

Towards Semantic Tagging of Segmented Holocaust Narratives

Christopher Brückner, Pavel Pecina

📅 January 23, 2025



Charles University
Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics



unless otherwise stated

Visual History Archive (VHA)

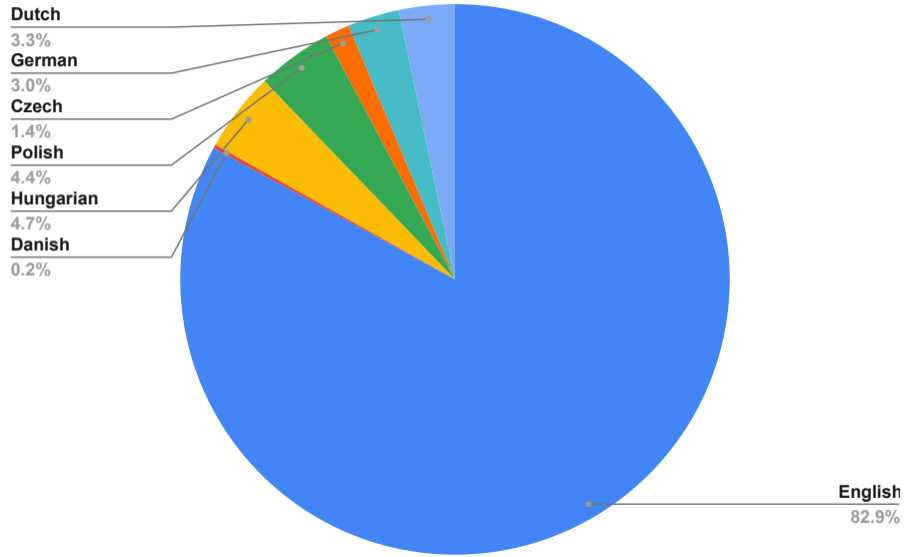
- ~50,000 testimonies of Holocaust survivors and witnesses in ~40 languages
- Hierarchical ontology
 - ~68,000 keywords in total
 - ~19,000 subjects

Subject APPLY Clear ^

- + beliefs 0 / 44,857 ⓘ
- + captivity 0 / 45,215 ⓘ
- + culture 0 / 40,984 ⓘ
- + daily life 0 / 54,377 ⓘ
- + discrimination 0 / 40,615 ⓘ
- discrimination responses 0 / 50,241 ⓘ
- adaptation and survival 0 / 46,566 ⓘ
 - adaptation 0 / 31,092 ⓘ
 - adaptation methods 14,277 / 27,627 ⓘ
 - camp adaptation methods 9,603 ⓘ
 - deportation adaptation methods 614 ⓘ
 - forced labor battalion adaptation methods 541 ⓘ

Semantic Tagging: Linking documents to concepts in the ontology

VHA: Languages



VHA Annotations: Semantic Segmentation

< **Eva Kor**
Jewish Survivor

USC Shoah Foundation Sponsor Cite Options

Metadata tabs below testimony player can be dragged into this sidebar.

SEGMENTS Indexing Date: Aug 29, 2005

Terms:

- Auschwitz II-Birkenau (Poland : Death Camp)
- camp adaptation methods
- camp children
- camp hospitals
- camp medical experiments
- health deceptions
- König, Hans-Wilhelm
- Mengele, Josef
- Poland 1941 (June 21) - 1945 (May 7)
- twins in the camps

SEGMENT 14

Terms:

BIO PROFILE SUBJECT TERMS PEOPLE MAP FIND IMAGES

Initial Experiments

Goal: Train a language model to classify VHA segments in different languages

- 942 partial English transcripts (Malach Speech Recognition data)
- Uniform segmentation with sparse annotations only
- 2115 segments → 1063 subject keywords

Input: Transcript

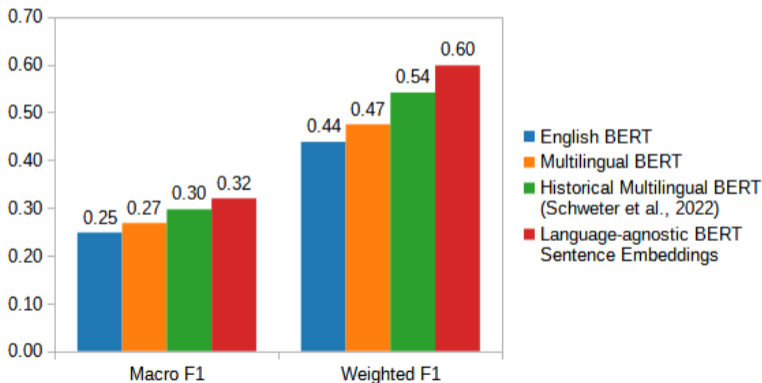
I had to help my sister with her homework if I could, if she would let me. And my brother. He wouldn't let me near him. I did like to follow him, although he was two years younger. He says, you're a girl, go away. [...]

Output: Binary-encoded keywords

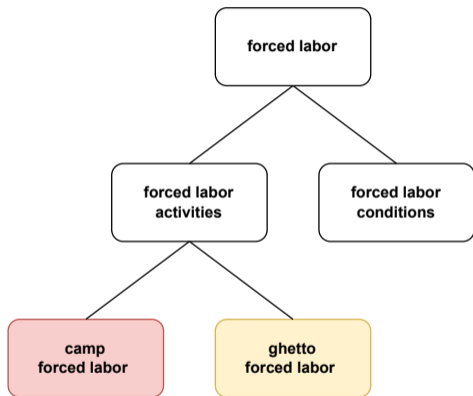
$(001 \dots 010 \dots 000)$
1063

Initial Experiments

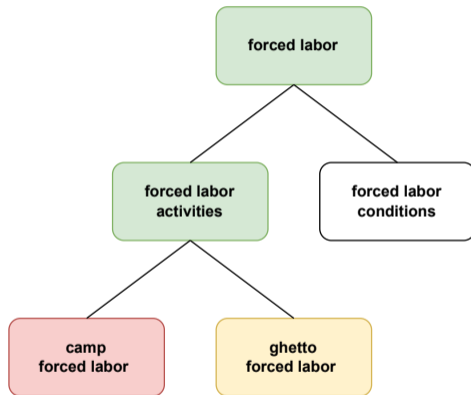
- 1-minute segments extended to 3 minutes (\sim avg. length of semantic segments)
- Unweighted and weighted class-average scores of 4 models:



Current Experiments: Hierarchical Scores

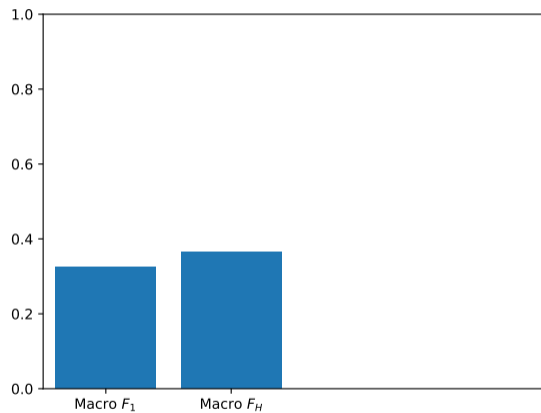


Current Experiments: Hierarchical Scores

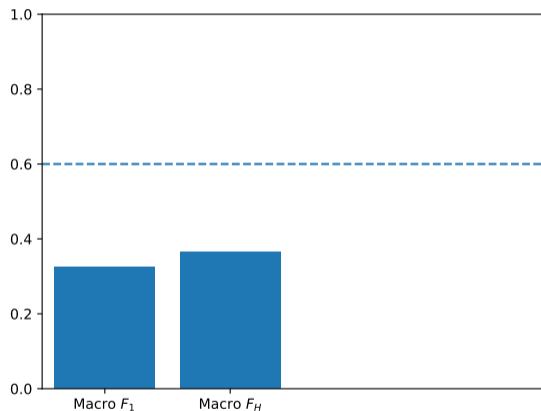


This increases the number of keywords by 30%

Current Experiments: Hierarchical Scores

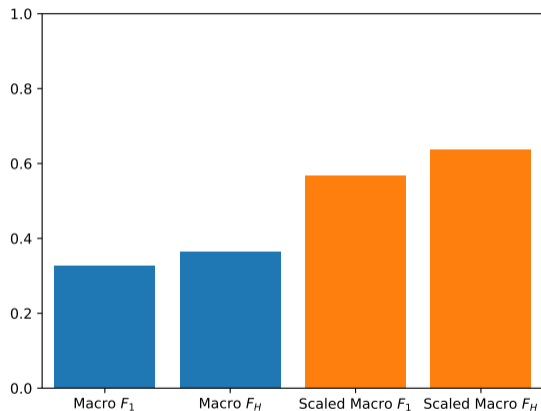


Current Experiments: Hierarchical Scores



⇒ The best possible F_1 score is ~ 0.6

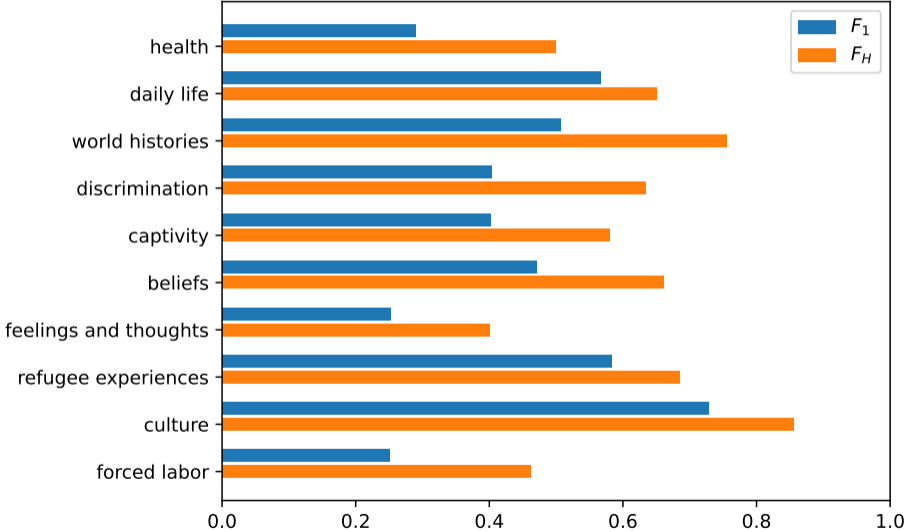
Current Experiments: Hierarchical Scores



⇒ The best possible F_1 score is ~ 0.6

⇒ Scale scores to "ignore" missing keywords

Current Experiments: Hierarchical Scores



Multilingual Experiments: DEGOB

There were about 12,000-13,000 Jews living in **Munkács**; tradesmen, physicians, lawyers, officials, craftsmen, etc. Most of them were wealthy people. **I lived with my parents**. I had a shop of my own. In the spring of 1944 I had a stock of textile fabrics worth about 30,000 pengős. In the beginning of **May 1944**, German soldiers and unknown policemen appeared in the city and they announced that **all Jews had to leave their homes** and move within 24 hours to the part of the city they had marked out. **That place was too small to house about 10,000 Jews. Ten of us lived in one room.** We were chased and hurried to move to the ghetto, although, we were allowed to take with us whatever we wanted. **There was a soup kitchen in the ghetto.** The **men's task was to pull down the fences** separating the houses in the ghetto and take apart the gates, in order that they could freely walk among the Jewish houses and rob them. **There were robberies in one house or another almost every night.** They took away all the valuables they could find and they left one person with not more than 10 pengős. One day, under the pretext that the men had not gone to work, they collected almost every man, they took them into two synagogues and **beat everybody terribly**. The wealthier Jews were dragged into the Kohner castle and **they were horribly beaten** in order to disclose where or with whom they had hidden their money and jewellery. **They also seriously beat up a shopkeeper called Honig, and then they shot him dead**, probably because he would have not been able to go back to the ghetto by himself anyway. We spent three weeks in the ghetto. Later we ran out of our own food and began to starve; we suffered much during the water shortage. **Approximately 10,000 Jews were crammed there in very little space and it was very hot there.** A lot of people tried to escape from the ghetto and many of them managed. From the ghetto we were taken to a brickyard; what had remained of our possessions we were allowed to take with us. We stayed there only for some days. We were entrained at the brickyard.

Multilingual Experiments: DEGOB

Top 10 predictions on a testimony from the National Committee for Attending Deportees:

English translation¹

ghetto living conditions
ghetto-related aid giving
sustenance provision
ghetto forced labor
camp forced labor
ghetto housing conditions
anti-Jewish roundups
beatings
Poland 1941 (June 21) - 1945 (May 7)
loved ones' fates

Hungarian original²

ghetto-related aid giving
sustenance provision
ghetto forced labor
loved ones' fates
beatings
ghetto living conditions
anti-Jewish measures
Poland 1939 (Sept 1) - 1945 (May 7)
anti-Jewish roundups
Hungary 1944

¹<http://degob.org/index.php?showjk=2>

²<http://degob.hu/index.php?showjk=2>

Multilingual Experiments: DEGOB

Top 10 predictions on a testimony from the National Committee for Attending Deportees:

English translation¹

ghetto living conditions

ghetto-related aid giving

sustenance provision

ghetto forced labor

camp forced labor

ghetto housing conditions

anti-Jewish roundups

beatings

Poland 1941 (June 21) - 1945 (May 7)

loved ones' fates

Hungarian original²

ghetto-related aid giving

sustenance provision

ghetto forced labor

loved ones' fates

beatings

ghetto living conditions

anti-Jewish measures

Poland 1939 (Sept 1) - 1945 (May 7)

anti-Jewish roundups

Hungary 1944

¹<http://degob.org/index.php?showjk=2>

²<http://degob.hu/index.php?showjk=2>

Multilingual Experiments: DEGOB

Top 10 predictions on a testimony from the National Committee for Attending Deportees:

English translation¹

ghetto living conditions

ghetto-related aid giving

sustenance provision

ghetto forced labor

camp forced labor

ghetto housing conditions

anti-Jewish roundups

beatings

Poland 1941 (June 21) - 1945 (May 7)

loved ones' fates

Hungarian original²

ghetto-related aid giving

sustenance provision

ghetto forced labor

loved ones' fates

beatings

ghetto living conditions

anti-Jewish measures

Poland 1939 (Sept 1) - 1945 (May 7)

anti-Jewish roundups

Hungary 1944

¹<http://degob.org/index.php?showjk=2>

²<http://degob.hu/index.php?showjk=2>

Multilingual Experiments: DEGOB

Top 10 predictions on a testimony from the National Committee for Attending Deportees:

English translation¹

ghetto living conditions

ghetto-related aid giving

sustenance provision

ghetto forced labor

camp forced labor

ghetto housing conditions

anti-Jewish roundups

beatings

Poland 1941 (June 21) - 1945 (May 7)

loved ones' fates

Hungarian original²

ghetto-related aid giving

sustenance provision

ghetto forced labor

loved ones' fates

beatings

ghetto living conditions

anti-Jewish measures

Poland 1939 (Sept 1) - 1945 (May 7)

anti-Jewish roundups

Hungary 1944

¹<http://degob.org/index.php?showjk=2>

²<http://degob.hu/index.php?showjk=2>

Multilingual Experiments: DEGOB

Top 10 predictions on a testimony from the National Committee for Attending Deportees:

English translation¹

ghetto living conditions
ghetto-related aid giving
sustenance provision
ghetto forced labor
camp forced labor
ghetto housing conditions
anti-Jewish roundups
beatings
Poland 1941 (June 21) - 1945 (May 7)
loved ones' fates

Hungarian original²

ghetto-related aid giving
sustenance provision
ghetto forced labor
loved ones' fates
beatings
ghetto living conditions
anti-Jewish measures
Poland 1939 (Sept 1) - 1945 (May 7)
anti-Jewish roundups
Hungary 1944

¹<http://degob.org/index.php?showjk=2>

²<http://degob.hu/index.php?showjk=2>

Multilingual Experiments: DEGOB

Top 10 predictions on a testimony from the National Committee for Attending Deportees:

English translation¹

ghetto living conditions

ghetto-related aid giving

sustenance provision

ghetto forced labor

camp forced labor

ghetto housing conditions

anti-Jewish roundups

beatings

Poland 1941 (June 21) - 1945 (May 7)

loved ones' fates

Hungarian original²

ghetto-related aid giving

sustenance provision

ghetto forced labor

loved ones' fates

beatings

ghetto living conditions

anti-Jewish measures

Poland 1939 (Sept 1) - 1945 (May 7)

anti-Jewish roundups

Hungary 1944

¹<http://degob.org/index.php?showjk=2>

²<http://degob.hu/index.php?showjk=2>

Multilingual Experiments: DEGOB

Top 10 predictions on a testimony from the National Committee for Attending Deportees:

English translation¹

ghetto living conditions

ghetto-related aid giving

sustenance provision

ghetto forced labor

camp forced labor

ghetto housing conditions

anti-Jewish roundups

beatings

Poland 1941 (June 21) - 1945 (May 7)

loved ones' fates

Hungarian original²

ghetto-related aid giving

sustenance provision

ghetto forced labor

loved ones' fates

beatings

ghetto living conditions

anti-Jewish measures

Poland 1939 (Sept 1) - 1945 (May 7)

anti-Jewish roundups

Hungary 1944

¹<http://degob.org/index.php?showjk=2>

²<http://degob.hu/index.php?showjk=2>

Multilingual Experiments: DEGOB

Top 10 predictions on a testimony from the National Committee for Attending Deportees:

English translation¹

ghetto living conditions
ghetto-related aid giving
sustenance provision
ghetto forced labor
camp forced labor
ghetto housing conditions
anti-Jewish roundups
beatings
Poland 1941 (June 21) - 1945 (May 7)
loved ones' fates

Hungarian original²

ghetto-related aid giving
sustenance provision
ghetto forced labor
loved ones' fates
beatings
ghetto living conditions
anti-Jewish measures
Poland 1939 (Sept 1) - 1945 (May 7)
anti-Jewish roundups
Hungary 1944

¹<http://degob.org/index.php?showjk=2>

²<http://degob.hu/index.php?showjk=2>

Multilingual Experiments: DEGOB

Top 10 predictions on a testimony from the National Committee for Attending Deportees:

English translation¹

ghetto living conditions

ghetto-related aid giving

sustenance provision

ghetto forced labor

camp forced labor

ghetto housing conditions

anti-Jewish roundups

beatings

Poland 1941 (June 21) - 1945 (May 7)

loved ones' fates

Hungarian original²

ghetto-related aid giving

sustenance provision

ghetto forced labor

loved ones' fates

beatings

ghetto living conditions

anti-Jewish measures

Poland 1939 (Sept 1) - 1945 (May 7)

anti-Jewish roundups

Hungary 1944

¹<http://degob.org/index.php?showjk=2>

²<http://degob.hu/index.php?showjk=2>

Multilingual Experiments: DEGOB

Top 10 predictions on a testimony from the National Committee for Attending Deportees:

English translation¹

ghetto living conditions

ghetto-related aid giving

sustenance provision

ghetto forced labor

camp forced labor

ghetto housing conditions

anti-Jewish roundups

beatings

Poland 1941 (June 21) - 1945 (May 7)

loved ones' fates

Hungarian original²

ghetto-related aid giving

sustenance provision

ghetto forced labor

loved ones' fates

beatings

ghetto living conditions

anti-Jewish measures

Poland 1939 (Sept 1) - 1945 (May 7)

anti-Jewish roundups

Hungary 1944

¹<http://degob.org/index.php?showjk=2>

²<http://degob.hu/index.php?showjk=2>

Multilingual Experiments: DEGOB

Top 10 predictions on a testimony from the National Committee for Attending Deportees:

English translation¹

ghetto living conditions
ghetto-related aid giving
sustenance provision
ghetto forced labor
camp forced labor
ghetto housing conditions
anti-Jewish roundups
beatings
Poland 1941 (June 21) - 1945 (May 7)
loved ones' fates

Hungarian original²

ghetto-related aid giving
sustenance provision
ghetto forced labor
loved ones' fates
beatings
ghetto living conditions
anti-Jewish measures
Poland 1939 (Sept 1) - 1945 (May 7)
anti-Jewish roundups
Hungary 1944

¹<http://degob.org/index.php?showjk=2>

²<http://degob.hu/index.php?showjk=2>

Multilingual Experiments: DEGOB

Top 10 predictions on a testimony from the National Committee for Attending Deportees:

English translation¹

ghetto living conditions

ghetto-related aid giving

sustenance provision

ghetto forced labor

camp forced labor

ghetto housing conditions

anti-Jewish roundups

beatings

Poland 1941 (June 21) - 1945 (May 7)

loved ones' fates

Hungarian original²

ghetto-related aid giving

sustenance provision

ghetto forced labor

loved ones' fates

beatings

ghetto living conditions

anti-Jewish measures

Poland 1939 (Sept 1) - 1945 (May 7)

anti-Jewish roundups

Hungary 1944

¹<http://degob.org/index.php?showjk=2>

²<http://degob.hu/index.php?showjk=2>

Multilingual Experiments: DEGOB

Top 10 predictions on a testimony from the National Committee for Attending Deportees:

English translation¹

ghetto living conditions
ghetto-related aid giving
sustenance provision
ghetto forced labor
camp forced labor
ghetto housing conditions
anti-Jewish roundups
beatings
Poland 1941 (June 21) - 1945 (May 7)
loved ones' fates

Hungarian original²

ghetto-related aid giving
sustenance provision
ghetto forced labor
loved ones' fates
beatings
ghetto living conditions
anti-Jewish measures
Poland 1939 (Sept 1) - 1945 (May 7)
anti-Jewish roundups
Hungary 1944

¹<http://degob.org/index.php?showjk=2>

²<http://degob.hu/index.php?showjk=2>

Future Experiments

- Thanks to the University of West Bohemia, we now have access to a lot more data
- ~5000 English testimonies have semantic segmentation and rich annotation
 - We can create even more via machine translation
- We can use these for 2 tasks:
 - Train a better semantic tagging model on higher-quality data
 - Train a text segmentation model to re-segment the remaining VHA

Future Experiments: Text Re-Segmentation

SEGMENTS Indexing Date: Jul 22, 2004

- SEGMENT 26
- SEGMENT 27
- SEGMENT 28**
- SEGMENT 29
- SEGMENT 30
- SEGMENT 31

Terms:

- Czechoslovakia 1939 ⓘ
- flight decisions ⓘ
- Olomouc (Moravia-Silesia, Czechoslovakia) ⓘ 📄

People:

- Berta Heimova - mothers ⓘ
- Hanus Heim - brothers ⓘ



SEGMENTS Indexing Date: Aug 29, 2005

Terms:

- Auschwitz II-Birkenau (Poland : Death Camp) ⓘ 📄
- camp adaptation methods ⓘ 📄
- camp children ⓘ 📄
- camp hospitals ⓘ 📄
- camp medical experiments ⓘ 📄
- health deceptions ⓘ 📄
- König, Hans-Wilhelm ⓘ 📄
- Mengele, Josef ⓘ 📄
- Poland 1941 (June 21) - 1945 (May 7) ⓘ 📄
- twins in the camps ⓘ 📄

SEGMENT 14

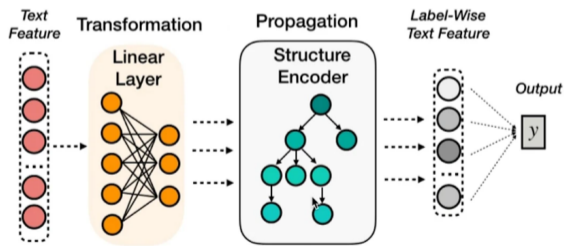
Terms:

Future Experiments: Hierarchy-Aware Models

Some of our problems will be solved by using more data, but:

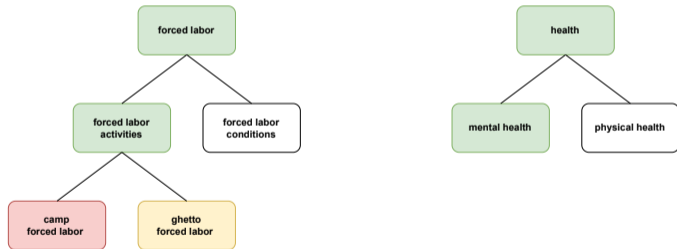
The current models consider all keywords to be independent

Hierarchy-Aware Global Model for Hierarchical Text Classification (Zhou et al., ACL 2020):



Future Experiments: Hierarchy-Aware Models

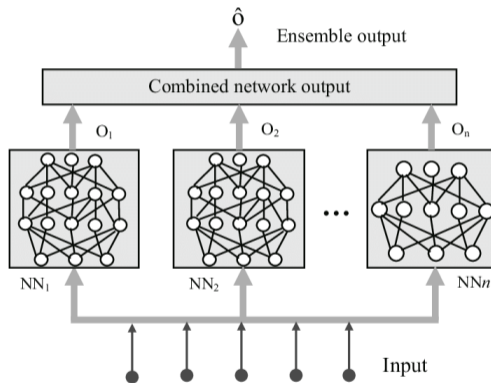
- Differences between specific keywords of the same topic are nuanced
- Our model must distinguish these nuances, but its weights are influenced by unrelated branches of the hierarchy



Idea: Train multiple more focused language models on distinct parts of the hierarchy

Future Experiments: Hierarchy-Aware Models

Topic-parallel ensemble learning: Train n models using n keyword masks



\Rightarrow each model only has to distinguish between smaller sets of related keywords

Summary

- 1 Goal: Tag documents related to Nazi persecution with concepts from the VHA ontology
- 2 First models are simple and trained on only $\sim 1\%$ of the VHA but already look promising
- 3 Future work: Text segmentation and hierarchy-aware semantic tagging models