

Part 4

Full text extraction

VITA extracts text from the following file types: .doc, .pdf, .txt, .xls, .rtf.

Upload accompanying .doc, .pdf, .txt, .xls, .rtf files to include full text content from digital objects such as JPGs, TIFFs, PDFs where no text is available (such as those from Google Books), or web pages. When a .doc, .pdf, .txt, .xls, or .rtf file type is uploaded, the text is automatically extracted and loaded into the data management Full Text input boxes at the Page Edit level.

As with any full text content, correct the text on the file/page edit screen and select Yes to corrected OCR and Yes if you want this version of the text from this object to be the content that is indexed for discovery.

Halls1871.txt


Put in public display?:
 Yes No

Is this corrected OCR?:
 Yes No

Index the text in this file?:
 Yes No
Turn this off if you have uploaded multi

For PDFs: Select Yes for “Put in the Public Display” so that users will be able to download the PDF object(s) using the download icon in the Public Display.

COBOURG CENTENNIAL OLD BOYS' RE-UNION
1837 - 1937
Thursday, July 29th to Monday, August 2nd CORONATION YEAR
OFFICIAL PROGRAM AND HISTORICAL SOUVENIR PRICE 25c.

Comment on this item
Create electronic postcard
Full Image
 Cobourg Centennial (.PDF)

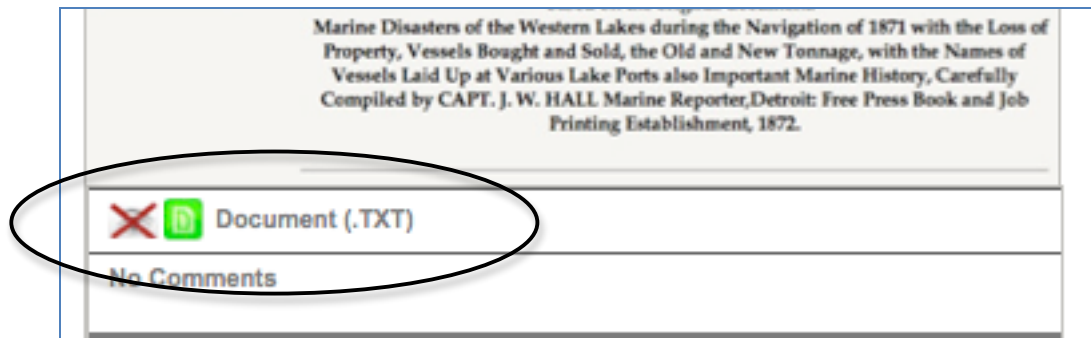
If a multipage PDF is uploaded as a single file, the text will be extracted and display in a single “toilet paper roll” on the Page Edit screen and not return users to page-level links for viewing the search terms in context.

Note: If the text from a text file other than PDF exceeds 64,000 characters, this friendly reminder will be returned in place of the full text edit box:

“The text associated with this file is too large to display for editing purposes. Some browsers can only handle about 64,000 characters in a text box. We have [75,047] in this instance. If you need to edit this file please do so in a desktop editing tool that can handle files of this size and then upload it again. **Note:** The number of characters should approach but may not necessarily match the file size noted above.”

If this is case, consider breaking the object text into pages in order to display and edit it in VITA; otherwise, be sure that the text is corrected before it is uploaded so that it can be indexed and considered “correct”.

For other files types: To include full text for discovery using **.doc, .pdf, .txt, .xls, or .rtf** file types where you do **not** want to expose the file itself, select No for “Put in the Public Display” on the File/Page edit screen. The file will display on the Data Management side, but with an icon that indicates it will not be seen on the public side:



Replacing Full Text:


If you need to replace full text in the index, delete the text you wish to replace and add the replacement text. Do not leave the full text input box empty or it will not overwrite the original. You must include something. To make deleted text unsearchable, enter some garbage like kesfhouuy4iuhnvmvmsdkjh and this will overwrite the original indexed text but not mislead people searching for real words.

Part 4.1 Full text discovery & display

Making full text content available for each text object and each page of multipart objects in a digital collection enables users to discover keywords at page level as they search.

With VITA 5, a user's keyword search returns the object results, with the keyword highlighted in a snippet from the first page hit as well as links to other pages within the object where the keyword is also found:

We found 1 matching item.
Keywords: flander Sof Relevance
Page 1 of 1




Cobourg Centennial 1837-1937
... McKenzie, C. C. 1.) In **Flanders** Fields.....Etheland Skill Last Post.....E. Cowin 23rd Speaker..... Col. J. W. Odell 7. Pageant, "May Day in Merrie England" Cobourg...

Pages/Parts: [2](#) [7](#)

Cobourg Centennial, 1837-1937, page 2
Pages/Parts: [2](#) [7](#)

Pages: Search within this title: Go



Lest We Forget

In Flanders' Fields
By CAPT. JOHN McRAE, of Guelph, Ont.

*In Flanders' fields the poppies blow
Between the crosses, row on row,
That mark our place, and in the sky
The larks still bravely singing fly,
Scarce heard amid the guns below.
We are the dead. Short days ago
We lived, felt down, saw sunset glow,
Loved and were loved, and now we lie
In Flanders' Fields.*

*Take up our quarrel with the foe,
To you from falling hands we throw
The Torch — be yours to hold it high;
If ye break faith with us who die
We shall not sleep though poppies grow.
In Flanders' Fields.*