Title: Identifying autopsy-reports in unstructured, digitized patient records:
A pilot-study from the Norwegian Health Archives Registry

Authors: Elin Bjørnstad-Tuveng[1] , Gina Helstad[1] , Ahmed Abouzeid[2], Ole-Christoffer Granmo[2], Kari Nytrøen[1]

Affiliations: [1]The Norwegian Health Archives Registry, [2]Anzyz Technologies AS

Introduction:
The Norwegian Health Archives Registry (NHAR) is the first of its kind to preserve a whole nation's patient health records (PHRs) from hospitals and establish a database for research purposes. PHRs are scanned, OCR-read and registered with a limited, pre-defined set of metadata, but in order to build statistics and search for a specific lab-result; a symptom and a prescribed medication, AI is needed. The main purpose of this pilot-study was to use AI-technology (natural language processing algorithm) in identifying PHRs which included autopsy-reports and to define the success rate based on manual control.

Method:
15.092 PHRs (sampled from a representative hospital trust) was drawn from NHAR. The data had words/phrases that often appear in autopsy-reports: f.ex. "post mortem rigidity". An AI-autopsy-concept was built based on Anzyz CCL™ Machine Learning Technology (1). Unlike typically sophisticated Machine Learning algorithms, the former provides fully transparent results (2, 3, 4). The technology involves interactive humans-in-the-loop for quality assurance. Figure 1 demonstrates how the Anzyz dashboard facilitates the retrieval of related documents when medical terms are queried. Moreover, the dashboard provides a semantic relatedness score to rank document relatedness (5) to the searched terms.

Results:
The purpose of the AI-autopsy-concept was to evaluate the capabilities of information retrieval in PHRs. Hence, it can be evaluated for its matching capabilities to identify relevant information based on a given query. An F-measure (6) performance metric is convenient for the evaluation (Figure 2). Among the 15,092 PHRs, the algorithm identified 669 records with autopsy-reports. Table 1 shows the result (F-measure) based on manual control of 124 of the AI-identified PHRs and 146 random samples drawn from the 15,092 PHRs. Of the 124, 86% were actual autopsy-reports. The 14% that were not actual autopsy-reports were false positives due to different types of standard forms in the PHRs. There were no false negatives.

Discussion:
Given the absence of false negatives in this pilot-study, the results demonstrate promising capabilities to investigate other medical concepts. We believe the Anzyz based-dashboards will be a valuable asset for NHAR in order to identify PHRs with the specific content that researchers request. For false positives, we believe it can be ruled out in the next project phase by adding more configurations. Typically, a recall metric indicates how good an AI-model is for not missing false negatives, while a precision indicates how good the model is for not getting false positives (Table 1) (7).

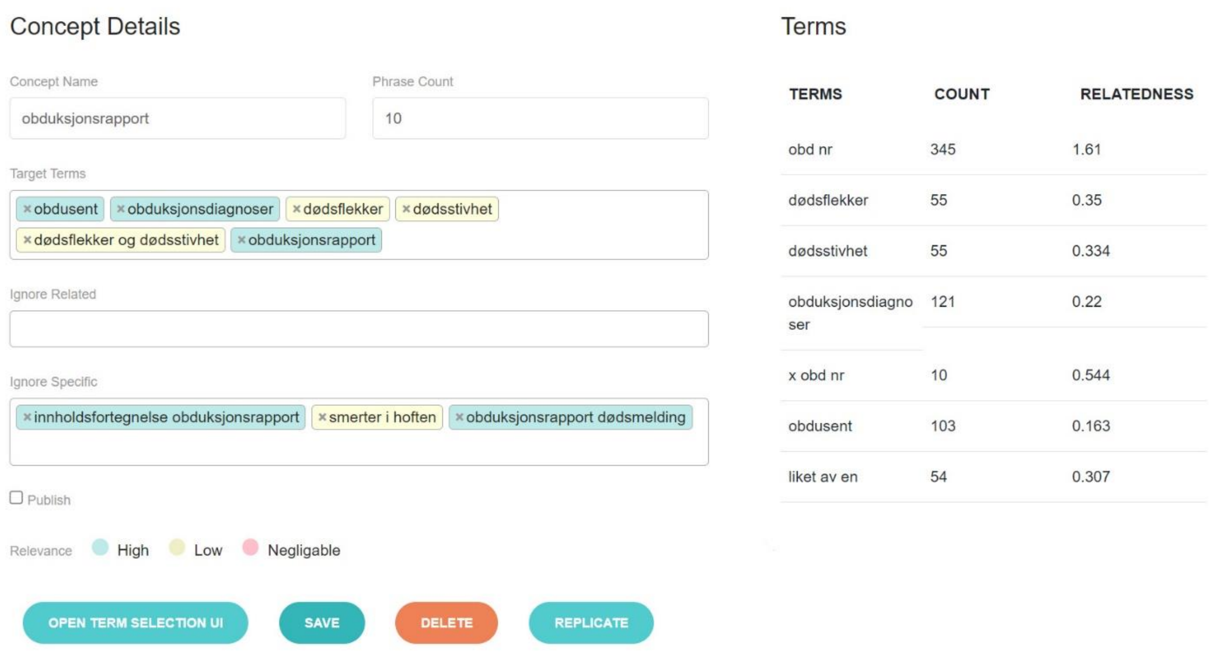

Figure 1: Example of an Anzyz dashboard: autopsy-report concept building and relatedness score

Type	***Recall (%)***	***Precision (%)***	***F_1-measure (%)***	***F_2-Measure (%)***
PHRs which included autopsy reports	*100*	*86*	*92*	*97*

Table 1: Success rate (F-measure) based on manual control of 124 of the 669 PHRs initially identified as having autopsy-reports (107 true positives, 17 false positives) and 146 random samples drawn from the 15,092 PHRs (139 true negatives, 0 false negatives)

$$F - measure \;= \frac{precision \; x \; recall}{precision \; + \; recall}$$

$$F_2 = \frac{(1 + 2_2) \; x \; precision \; x \; recall}{2_2 \; x \; precision + recall}$$

Figure 2: F-measure equation

References:

1. *Berge GT, Granmo O-C, Tveit TO. "Combining Unsupervised, Supervised, and Rule-Based Algorithms for Text Mining of Electronic Health Records-A Clinical Decision Support System for Identifying and Classifying Allergies of Concern for Anesthesia During Surgery".*
2. *Loyola-Gonzalez, Octavio. "Black-box vs. white-box: Understanding their advantages and weaknesses from a practical point of view." IEEE Access 7 (2019): 154096-154113.*
3. *Price, W. Nicholson. "Big data and black-box medical algorithms." Science translational medicine 10.471 (2018): eaao5333.*
4. *Abeyrathna, K. Darshana, Ole-Christoffer Granmo, and Morten Goodwin. "Extending the tsetlin machine with integer-weighted clauses for increased interpretability." IEEE Access 9 (2021): 8233-8248.*
5. *Desai, Digvijay, et al. "A Comparative Study of Information Retrieval Models for Short Document Summaries." Computer Networks and Inventive Communication Technologies. Springer, Singapore, 2022. 547-562.*
6. *Li, Yixuan, and Zixuan Chen. "Performance evaluation of machine learning methods for breast cancer prediction." Appl Comput Math 7.4 (2018): 212-216.*
7. *Hofer, Ira S., et al. "Integration of Feature Vectors from Raw Laboratory, Medication and Procedure Names Improves the Precision and Recall of Models to Predict Postoperative Mortality and Acute Kidney Injury." (2022).*