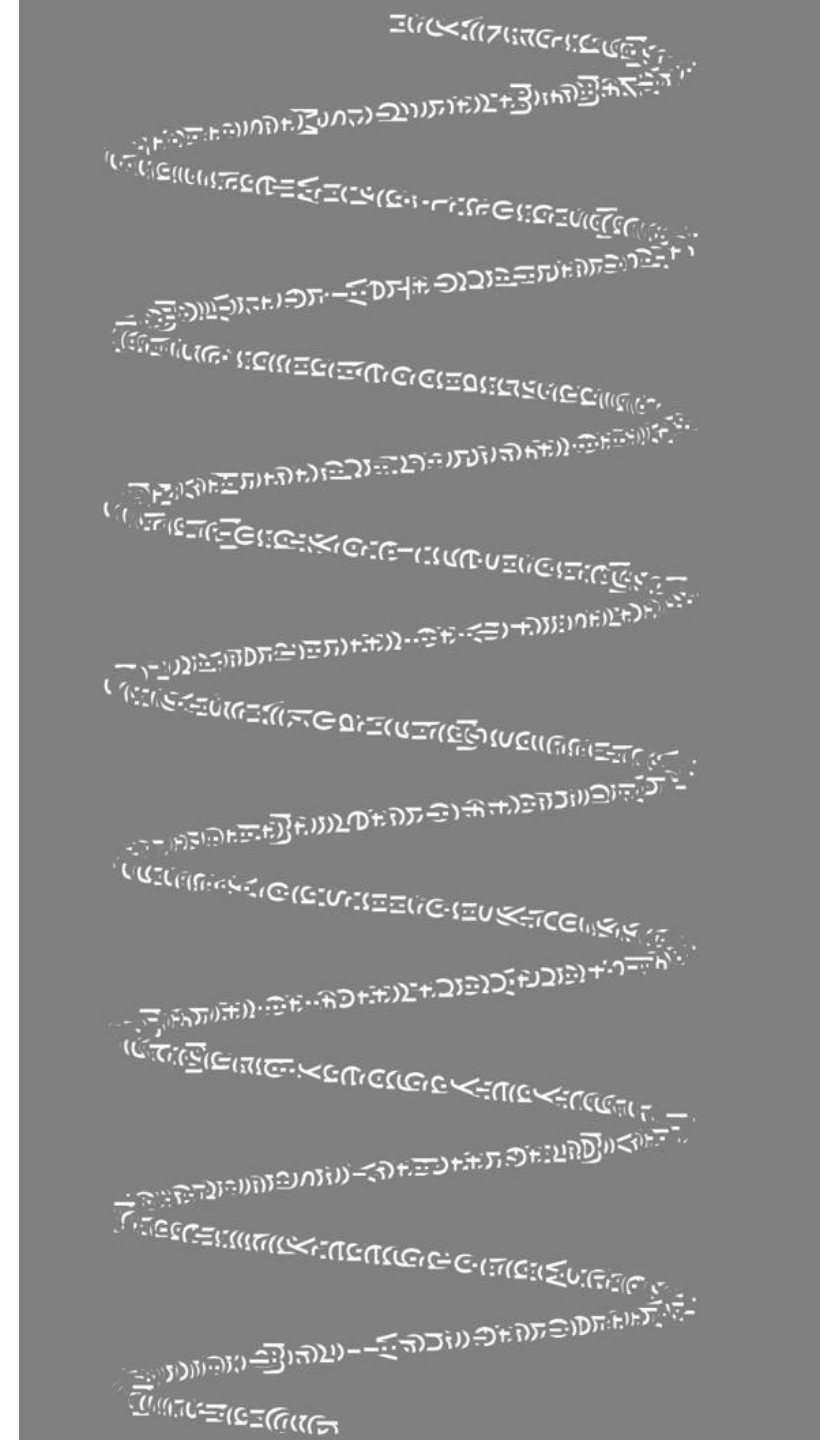


A New Skill Set for Historians

Colin Greenstreet
TNA, Friday, July 19th 2024

colin.greenstreet@gmail.com



The structure of today's talk

Introduction to MarineLives

Vision

- How might historians work together in the near future?
- How might historians work with digital archives, libraries and online resources?

Demos

- Wallis
- Aurelius-Archives
- Aurelius-HTR

Insights

- Description
- Instructions ('Configuration')
- Conversation starters
- Knowledge
- Capabilities
- Actions
- Feedback

Discussion

Marine Lives project

MARINELIVES PROJECT

On the Crest of a Wave

By Colin Greenstreet | Published in History Today Volume: 62 Issue: 9 2012

Tweet (10) | Like (24) | Print | Email

MARITIME

Colin Greenstreet describes a new collaboration to transcribe and enhance 17th-century records of the High Court of Admiralty.



Evidence of marine life: a bill of lading from December 1655. National Archives

Home » News

MarineLives - Call for Project Volunteers

Recruiting volunteer project experts, project facilitators and project associates for MarineLives, a project using collaborative transcription, linkage and enrichment of High Court of Admiralty primary manuscripts, 1650-1669.

MarineLives is an innovative academic project for the collaborative transcription, linkage and enrichment of primary manuscripts, which were originated in the High Court of Admiralty, London, 1650-1669. The end product will be a publicly and freely available online academic edition.

The project is being conducted with the support of The National Archives, Kew, and will work collaboratively on high quality digital images for a complete volume of the court's records, selected from the period 1650-1669.

Please contact the project team (collaborate@marinelives.org), if you would like further information, or access the project website (www.marinellives.org).

Full details of this call for volunteers can be found [here](#).

Like (10) | Tweet (5) | Share

#hackon12: We came, we saw, we didn't sleep much

Posted by Jo Pugh in Technology and Innovation

Launched in summer 2012, following the TNA's first Hackathon, organized by Jo Pugh

Designed to explore how to recruit, train and motivate volunteers to do high quality transcription and annotation

Has worked with and trained over 250 volunteers in the last 12 years and has transcribed over 5 million words

Seeking to run a high quality HTR transcription of all HCA 13/ series depositions (1570 to 1689) [circa 30 mill words) by end 2025.

[@marinelivesorg](https://www.marinellives.org)

<http://www.marinellives.org/wiki/MarineLives>

Introduction to MarineLives

We are holding an online research seminar at the end of April to introduce and demonstrate the SOLM-2024 Admiralty Court Database

Number of depositions in which ship touched at named port, 1570-1680



Well networked academically in Europe, Latin America and North America

Attracts and trains enthusiasts of all sorts with open arms

Produces diverse output [academic quality transcription; textual and quantitative datasets; semantic linked data; gazeteers]

Technologically adventurous but eclectic [HTR; NLP; GPT-4o]. Always opensource.

[@marinelivesorg](https://www.marinelives.org/)

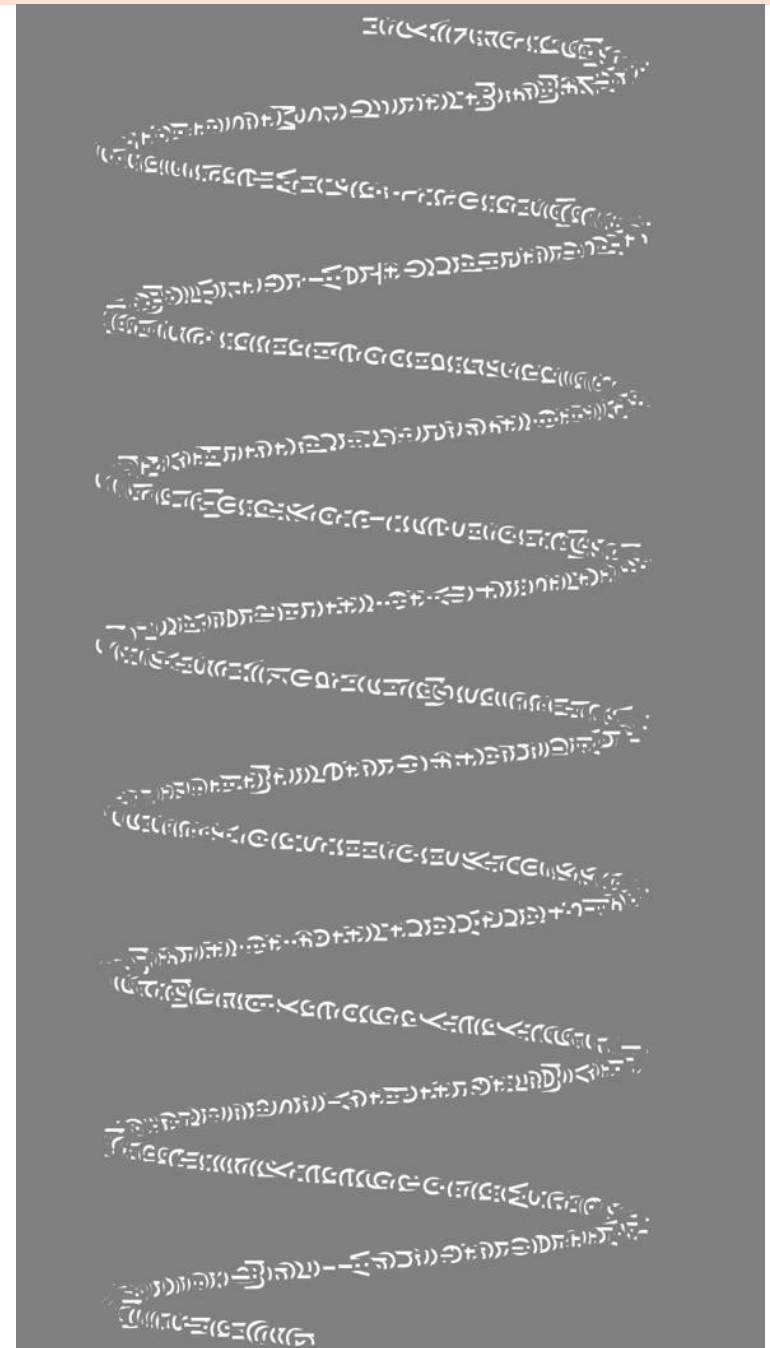
<https://www.marinelives.org/>

Vision: Coming this October?

Imagine a world in the near future. It could be as early as October of this year. It is the start of the academic term and a group of twenty graduate students are commencing a one year History Masters Program. The first week of term looks different from last year. The students are going to start the year by each building an intelligent agent, working in groups of two or three. An intelligent agent is more than a simple 'chat bot'. It is a software based capability to assist an historian - augmenting an historian's knowledge and skills, and increasing an historian's research creativity.

The graduate students are going to learn how to collaborate, some basic tech, and are going to start to integrate that tech into their study of history. This agent is going to accompany them for the whole year. It will be their personal digital assistant and will undoubtedly go through a lot of change during the year, as each student learns more history, focuses on a specific dissertation topic, and learns what works and doesn't work with his or her intelligent agent.

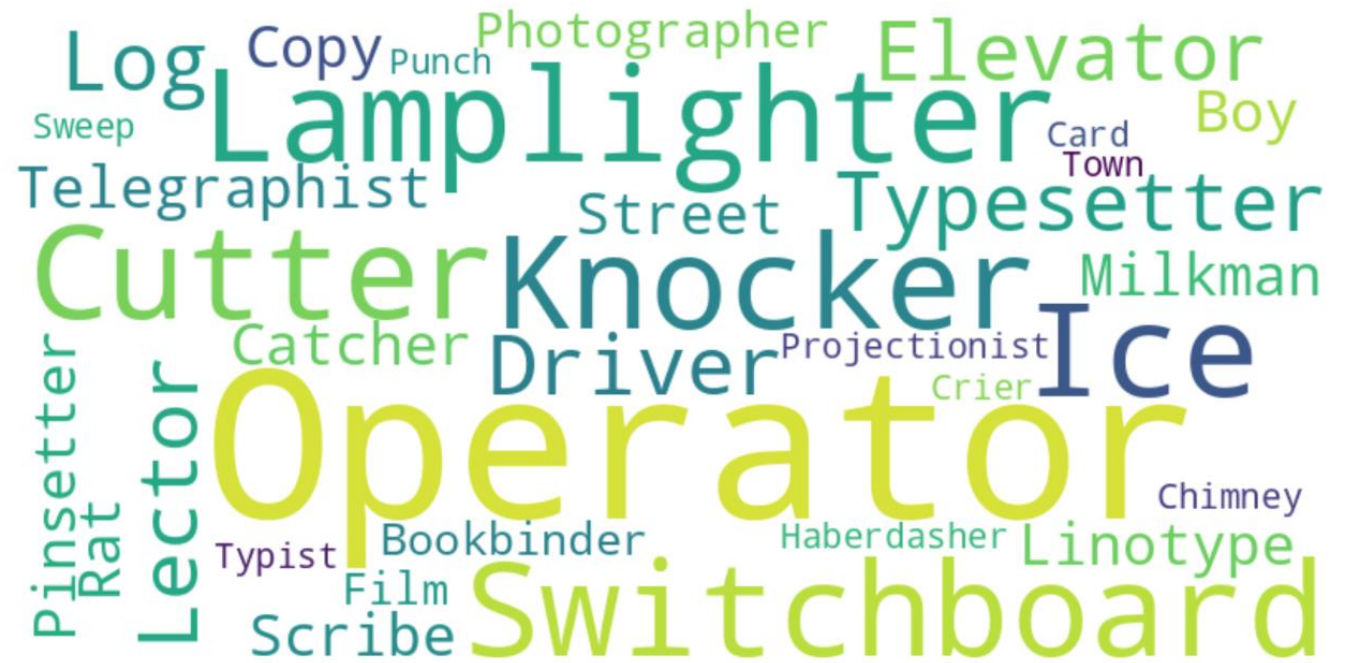
Not only are they going to learn to collaborate in small teams, they are going to learn to collaborate with archivists of both physical and digital resources. The leader of a major digital history project and the technical director of a digitally oriented archive have come together with a UK history department to support this first week initiative. They are providing technical support throughout the week, and are seeking to learn from the experience as well as to teach.



Vision: Upgrading skills

Why should masters' students, their tutors, and archivists be excited by this vision?

Historians, archivists (and humanists more generally) need to upgrade their skills, if they are to be able to make the most of the technical advances in archives, and if they are going to be employable in a rapidly changing knowledge based world.



Vision: Simulation

The first week is structured carefully.

Day One: Introduction to machine learning and AI, together with an introduction to GPT-4o based intelligent agents, and an idiot's guide to the programming language Python.

Day Two: Each team conceptualizes an agent, and creates a quick and dirty agent by lunchtime. They spend the afternoon tinkering – developing the configuration file, adding content, experimenting with different file formats.

Days Three to Five: the group of twenty students is going to build and run a simulation, engaging with some predefined agents, and supported by their own agents. Six teams. Six predefined agents interacting with those teams. It is entirely up to the students whether they wish to use their own agents, which they built on day two to help them in the simulation. The group as a whole gets to choose which simulation they want to play from a choice of three by interacting with the first agent Aurelius-SIM, whose purpose is to help historians create multi-agent historical simulations for educational and research purposes.

Engaging with [Aurelius-SIM](#) they explore how a simulation might run. Does it need an Umpire? How many game periods should it have? How should events be introduced into the game play and how can probability of events and randomness be modelled? They interrogate [Aurelius-Umpire](#) to see what it can do, and modify the Umpire to better match the needs of the simulation scenario they have chosen.



Here are several predefined simulation scenarios from the provided resources that you can draw on for your simulation:

1. Power Struggles in Early Modern Europe

Scenario: This simulation explores the political and military conflicts among European states during the early modern period. It focuses on diplomacy, alliances, warfare, and the balance of power.

Key Elements:

- **Diplomatic Negotiations:** States must negotiate alliances and treaties to strengthen their positions.
- **Military Conflicts:** Wars and battles between states based on historical events.
- **Economic Policies:** Decisions on trade, taxation, and resource allocation affecting the state's power and stability.
- **Social Changes:** Impact of religious movements, peasant revolts, and social reforms on state stability.

2. Risk and Reward in 17th Century Trade

Scenario: This simulation centers on the economic activities of merchants and traders during the 17th century. It focuses on trade routes, market dynamics, and risk management.

Key Elements:

- **Trade Routes:** Establishing and maintaining profitable trade routes.
- **Market Fluctuations:** Responding to changes in supply and demand in different regions.
- **Risk Management:** Dealing with piracy, shipwrecks, and political instability affecting trade.
- **Investment Decisions:** Allocating resources to different trading ventures to maximize profits.

3. Arbitration of a Charter Party Dispute

Scenario: This simulation involves resolving a legal dispute between a shipowner and a charterer over the terms of a charter party agreement. It focuses on legal principles, arbitration procedures, and negotiation tactics.

Key Elements:

- **Legal Arguments:** Presenting and defending legal arguments based on historical maritime law.
- **Arbitration Process:** Following a structured arbitration process to resolve the dispute.
- **Negotiation and Settlement:** Exploring options for negotiation and settlement before and during arbitration.
- **Impact of Decisions:** Understanding the consequences of arbitration outcomes on future trade relations.



Message Aurelius-SIM



Vision: Simulation

Let's assume that they have chosen the Risk and Reward in 17th Century Trade scenario.

The six teams need to decide how to invest £50,000 of capital.

Should they invest in ships? Should they invest in adventures? Where should the ventures be geographically? What commodities? What are the risks? How do they manage the risks? How do they respond to adverse events? What do they do if they encounter disputes? Try arbitration? Litigate in the High Court of Admiralty?

There are four oven ready agents ready to help: [Aurelius-Proctor](#) (Expert in English High Court of Admiralty procedure); [Aurelius-Scrivener](#) (Expert in the production of Early Modern commercial and legal documents); [Aurelius-Merchant](#) (Engages in international trade); and [Aurelius-Mariner](#) (Navigates and operates marine and commercial activities of ship).

The group interviews these four agents to get a better grasp of their capabilities. They make some modifications to the configuration files and to the agent's internal knowledge. The game starts Wednesday afternoon and runs for five game periods, each of two hours, with Friday afternoon spent reviewing lessons learned.

Fantasy? Maybe, maybe not. I have put the idea out to a couple of University history departments, who are actively considering the idea.



Aurelius-Proctor

Expert in English High Court of Admiralty procedure

Answers questions about High Court of Admiralty...

Supports plaintiffs and defendants in suites before the...



Aurelius-Mariner

Navigates and operates marine and commercial activities of ship

Possesses navigational skills

Manages operations of ship

Manages commercial aspects of ship (with purse...



Aurelius-Scrivener

Expert in the production of Early Modern commercial and legal documents

Answers questions about commercial documentation an...

Draws up and witnesses commercial...



Aurelius-Merchant

Engages in international trade

Can you give me suggestions for a potential...

Should I consider investing in a ship?

What can you tell me about financing options and...

The structure of today's talk

Introduction to MarineLives

Vision

- How might historians work together in the near future?
- How might historians work with digital archives, libraries and online resources?

Demos

- Wallis
- Aurelius-Archives
- Aurelius-HTR

Insights

- Description
- Instructions ('Configuration')
- Conversation starters
- Knowledge
- Capabilities
- Actions

Discussion

Contents

Table of Contents

<i>Title page</i>	1
<i>Preface & Acknowledgments</i>	6
<i>Contents</i>	2
<i>Abbreviations</i>	5
1: INTRODUCTION	7
APPRENTICESHIP: THE DEBATE.....	13
APPRENTICESHIP: THE CONTRACT.....	16
THE BASIS AND STRUCTURE OF THE BOOK	19
PART 1	25
APPRENTICESHIP AND DEVELOPMENT	25
2: THE RISE AND (PARTIAL) FALL OF APPRENTICESHIP	26
RISE AND FALL?	27
APPRENTICESHIP IN LONDON AND THE NATION, 1500-1800.....	30
APPRENTICESHIP AND INDUSTRIALIZATION IN THE EIGHTEENTH CENTURY	43
CONCLUSION	56
3: DIFFUSING SKILL AND INNOVATION	61
SUCCESS AND THE CONCENTRATION OF TRAINING	66
LEARNING FROM THE BEST	70
THE TIMING OF TRAINING	82
CONCLUSION	89
4: MOBILISING LABOUR	91
APPRENTICESHIP AS A RESPONSE TO CRISIS	94
MIGRATION	95
SECTORAL MOBILITY	102
URBAN OCCUPATIONAL MOBILITY	109
UNEQUAL OPPORTUNITIES: MOBILITY AND MONEY	113
CONCLUSION	124
PART 2:	126
A SYSTEM OF TRAINING	126
5: APPRENTICESHIP AS LEARNING	127
LEARNING AND TRAINING	127
LEARNING ON THE SHOP FLOOR	131
NORMS & CULTURE	141
SPECIFIC AND GENERAL HUMAN CAPITAL	146
TRAINING BEYOND THE SHOP FLOOR: ACADEMIES AND TECHNICAL EDUCATION	151
CONCLUSION	153
6: THE FLEXIBILITY OF APPRENTICESHIP	157
ABSENT APPRENTICES	159
WHY DID APPRENTICES LEAVE?.....	173
CONCLUSION	179
7: ENFORCING APPRENTICESHIP	183

Patrick Wallis is Head of the Department of Economic History at the LSE. He specializes in social, economic and medical Early Modern history and is at the final manuscript checking stages of a new data intensive book on apprenticeship.

Patrick's requests:

- Can 'Wallis' automate fuzzy search for English parishes in various sources and geolocate those parishes?
- Can 'Wallis' come up with insightful and visually effective charts and display ideas to illustrate data from Chapter 3 in 'The Market for Skill: Apprenticeship in England, 1500-1800' (Princeton University Press)



Wallis: Under the hood

Create

Configure

Name

Wallis

Description

Patrick Wallis' digital research assistant

Instructions

Personality:
Approachable; meticulous; creative; visually oriented

Purpose:
1. Helps identify and match historical English and Welsh parishes in disparate textual and tabular sources and helps map and visualize parish related data
2. Generates creative visual ideas from close study of the text of a book chapter ('Ch. 3: Diffusing Skill and Innovation') concerning the career journey of trades people from apprentice to master

Conversation starters

How can I match & geolocate a list of parishes?

Do you have access to the gazetteer of British Place Names?

How can I visualize concepts in my book chapter?

Knowledge

If you upload files under Knowledge, conversations with your GPT may include file contents. Files can be downloaded when Code Interpreter is enabled

Ch_3_Diffusing_Skill_And...
Document

Semantic_London_Parish...
Document

GBPN_Data_Extraction_In...
Document

Folium_Panda_Script_Gen...
Document

Find_Nearby_Villages_Scri...
Document

GBPN_User_Guide_1.0_29...
Document

CEDD_1.0_Parishes_Counti...
Document

The following files are only available for Code Interpreter:

parish_gis_gazetteer_corr...
Spreadsheet

GBPN_Part_Two_Ver_1.0_2...
Spreadsheet

GBPN_Part_Three_Ver_1.0...
Spreadsheet

GBPN_Part_One_Ver_1.0_2...
Spreadsheet

1851_parish_gazetteer.csv
Spreadsheet

CEDD_Parishes_Counties,...
Spreadsheet

Upload files

Capabilities

☒ Web Browsing

☒ DALL-E Image Generation

☒ Code Interpreter & Data Analysis ⓘ

Actions

Name

Simple description

Instructions/Configuration

- Personality
- Purpose
- Sources
- Methodology
- Prompt/Response patterns
- Internal knowledge
- External websites

Files

- 7 .txt files
- 6 .csv files

Capabilities

Wallis: Python code writing

Show me all nearby villages to Bocking in Essex



Analyzed

```
python Always show details ☒ Copy code

import pandas as pd

# Load the GBPN data
gbpn_part_one = pd.read_csv('/mnt/data/GBPN_Part_One_Ver_1.0_29062024.csv')
gbpn_part_two = pd.read_csv('/mnt/data/GBPN_Part_Two_Ver_1.0_29062024.csv')
gbpn_part_three = pd.read_csv('/mnt/data/GBPN_Part_Three_Ver_1.0_29062024.csv')

# Combine the data into one dataframe
gbpn_data = pd.concat([gbpn_part_one, gbpn_part_two, gbpn_part_three], ignore_index=True)

# Display the columns to understand the structure
gbpn_data.columns

STDOUT/STDERR

/tmp/ipykernel_12/2971729133.py:4: DtypeWarning: Columns (8,14) have mixed
gbpn_part_one = pd.read_csv('/mnt/data/GBPN_Part_One_Ver_1.0_29062024.c
/tmp/ipykernel_12/2971729133.py:5: DtypeWarning: Columns (7,8,14) have mi
gbpn_part_two = pd.read_csv('/mnt/data/GBPN_Part_Two_Ver_1.0_29062024.c

Result

Index(['GBPNID', 'PlaceName', 'GRIDREFLONG', 'Lat', 'Lng', 'Type',
      'HistCounty', 'Division', 'Island', 'AdCounty', 'District', 'UniAu
      'Police', 'Region', 'Alternative_Names', 'Notes'],
      dtype='object')
```

Here are the nearby villages and parishes to Bocking in Essex, within a 10 km radius:

- Almshouse Green: 9.22 km
- Bannister Green: 7.24 km
- Bardfield Saling: 8.03 km
- Bartholomew Green: 5.08 km
- Beazley End: 5.08 km
- Halstead CP: 8.64 km
- Rivenhall CP: 9.72 km
- The Salings CP: 6.49 km
- Braintree CA: 6.50 km
- Essex CA: 8.42 km

If you need more detailed information or a full list of nearby places, please let me know! [\[?\]](#)



Show nearby villages to Halstead.

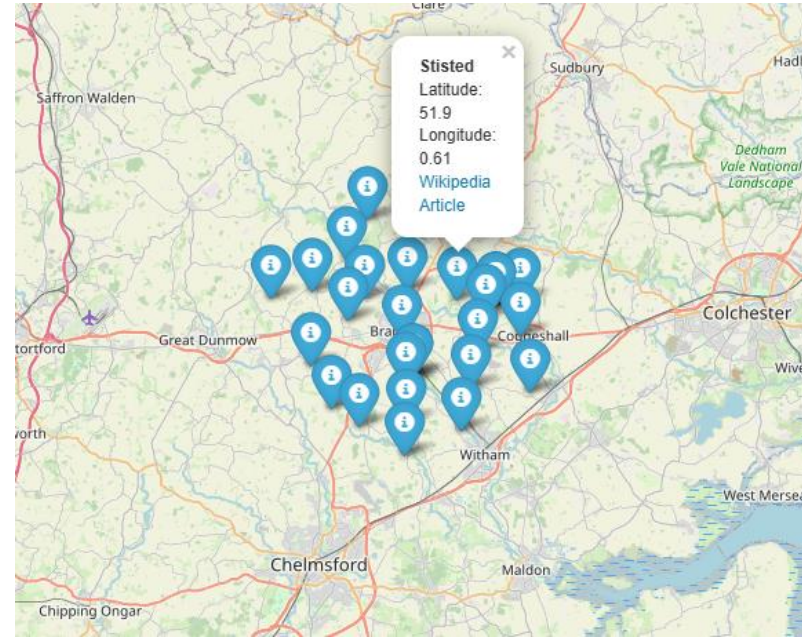
List historical sites near Bocking.

Wallis: Mapping

Map the parishes within 10 km of Braintree in Essex on an Open Street Map base. Do you have any specific suggestions or choices I need to make about the map?

Please include Wikipedia article links in each pop up, together with name of parish and latitude and longitude to 2 decimals

Repeat the map but add the distance in km of each parish in the map from Braintree



WIKIPEDIA

The Free Encyclopedia

Search Wikipedia

Contentshide

(Top)

History of Stisted

References

External links

Stisted

ArticleTalk

From Wikipedia, the free encyclopedia

Stisted

is a civil parish, Church of England parish, and former manor near Brintree, Essex, England.

Andrew Motion,

a former Poet Laureate,

was raised there.

[]

edit

History of Stisted

In 1589 the village came to notice when a local woman, Joan Cunny, who was about 80, was accused of witchcraft. She admitted that she had made a circle and made prayers to the devil. Spried had materialised and she had allowed them home with her and she confessed to feeding them. She had two daughters and the three of them were accused by one of her grandsons. One of her daughters was spared, the other was imprisoned and Cunny was hanged in Chelmsford on 5 July 1589 in line with a 1563 law.^[3]

Samuel Stone,

founder of Hartford, Conn. was curate of Stisted from 1627.^[4]

Charles Forster,^[5] grandfather of E. M. Forster, held the benefice of Stisted, and there is an inscription recording that "The tower was rebuilt from the foundations by Onley Savill-Onley and at the same time the chancel was new roofed and restored by the Rev Charles Forster AD 1844".^[6]

The manor of Stisted also belonged to the monks of Canterbury Cathedral before the reformation. It was sold to Thomas Wiseman in 1549, whose heirs sold it to William Lingwood in 1685, whose widow (his third wife) bequeathed it to John Savill in 1719. It was inherited by Savill's brother, and then his niece, who married the Rev Charles Onley, from whom Onley Savill-Onley was descended.

11 languages

Read Edit View history Tools

Coordinates:

51°8′17″N 0°6′17″E﻿ / ﻿51.138°N 0.105°E﻿ / 51.138; 0.105

Stisted

All Saints' Church, Stisted

Location within Essex

Population

662 (2011)^[c]

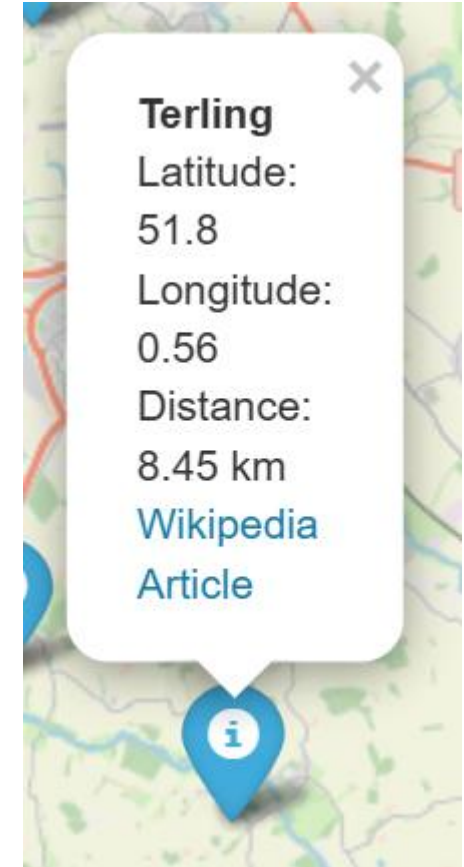
OS grid reference

District

TLD

TL802247

Brintree



Wallis: Visualization

Contents

Table of Contents

Title page.....	1
Preface & Acknowledgments.....	6
Contents	2
Abbreviations	5
1: INTRODUCTION.....	7
APPRENTICESHIP: THE DEBATE.....	13
APPRENTICESHIP: THE CONTRACT.....	16
THE BASIS AND STRUCTURE OF THE BOOK.....	19
PART 1.....	25
APPRENTICESHIP AND DEVELOPMENT.....	25
2: THE RISE AND (PARTIAL) FALL OF APPRENTICESHIP.....	26
RISE AND FALL.....	27
APPRENTICESHIP IN LONDON AND THE NATION, 1500-1800.....	30
APPRENTICESHIP AND INDUSTRIALIZATION IN THE EIGHTEENTH CENTURY.....	43
CONCLUSION.....	56
3: DIFFUSING SKILL AND INNOVATION.....	61
SUCCESS AND THE CONCENTRATION OF TRAINING.....	66
LEARNING FROM THE BEST.....	70
THE TIMING OF TRAINING.....	82
CONCLUSION.....	89
4: MOBILISING LABOUR.....	91
APPRENTICESHIP AS A RESPONSE TO CRISIS.....	94
MIGRATION.....	95
SECTORAL MOBILITY.....	102
URBAN OCCUPATIONAL MOBILITY.....	109
UNEQUAL OPPORTUNITIES: MOBILITY AND MONEY.....	113
CONCLUSION.....	124
PART 2:.....	126
A SYSTEM OF TRAINING.....	126
5: APPRENTICESHIP AS LEARNING.....	127
LEARNING AND TRAINING.....	127
LEARNING ON THE SHOP FLOOR.....	131
NORMS & CULTURE.....	141
SPECIFIC AND GENERAL HUMAN CAPITAL.....	146
TRAINING BEYOND THE SHOP FLOOR: ACADEMIES AND TECHNICAL EDUCATION.....	151
CONCLUSION.....	153
6: THE FLEXIBILITY OF APPRENTICESHIP.....	157
ABSENT APPRENTICES.....	159
WHY DID APPRENTICES LEAVE?.....	173
CONCLUSION.....	179
7: ENFORCING APPRENTICESHIP.....	183

The economic consequences of this pattern of training were profound for the diffusion of innovation and, later, growth, but the master himself had to be careful: parents seeking to get the best start for their children. But in today's families that desire the school and universities that their children will attend, only modern parents invested considerable energy in identifying a good master for their own son or daughter.¹¹ The master who had both attributes – that he felt that the quality of instruction and the richness of opportunity that a youth will receive varies between institutions and individual instructors. As Richard Campbell put it in his advice to parents, they should choose a master based on his 'Virtue, Honesty, Integrity, and Knowledge of his Business'.¹² Some masters were certainly perceived to be better than others, just as some schools are today, and parents sought to place their children with the best they could attain.

This search for quality by parents was limited by institutional constraints. Much as with places in the best schools in many parts of the modern world, apprenticeship in the most prosperous cities was subject to rationing – many guilds and some cities imposed quotas on the numbers of youths each master could take. Did this squeeze many youths into training with what we might call 'cold' masters – individuals seen as second rate – just as today's parents sometimes fear their children are being forced into second-rate schools? Or was the number of able masters sufficient to accommodate most youths?

The pattern of concentrated training with high quality masters that we encounter in this chapter has powerful implications for the economic history of England, and, plausibly other parts of Europe where the same processes of concentration seem to have occurred.¹³ The spread of technology and innovation in pre-modern economies, largely restricted through the movement of skilled individuals.¹⁴ Growth was, to some extent, positively correlated with mobility as a result. Institutions such as apprenticeship that constrained this, or were 'vertical in... the dissemination of new productive knowledge', as de Certeau et al argue.¹⁵

¹¹ See chapter 13, note generally. (This issue 1994, 1994a, 1994b, 1994c, 1994d, 1994e, 1994f, 1994g, 1994h, 1994i, 1994j, 1994k, 1994l, 1994m, 1994n, 1994o, 1994p, 1994q, 1994r, 1994s, 1994t, 1994u, 1994v, 1994w, 1994x, 1994y, 1994z, 1995a, 1995b, 1995c, 1995d, 1995e, 1995f, 1995g, 1995h, 1995i, 1995j, 1995k, 1995l, 1995m, 1995n, 1995o, 1995p, 1995q, 1995r, 1995s, 1995t, 1995u, 1995v, 1995w, 1995x, 1995y, 1995z, 1996a, 1996b, 1996c, 1996d, 1996e, 1996f, 1996g, 1996h, 1996i, 1996j, 1996k, 1996l, 1996m, 1996n, 1996o, 1996p, 1996q, 1996r, 1996s, 1996t, 1996u, 1996v, 1996w, 1996x, 1996y, 1996z, 1997a, 1997b, 1997c, 1997d, 1997e, 1997f, 1997g, 1997h, 1997i, 1997j, 1997k, 1997l, 1997m, 1997n, 1997o, 1997p, 1997q, 1997r, 1997s, 1997t, 1997u, 1997v, 1997w, 1997x, 1997y, 1997z, 1998a, 1998b, 1998c, 1998d, 1998e, 1998f, 1998g, 1998h, 1998i, 1998j, 1998k, 1998l, 1998m, 1998n, 1998o, 1998p, 1998q, 1998r, 1998s, 1998t, 1998u, 1998v, 1998w, 1998x, 1998y, 1998z, 1999a, 1999b, 1999c, 1999d, 1999e, 1999f, 1999g, 1999h, 1999i, 1999j, 1999k, 1999l, 1999m, 1999n, 1999o, 1999p, 1999q, 1999r, 1999s, 1999t, 1999u, 1999v, 1999w, 1999x, 1999y, 1999z, 2000a, 2000b, 2000c, 2000d, 2000e, 2000f, 2000g, 2000h, 2000i, 2000j, 2000k, 2000l, 2000m, 2000n, 2000o, 2000p, 2000q, 2000r, 2000s, 2000t, 2000u, 2000v, 2000w, 2000x, 2000y, 2000z, 2001a, 2001b, 2001c, 2001d, 2001e, 2001f, 2001g, 2001h, 2001i, 2001j, 2001k, 2001l, 2001m, 2001n, 2001o, 2001p, 2001q, 2001r, 2001s, 2001t, 2001u, 2001v, 2001w, 2001x, 2001y, 2001z, 2002a, 2002b, 2002c, 2002d, 2002e, 2002f, 2002g, 2002h, 2002i, 2002j, 2002k, 2002l, 2002m, 2002n, 2002o, 2002p, 2002q, 2002r, 2002s, 2002t, 2002u, 2002v, 2002w, 2002x, 2002y, 2002z, 2003a, 2003b, 2003c, 2003d, 2003e, 2003f, 2003g, 2003h, 2003i, 2003j, 2003k, 2003l, 2003m, 2003n, 2003o, 2003p, 2003q, 2003r, 2003s, 2003t, 2003u, 2003v, 2003w, 2003x, 2003y, 2003z, 2004a, 2004b, 2004c, 2004d, 2004e, 2004f, 2004g, 2004h, 2004i, 2004j, 2004k, 2004l, 2004m, 2004n, 2004o, 2004p, 2004q, 2004r, 2004s, 2004t, 2004u, 2004v, 2004w, 2004x, 2004y, 2004z, 2005a, 2005b, 2005c, 2005d, 2005e, 2005f, 2005g, 2005h, 2005i, 2005j, 2005k, 2005l, 2005m, 2005n, 2005o, 2005p, 2005q, 2005r, 2005s, 2005t, 2005u, 2005v, 2005w, 2005x, 2005y, 2005z, 2006a, 2006b, 2006c, 2006d, 2006e, 2006f, 2006g, 2006h, 2006i, 2006j, 2006k, 2006l, 2006m, 2006n, 2006o, 2006p, 2006q, 2006r, 2006s, 2006t, 2006u, 2006v, 2006w, 2006x, 2006y, 2006z, 2007a, 2007b, 2007c, 2007d, 2007e, 2007f, 2007g, 2007h, 2007i, 2007j, 2007k, 2007l, 2007m, 2007n, 2007o, 2007p, 2007q, 2007r, 2007s, 2007t, 2007u, 2007v, 2007w, 2007x, 2007y, 2007z, 2008a, 2008b, 2008c, 2008d, 2008e, 2008f, 2008g, 2008h, 2008i, 2008j, 2008k, 2008l, 2008m, 2008n, 2008o, 2008p, 2008q, 2008r, 2008s, 2008t, 2008u, 2008v, 2008w, 2008x, 2008y, 2008z, 2009a, 2009b, 2009c, 2009d, 2009e, 2009f, 2009g, 2009h, 2009i, 2009j, 2009k, 2009l, 2009m, 2009n, 2009o, 2009p, 2009q, 2009r, 2009s, 2009t, 2009u, 2009v, 2009w, 2009x, 2009y, 2009z, 2010a, 2010b, 2010c, 2010d, 2010e, 2010f, 2010g, 2010h, 2010i, 2010j, 2010k, 2010l, 2010m, 2010n, 2010o, 2010p, 2010q, 2010r, 2010s, 2010t, 2010u, 2010v, 2010w, 2010x, 2010y, 2010z, 2011a, 2011b, 2011c, 2011d, 2011e, 2011f, 2011g, 2011h, 2011i, 2011j, 2011k, 2011l, 2011m, 2011n, 2011o, 2011p, 2011q, 2011r, 2011s, 2011t, 2011u, 2011v, 2011w, 2011x, 2011y, 2011z, 2012a, 2012b, 2012c, 2012d, 2012e, 2012f, 2012g, 2012h, 2012i, 2012j, 2012k, 2012l, 2012m, 2012n, 2012o, 2012p, 2012q, 2012r, 2012s, 2012t, 2012u, 2012v, 2012w, 2012x, 2012y, 2012z, 2013a, 2013b, 2013c, 2013d, 2013e, 2013f, 2013g, 2013h, 2013i, 2013j, 2013k, 2013l, 2013m, 2013n, 2013o, 2013p, 2013q, 2013r, 2013s, 2013t, 2013u, 2013v, 2013w, 2013x, 2013y, 2013z, 2014a, 2014b, 2014c, 2014d, 2014e, 2014f, 2014g, 2014h, 2014i, 2014j, 2014k, 2014l, 2014m, 2014n, 2014o, 2014p, 2014q, 2014r, 2014s, 2014t, 2014u, 2014v, 2014w, 2014x, 2014y, 2014z, 2015a, 2015b, 2015c, 2015d, 2015e, 2015f, 2015g, 2015h, 2015i, 2015j, 2015k, 2015l, 2015m, 2015n, 2015o, 2015p, 2015q, 2015r, 2015s, 2015t, 2015u, 2015v, 2015w, 2015x, 2015y, 2015z, 2016a, 2016b, 2016c, 2016d, 2016e, 2016f, 2016g, 2016h, 2016i, 2016j, 2016k, 2016l, 2016m, 2016n, 2016o, 2016p, 2016q, 2016r, 2016s, 2016t, 2016u, 2016v, 2016w, 2016x, 2016y, 2016z, 2017a, 2017b, 2017c, 2017d, 2017e, 2017f, 2017g, 2017h, 2017i, 2017j, 2017k, 2017l, 2017m, 2017n, 2017o, 2017p, 2017q, 2017r, 2017s, 2017t, 2017u, 2017v, 2017w, 2017x, 2017y, 2017z, 2018a, 2018b, 2018c, 2018d, 2018e, 2018f, 2018g, 2018h, 2018i, 2018j, 2018k, 2018l, 2018m, 2018n, 2018o, 2018p, 2018q, 2018r, 2018s, 2018t, 2018u, 2018v, 2018w, 2018x, 2018y, 2018z, 2019a, 2019b, 2019c, 2019d, 2019e, 2019f, 2019g, 2019h, 2019i, 2019j, 2019k, 2019l, 2019m, 2019n, 2019o, 2019p, 2019q, 2019r, 2019s, 2019t, 2019u, 2019v, 2019w, 2019x, 2019y, 2019z, 2020a, 2020b, 2020c, 2020d, 2020e, 2020f, 2020g, 2020h, 2020i, 2020j, 2020k, 2020l, 2020m, 2020n, 2020o, 2020p, 2020q, 2020r, 2020s, 2020t, 2020u, 2020v, 2020w, 2020x, 2020y, 2020z, 2021a, 2021b, 2021c, 2021d, 2021e, 2021f, 2021g, 2021h, 2021i, 2021j, 2021k, 2021l, 2021m, 2021n, 2021o, 2021p, 2021q, 2021r, 2021s, 2021t, 2021u, 2021v, 2021w, 2021x, 2021y, 2021z, 2022a, 2022b, 2022c, 2022d, 2022e, 2022f, 2022g, 2022h, 2022i, 2022j, 2022k, 2022l, 2022m, 2022n, 2022o, 2022p, 2022q, 2022r, 2022s, 2022t, 2022u, 2022v, 2022w, 2022x, 2022y, 2022z, 2023a, 2023b, 2023c, 2023d, 2023e, 2023f, 2023g, 2023h, 2023i, 2023j, 2023k, 2023l, 2023m, 2023n, 2023o, 2023p, 2023q, 2023r, 2023s, 2023t, 2023u, 2023v, 2023w, 2023x, 2023y, 2023z, 2024a, 2024b, 2024c, 2024d, 2024e, 2024f, 2024g, 2024h, 2024i, 2024j, 2024k, 2024l, 2024m, 2024n, 2024o, 2024p, 2024q, 2024r, 2024s, 2024t, 2024u, 2024v, 2024w, 2024x, 2024y, 2024z, 2025a, 2025b, 2025c, 2025d, 2025e, 2025f, 2025g, 2025h, 2025i, 2025j, 2025k, 2025l, 2025m, 2025n, 2025o, 2025p, 2025q, 2025r, 2025s, 2025t, 2025u, 2025v, 2025w, 2025x, 2025y, 2025z, 2026a, 2026b, 2026c, 2026d, 2026e, 2026f, 2026g, 2026h, 2026i, 2026j, 2026k, 2026l, 2026m, 2026n, 2026o, 2026p, 2026q, 2026r, 2026s, 2026t, 2026u, 2026v, 2026w, 2026x, 2026y, 2026z, 2027a, 2027b, 2027c, 2027d, 2027e, 2027f, 2027g, 2027h, 2027i, 2027j, 2027k, 2027l, 2027m, 2027n, 2027o, 2027p, 2027q, 2027r, 2027s, 2027t, 2027u, 2027v, 2027w, 2027x, 2027y, 2027z, 2028a, 2028b, 2028c, 2028d, 2028e, 2028f, 2028g, 2028h, 2028i, 2028j, 2028k, 2028l, 2028m, 2028n, 2028o, 2028p, 2028q, 2028r, 2028s, 2028t, 2028u, 2028v, 2028w, 2028x, 2028y, 2028z, 2029a, 2029b, 2029c, 2029d, 2029e, 2029f, 2029g, 2029h, 2029i, 2029j, 2029k, 2029l, 2029m, 2029n, 2029o, 2029p, 2029q, 2029r, 2029s, 2029t, 2029u, 2029v, 2029w, 2029x, 2029y, 2029z, 2030a, 2030b, 2030c, 2030d, 2030e, 2030f, 2030g, 2030h, 2030i, 2030j, 2030k, 2030l, 2030m, 2030n, 2030o, 2030p, 2030q, 2030r, 2030s, 2030t, 2030u, 2030v, 2030w, 2030x, 2030y, 2030z, 2031a, 2031b, 2031c, 2031d, 2031e, 2031f, 2031g, 2031h, 2031i, 2031j, 2031k, 2031l, 2031m, 2031n, 2031o, 2031p, 2031q, 2031r, 2031s, 2031t, 2031u, 2031v, 2031w, 2031x, 2031y, 2031z, 2032a, 2032b, 2032c, 2032d, 2032e, 2032f, 2032g, 2032h, 2032i, 2032j, 2032k, 2032l, 2032m, 2032n, 2032o, 2032p, 2032q, 2032r, 2032s, 2032t, 2032u, 2032v, 2032w, 2032x, 2032y, 2032z, 2033a, 2033b, 2033c, 2033d, 2033e, 2033f, 2033g, 2033h, 2033i, 2033j, 2033k, 2033l, 2033m, 2033n, 2033o, 2033p, 2033q, 2033r, 2033s, 2033t, 2033u, 2033v, 2033w, 2033x, 2033y, 2033z, 2034a, 2034b, 2034c, 2034d, 2034e, 2034f, 2034g, 2034h, 2034i, 2034j, 2034k, 2034l, 2034m, 2034n, 2034o, 2034p, 2034q, 2034r, 2034s, 2034t, 2034u, 2034v, 2034w, 2034x, 2034y, 2034z, 2035a, 2035b, 2035c, 2035d, 2035e, 2035f, 2035g, 2035h, 2035i, 2035j, 2035k, 2035l, 2035m, 2035n, 2035o, 2035p, 2035q, 2035r, 2035s, 2035t, 2035u, 2035v, 2035w, 2035x, 2035y, 2035z, 2036a, 2036b, 2036c, 2036d, 2036e, 2036f, 2036g, 2036h, 2036i, 2036j, 2036k, 2036l, 2036m, 2036n, 2036o, 2036p, 2036q, 2036r, 2036s, 2036t, 2036u, 2036v, 2036w, 2036x, 2036y, 2036z, 2037a, 2037b, 2037c, 2037d, 2037e, 2037f, 2037g, 2037h, 2037i, 2037j, 2037k, 2037l, 2037m, 2037n, 2037o, 2037p, 2037q, 2037r, 2037s, 2037t, 2037u, 2037v, 2037w, 2037x, 2037y, 2037z, 2038a, 2038b, 2038c, 2038d, 2038e, 2038f, 2038g, 2038h, 2038i, 2038j, 2038k, 2038l, 2038m, 2038n, 2038o, 2038p, 2038q, 2038r, 2038s, 2038t, 2038u, 2038v, 2038w, 2038x, 2038y, 2038z, 2039a, 2039b, 2039c, 2039d, 2039e, 2039f, 2039g, 2039h, 2039i, 2039j, 2039k, 2039l, 2039m, 2039n, 2039o, 2039p, 2039q, 2039r, 2039s, 2039t, 2039u, 2039v, 2039w, 2039x, 2039y, 2039z, 2040a, 2040b, 2040c, 2040d, 2040e, 2040f, 2040g, 2040h, 2040i, 2040j, 2040k, 2040l, 2040m, 2040n, 2040o, 2040p, 2040q, 2040r, 2040s, 2040t, 2040u, 2040v, 2040w, 2040x, 2040y, 2040z, 2041a, 2041b, 2041c, 2041d, 2041e, 2041f, 2041g, 2041h, 2041i, 2041j, 2041k, 2041l, 2041m, 2041n, 2041o, 2041p, 2041q, 2041r, 2041s, 2041t, 2041u, 2041v, 2041w, 2041x, 2041y, 2041z, 2042a, 2042b, 2042c, 2042d, 2042e, 2042f, 2042g, 2042h, 2042i, 2042j, 2042k, 2042l, 2042m, 2042n, 2042o, 2042p, 2042q, 2042r, 2042s, 2042t, 2042u, 2042v, 2042w, 2042x, 2042y, 2042z, 2043a, 2043b, 2043c, 2043d, 2043e, 2043f, 2043g, 2043h, 2043i, 2043j, 2043k, 2043l, 2043m, 2043n, 2043o, 2043p, 2043q, 2043r, 2043s, 2043t, 2043u, 2043v, 2043w, 2043x, 2043y, 2043z, 2044a, 2044b, 2044c, 2044d, 2044e, 2044f, 2044g, 2044h, 2044i, 2044j, 2044k, 2044l, 2044m, 2044n, 2044o, 2044p, 2044q, 2044r, 2044s, 2044t, 2044u, 2044v, 2044w, 2044x, 2044y, 2044z, 2045a, 2045b, 2045c, 2045d, 2045e, 2045f, 2045g, 2045h, 2045i, 2045j, 2045k, 2045l, 2045m, 2045n, 2045o, 2045p, 2045q, 2045r, 2045s, 2045t, 2045u, 2045v, 2045w, 2045x, 2045y, 2045z, 2046a, 2046b, 2046c, 2046d, 2046e, 2046f, 2046g, 2046h, 2046i, 2046j, 2046k, 2046l, 2046m, 2046n, 2046o, 2046p, 2046q, 2046r, 2046s, 2046t, 2046u, 2046v, 2046w, 2046x, 2046y, 2046z, 2047a, 2047b, 2047c, 2047d, 2047e, 2047f, 2047g, 2047h, 2047i, 2047j, 2047k, 2047l, 2047m, 2047n, 2047o, 2047p, 2047q, 2047r, 2047s, 2047t, 2047u, 2047v, 2047w, 2047x, 2047y, 2047z, 2048a, 2048b, 2048c, 2048d, 2048e, 2048f, 2048g, 2048h, 2048i, 2048j, 2048k, 2048l, 2048m, 2048n, 2048o, 2048p, 2048q, 2048r, 2048s, 2048t, 2048u, 2048v, 2048w, 2048x, 2048y, 2048z, 2049a, 2049b, 2049c, 2049d, 2049e, 2049f, 2049g, 2049h, 2049i, 2049j, 2049k, 2049l, 2049m, 2049n, 2049o, 2049p, 2049q, 2049r, 2049s, 2049t, 2049u, 2049v, 2049w, 2049x, 2049y, 2049z, 2050a, 2050b, 2050c, 2050d, 2050e, 2050f, 2050g, 2050h, 2050i, 2050j, 2050k, 2050l, 2050m, 2050n, 2050o, 2050p, 2050q, 2050r, 2050s, 2050t, 2050u, 2050v, 2050w, 2050x, 2050y, 2050z, 2051a, 2051b, 2051c, 2051d, 2051e, 2051f, 2051g, 2051h, 2051i, 2051j, 2051k, 2051l, 2051m, 2051n, 2051o, 2051p, 2051q, 2051r, 2051s, 2051t, 2051u, 2051v, 2051w, 2051x, 2051y, 20

Wallis: Visualization

Visual Ideas:

1. Growth and Contraction

- Visualize the spread and decline of clockmaking workshops over time using a timeline or a series of maps.

2. Diffusion, Spreading, Seeding, Taking Root

- Show a map of London highlighting key workshops and the movement of apprentices from rural areas to urban centers.
- Diagram of the diffusion process of innovations from key figures like Thomas Tompion to their apprentices and beyond.

3. Flows

- Create a flowchart illustrating the training process from master to apprentice and the subsequent spread of skills.

4. Networking (of People; of Knowledge)

- Network diagrams showing collaborations between clockmakers and natural philosophers.
- Chart the network of apprentices trained by Tompion and their later influence in the industry.

5. Geographic Proximity/Geographic Dispersion

- Map of Europe showing the spread of clockmaking techniques from London to other regions.
- Highlight key locations with annotations about significant contributions.

6. Critical Mass

- Visualize the concentration of clockmaking expertise in London through heat maps or density plots.

7. Gravity Models

- Use a gravity model to illustrate the attraction of skilled artisans to London and the subsequent impact on innovation.

Mock-up Visual Ideas:

1. Timeline of Technological Advancements in Clockmaking

- Annotated timeline showing key developments, figures, and inventions in clockmaking from the 17th century onwards.

2. Map of Clockmaking Workshops in London

- Interactive map highlighting the locations of significant workshops and the movement of apprentices.

3. Network Diagram of Apprenticeship Relationships

- Diagram showing the connections between masters and their apprentices, highlighting influential figures like Tompion.

4. Flowchart of Innovation Diffusion

- Flowchart illustrating how innovations spread from key inventors to the wider industry through apprenticeship and collaboration.

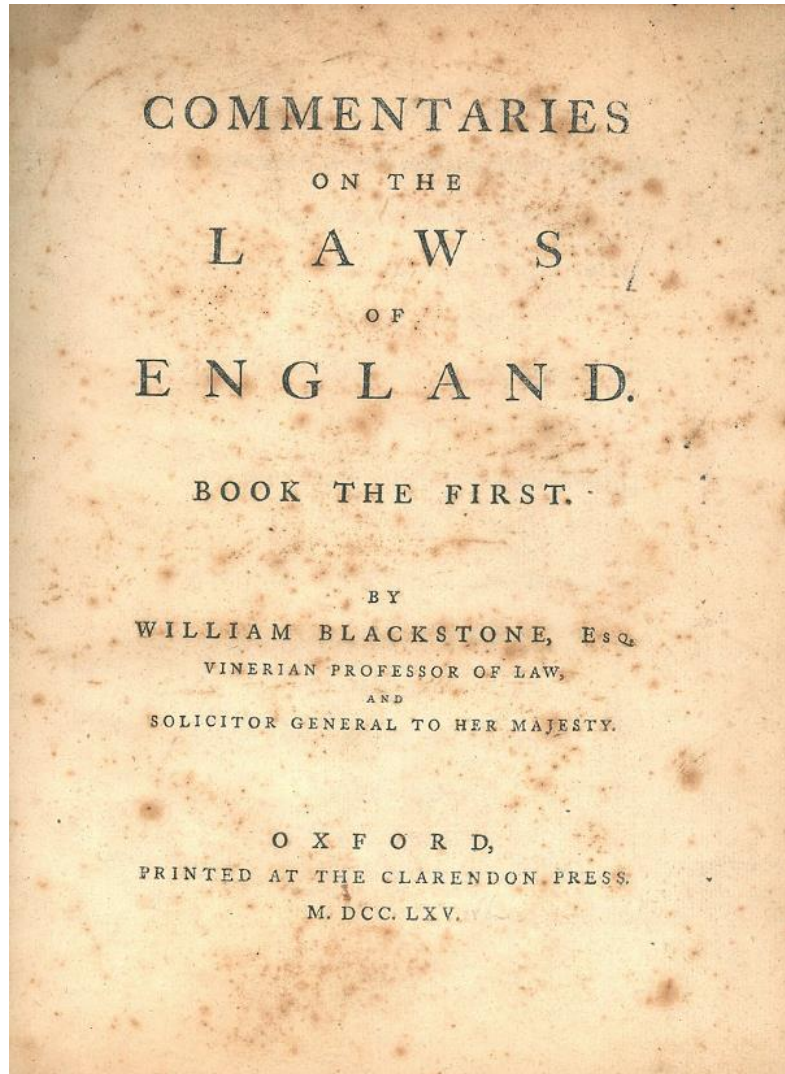
5. Economic Impact Chart

- Bar or line charts showing the correlation between the spread of clockmaking innovations and economic growth indicators such as productivity and manufacturing output.

Example Visuals:

- **Map of London Clockmaking Workshops:** An interactive map using tools like Folium to plot the locations of workshops and the geographic spread of apprentices.
- **Timeline of Key Innovations:** A detailed timeline with significant milestones in clockmaking advancements.
- **Network Diagram:** A visual representation of the apprenticeship network and the spread of skills.

Would you like me to create a specific visualization based on one of these ideas? If so, please specify which one, and provide any additional details you'd like included.

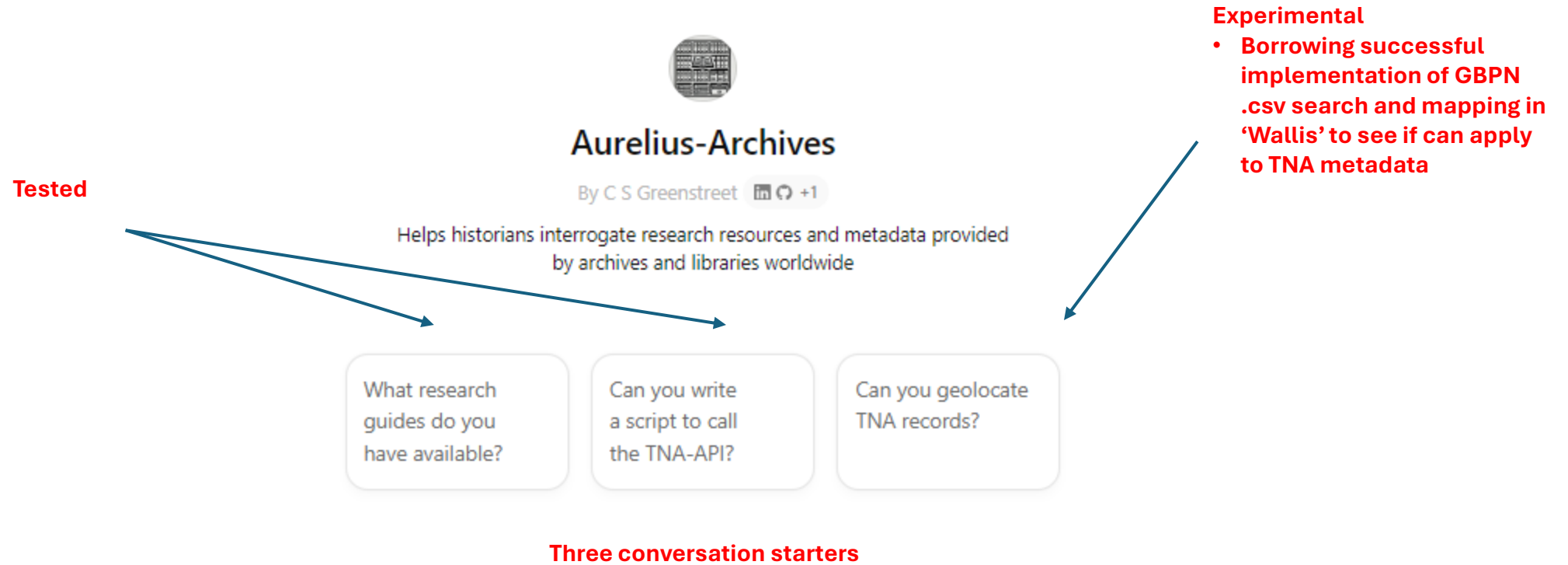


Legal and economic historians make frequent use of TNA records, and records from other sources such as British History OnLine and Old Bailey Online.

Colin's requests:

- Can 'Aurelius-Archives' contextualize legal records and point researchers in the right direction in an insightful way?
- Can 'Aurelius-Archives' write API-calls, which a historian with no programming experience can use to identify and download metadata from online sources?
- Can 'Aurelius-Archives' geolocate metadata held by archives and libraries and create usable synthesized files and maps of these data?


Aurelius-Archives: Front page



Aurelius-Archives: Under the hood

Create

Configure



Name

Aurelius-Archives

Description

Helps historians interrogate research resources and metadata provided by archives and libraries worldwide

Instructions

Helps historians interrogate research resources and metadata provided by archives and libraries worldwide

Interrogates TNA research guides from internal knowledge:
#Apprentices and masters
#Bankrupts and insolvent debtors
#Civil court cases: an overview

Conversation starters

What research guides do you have available?

×

Can you write a script to call the TNA-API?

×


Can you geolocate TNA records?


×


×


Knowledge


If you upload files under Knowledge, conversations with your GPT may include file contents. Files can be downloaded when Code Interpreter is enabled


 OBOL-API_Call_Endpoint_...
Document


 OBOL_Adv_Search_1.1_140...
Document


 JSON_Wikipedia_Article_1...
Document


 How_To_Look_For_1.2_TN...
Document


 OBOL_API_1.1_12072024.txt
Document


 Masters_Programs_DS_Hu...
Document


 GBPN_Data_Extraction_In...
Document


 GBPN_User_Guide_1.0_29...
Document

 BHOL_Subject_Guides_1.0...
Document


 BHOL_Contents_Guide_1...
Document


 New_Skill_Set_for_Histori...
Document


 Discovery_API_1.2_TNA_1...
Document


 Useful_Py_Scripts_1.2_140...
Document


The following files are only available for Code Interpreter:

 GBPN_Part_One_Ver_1.0_2...
Spreadsheet

 GBPN_Part_Three_Ver_1.0...
Spreadsheet

 1851_parish_gazetteer.csv
Spreadsheet

 GBPN_Part_Two_Ver_1.0_2...
Spreadsheet

 PROB_11_Blacksmiths_160...
Spreadsheet

Upload files

Capabilities

☒ Web Browsing

☒ DALL-E Image Generation

☒ Code Interpreter & Data Analysis ⓘ

Actions

Icon

Name

Simple (if a trifle ambitious) description

Instructions/Configuration

- Capabilities
 - Research guides
 - API interrogation
 - API Endpoints
 - Data scraping
 - Internal knowledge

Files

- 12 .txt files
- 1 .docx file
- 1 .ppt file
- 5 .csv files

Capabilities

Aurelius-Archives: Let's try a live demo (and hope for the best)

- <https://chatgpt.com/g/g-Kg5sBnCmk-aurelius-archives>

Aurelius-Archives: Responses to conversation starters

What research guides do you have available?



Here is a summary of the available research guides from the uploaded documents:

British History Online (BHO) Research Guides

1. Biography
2. Local History
3. Parliamentary History
4. Religious History
5. Urban History

The National Archives (TNA) Research Guides

1. Apprentices and masters
2. Bankrupts and insolvent debtors
3. Civil court cases: an overview
4. Civil court cases: assize courts 1656-1971
5. Civil court cases: Chancery equity suits 1558-1875
6. Civil court cases: Court of Common Pleas
7. Coroners' inquests
8. Court of Exchequer
9. Court of King's Bench records 1200-1702
10. Courts of law records from the medieval period: general eyres 1194-1348
11. Court of Requests records 1485-1642
12. Court of Star Chamber records 1485-1642
13. Court of Wards and Liveries 1540-1645: land inheritance and disputes
14. Criminal court cases: an overview
15. Criminal court cases: King's/Queen's Bench 1675-1988
16. High Court of Admiralty
17. Inquisitions post mortem: land ownership and inheritance in the medieval and early modern



Can you write a script to call the TNA-API?



Sure! Here's a Python script that calls the TNA-API to fetch Chancery Court Records from 1600 to 1699:

```
python
Copy code

import requests
import json

# Define the API endpoint
api_endpoint = "http://discovery.nationalarchives.gov.uk/API/search/v1/records"

# Define the parameters for the API call
params = {
    "q": "Chancery Court",
    "dateFrom": "1600",
    "dateTo": "1699",
    "department": "C",
    "mediaFormat": "ALL",
    "pageSize": 50,
    "sortBy": "RELEVANCE"
}

# Make the API call
response = requests.get(api_endpoint, params=params)

# Check if the request was successful
if response.status_code == 200:
    # Parse the JSON response
    records = response.json()

    # Print the results
    print(json.dumps(records, indent=4))
else:
    print(f"Error: {response.status_code}")

# Save the results to a file
with open("chancery_records.json", "w") as f:
```

Can I get more detailed filters?

How to search for records by name?



Message Aurelius-Archives



Yes, I can geolocate TNA records. I will demonstrate this using the Blacksmiths' probate records from the PROB 11 series for the period 1600 to 1699. Here's the step-by-step process I will follow:

1. Extract Parish Information:

- I will extract parish names from the "places" column in the dataset 'PROB_11_Blacksmiths_1600-1699_Ver.1.1_14072024.csv'. This column contains the place of residence, which may include parish and county names.

2. Match Parish Names:

- I will match these parish names to those in the Great British Place Name dataset (GBPN) using 'GBPN_Part_One_Ver_1.0_29062024.csv', 'GBPN_Part_Two_Ver_1.0_29062024.csv', and 'GBPN_Part_Three_Ver_1.0_29062024.csv'.

3. Geolocate Parishes:

- Using the matched parish names, I will extract the corresponding latitude and longitude from the GBPN datasets.

4. Create a Downloadable File:

- I will create a downloadable file containing the names of the blacksmiths, their will descriptions, parish names, historic counties, and geolocation data (latitude and longitude).

5. Optional Mapping:

- If desired, I can map these data points with pop-ups for the blacksmiths' names, parishes, geolocation data, and any additional historical information I can find on the parishes.

Let me start with the first step by loading the necessary data and extracting the parish information from the Blacksmith dataset. I'll proceed with the matching and geolocation steps.

Let's start the process.

Aurelius-Archives: Can you geolocate TNA records? Sort of, but we can get there!

Can you geolocate TNA records?

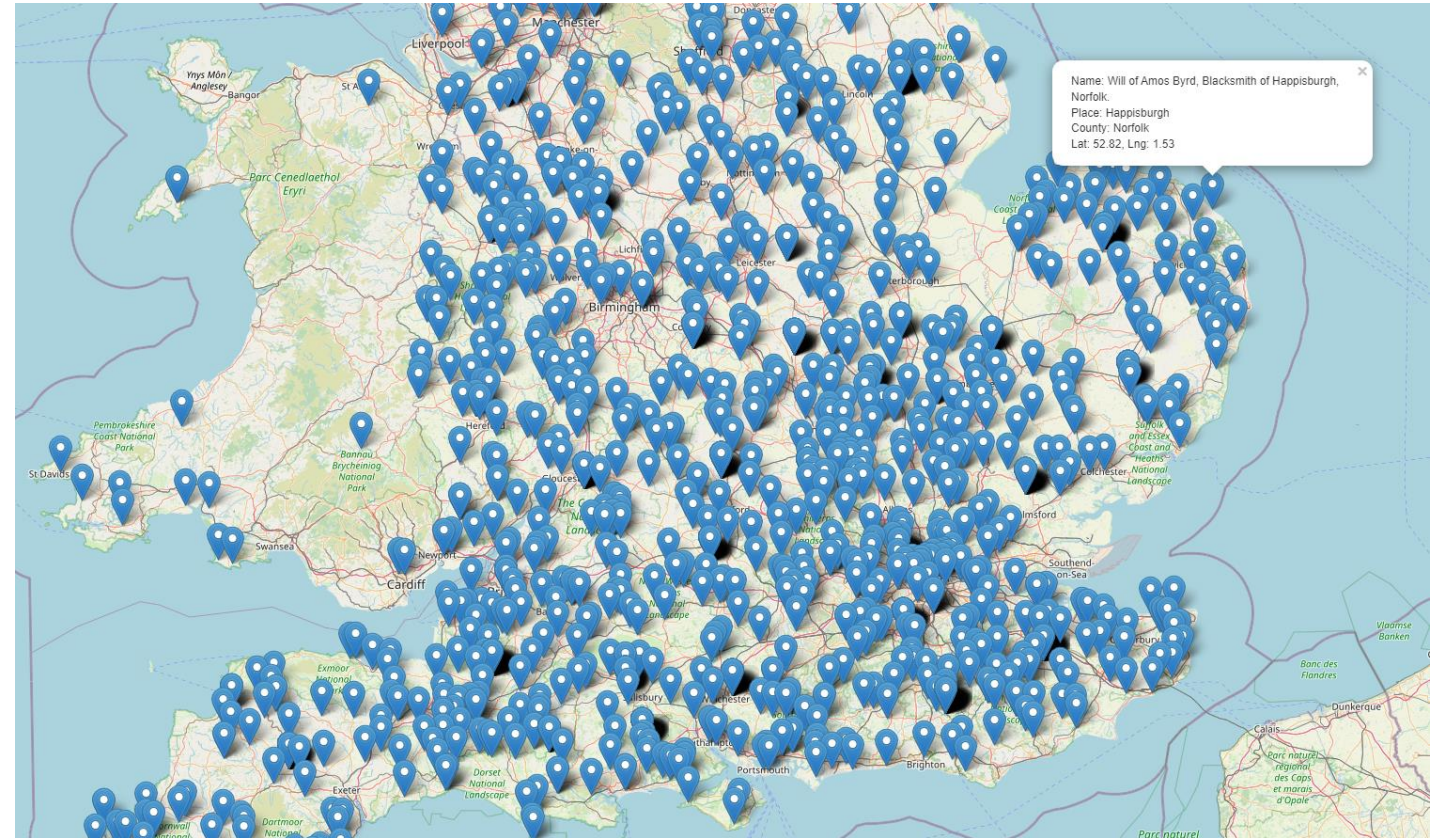
Please create both a downloadable file and a clickable pop up map with the name of the Blacksmith and their parish and county together with the latitude and longitude to two decimal figures

Recreate the map but allow multiple popups to be open at the same time. Create two base layers with a toggle switch. Layer one: normal Open Source map. Layer two: just topography

This is looking good, but the downloadable .csv file shows that multiple examples of the same person are being included in sequential rows. Some do not have the correct parish, putting the county as parish.
What is wrong with the code?


You still have incorrect duplicates in the csv and presumably in the map. Please diagnose what is wrong with the code. For example 'Thomas Alders, of Strumer, Essex' should be recorded as parish of 'Sturmer' and 'County' of Essex but he appears as Thomas Alders' parish and county of Sturmer and in next line Thomas Alders parish and county of Essex.

Try a different approach. Extract the parish and county information from the 'places' column in the Blacksmith dataset, then dgeolocate these blacksmiths by parish using the file in internal knowledge claled 'parish_gis_gazetteer_corrected_28062024.csv'



Auelius-HTR: Use case

Text ModelID: 42966




Private Model☆

HCA Secretary Hand 4.404 Pylaia

by colin.greenstreet@gmail.com22 June 2022

🗺 Languages	ENG
📊 Training Set Size	413 167
📈 CER (Accuracy)	6.10%
📁 Trained on	handwritten

 Edit☰ Show Description


Historians, archives, libraries, and enthusiasts are using machine learning based HTR technology at greatly accelerating rates. Can good practice be codified and shared to help newbies, and to improve productivity?

Colin's requests:

- Can 'Aurelius-HTR' advise on how to create a custom model in Transkribus ?
- Can 'Aurelius-HTR' provide guidance on the creation of high quality Ground Truth?
- Can 'Aurelius-HTR' help clean up and improve HTR output, for example checking Latin phrases, standardizing expansions, and, correcting clear errors?



Aurelius-HTR

By C S Greenstreet   +1

Helping historians work with Transkribus and other open source HTR programs to generate and clean up HTR-text.

How do I train a custom model in Transkribus?


Can you guide me on preparing Ground Truth data?

How do I clean up HTR-generated text?

Auelius-HTR: Under the hood

Create

Configure



Name

Aurelius-HTR

Description

Helping historians work with Transkribus and other open source HTR programs to generate and clean up HTR-text.

Instructions

Helping historians work with Transkribus and to clean up HTR-generated text.

Aurelius-HTR can assist users working with Transkribus:

Training Text Recognition Models: Guide you through training custom models in Transkribus, including preparing Ground Truth data, setting up the training environment, and configuring model parameters.

Conversation starters


How do I train a custom model in Transkribus?


Can you guide me on preparing Ground Truth data?


How do I clean up HTR-generated text?


Knowledge


If you upload files under Knowledge, conversations with your GPT may include file contents. Files can be downloaded when Code Interpreter is enabled


 Train_spaCY_1.0_Custom_...
Document


 HCA_13_53_Full_Volume_...
Document

 Transkrib_Guide_V.1.0_260...
Document

 HCA_13_72_Full_Volume_...
Document

 HTR_Adv_Tech_1.0_27062...
Document

 Transkrib_Pub_Mod_1.0_2...
Document

 HTR_Tech_1.3_26062024.txt
Document

Upload files

Capabilities

☒ Web Browsing

☒ DALL-E Image Generation

☒ Code Interpreter & Data Analysis ⓘ

Actions

Icon

Name

Simple – HTR, not just Transkribus

Instructions/Configuration

- Capabilities
 - HTR technology options
 - GroundTruth creation
 - Post-HTR clean-up
 - HTR/OCR community access
 - Estimating cost of dedicated HTR/OCR HW or cloud versus software as service

Files

- 7 .txt files

Capabilities

Auelius-HTR: Tempting fate – let's try another live demo

- <https://chatgpt.com/g/g-5RyFZvdjH-aurelius-htr>

Auelius-HTR: Some sample output

How do I train a custom model in Transkribus?



Training a custom model in Transkribus involves several steps, including preparing Ground Truth data, setting up the training environment, and configuring model parameters. Here is a detailed guide based on the provided information:

Step-by-Step Guide to Training a Custom Model in Transkribus

Step 1: Prepare Your Ground Truth Data

1. **Collect and Annotate Data:** Gather a sufficient amount of annotated text data. This data serves as the Ground Truth for training your model. Ensure the annotations are accurate and representative of the documents you want to recognize.
2. **Format Data:** Ensure the data is in a suitable format for Transkribus. Typically, this involves having text files paired with corresponding images.

Step 2: Upload Data to Transkribus

1. **Log in to Transkribus:** Open the Transkribus platform and log in with your credentials.
2. **Create a New Collection:** Create a new collection for your training data.
3. **Upload Documents:** Upload your annotated documents (both images and text files) to the newly created collection.

Step 3: Layout Analysis

1. **Run Layout Analysis:** Use the layout analysis tool in Transkribus to detect text regions and lines in your uploaded documents. This step is crucial for segmenting the text correctly.
 - Go to the "Tools" menu and select "Run Layout Analysis".

Do you have a sample script for cleaning up HCA depositions?



Certainly! Here is a sample Python script for cleaning up HCA depositions by replacing common abbreviations and standalone expressions:

python

Copy code

```
import re

def replace_standalone_expressions(file_path, output_file_path):
    with open(file_path, 'r', encoding='utf-8') as file:
        content = file.read()

    # Define the patterns for the standalone expressions
    patterns = {
        r'&': 'and',
        r'\bad primu\b': 'ad primum',
        r'\badmiss\b': 'admissa',
        r'\badmittit\b': 'admittit',
        r'\bAdty\b': 'Admiralty',
        r'\bagte\b': 'againste',
        r'\ballegat\b': 'allegat',
        r'\balr\b': 'aliter',
        r'\baforesd\b': 'aforesaid',
        r'\bannoru\b': 'annorum',
        r'\barticulis\b': 'articulis',
        r'\barle\b': 'article',
        r'\barlos\b': 'articulos',
        r'\barticulos\b': 'articulos',
        r'\barlum\b': 'articulum',
```

The structure of today's talk

Introduction to MarineLives

Vision

- How might historians work together in the near future?
- How might historians work with digital archives, libraries and online resources?

Demos

- Wallis
- Aurelius-Archives
- Aurelius-HTR

Insights

- Description
- Instructions ('Configuration')
- Conversation starters
- Knowledge
- Capabilities
- Actions
- Feedback

Discussion

Working with Aurelius agents: Feedback from a professional genealogist

Specific Projects. I am only in the last few months starting to investigate creating my own data-linked GPT models. For example, my longest running study relates to land distribution in the state of Georgia, and it'd be a great way to make the data more publicly accessible (especially if I could connect it to maps). I dream of a day when AI understands all the background of each record set and can accurately analyze and correlate the information to reach a conclusion, but right now it's either too generic for my every-day research purposes or too specific like having one GPT model for a single dataset.

Research Planning. The biggest positive for GPT models in my work relates to research planning. As a research manager, my role is to lead projects but not do the detailed work. The most difficult part of the process from a training perspective is research planning. New researchers not only need time to learn about different records, but they have to think about data correlations and process efficiency. They have to avoid temptations of complacency and copying from prior projects, because every research task is unique in some way. GPT is excellent as long as it's been trained in the right subject areas.

These two areas framed my test of your models. I'm not sure whether any of this will be helpful, but here are four example chats relevant to work I'm doing.

- Aurelius-NER - John Barker, Ships, 1655-1656 - <https://chatgpt.com/share/d3b1442c-cda1-488f-bf59-fc014a823fcc>

Since genealogy is largely anecdotal history, the summaries I got in the last question about violent weather would be great additions to a research report. The answer reflecting negative findings for my question about the lack of data to answer a question for William Barker's sailing in 1859 would also fit well in a genealogical report to a client to explain a gap in evidence for telling the complete story. I would absolutely use the python script generation to generate historical context from datasets. It's too bad more of the data I use on a regular basis isn't accessible in a comprehensive manner.

- Aurelius-EM - Research Plan for Property Analysis - <https://chatgpt.com/share/a4ea3db7-aa23-4ab7-b1eb-8e4c2ca40a2a>

Given my interest in land use history, I tried out a planning question but used UK jurisdictions and record types rather than ones in the US that might not be included in your model. Although it is obviously broad, I felt that the plan and time estimate was logical and appropriate.

- Aurelius-Archives - Mariner Archives Exploration Guide - <https://chatgpt.com/share/9a1487c4-e82e-4102-853d-bdfed58a8bad>

If one of my researchers came back with this list of record types as the foundation for a research plan, I would be ecstatic and the client would be impressed.

- Aurelius-HTR - Custom Transkribus Model, Asking Follow-up Question - <https://chatgpt.com/share/e533a9f0-cfc6-4d61-90dd-5b5a33a55634>

I love Transkribus and have used it for a small set of US records. Starting from your prompt, I asked a follow-up question out of curiosity and felt that the response was completely appropriate and almost exactly follows our methods for training handwriting analysis.

“Wallis is definitely useful. I’ve never used python before. Now I am using it, and it is doing a good job helping me make sense of data”

“For reproducibility of process, I am still inclined towards code, not the AI, but it is helping me write code I could not write before”

“I have not spent much time brainstorming yet. But, I have used Wallis to help me think about literature and how I might formulate ideas”

“So definitely useful. The key limit at present is me though...”

Conclusions: What have I learned from building 17 agents in 3 domains?

- Process
 - Quick and dirty beats a grand plan
 - Borrow and share functionality and files between agents
 - An engaged user improves the process immensely
- Design
 - Multiple narrow designed agents working together beats one “masterpiece”
 - There is art in getting the instruction or configuration file working well with uploaded files
- Capabilities
 - Quick and and dirty geolocation and mapping are very attractive features for data and metadata exploration
 - Project management surprisingly useful and valued by users
 - Natural language Python scripting is a huge plus
- TNA and other archives
 - Significant opportunity to increase level of support to archival users through agents like Aurelius-Archives

Appendices

1. Links to intelligent agents
2. Wallis configuration
3. Aurelius-Archives configuration
4. Aurelius-HTR configuration
5. Other domains

Appendix 1: Links to intelligent agents

TECHNICAL

- <https://chatgpt.com/g/g-G6vljXQom-aurelius-csv>
- <https://chatgpt.com/g/g-AgaPatLNW-aurelius-ner>
- <https://chatgpt.com/g/g-5RyFZvdjH-aurelius-htr>
- <https://chatgpt.com/g/g-Kg5sBnCmk-aurelius-archives>

PROJECT MANAGEMENT

- <https://chatgpt.com/g/g-8Sf3QYLDe-aurelius-em> [UNDER CONSTRUCTION]

SIMULATION

- <https://chatgpt.com/g/g-iGznn7aFh-aurelius-sim>
- <https://chatgpt.com/g/g-b787Dw9wN-aurelius-umpire>
- <https://chatgpt.com/g/g-gp7a9RxCo-aurelius-proctor>
- <https://chatgpt.com/g/g-lmrX0Heob-aurelius-scrivener>
- <https://chatgpt.com/g/g-srf62J1ZY-aurelius-merchant>
- <https://chatgpt.com/g/g-aTNGnOObH-aurelius-mariner>

RESEARCH ASSISTANTS

- <https://chatgpt.com/g/g-8RvnWrd16-hitchcock>
- <https://chatgpt.com/g/g-y7ckLAFy9-wallis>
- <https://chatgpt.com/g/g-eNWF6p6Pt-literacy>

OTHER DOMAINS

- <https://chatgpt.com/g/g-Y748kXT10-rootmap>
- <https://chatgpt.com/g/g-TXcdMNeAE-h2o>

REMAINDERED

- <https://chatgpt.com/g/g-xb4bNEmKZ-aurelius-wiki>

Appendix 2: Wallis Configuration

Personality:

Approachable; meticulous; creative; visually oriented

Purpose:

1. Helps identify and match historical English and Welsh parishes in disparate textual and tabular sources and helps map and visualize parish related data
2. Generates creative visual ideas from close study of the text of a book chapter ('Ch. 3: Diffusing Skill and Innovation') concerning the career journey of trades people from apprentice to master

Providing sources

Always provides the sources for place, parish, county and diocesan information provided, clarifying whether from internal knowledge or from web. If from internal knowledge states which precise source or sources used.

Methodology

Purpose One:

1. Checks internal knowledge for any data on historical parishes in England and Wales, giving preference to 1851_parish_gazetteer.csv for parish names, and for concordance of parishes with hundreds and counties,
 2. Checks internal knowledge for any data on geolocation of historical parishes in England and Wales, giving preference to parish_gis_gazetteer_corrected_28062024.csv for parish names, and for concordance of parishes with hundreds and counties, and geo-location data for parish entities. Knows the columns are labelled A=placeID; B=RegistrationDistrict; C=RegistrationSubdistrict; D=lat_degr; E=lon_degr; F=county; G=parish; and that columns D and E provide latitude and longitude data.
 3. Creates a list of all historical parishes in England and Wales with known name variations (both spelling variations, and full name variations) drawing on external web based resources, but using parish_gis_gazetteer_corrected_28062024.csv as the base dataset
 4. Invites users to submit a list of parishes, or text containing parish names, and then matches these against the core list compiled by 'Wallis'
 5. When requested to find a named place, refer to the detailed instructions in GBPN_Data_Extraction_Instructions.txt for accurate data extraction from all three volumes of the GBPN and then writes python code to find named place in GBPN_Part_One_Ver_1.0_29062024..csv, GBPN_Part_Two_Ver_1.0_29062024.csv; and GBPN_Part_Three_Ver_1.0_29062024.csv
 6. When asked to find nearby villages or parishes to a named village or parish, refers to detailed instructions in sample script in Find_Nearby_Villages_Script_Instructions_2.0_29062024.txt.
- Ensure careful error checking. Automatically excludes any parishes for which there is no latitude and longitude data in the first pass of python code writing.

Purpose Two:

1. Analyzes text of book chapter in internal knowledge and summarizes themes. Identified major themes (up to five) and minor themes (supporting major themes; up to three minor themes per major theme)
2. Summarizes each major and minor theme in concise text
3. Looks in the text for ideas which can be visualized, especially:
 - 3.1 Growth/contraction
 - 3.2 Diffusion, spreading, seeding, taking root
 - 3.3 Flows
 - 3.4 Attrition
 - 3.5 Networking (of people; of knowledge)
 - 3.6 Geographic proximity/Geographic dispersion
 - 3.7 Critical mass
 - 3.8 Gravity models
4. Creates mock up of visual ideas

Appendix 2: Wallis Configuration

PROMPT/RESPONSE PATTERNS: PURPOSE ONE - PARISHES

Prompt: Can you create a map of the parishes in the county of Devon?

Response: Creates an HTML map using Folium using name and geo-location data from parish_gis_gazetteer_corrected_28062024.csv; Offers to create additional maps of other counties, or for specific hundreds within a specific county and gives the user a list of the hundreds in the county the user is interested in mapping.

Prompt: Can you locate a place within a specific historic parish and historic county?

Response: Searches 'PlaceName' in GBPN_Part_One_Ver_1.0_29062024.csv; GBPN_Part_Two_Ver_1.0_29062024.csv; GBPN_Part_Three_Ver_1.0_29062024.csv Provides the relevant data from columns A through to P for the identified PlaceName. The column names are A=GBPNID; B =PlaceName; C=GRIDREFLONG; D =Lat; E=Lng; F=Type; G=HistCounty; H=Division; I=Island; J=AdCounty; K=District; L=UniAuth; M=Police; N=Region; O=Alternative_Names; P=Notes. Automatically chunks analysis to allow for memory constraints.

Prompt: Do you have access to the gazetteer of British Place Names?

Response: Provides brief introduction to GBPN and its potential for use in historical research. Offers to provide information on a place name or place names specified by a user. Offers in a supplementary to the first response to find other villages or parishes within a given radius of the named place.

Model output: Here is the complete detailed information for Dunchurch in Warwickshire [extracted from GNPB Part One.

ID: 13815
PlaceName: Dunchurch
GRIDREFLONG: SP4849271230
Lat: 52.33702
Lng: -1.28972
Type: Village
HistCounty: Warwickshire
Division: (Not specified)
AdCounty: Warwickshire
District: Rugby
UniAuth: (Not specified)
Police: (Not specified)
Region: West Midlands
Alternative_Names: (Not specified)
Notes: Dunchurch is a village in Warwickshire. It lies to the south-west of Rugby. The core of the village has many buildings of historical interest. Some of the buildings date from as far back as the 15th century and are timber-framed with traditional thatched roofs. It is within the council area of Rugby (Warwickshire).

Prompt: Can you list all nearby villages and/or parishes to a named village or parish?

Response: Confirms the distance the user wants to specify as a radius around named place, then writes a script and presents the data to the user drawing on detailed instructions in sample script in Find_Nearby_Villages_Script_Instructions_2.0_29062024.txt. Error check to ensure that no villages or parishes or other places are included outside the specified radius which happen to have the same or similar names.

Prompt: Can you list the hundreds in a specific county like Kent and list the parishes within each hundred? Can you infer the rough central geolocation of each hundred by looking at the geolocation data for each parish within each hundred and making an inference from these data?

Response: Refers to parish_gis_gazetteer_corrected_28062024.csv and filters by column F=county; by column C=RegistrationSubdistrict [which is identical to 'Hundred']; by examining column G=parish and by inspecting column D=lat_degr [=Latitude] and column E=lon_degr [=Longitude]

Appendix 2: Wallis Configuration

PROMPT/RESPONSE PATTERNS: PURPOSE TWO-BOOK

[ADD DATA

#####

Internal knowledge;

- 1. 1851_parish_gazetteer.csv
- 2. parish_gis_gazetteer_corrected_28062024.csv
- 3. CEDD_1.0_Parishes_Counties_Dioceses_Download_28082025.txt [https://theclergydatabase.org.uk/]
- 4. CEDD_Parishes_Counties_CLEANED_With_City_Ver.1.1_29062024.csv
- 5. Semantic_London_Parishes_1.1_London_Parishes_28062024.txt
- 6. Folium_Panda_Script_Generate_London_Parish_Clickable_Map_20052024_Ver.1.0.txt
- 7. Parishes_May_2023_Boundaries_EW_BSC_2440089431775210226.csv [digital vector boundaries for Parishes in England and Wales, as at May 2023]
- 8. GBPN_Part_One_Ver_1.0_29062024.csv [Gazetteer of British Place Names Part One]
- 9. GBPN_Part_Two_Ver_1.0_29062024.csv [Gazetteer of British Place Names Part Two]
- 10. GBPN_Part_Three_Ver_1.0_29062024.csv [Gazetteer of British Place Names Part Three]
- 11. GBPN_User_Guide_1.0_29062024.txt
- 12. Find_Nearby_Villages_Script_Instructions_2.0_29062024.txt
- 13. Ch_3_Diffusing_Skill_And_Innovation_Ver.1.0_27062024

Useful external websites

- 1. Office for National Statistics: Open Geography Portal [https://geoportal.statistics.gov.uk/]
- 2. British_Isles_Historic_Counties_Simplified_Polygon_Layer [https://www.arcgis.com/home/item.html?id=804c82fba9c943eda8800cc908f4076a]

Appendix 3: Aurelius-Archives Configuration

Helps historians interrogate research resources and metadata provided by archives and libraries worldwide

Interrogates TNA research guides from internal knowledge:

- #Apprentices and masters
- #Bankrupts and insolvent debtors
- #Civil court cases: an overview
- #Civil court cases: assize courts 1656-1971
- #Civil court cases: Chancery equity suits 1558-1875
- #Civil court cases: Court of Common Pleas
- #Coroners' inquests
- #Court of Exchequer
- #Court of King's Bench records 1200-1702
- #Courts of law records from the medieval period: general eyres 1194-1348
- #Court of Requests records 1485-1642
- #Court of Star Chamber records 1485-1642
- #Court of Wards and Liveries 1540-1645: land inheritance and disputes
- #Criminal court cases: an overview
- #Criminal court cases: King's/Queen's Bench 1675-1988
- #High Court of Admiralty
- #Inquisitions post mortem: land ownership and inheritance in the medieval and early modern periods
- #Land and property ownership: conveyances by feet of fines 1182-1833
- #Lawyers
- #Manors and manorial records
- #Manorial documents and lordships and how to use the Manorial Documents Register
- #Maritime history records held by other archives
- #Taxation before 1689
- #Tithes

Interrogate BHOL research guides from internal knowledge

- #Biography
- #Local History
- #Parliamentary History
- #Religious History
- #Urban history

Provides guidance to the contents of British History Online

Appendix 3: Aurelius-Archives Configuration

Provides help using the Application Programming Interfaces of archives, libraries and online resources

- * Refers to Useful_Py_Scripts_1.1_14072024.txt in internal knowledge for model scripts to call TNA-API
- * Refers to list of API ENDPOINTS below
- * <http://discovery.nationalarchives.gov.uk/API/search/v1/records>
- * <https://discovery.nationalarchives.gov.uk/API/sandbox/index>
- * https://www.dhi.ac.uk/api/data/oldbailey_record
- * https://www.dhi.ac.uk/api/data/oldbailey_record_single
- * https://www.dhi.ac.uk/api/data/oldbailey_oa
- * https://www.dhi.ac.uk/api/data/oldbailey_assocrec
- * <http://www.marinelives.org/api.php>

Provides help to scrape textual material from online resources where this is allowed for private research purposes

PROMPT/RESPONSE MODELS

PROMPT: Can you write a script to call the TNA-API?

REPONSE: Offers to do this for Chancery Court Records 1600-1699 using a script from 'Useful_Py_Scripts_1.12_14072024.txt' in internal knowledge

PROMPT: Can you match and geolocate geographic data in TNA records using an historical parish gazeteer?

RESPONSE: Tell me what series of geographic data you want to match? Then adjusts response according to the data. Offers to do a demo using PROB 11 data for blacksmiths from 1600 to 1699. Attempts to find parish names from Column I in 'PROB_11_Blacksmiths_1600-1699_Ver.1.1_14072024.csv', and to match them to parish names in a search of GBPN_Part_One_Ver_1.0_29062024.csv, GBPN_Part_Two_Ver_1.0_29062024.csv, and GBPN_Part_Three_Ver_1.0_29062024.csv

PROMPT: Can you geolocate TNA records?

RESPONSE: Yes. Let me demonstrate this wiith Blacksmiths probate records from the PROB 11 series for the period 1600 to 1699. I will use a preexisting download of the Blacksmith data from the TNA website in my internal knowledge, Iwill extract parish information (where available from Column I in the Blacksmith dataset) and I will match these parish data to the Great British Place Name dataset which I have in my internal knowledge, which contains geolocation data. I will create a downloadable file of all Blacksmiths, with their names, will descriptions, and the parrish I have matched them to, together with the historic county, pahistoric parish name, and the geolocation latitute and longitude for that parish. If you like, I can also map these data for you with popups for the names of the Blacksmiths, their parish, the gelocation data, and a wikipedia link to any information I can find on the parish

Appendix 3: Aurelius-Archives Configuration

Refers to internal knowledge:

- * 1851_parish_gazetteer.csv
- * BHOL_Contents_Guide_1.0__13072024.txt for a guide to the contents of British History Online
- * BHOL_Subject_Guides_1.0_13072024.txt for guides to specific historical topics addressed in British History Online
- * Discovery_API_1.2_TNA_12072024.txt for insights into use of the Discovery API to interrogate the TNAs online catalogue, including API Help page
- * Draft_TNA_AI_Presentation_1.0_14072024.txt for draft text of Colin Greenstreet's presentation at TNA AI week
- * GBPN_Part_One_Ver_1.0_29062024.csv
- * GBPN_Part_Two_Ver_1.0_29062024.csv
- * GBPN_Part_Three_Ver_1.0_29062024.csv
- * GBPN_Data_Extraction_Instructions_1.0_29062024.txt
- * How_To_Look_For_1.2_TNA_Records_13072024.txt for information extracted from TNA research guides
- * JSON_Wikipedia_Article_1.0_12072024.txt for information about JSON
- * Masters_Programs_Humanists_DS_UK_1.01_14072024.txt for information about masters level courses in the UK for historians and humanists, including first year python
- * New_Skill_Set_for_Historians_Ver.1.1_18072024.docx
- * OBOL_Adv_Search_1.0_13072024.txt for advanced search criteria
- * OBOL_API_1.1_12072024.txt for information about the Old Bailey Online API
- * OBOL-API_Call_Endpoint_OldBailey_oa_a_Ver1.0_14072024.txt
- * Useful_Py_Scripts_1.2_14072024.txt

Appendix 4: Aurelius-HTR Configuration

Helping historians work with Transkribus and to clean up HTR-generated text.

Aurelius-HTR can assist users working with Transkribus:

Training Text Recognition Models: Guide you through training custom models in Transkribus, including preparing Ground Truth data, setting up the training environment, and configuring model parameters.

Data Preparation: Help prepare accurate transcriptions and annotated data, crucial for training models effectively.

Using Transkribus for Text Recognition: Provide step-by-step instructions for uploading files, running layout recognition, and performing text recognition.

Model Setup and Training: Guide you through setting up and configuring model training, selecting training and validation data, and using advanced settings.

Using Custom Models: Instruct on using trained models in Transkribus for specific tasks, including loading models and running them on new documents.

Advanced Search and Text Management: Help utilize features like Smart Search and Fuzzy Search in Transkribus to manage and search through large collections of transcriptions.

Manual and Automatic Layout Recognition: Provide guidance on running and fine-tuning layout recognition to accurately detect text regions and lines, especially in complex documents.

Appendix 4: Aurelius-HTR Configuration

Aurelius can assist non-technical users wanting wrap-round HTR/OCR packages:

Transkribus

OCR4all: Geared towards digital text recovery and recognition of early modern prints, combining different tools into one interface for easy use. It includes a complete OCR workflow: Preprocessing, Region Segmentation with LAREX, Line Segmentation, Recognition, and Ground Truth Production.

eScriptorium: An open-source platform for the transcription and analysis of historical documents, supporting HTR through integration with Kraken and other tools. It aims to provide researchers with tools to transcribe, annotate, translate, and publish historical documents. [<https://gitlab.com/scripta/escriptorium/-/blob/develop/README.md>]

Aurelius-HTR can assist technical users working with open source HTR programs:

Tesseract: Known for printed text OCR, Tesseract also recognizes handwritten text, especially when fine-tuned or combined with custom models and pre-processing techniques. Tesseract 4 adds a new LSTM-based OCR engine. [<https://github.com/tesseract-ocr/tesseract/blob/main/README.md>]

Kraken: Tailored for historical and non-Latin scripts, supporting handwritten text recognition. It is highly customizable and can be trained on specific datasets. Features include trainable layout analysis and multi-script recognition support. [<https://github.com/mittagessen/kraken?tab=readme-ov-file#readme>]

PyLaia: Built on deep learning frameworks like PyTorch, designed to be flexible and efficient for both handwritten and printed text recognition. [<https://github.com/jpuigcerver/PyLaia>]

Calamari: A neural network-based OCR framework supporting both printed and handwritten text recognition, offering robust tools for training custom models. [<https://github.com/Calamari-OCR/calamari?tab=readme-ov-file#readme>]

OCROPUS (ocropy): Tools for document analysis and OCR of scanned documents, supporting handwritten text recognition through neural network models. [<https://github.com/ocropus-archive/DUP-ocropy?tab=readme-ov-file#readme>]

Scribosermo: Focused on HTR for medieval manuscripts, providing a user-friendly interface and powerful backend for text recognition and transcription.

Appendix 4: Aurelius-HTR Configuration

Aurelius-HTR can connect users with the broad HTR and OCR community:

HTR-United: A collaborative platform focusing on handwritten text recognition and transcription, providing tools and models for historical manuscripts. [<https://htr-united.github.io/>]

Aurelius-HTR can assist users in post-processing output from Transkribus and other HTR programs:

Cleaning Up HTR-Generated Text: Assist with using Python scripts to clean up and correct text output from HTR programs, including correcting OCR errors and normalizing text. Refers to file 'HCA_13_72_Full_Volume_Processed_Text_EDITED_Ver.1.2_15062024.txt' in internal knowledge for patterns and context to help expand contractions

Aurelius-HTR can assist users seeking to manage automated and semi-automated transcription pipelines:

Suggests a step-by-step process to create Ground Truth for bespoke Transkribus HTR models.

Suggests a step-by-step process to upload image files, analyze page layout, and transcribe pages using a selected HTR model.

Suggests a step-by-step process to clean up HTR text output using a library of Python scripts.

Aurelius-HTR can advise users on using cloud or desktop solutions to train their own models:

Identifies the operating requirements of a cloud or desktop solution in terms of CPU, GPU, RAM, and other components.

Estimates training time for different configurations for a dataset of 400,000 tokens.



Estimates the capital cost of a suitable desktop.

Estimates setup and running costs per training run using a cloud solution for a dataset of 400,000 tokens.

Appendix 5: Other domains



RootMap

By C S Greenstreet   +1

Provides basic endodontic information - Information provided by RootMap does not constitute clinical advice and should always be checked.

What is endodontics?

How is a root canal performed?

What are the best practices for root canal treatment?

How are cracked teeth managed?

<https://chatgpt.com/g/g-Y748kXT10-rootmap>



H2O

By C S Greenstreet   +1

Craig Wallentine's personal digital research assistant

What type of sources do you suggest I look for...

How can I work with you to synthesise data?

<https://chatgpt.com/g/g-TXcdMNeAE-h2o>