PLIN064 Úvod do *digital humanities*

Zuzana Nevěřilová
`xpopelk@fi.muni.cz`

Centrum zpracování přirozeného jazyka, B203
Fakulta informatiky, Masarykova univerzita

25. září 2019

Practical Hands On: Text Analysis

DH Activities

DH Examples

Further Steps in this Course

Analog and Digital Documents

Activities I

Fundamental Activities (not only in DH):

- curation – selection and organization in an interpretive framework
- analysis – processing of text or data & visualization
- editing – keeping track of authenticity, origin, transmission, production
- modeling – assumption about knowledge specific to the subject (e.g. in case of analyzing correspondence, a *timeline*)

Activities I

Fundamental Activities (not only in DH):

- curation – selection and organization in an interpretive framework
- analysis – processing of text or data & visualization
- editing – keeping track of authenticity, origin, transmission, production
- modeling – assumption about knowledge specific to the subject (e.g. in case of analyzing correspondence, a *timeline*)

Activities II

Digital Activities:

- digitization,
- classification,
- description,
- metadata organization,
- navigation

[Burdick et al., 2012]

DH Examples I

Text

- linguistics: synchrony and diachrony
- language variations: dialects, slang, pidgin ...
- language and social groups
- literary texts: authors, eras, areas
- cross-lingual discourse
- discourse analysis: event extraction, communication studies
- names: toponymy, ethnonymy, demonymy ...
- information science
- history
- cultural analytics: trends in cultural change
- social studies

DH Examples II

Visual art

- new media
- virtual reality
- museology: curation, preservation, access
- archaeology: 3D modeling
- architecture

Multi-modal

- anthropology: sound + video of human activities
- analysis of video-games and other born-digital objects

Internet of Things

- sensor data

DH in Czechia

LINDAT/CLARIAH-CZ at FF

<https://it.muni.cz/phil/lindat-clariah>

Czech Association for Digital Humanities

<https://www.czadh.cz/>

Czech Academy of Sciences

<https://digitalhumanities.cz/>

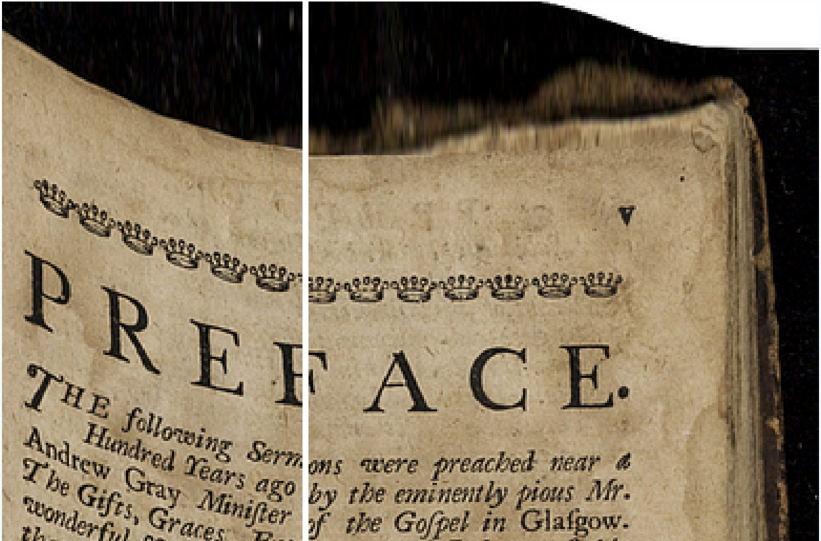
LINDAT/CLARIN, DARIAH-CZ

Further Steps in this Course

- Digital documents
- Analogue documents and digitization
- “Raw” data, enriched data, metadata
- Data sources
- Data science projects
- Data processing in Python
- Evaluation, presentation of results, visualization

Analog and Digital Documents: Preprocessing

Making (historical) analog document digital

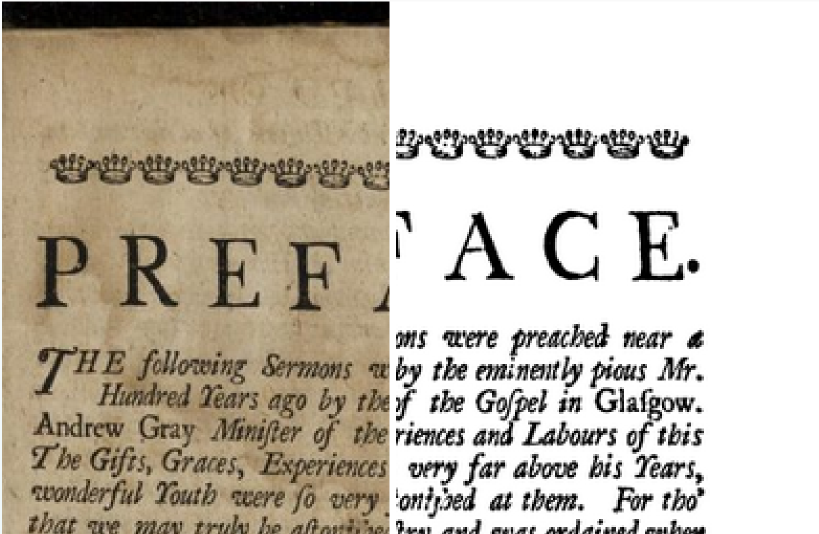


PREFACE.

THE following Sermons were preached near a
Hundred Years ago by the eminently pious Mr.
Andrew Gray Miniſter of the Goſpel in Glaſgow.
The Gifts, Graces,
wonderful

Analog and Digital Documents: Preprocessing

Making (historical) analog document digital



PREF

THE following Sermons w
Hundred Years ago by the
Andrew Gray Miniſter of the
The Gifts, Graces, Experiences
wonderful Youth were ſo very
that we may truly be aſtoniſhe

ACE.

ons were preached near a
by the eminently pious Mr.
of the Goſpel in Glaſgow.
riences and Labours of this
very far above his Years,
oniſhed at them. For tho'
ears and was ordained when

Optical Character Recognition (OCR)

1. Recognition of the area with text
2. Recognition of character shapes
3. Probability of a character in a context or other characters
4. Issues: different fonts, language-specific characters, noise

multi-layer PDFs

Software

Preprocessing: BIQE, Book Restorer

OCR: ABBY, InftyReader, Tesseract

neural network based OCR (Cloud service)

Analog and Digital Documents: Overview

- analog documents: historical artifacts
- digital documents: digital copy of analog documents
- retro-digital documents: originally analog document that were scanned and OCRed
- born-digital documents: artifacts that were created originally as digital documents

Czech experts: Digitalizační centrum Knihovny AV ČR v.v.i.

Burdick, A., Drucker, J., Lunenfeld, P., Presner, T., and Schnapp, J. (2012).
Digital_Humanities.
The MIT Press.

Terras, M., Nyhan, J., and Vanhoutte, E. (2013).
Defining Digital Humanities: A Reader.
Ashgate Publishing Company, Brookfield, VT, USA.