

Meaningful Memories: Navigating oral history interviews using LLM-based transcription and annotation

Simone van Bruggen (SURF), Boudewijn Koopmans (UvA), Ingeborg Verheul (UvA),
Leon van Wissen (UvA), Annette Langedijk (SURF)

Goals

- Extending the Amsterdam Diaries Time Machine project to include oral histories
- Reducing the amount of manual work required to transcribe, annotate and link these interviews
- Generate fragment-level annotations that can be used to find connections between interviews

Data

Portrait of Amsterdam interviews conducted by students (master Publieksgeschiedenis). All data is in Dutch.

- Video (177 files)
- Blogs (111 files, text only)
- <https://hart.amsterdam/nl/page/392478/biografie-van-amsterdam>

 Hart
Amsterdam
Diaries
Time Machine



Interviews



Database



Transcription
(speech-to-text)

Automatic annotation pipeline



External
knowledge
(knowledge
graph)



Named entity
recognition
(LLM)

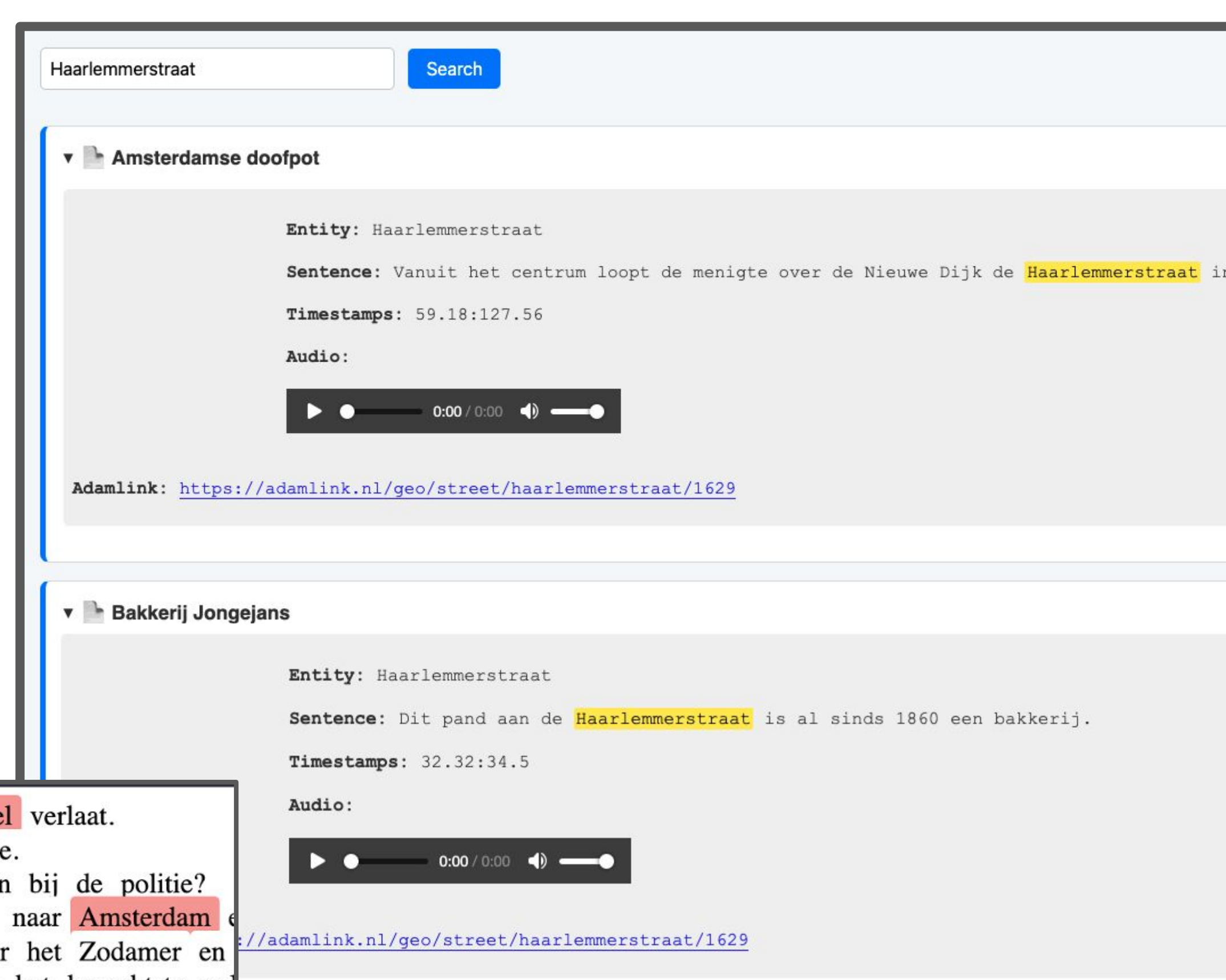


Human
evaluation

Set-up

- Transcription: Whisper / WhisperX
- NER for finding exact locations: Gliner
- Structured LLM output for finding non-literal matches:
 - Llama3 70B
 - Substitute with any local or HuggingFace-hosted LLM
 - Requires Dutch or multilingual model for acceptable results
- Using **local** models: on-premise or on local device
- Follow-up extensions:
 - RAG: use knowledge bases as input
 - Correct transcriptions using LLM + external knowledge (e.g. Warmoesstraat -> Warmoesstraat)

Het is 1971 als Gert van Beek Texel verlaat.
Hij is 19 en gaat op zoek naar actie.
En waar kun je dat beter vinden dan bij de politie?
Na drie jaar politieacademie mag hij naar Amsterdam.
En ik kwam in Amsterdam nog weer het Zodermer en
Bureau Warmoesstraat. In de jaren 70 het beruchtste po
buurt waar drugs, geweld en criminaliteit regeren. Voor
Gert van Beek is er geen betere plek om het vak te
Het was het hele pand tot aan water van Damrak en
Maar er waren vier panden die intern ook met elkaar



Do you want to learn more or
apply this tool to your data?
Join us at the
**ToekomstTienDaagse
Verhalen van Amsterdam
datasprint!**



Code available at
[https://github.com/
SURF-ML/meaningful-
memories](https://github.com/SURF-ML/meaningful-memories)



SURF

Digital Humanities Benelux
3-6 juni 2025
VU Amsterdam