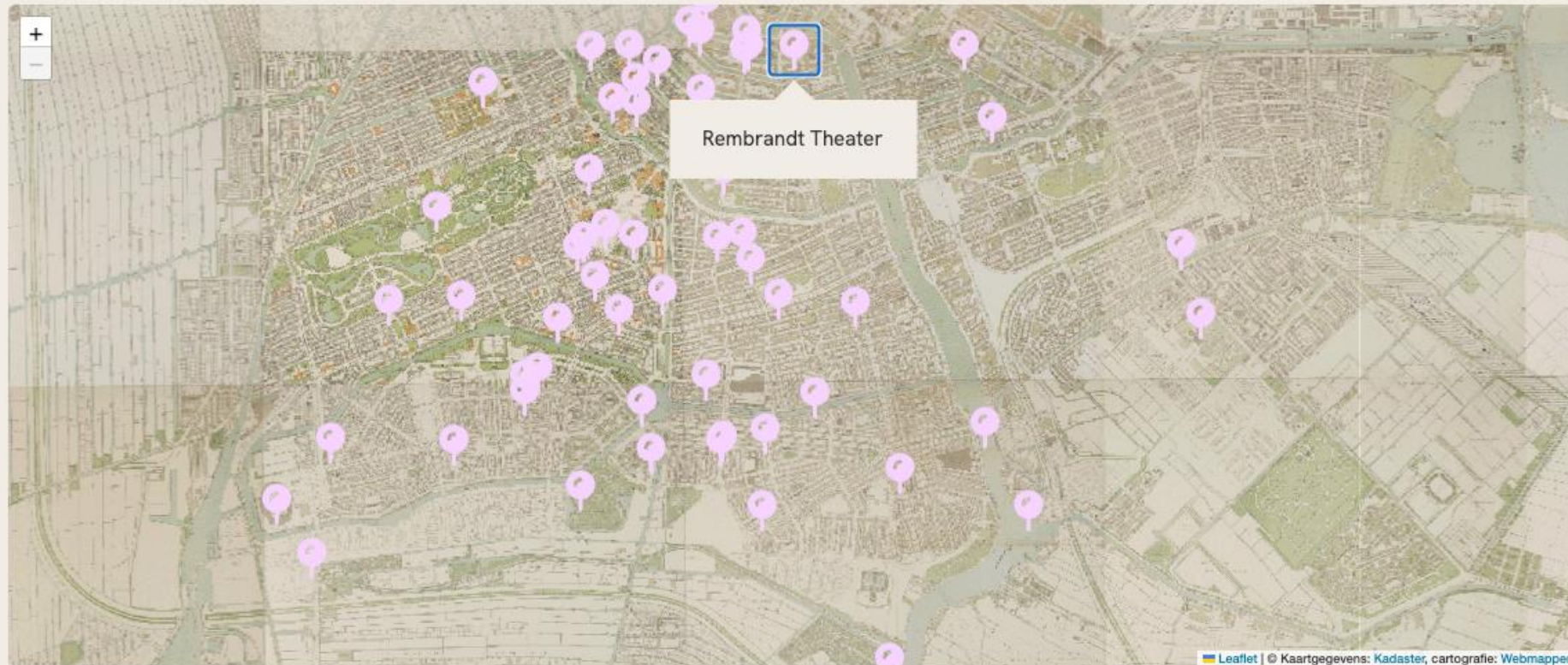


Meaningful Memories: an oral | history annotation pipeline

Simone van Bruggen (SURF)

| Amsterdam Diaries Time Machine





Dagboekteksten over Rembrandt Theater

Dagboek Berdi, 1942-1945 1942/1945

desondanks zijn we vanmiddag
of liever gezegd tegen de avond
gaan wandelen. Greet had nog
nooit gezien hoe of het

Dagboek Berdi, 1942-1945 1942/1945

mijn buikje liggen en dan til ik
ieder ogenblik mijn hoofdje op.
[Het Rembrandt theater](#) hier in
Amsterdam is vannacht

Dagboek Berdi, 1942-1945 1942/1945

laden. En nu is de tram nog wel
omgelegd met het oog op de
brand van het [Rembrandt
theater](#). De mensen hangen

| Goal 120 hr innovation project: a Proof-of-Concept

- Extending the Amsterdam Diaries Time Machine with oral history
- Reducing the amount of manual work
- Generate fragment-level annotations



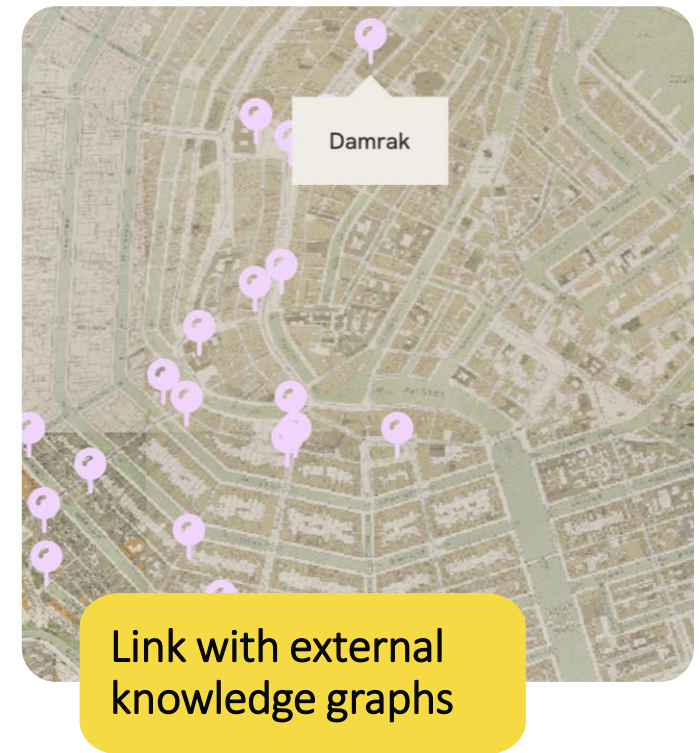
| Modular pipeline



5 augustus 1943 donderda

Het weer leek iets heter vanmorgen, maar het he
geonweerd en het regent dat het giet. Greet zit a
kopje thee surrogaat ; echte thee is niet meer te l
We hebben een fles koolzaadolie cadeau ge- kre
er zelfs flensjes in, ja, daar krijg ik nog wel niets v
geval en het ruikt niet slecht. Cor is even naar tai
ziek. Oom Antoon heeft hier gegeten. Ik ben 's na
trap

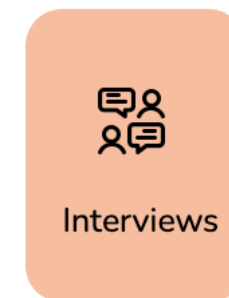
Extract concepts
and entities



| Data

Amsterdam documentaries
by students University of Amsterdam
(master Public History)

Videos and blogs

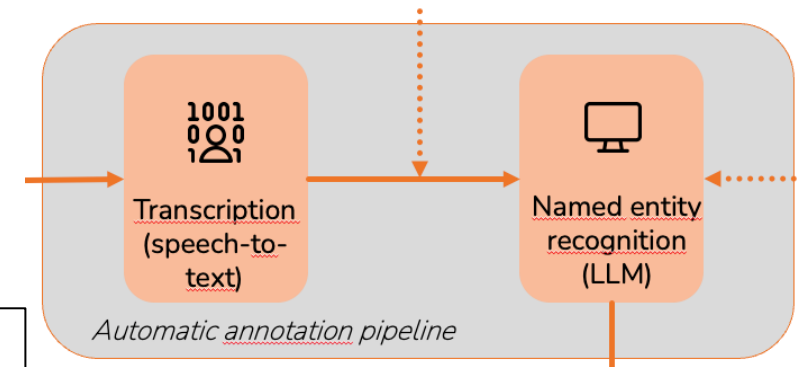


| Automatic transcription and annotation

Step 1: Speech-to-Text



"Dus die gingen... Voor studenten was het standaard, als je naar Amsterdam komt, dan ga je naar de eerste en de beste junk, dan kan je voor 10 gulden een fiets. Dat was in de studentenflat normaal. Gekke Cor was onze huisjunkie in de Weespestraat. Maar Gekke Cor hoorde erbij."



| Automatic transcription and annotation

Step 1: Speech-to-Text

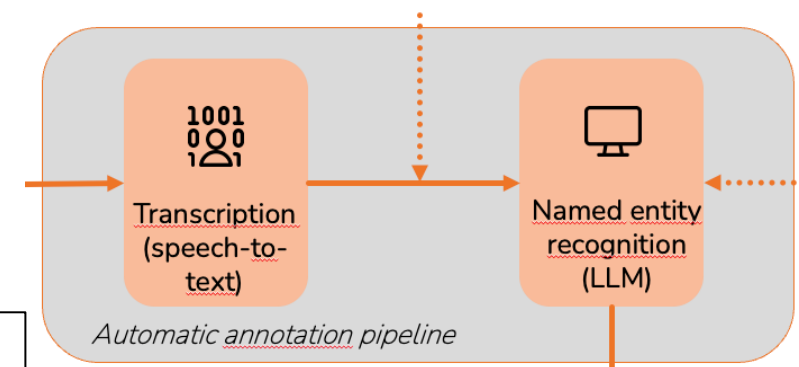


"Dus die gingen... Voor studenten was het standaard, als je naar Amsterdam komt, dan ga je naar de eerste en de beste junk, dan kan je voor 10 gulden een fiets. Dat was in de studentenflat normaal. Gekke Cor was onze huisjunkie in de Weespestraat. Maar Gekke Cor hoorde erbij.

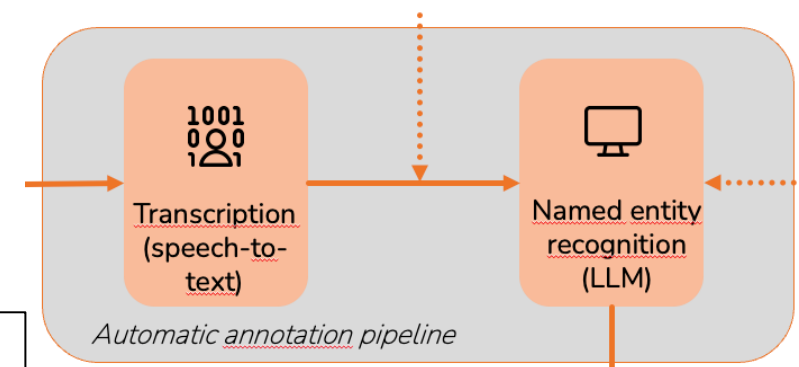
Step 2: Named Entity Recognition

GLiNER

Dus die gingen... Voor studenten was het standaard, als je naar Amsterdam komt, dan ga je naar de eerste en de beste junk, dan kan je voor 10 gulden een fiets. Dat was in de studentenflat normaal. MUZIEK Gekke Cor was onze huisjunkie in de Weespestraat. Maar Gekke Cor hoorde erbij.



| Automatic transcription and annotation



Step 1: Speech-to-Text



"Dus die gingen... Voor studenten was het standaard, als je naar Amsterdam komt, dan ga je naar de eerste en de beste junk, dan kan je voor 10 gulden een fiets. Dat was in de studentenflat normaal. Gekke Cor was onze huisjunkie in de Weespestraat. Maar Gekke Cor hoorde erbij.

Step 2: Named Entity Recognition

GLiNER

Dus die gingen... Voor studenten was het standaard, als je naar Amsterdam komt, dan ga ie naar de eerste en de beste junk, dan kan je voor 10 gulden een fiets. Dat was in de studentenflat normaal. MUZIEK Gekke Cor was onze huisjunkie in de Weespestraat. Maar Gekke Cor ho erbij.

Step 3: LLM based theme / concept extraction



Topics

1. Vrijheid
2. Studenten
3. Muziek
4. Cultuur
5. Identiteit

Using local models: on laptop or on-premise SURF AI-hub

SURF

| Link to external resources

[Wikidata](#)

And Dutch-specific:

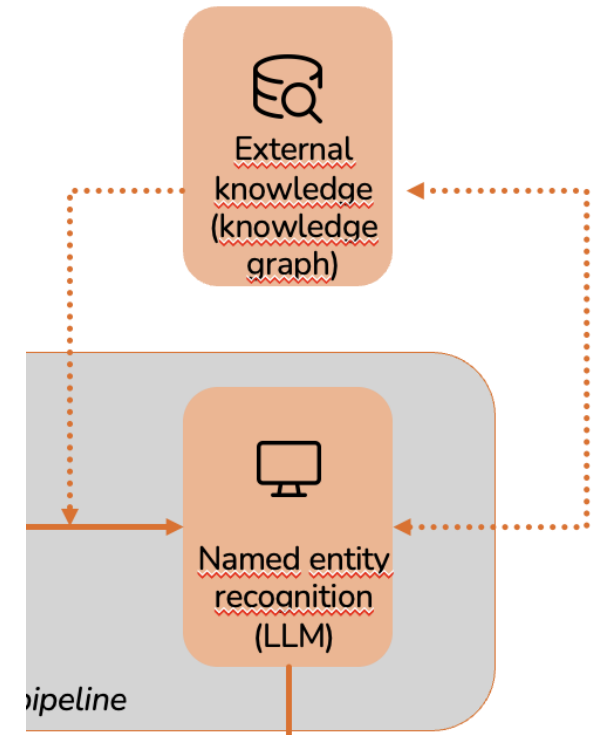
[Adamlink.nl](#)

[Network of Terms](#)

[GTAA: Gemeenschappelijke Thesaurus voor Audiovisuele Archieven](#)



SURF



Output and human evaluation

Dus die gingen... Voor studenten was het standaard, als je naar Amsterdam komt, dan ga je naar de eerste en de beste junk, dan kan je voor 10 gulden een fiets.

Dat was in de studentenflat normaal.

AnnoMap

MUZIEK **Gekke Cor** was onze huisjunkie in de **Weesperstraat**. Maar **Gekke Cor** hoorde erbij.



SURF



Human
evaluation



Label Studio



| Quality

- How to estimate quality?
- Is probability a good measure of quality?
- No gold standard: no ground truth annotations for this dataset
- Context is hard for machine learning models and a lot easier for humans 😊
- Best of both worlds?

Thanks to:

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

Project team Meaningful
Memories



Code available at: <https://github.com/SURF-ML/meaningful-memories>



WOII door de ogen van Amsterdamse vrouwen Amsterdam Diaries Time Machine

