

Background

Allowing patients to access their own electronic health record (EHR) notes through online patient portals has the potential to improve patient-centered care. However, EHR notes contain abundant medical jargon that can be difficult for patients to comprehend. One way to help patients is to reduce information overload and help them focus on medical terms that matter most to them. Targeted education can then be developed to improve patient EHR comprehension and the quality of care. Objective: The aim of this work was to develop FIT (Finding Important Terms for patients), an unsupervised natural language processing (NLP) system that ranks medical terms in EHR notes based on their importance to patients.

Methods: We built FIT on a new unsupervised ensemble ranking model derived from the biased random walk algorithm to combine heterogeneous information resources for ranking candidate terms from each EHR note.
Specifically, FIT integrates four single views (rankers) for term importance: patient use of medical concepts, document-level term salience, word co-occurrence based term relatedness, and topic coherence.
It also incorporates partial information of term importance as conveyed by terms' unfamiliarity levels and semantic types. We evaluated FIT on 90 expert-annotated EHR notes and used the four single-view rankers as baselines. In addition, we implemented three benchmark unsupervised ensemble ranking methods as strong baselines.
Results: FIT achieved 0.885 AUC-ROC for ranking candidate terms from EHR notes to identify important terms. When including term identification, the performance of FIT for identifying important terms from EHR notes was 0.813 AUC-ROC. Both performance scores significantly exceeded the corresponding scores from the four single rankers ( $P<0.001$ ).

FIT also outperformed the three ensemble rankers for most metrics. Its performance is relatively insensitive to its parameter. Conclusions: FIT can automatically identify EHR terms important to patients. It may help develop future interventions to improve quality of care. By using unsupervised learning as well as a robust and flexible framework for information fusion, FIT can be readily applied to other domains and applications.