COURSE OVERVIEW

CSC 105 - DIGITAL HUMANITIES - SPRING 23

Marcus Birkenkrahe

January 10, 2023

What are "Digital Humanities" about?

- What do **you** think "Digital Humanities" are about? (*Why're you here?*)
- Compare: TOC of The Digital Humanities Cookbook (Drucker, 2021)
- Compare: Curriculum suggestion by ChatGPT (01-06-2023)

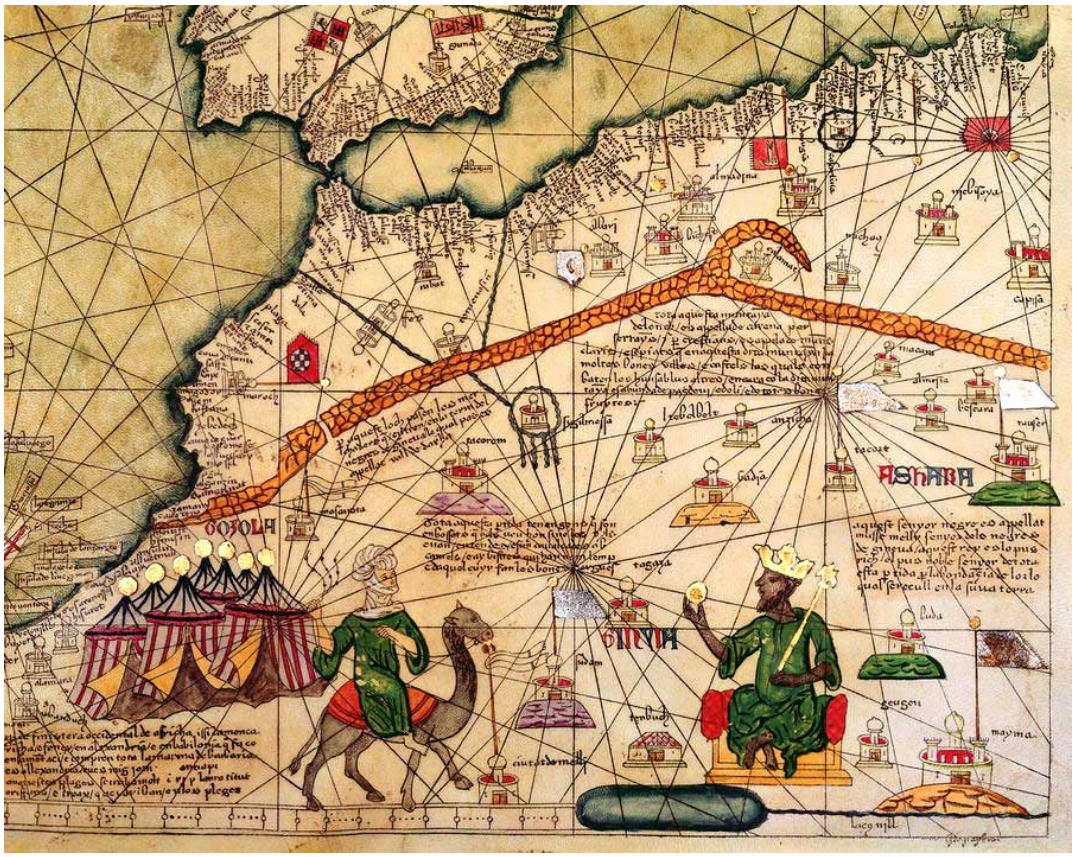


Figure 1: Map of West Africa in the Catalan Atlas (1375)

What will we do in this course?

- We're going to work with **text data** only (no maps[1])
- We will complete 50% of the "Text mining with R" DataCamp track
- You will complete one DataCamp chapter per fortnight
- In class, we will prepare and review the online material
- You will cover some of the most important AI/ML applications

How will you be evaluated?

- All course requirements have deadlines
- 8/10 home assignments are DataCamp lessons
- Late submissions will be penalized (loss of points)
- No final exam! But weekly tests are graded
- The project topic can come from any of the course sub-topics
- The project deliverable is an essay of at least 5,000 words

Which tools are you going to use?

- DataCamp courses (15 weekly home assignments)
- GitHub repository (all course materials except tests)
- GNU Emacs + ESS + R (literate programming environment)
- Canvas (Lyon's learning management system)

How can you register at DataCamp?

- You find the invitation link to the group for Spring 23 in Canvas.

[1] Maps and GIS (Geographic Information Systems) are super cool but deserve an introductory course on their own. The focus of my teaching is to get you practical experience that you can use on the job instead of a broad conceptual overview (though the latter would be easier for all of us). There is also a spring'23 GIS course in anthropology @Lyon.

WEEK	DATE	TOPIC / DATACAMP ASSIGNMENT	TESTS
1	Jan 10,12	Digital humanities & text mining	
2	Jan 17,19	**Text mining with Bag-of-Words**	Test 1
3	Jan 24,26	Syntactic and semantic parsing	Test 2
4	Jan 31, Feb 2	**Word clouds & other visuals**	Test 3
5	Feb 7,9	Graphics in R for text mining	
6	Feb 14,16	**Adding to your TM Skills**	Test 4
7	Feb 21,23	String manipulation	Test 5
8	Mar 2	**HR analytics case study**	Test 6
9	Mar 7,9	Review: Bag of Words technique	
10	Mar 14,16	**Fast & dirty: polarity scoring**	Test 7
11	Mar 28,30	Introduction to the tidyverse	Test 8
12	Apr 4,6	**Sentiment analysis with tidytext**	Test 9
13	Apr 11,13	term frequency vs inverse df	
14	Apr 18,20	**Visualizing sentiment**	Test 10
15	Apr 25,27	Machine learning and text mining	
16	May 2	**Case study: Airbnb reviews**	

Figure 2: Source: syllabus, Canvas (lyon.instructure.com) or GitHub (github.com/birkenkrahe/ml)

REQUIREMENT	UNITS	PPU	TOTAL	% of TOTAL
Final exam	0	0	0	0
Home assignments	10	15	150	30.
Class assignments	10	10	100	20.
Final project	1	150	150	30.
Multiple-choice tests	10	10	100	20.
TOTAL			500	100.

Figure 3: Source: syllabus, Canvas (lyon.instructure.com) or GitHub (github.com/birkenkrahe/ml)

Figure 4: Unsplash

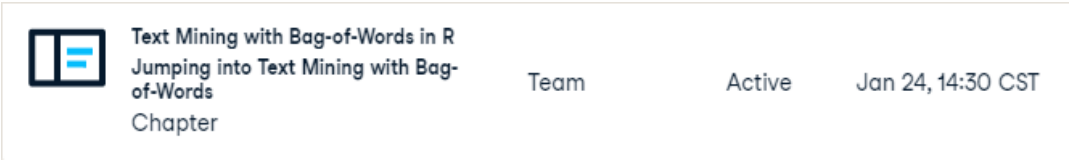


Figure 5: DataCamp assignment for January

- You will automatically be subscribed to the Digital Humanities team
- If you are in more than one course, I will add you later manually
- These accounts will be valid until July 8, 2023 only

When is the first assignment due?

- The first DataCamp home assignment is due on **January 19**. For late submissions, you lose 1 point per day (out of 10 possible points)
- The first in-class assignment is due on January 19. For late submissions, you lose 1 point per day (out of 10 possible points)
- We'll write the first weekly multiple-choice test on January 19.

What should we do as a project?

Figure 6: Source: learning.edanz.com

- The **final essay** (5,000+ words) should be about one of the areas of Digital Humanities that we did **not** cover in our course
- Here is the set of available essay topics (generated by an AI)
- Also possible: "**Topic modeling**" (available as DataCamp lesson), full chapter in the book by Kwartler - ML approach with clustering
- Technical, ethical, societal, and personal aspects of ChatGPT (here are some use cases):
    1. A psychology major could contrast "Eliza" and "ChatGPT" or test "Butter", a chatbot for individuals that stutter
    2. An English major could check ChatGPTs writing abilities
    3. A CompSci major could look at ChatGPT's technical implementation
- You need to do your own research, including a literature review, and adhere to the IMRaD framework (see video):
    1. **Introduction** (what did you want to research?) with abstract
    2. **Method** (what did you do?) with literature review
    3. **Results** (what did you find out?) with examples, illustrations
    4. **Discussion** (what does it mean?) with limitations and outlook

What else could you do for a good start?

Figure 7: Photo by Braden Collum on Unsplash

R proficiency

Complete/review introductory R or statistics courses:

- Introduction to R" in DataCamp (data structures)
- Introduction to statistics (no code)
- fasteR by Norman Matloff (GitHub) - fast lane to R

Literate Programming

If you do not have any experience with Emacs, work through the **online tutorial** (open it in Emacs with `CTRL + h t`) - ca. 1 hour.

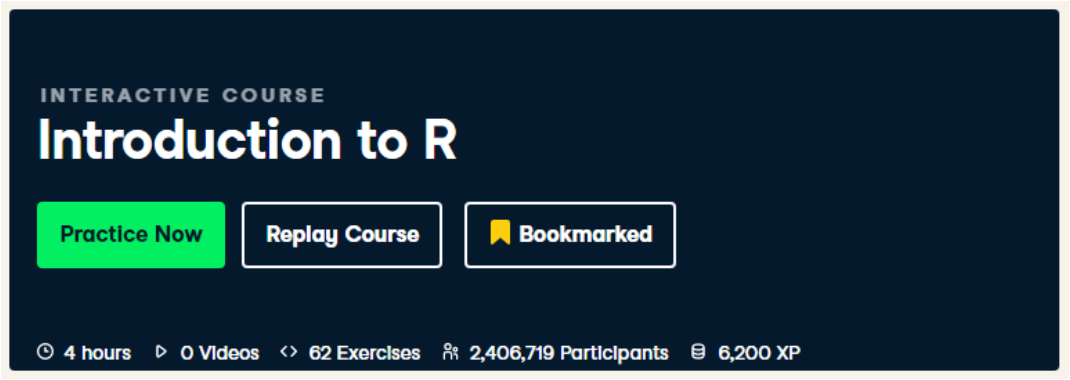


Figure 8: DataCamp course dashboard banner

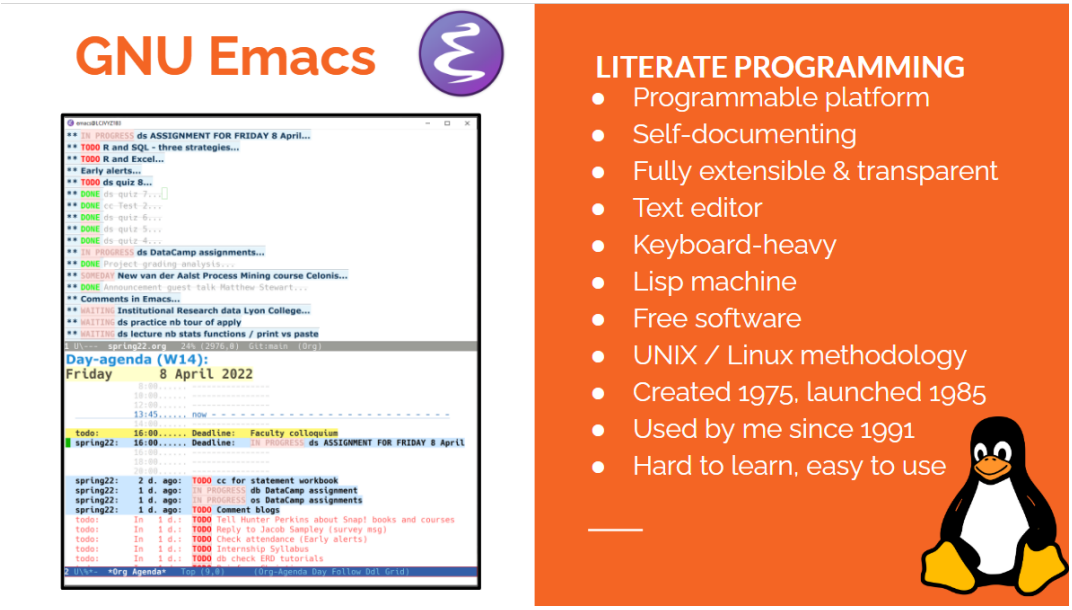
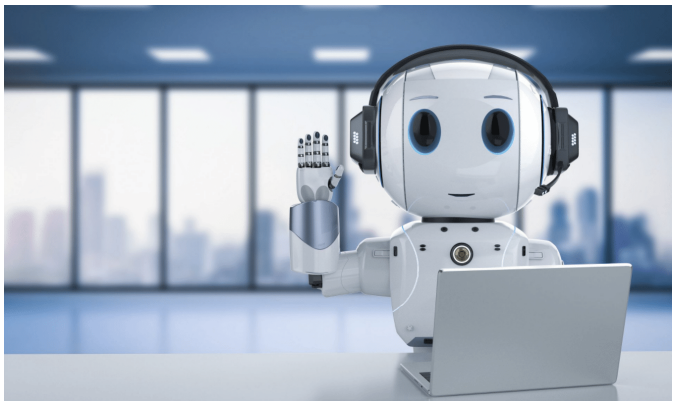


Figure 9: Literate Programming with GNU Emacs (illustration)

- Learn to open/close the editor
- Learn basic cursor control (moving around)
- Learn basic file management (open/close/find/save files)
- Learn basic windows (buffer) management

See my during my office hours for a personal tutorial!

What are you looking forward to?



- Learning more about text mining and NLP
- Having fun with R programming and real data sets
- Helping you on your own "digital humanities" journey

Next topic

- Text mining in practice: Bag of Words and Syntactic Parsing
- Base R data structures, functions and packages, importing data

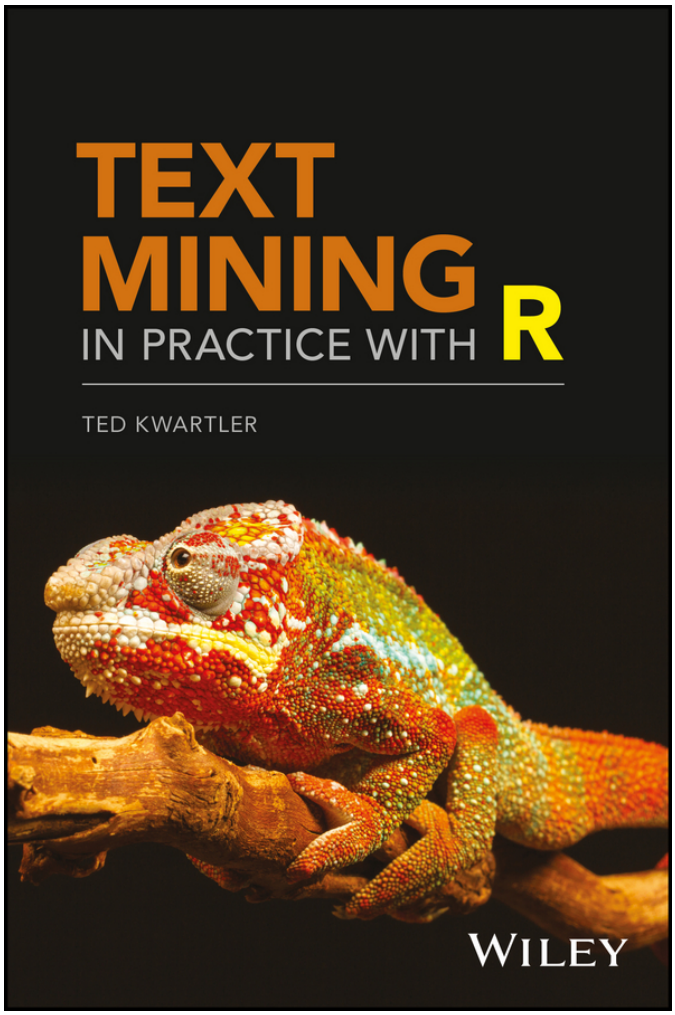


Figure 10: Cover of Text Mining In Practice With R by Ted Kwartler (Wiley, 2010)