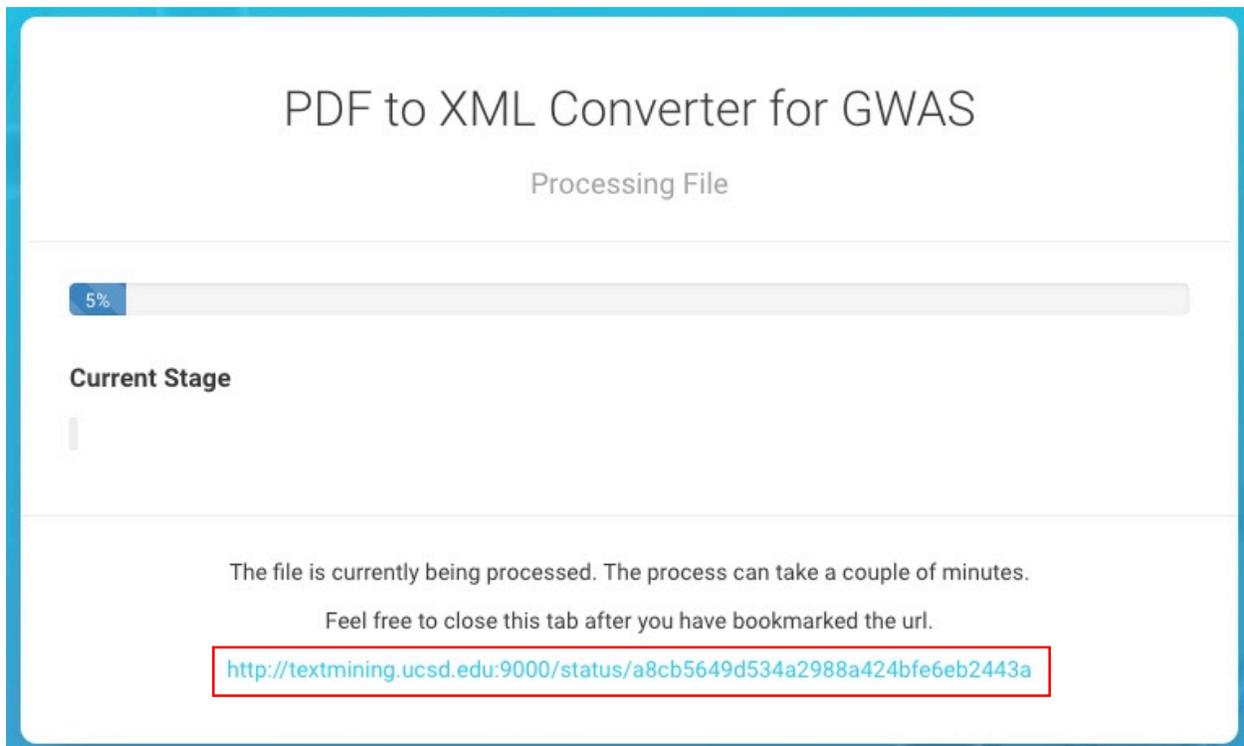
21

22

23

24 **Step 6.** Start bioPDFX by clicking “Upload File & Start Processing” button. A progress bar will  
25 show the percentage of the conversion process. Copy or bookmark the URL so the process can be  
26 checked later, even if the browser is closed.

27



The screenshot displays a web interface for a "PDF to XML Converter for GWAS". The main heading is "PDF to XML Converter for GWAS" with a sub-heading "Processing File". Below this is a progress bar showing 5% completion. Underneath the progress bar is a section titled "Current Stage" with a small vertical bar. At the bottom of the interface, there is a message: "The file is currently being processed. The process can take a couple of minutes. Feel free to close this tab after you have bookmarked the url." Below this message is a URL: <http://textmining.ucsd.edu:9000/status/a8cb5649d534a2988a424bfe6eb2443a>, which is highlighted with a red rectangular border.

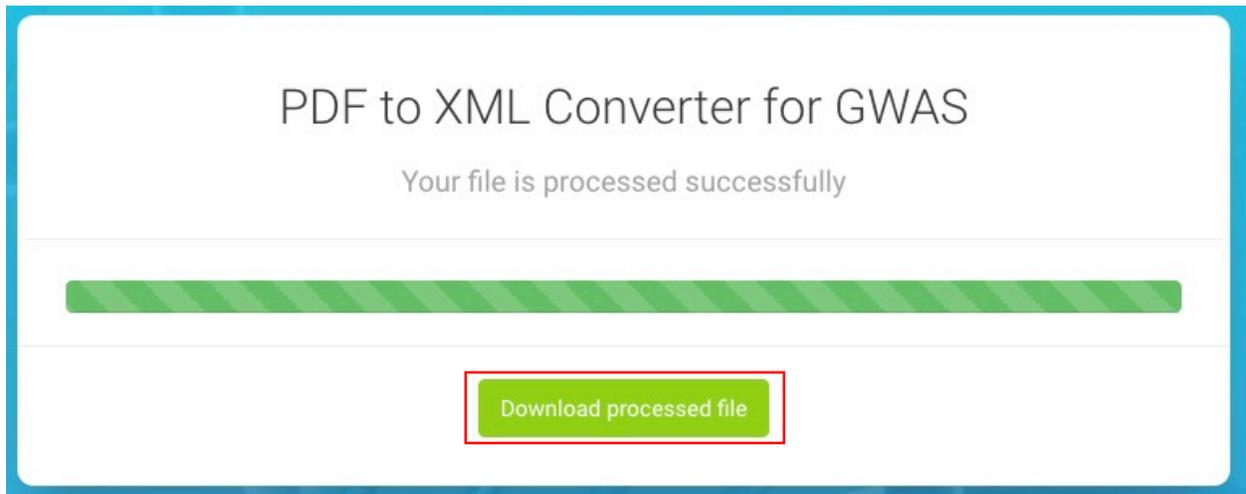
28

29

30

31 **Step 7.** After the process is completed, click the “Download processed file” to obtain converted  
32 XML file.

33



34

35