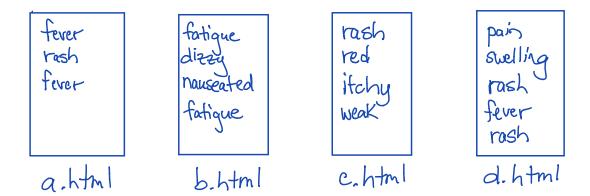
## C4M: Medical Document Retrieval Worksheet

Sample documents:



Term frequency / inverse document frequency:

$$TF(k,d) = \begin{cases} 0 & \text{if } k \text{ does not appear in } d \\ 1 & \text{otherwise} \end{cases}$$
 
$$IDF(k,D) = \log \frac{\text{No. of does in } D}{\text{No. of does in } D \text{ that contain } k}$$
 
$$TF\text{-}IDF(k,d,D) = TF(t,d) * IDF(k,D)$$

k represents a keyword, d represents a document, and D represents the set of all documents.

- 1. Query: nauseated
  - (a) IDF(k, D) =
  - (b) Complete the doc\_to\_score dictionary, where each key represents a document and each value represents that document's TF-IDF score.

- (c) Best match:
- 2. Query: fever rash
  - (a)  $IDF(k_1, D) =$

$$IDF(k_2, D) =$$

(b) Complete the doc\_to\_score dictionary, where each key represents a document and each value represents that document's TF-IDF score.

(c) Best match: