

IDENTIFYING HISTORICAL OCCUPATIONS IN UNSTRUCTURED TEXTS: A USER CENTRIC APPROACH

OBJECTIVE

Standardize the process of matching occupation names found in biographical texts to occupational codes.

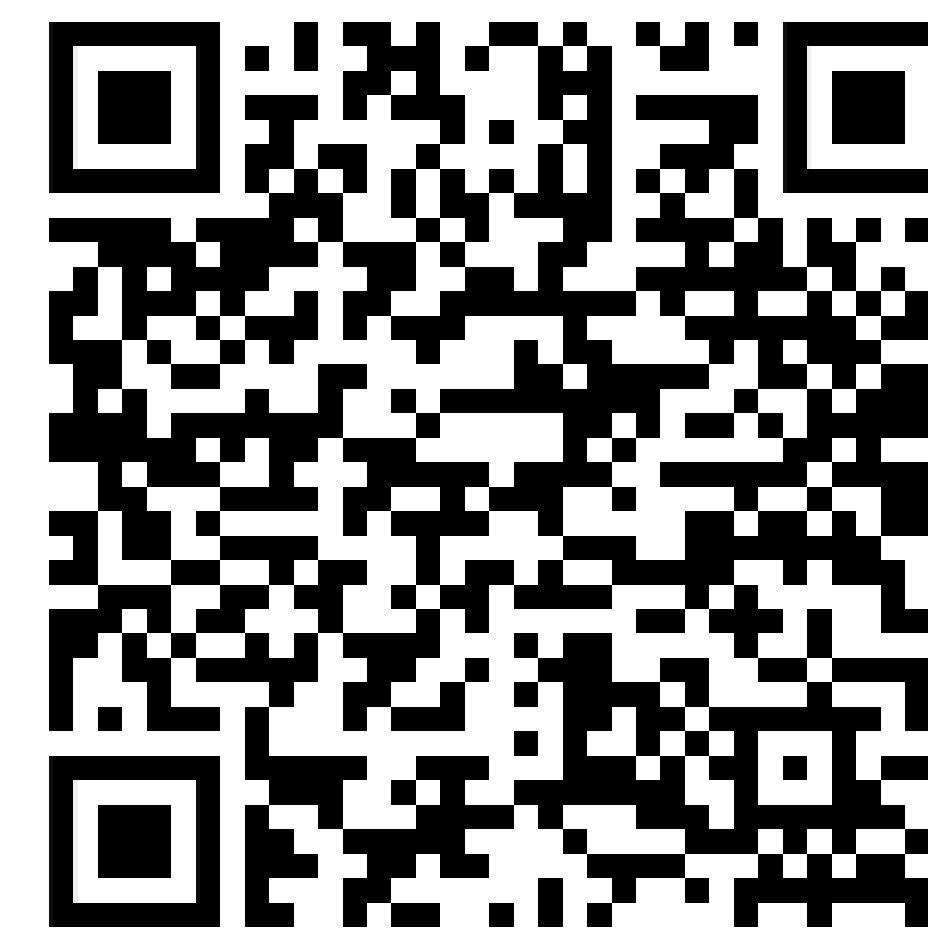
Antoine Louis des Tombe

1907-1987, *Burgemeester*

Antoine Louis des Tombe werd op 19 februari 1907 in De Bilt geboren. Hij was een zoon van Jacob Willem des Tombe, wethouder te De Bilt (1861-1921), en Antoinette Louise Baronesse van Boetzelaer (1873-1965). Op 2 juli 1954 trad hij te Blaricum in het huwelijk met Alice de Vries Robbé (geb. 1933). Zij kregen een dochter. Antoine Louis des Tombe overleed op 12 juni 1987 in Apeldoorn.

Antoine Louis des Tombe stamde uit een familie die van oorsprong afkomstig was uit de Noord-Franse textielstad Tourcoing. Tijdens de Hugenotenoerlog vluchtten zijn voorouders uit Frankrijk en vestigden zich eind 16de eeuw in Leiden waar ze via de lakenindustrie tot aanzien kwamen. Een van de nazaten, Jacob des Tombe, was in de 18de eeuw drost van Amerongen, en zijn achterkleinzoon Jacob Willem was wethouder te De Bilt en de vader van Antoine Louis.

Antoine Louis was 14 jaar toen zijn vader stierf. Hij bezocht op dat moment het gymnasium te Utrecht. Na zijn rechtenstudie aldaar vertrok hij in 1932 als volontair naar de gemeentesecretarie van Maartensdijk en in juli 1933 werd hij benoemd tot **adjunct-commies/redacteur** in de voormalige gemeente Zuylen. Des Tombe, lid van de Christelijk Historische Unie, werd in oktober 1934 benoemd tot **burgemeester** van Abcoude. Zijn benoeming tot **burgemeester** betekende voor de hele familie een verhuizing. Zijn



ACCESS THE QR CODE FOR AN IN-DEPTH EXPLANATION OF THE PROJECT!

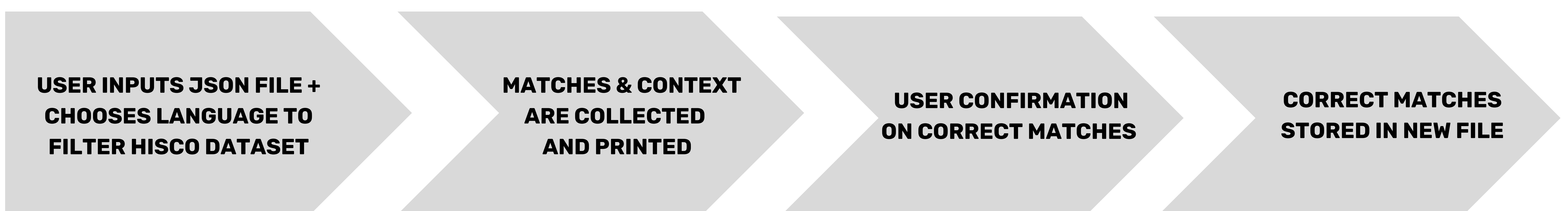
OCCUPATIONAL CATEGORY CODES

2 : Administrative and managerial workers	→	Major Group
2-1 : Managers	→	Minor Group
2-0 : Legislative Officials And Government Administrators	→	Minor Group
2-01 : Legislative Officials	→	Unit Group
2-12 : Production Managers	→	Unit Group
2-01.10 : Legislative Official (Burgemeester)	→	Occupation

METHOD

Produce a workflow to establish entity recognition between HISCO occupational codes and user chosen biographical text

APPROACH WORKFLOW



Match found!

Currently analyzing the biography of: Antoine Louis des Tombe

Word with potential occupational match: burgemeester

Potential HISCO occupation match: Burgemeester

In-text contextualization of the word:

"Zijn benoeming tot burgemeester betekende voor de hele familie een verhuizing "

HISCO code for potential occupation: 20110

Levenshtein distance between word and potential HISCO occupation: 1

Do you agree with this matching? y/n

CHALLENGES

- Slow run time creates hiccups
- Word repetition leads to unnecessary instances of the same token
- Random words mistaken for typos are matched to occupations
- Tokens ignored when in proximity of another token with a small Levenshtein distance.

FUTURE WORK

- Option for user to max out number of exact repetitions
- Automatically reject matches based on small contextual cues around the given word
- User interface