

The background is a solid dark blue. On the left side, there are several parallel white lines that form a corner-like shape, extending from the top left towards the center. In the bottom right corner, there are several parallel teal lines that form a diagonal shape, extending from the bottom left towards the top right. The main title is centered in the upper half of the image.

Digital Humanities Abstracts

ISSUES OF A YOUNG DISCIPLINE

Agenda

- Dataframe
- Possible Questions
- Implementation and Difficulties
- Results
- Outlook
- Q&A

Dataframe

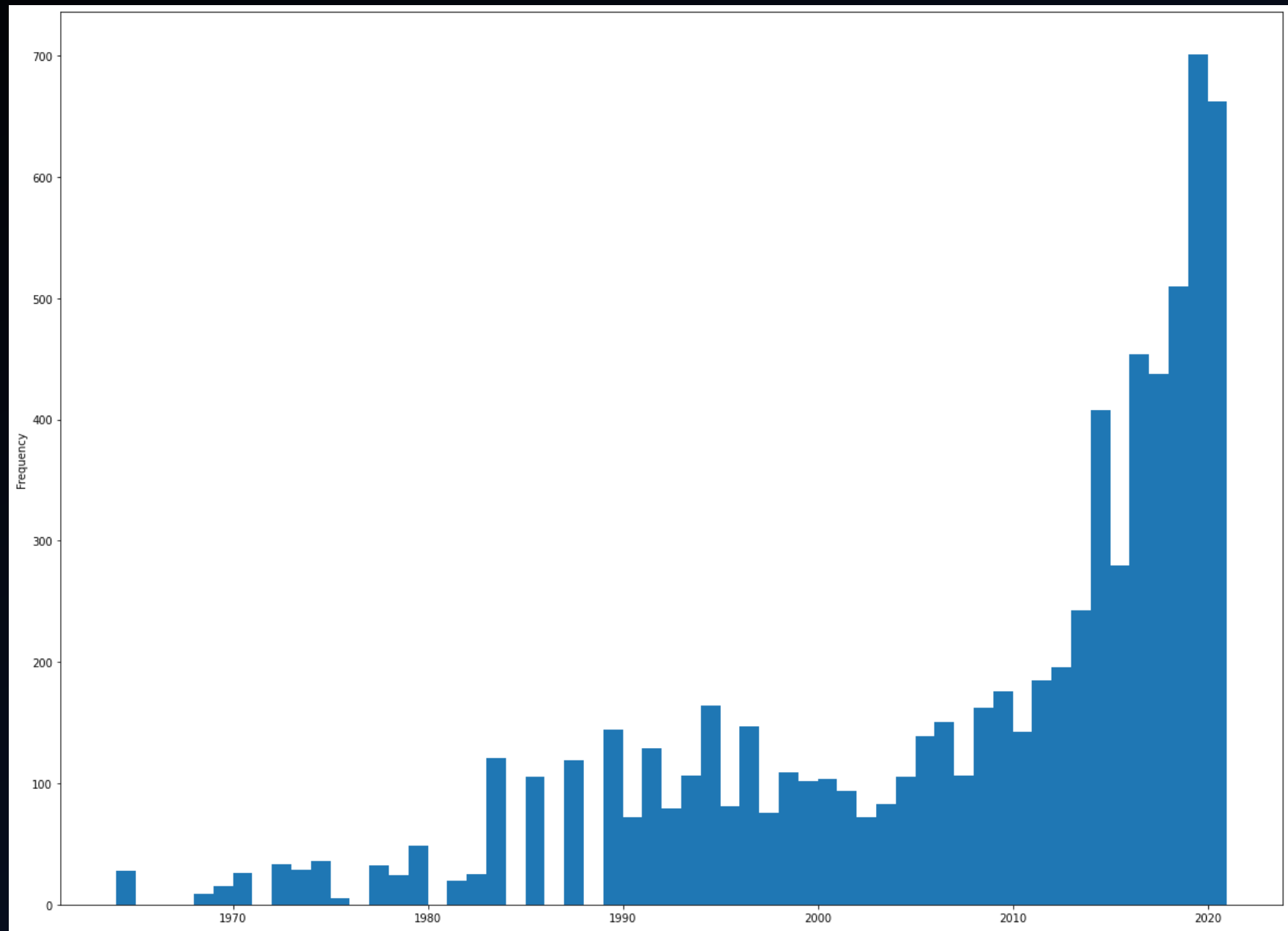
	work_id	conference_label	conference_short_title	conference_theme_title	conference_year	conference_organizers	conference_series	conference_hosting
0	1	2000 - Glasgow	Glasgow	NaN	2000	ACH;ALLC	ACH/ALLC;ACH /ICCH;ALLC/EADH	University of Glasgow
1	2	2000 - Glasgow	Glasgow	NaN	2000	ACH;ALLC	ACH/ALLC;ACH /ICCH;ALLC/EADH	University of Glasgow
2	3	2000 - Glasgow	Glasgow	NaN	2000	ACH;ALLC	ACH/ALLC;ACH /ICCH;ALLC/EADH	University of Glasgow
3	4	2000 - Glasgow	Glasgow	NaN	2000	ACH;ALLC	ACH/ALLC;ACH /ICCH;ALLC/EADH	University of Glasgow
4	5	2000 - Glasgow	Glasgow	NaN	2000	ACH;ALLC	ACH/ALLC;ACH /ICCH;ALLC/EADH	University of Glasgow

Possible Questions

- Predict keywords from title
- Predict keywords from time series
- Cluster titles and keywords to find „topics“

Implementation and Difficulties

- Little data
- Unequally distributed



Implementation and Difficulties

- Little data
- Unequally distributed
- No discernible correlation between keywords and titles
- Different number of topics per paper

Approach: Predict the topic from work_title

- Select all topics
- Count the occurrence of each topic
- Re-assign the most frequent topic to each row (work_title)
- Divide the data into train and test set
- Calculate the confusion matrix

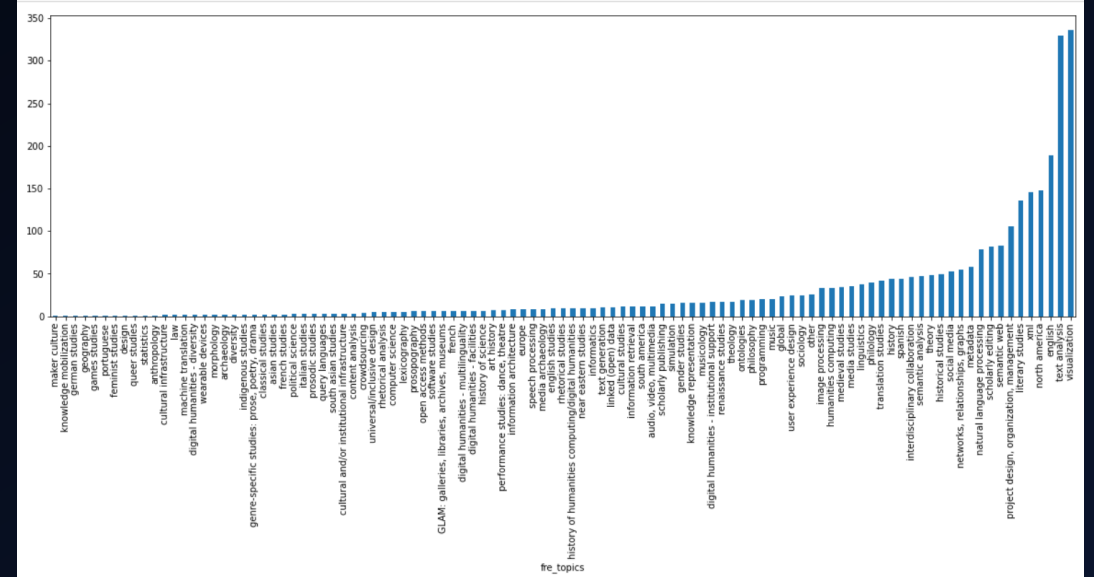
Results: Each row with most frequent title

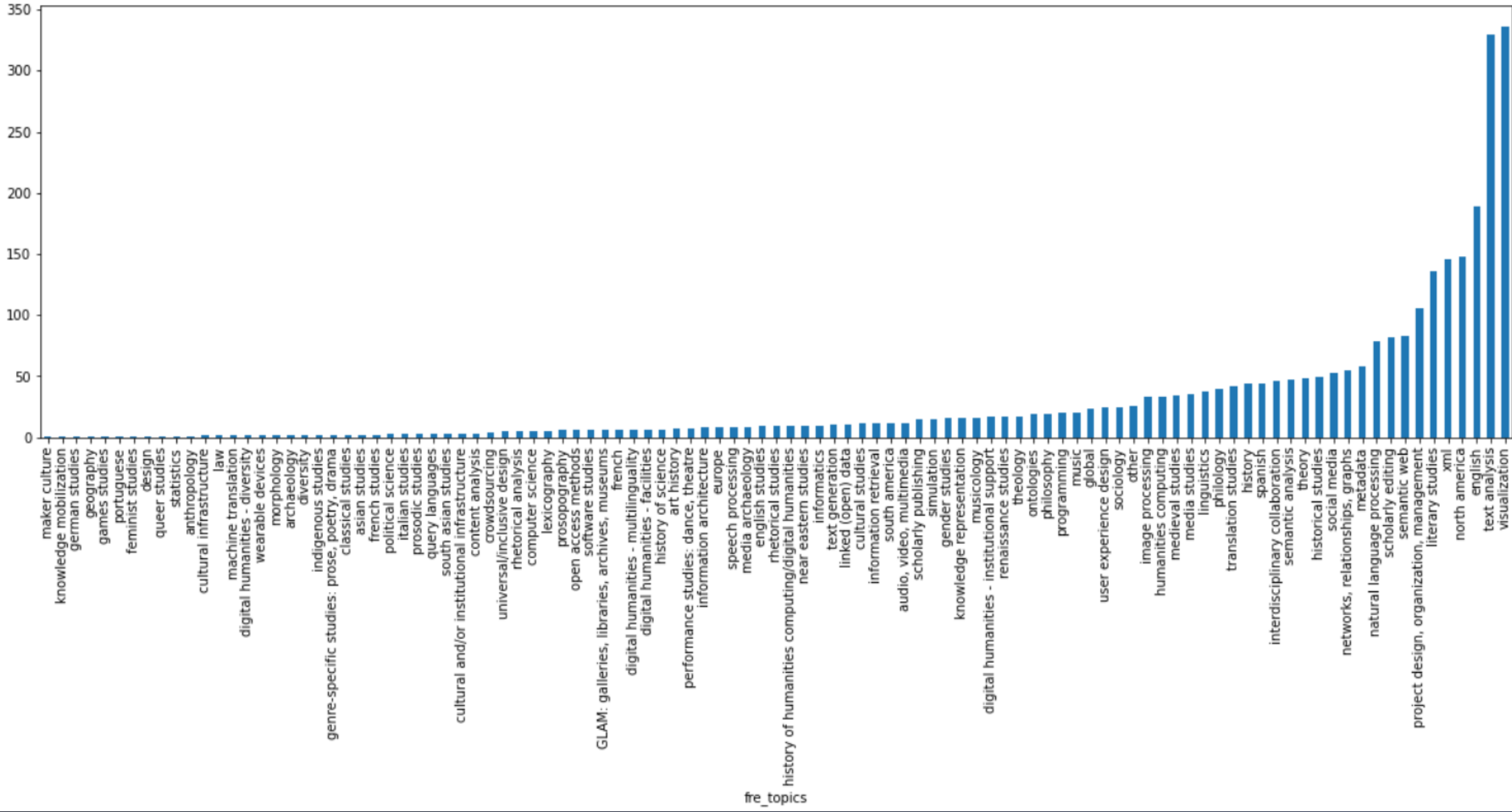
work_title	work_url	work_authors	work_type	parent_work_id	keywords	languages	topics	fre_topics
A 3D Common Ground: Bringing Humanities Data T...	NaN	James Joel Coltrain	paper, specified "short paper"	NaN	3d;architecture;game engine;historical reconst...	English	anthropology;archaeology;art history;games and...	visualization
4Humanities: Designing Digital Advocacy	NaN	Lindsay Thomas;Alan Liu;Geoffrey Rockwell;Stéf...	paper, specified "long paper"	NaN	4humanities;digital humanities;humanities advo...	English	crowdsourcing;digital humanities - institution...	text analysis
Voyant Notebooks: Literate Programming and Pro...	NaN	Stéfan Sinclair;Geoffrey Rockwell	poster / demo / art installation	NaN	literate computing;text analysis;voyant	English	text analysis	text analysis
Textal: a text analysis smartphone app for Dig...	NaN	Melissa Terras;Steven James Gray;Rudolf Ammann	poster / demo / art installation	NaN	app;mobile;smartphone;text analysis;wordcloud	English	data mining / text mining;mobile applications ...	text analysis
Academic Migrants: A Digital Discussion of Tra...	NaN	Olivia Donaldson	paper, specified "short paper"	NaN	education;migration;storytelling	English	audio, video, multimedia;cultural studies;lite...	other

Results

Most frequent topics:

- Visualization
- Text analysis
- English





Problems with Current Approach

- Most frequent topic is not always representative for title
- Only one topic per title is predicted
- Quality of model is not satisfactory

Workflow- and organizational Difficulties

- Local hosting of jupyter notebooks
- Chaotic structure of notebook
- Uncommented code
- Missing schedule (respectively communication channels)

Outlook

- Model with „single topic approach“
- Model with many outout topics
- Correlations between topics
- Time series



Questions & Answers

THANK YOU FOR YOUR ATTENTION