



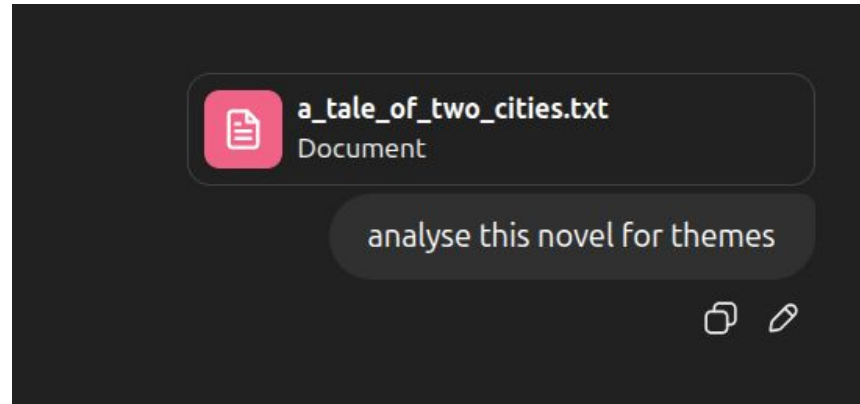
Hands on Session in AI

Kushal Tatariya
kushaljayesh.tatariya@kuleuven.be



But I can just use ChatGPT, right?

But I can just use ChatGPT, right?



A Tale of Two Cities by Charles Dickens is a powerful historical novel set during the French Revolution, and it presents a rich collection of interwoven themes. Here's an overview of the central themes explored in the novel:

1. Resurrection and Transformation

...

Would you like a deeper analysis of how one or two of these themes appear through specific characters or chapters?

A Tale of Two Cities by Charles Dickens is a powerful historical novel set during the French Revolution, and it presents a rich collection of interwoven themes. Here's an overview of the central themes explored in the novel:

1. Resurrection and Transformation

...

Would you like a deeper analysis of how one or two of these themes appear through specific characters or chapters?

Sure!

A Tale of Two Cities by Charles Dickens is a powerful historical novel set during the French Revolution, and it presents a rich collection of interwoven themes. Here's an overview of the central themes explored in the novel:

1. Resurrection and Transformation

...

Would you like a deeper analysis of how one or two of these themes appear through specific characters or chapters?

Sure!

Great! Let's explore **two major themes**—**Resurrection** and **Violence/Revenge**—through **key characters** and moments in *A Tale of Two Cities*:

...

🔥 Theme 2: Violence, Revenge, and Revolutionary Justice

Key Characters: Madame Defarge & The Revolutionaries

She represents the **corruption of justice into vengeance**:

"Tell the Wind and Fire where to stop; not me!"

A Tale of Two Cities by Charles Dickens is a powerful historical novel set during the French Revolution, and it presents a rich collection of interwoven themes. Here's an overview of the central themes explored in the novel:

1. Resurrection and Transformation

...

Would you like a deeper analysis of how one or two of these themes appear through specific characters or chapters?

Sure!

Great! Let's explore **two major themes**—**Resurrection** and **Violence/Revenge**—through **key characters** and moments in *A Tale of Two Cities*:

...

🔥 Theme 2: Violence, Revenge, and Revolutionary Justice

Key Characters: Madame Defarge & The Revolutionaries

She represents the **corruption of justice into vengeance**:

"Tell the Wind and Fire where to stop; not me!"

"Then tell Wind and Fire where to stop," returned madame; "but don't tell me."

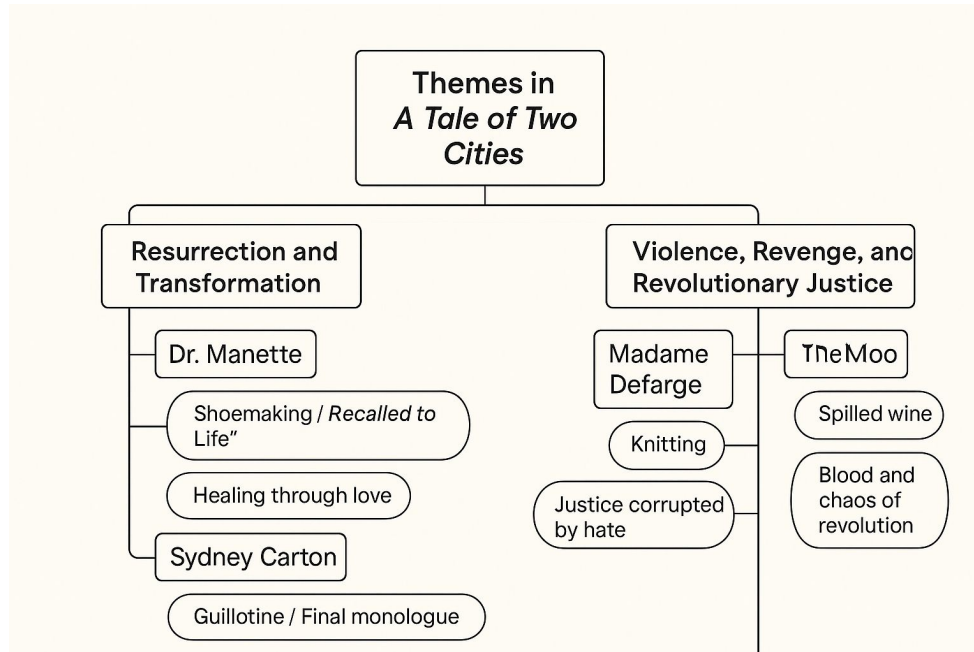
Would you like a visual chart or mind map of these themes and character connections? Or should we dive into another theme, like **Duality** or **Class Conflict**?

Would you like a visual chart or mind map of these themes and character connections? Or should we dive into another theme, like **Duality** or **Class Conflict**?

Show me the map!

Would you like a visual chart or mind map of these themes and character connections? Or should we dive into another theme, like **Duality** or **Class Conflict**?

Show me the map!



But soon you encounter problems:

But soon you encounter problems:

1. What if you have a large corpus with a lot of documents?

But soon you encounter problems:

1. What if you have a large corpus with a lot of documents?
2. Using it for something it was not designed to do - for eg. transcribing handwritten letters.

According to Cameron Blevins: "...even with an "ideal" source like Benjamin Curtis's letter. To take one example: towards the end of his letter, Benjamin revealed that he and his wife Mary had decided to name their newborn daughter "Delia Etta". ChatGPT, however, transcribed this as "Selia Etta"...

...a tiny transcription mistake - switching an "S" for a "D" - meant that **ChatGPT missed an important part of this letter and its larger significance.**"

But soon you encounter problems:

1. What if you have a large corpus with a lot of documents?
2. Using it for something it was not designed to do - for eg. transcribing handwritten letters.
3. Is it trained on data in your domain?

But soon you encounter problems:

1. What if you have a large corpus with a lot of documents?
2. Using it for something it was not designed to do - for eg. transcribing handwritten letters.
3. Is it trained on data in your domain?
4. It is a black box, after all.

But soon you encounter problems:

1. What if you have a large corpus with a lot of documents?
2. Using it for something it was not designed to do - for eg. transcribing handwritten letters.
3. Is it trained on data in your domain?
4. It is a black box, after all.
5. There are specialised tools for specific tasks.

But soon you encounter problems:

1. What if you have a large corpus with a lot of documents?
2. Using it for something it was not designed to do - for eg. transcribing handwritten letters.
3. Is it trained on data in your domain?
4. It is a black box, after all.
5. There are specialised tools for specific tasks.

And more...

What can AI do for your research?

What can AI do for your research?

1. Find clusters of ideas in your data (topic modelling)

What can AI do for your research?

1. Find clusters of ideas in your data (topic modelling)

Gillings, M., & Hardie, A. (2023). The interpretation of topic models for scholarly analysis: An evaluation and critique of current practice. *Digital Scholarship in the Humanities*, 38(2), 530–543.

Schöch, C. (2017). Topic Modeling Genre: An Exploration of French Classical and Enlightenment Drama. *Digital Humanities Quarterly*, 11(2). ²¹

What can AI do for your research?

1. Find clusters of ideas in your data (topic modelling)
2. Understand sentiment (sentiment analysis)

What can AI do for your research?

1. Find clusters of ideas in your data (topic modelling)
2. Understand sentiment (sentiment analysis)
3. Explore different word senses (embeddings)

What can AI do for your research?

1. Find clusters of ideas in your data (topic modelling)
2. Understand sentiment (sentiment analysis)
3. Explore different word senses (embeddings)
4. Look at patterns in named entity mentions (NER)

What can AI do for your research?

1. Find clusters of ideas in your data (topic modelling)
2. Understand sentiment (sentiment analysis)
3. Explore different word senses (embeddings)
4. Look at patterns in named entity mentions (NER)
5. Look into the black box to understand patterns in your data (interpretability)

What can AI do for your research?

1. Find clusters of ideas in your data (topic modelling)
2. Understand sentiment (sentiment analysis)
3. Explore different word senses (embeddings)
4. Look at patterns in named entity mentions (NER)
5. Look into the black box to understand patterns in your data (interpretability)
6. Combine modalities (image + text)

What can AI do for your research?

1. Find clusters of ideas in your data (topic modelling)
2. Understand sentiment (sentiment analysis)
3. Explore different word senses (embeddings)
4. Look at patterns in named entity mentions (NER)
5. Look into the black box to understand patterns in your data (interpretability)
6. Combine modalities (image + text)

...the possibilities are endless!

All Purpose

Corpus Sense: <https://corpus-sense.uma.es/>



Corpus Sense

MAKING SENSE OF TEXT THE EASY WAY

Sentiment Analysis

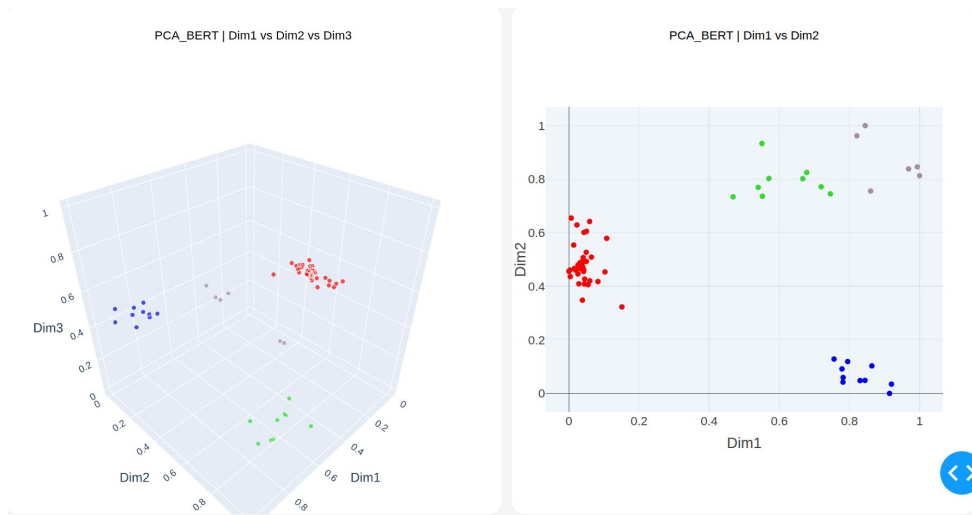
Lingmotif: <https://ltl.uma.es/>

Lingmotif: Sentiment Analysis for the Digital Humanities

Antonio Moreno-Ortiz
University of Málaga
Spain
`amo@uma.es`

Word Embeddings

ContextLens: <https://contextlens.cls.ru.nl/>



Topic Modelling

BERTopic



MALLET



Machine learning for language toolkit

Named Entity Recognition

spaCy: <https://spacy.io/>



Interpretability

LIT: <https://pair-code.github.io/lit/>



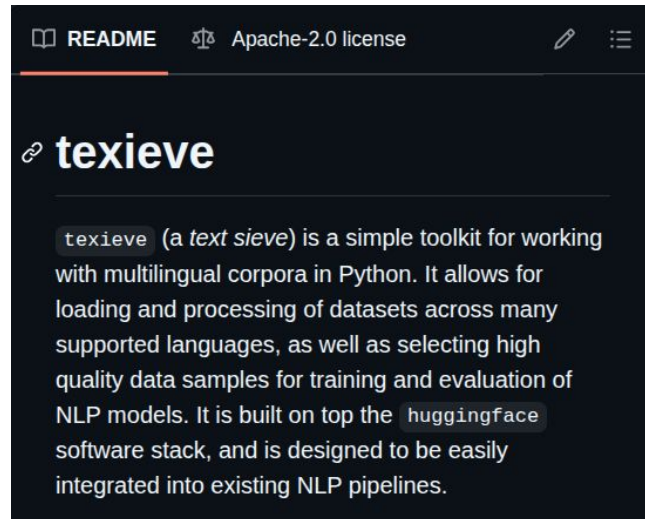
Learning Interpretability Tool

Train it yourself!

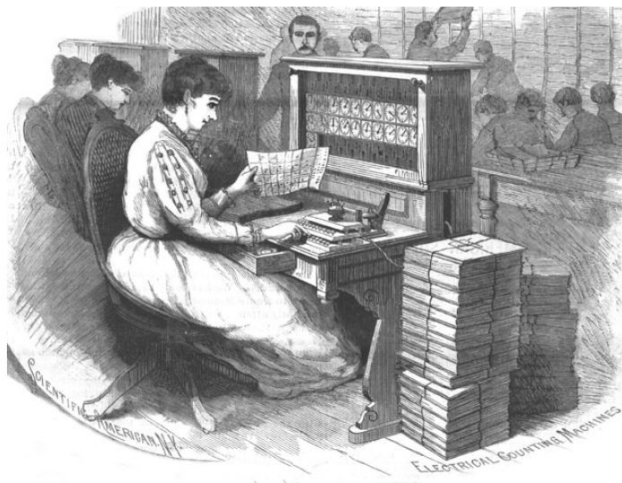
Huggingface
Transformers



texieve



Find some tutorials at The Programming Historian!



The Programming Historian

We publish novice-friendly, peer-reviewed tutorials that help humanists learn a wide range of digital tools, techniques, and workflows to facilitate research and teaching. We are committed to fostering a diverse and inclusive community of editors, writers, and readers.

Find some tutorials at The Programming Historian!



AVERY BLANKENSHIP, SARAH CONNELL, AND QUINN DOMBROWSKI

Understanding and Creating Word Embeddings

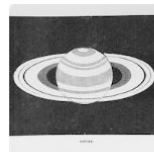
Word embeddings allow you to analyze the usage of different terms in a corpus of texts by capturing information about their contextual usage.

Through a primarily theoretical lens, this lesson will teach you how to prepare a corpus and train a word embedding model. You will explore how word vectors work, how to interpret them, and how to answer humanities research questions using them.



SHAWN GRAHAM, SCOTT WEINGART, AND IAN MILLIGAN

Getting Started with Topic Modeling and MALLET



MEGAN S. KANE

Corpus Analysis with spaCy

This lesson demonstrates how to use the Python library spaCy for analysis of large collections of texts. This lesson details the process of using spaCy to enrich a corpus via lemmatization, part-of-speech tagging, dependency parsing, and named entity recognition. Readers will learn how the linguistic annotations produced by spaCy can be analyzed to help researchers explore meaningful trends in language patterns across a set of texts.



ZOË WILKINSON SALDAÑA

Sentiment Analysis for Exploratory Data Analysis

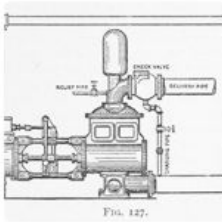
In this lesson you will learn to conduct 'sentiment analysis' on texts and to interpret the results. This is a form of exploratory data analysis based on natural language processing. You will learn to install all appropriate software and to build a reusable program that can be applied to your own texts.

Working with images?



DANIEL VAN STRIEN, KASPAR BEELEN, MELVIN WEVERS, THOMAS SMITS, AND KATHERINE MCDONOUGH

Computer Vision for the Humanities: An Introduction to Deep Learning for Image Classification (Part 1)



DANIEL VAN STRIEN, KASPAR BEELEN, MELVIN WEVERS, THOMAS SMITS, AND KATHERINE MCDONOUGH

Computer Vision for the Humanities: An Introduction to Deep Learning for Image Classification (Part 2)

Working with images/scans of text

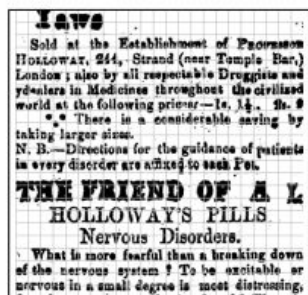
Have a look at PIXEL!

PHD: Pixel-Based Language Modeling of Historical Documents

Nadav Borenstein Phillip Rust Desmond Elliott Isabelle Augenstein

Department of Computer Science, University of Copenhagen

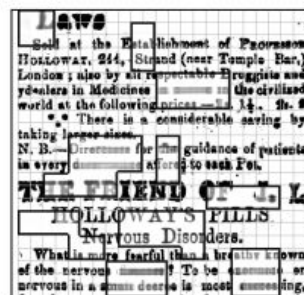
{nadav.borenstein, p.rust, de, augenstein}@di.ku.dk



(a) Input example.



(b) Masking the input.



(c) Model predictions.

See you at the Q&A for questions!

Get in touch: kushaljayesh.tatariya@kuleuven.be

Slides available: https://github.com/kushaltatariya/hands_on_session_in_ai