

Structural Methods for Handwriting Analysis

Lab Session: Word Spotting

Andreas Fischer (andreas.fischer@unifr.ch)

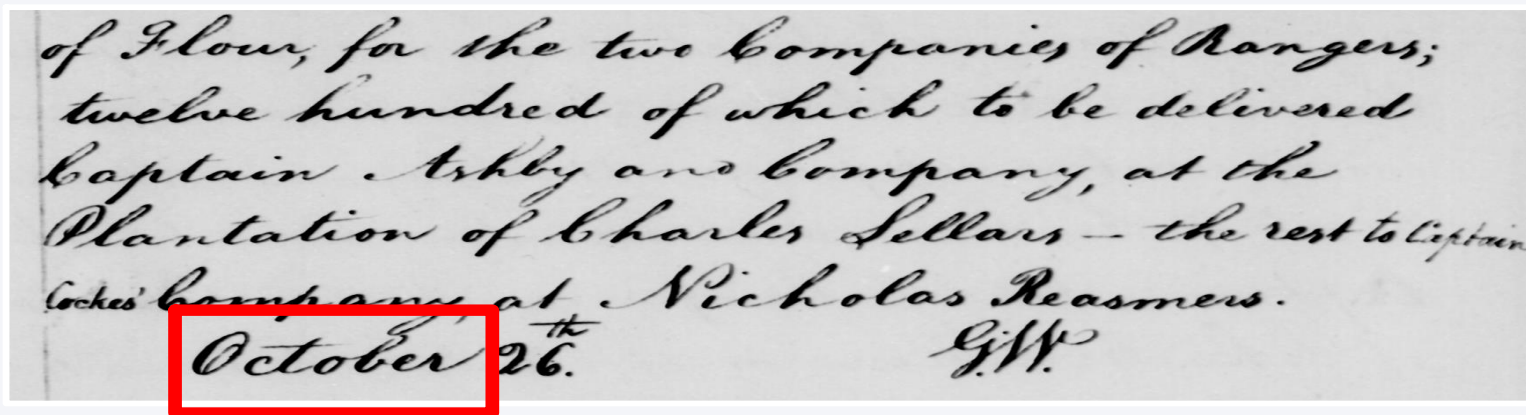
Marcel Würsch (marcel.wuersch@unifr.ch)

Digitizing historical manuscripts for cultural heritage preservation

Textual content: searching and browsing scanned page images

Widely unsolved for historical handwriting
too many writing styles and languages

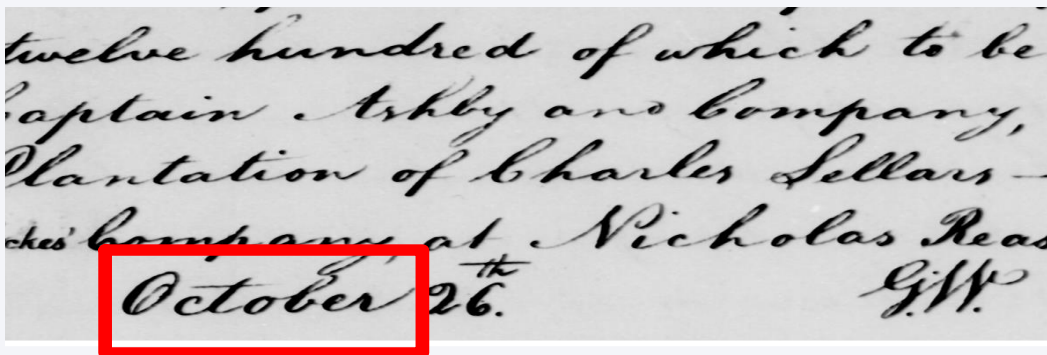
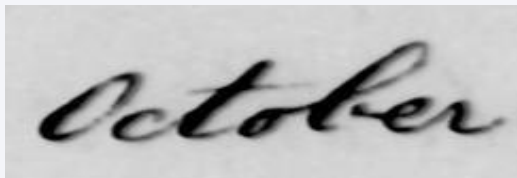
Keyword spotting is a “shortcut”: identify individual search terms



“one-shot learning”: provide one example word image

Goal: find similar word images in the manuscript

Usually constrained to a single-writer scenario
(sample from the same manuscript)



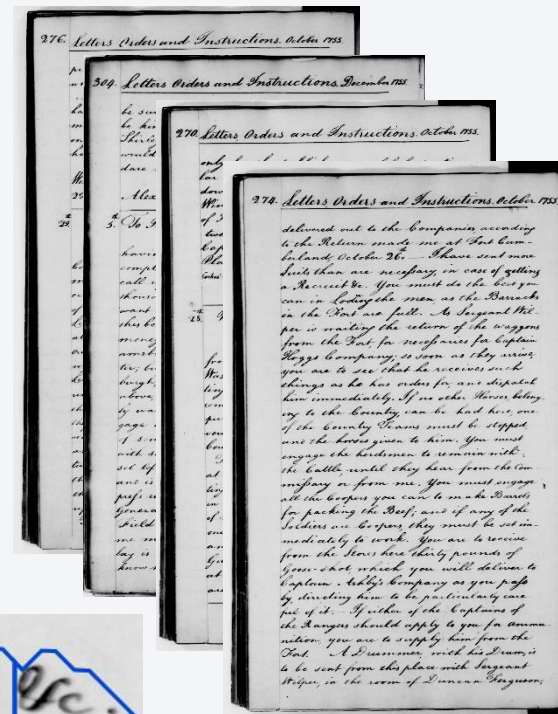
WashingtonDB

Letters of George Washington

Library of Congress

18th century, longhand script

Given under my Hand &c.
October G.W.

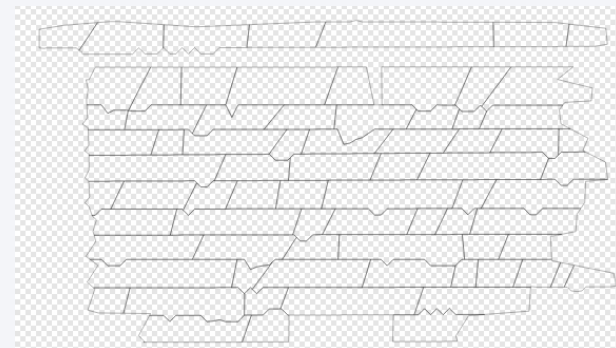


/data/ground-truth/transcription.txt

Character based transcription

/data/ground-truth/locations/*.svg

Polygons of word segments



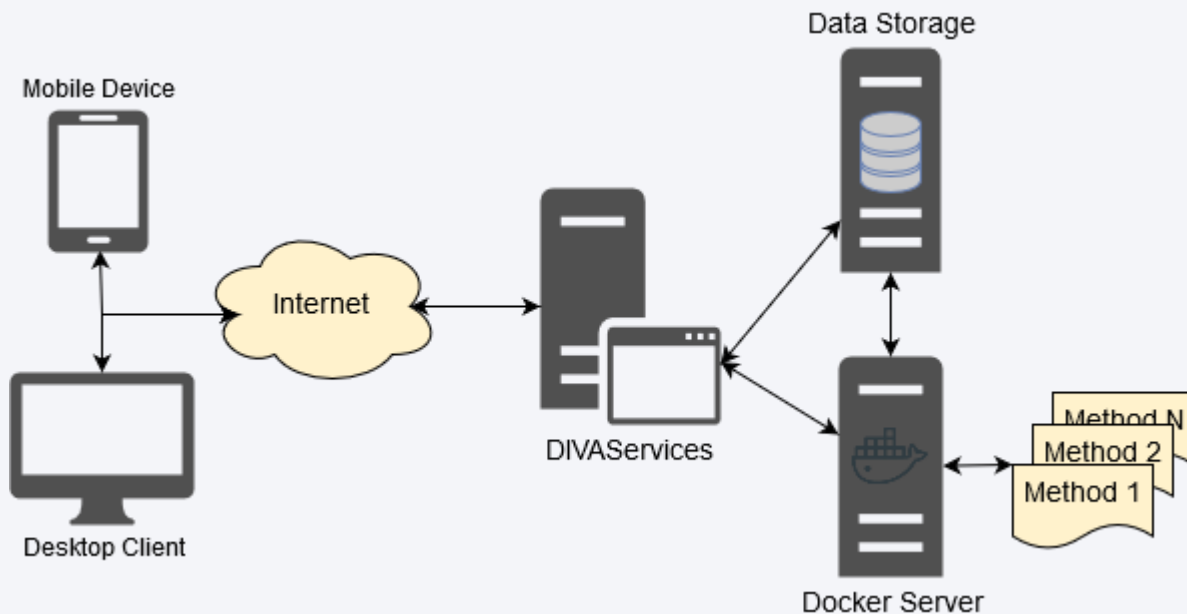
/data/images/*.jpg

The page images

Web Services

Unified API

Extensible



Download sources from Github

[http://bit.ly/SS LAB CODE](http://bit.ly/SS_LAB_CODE)

(all capitals!)



Questions?