# INFO/CS 4300: Language and Information
# The Inverted Index Algorithm for Cosine Similarity

Name: _____

NetID: _____

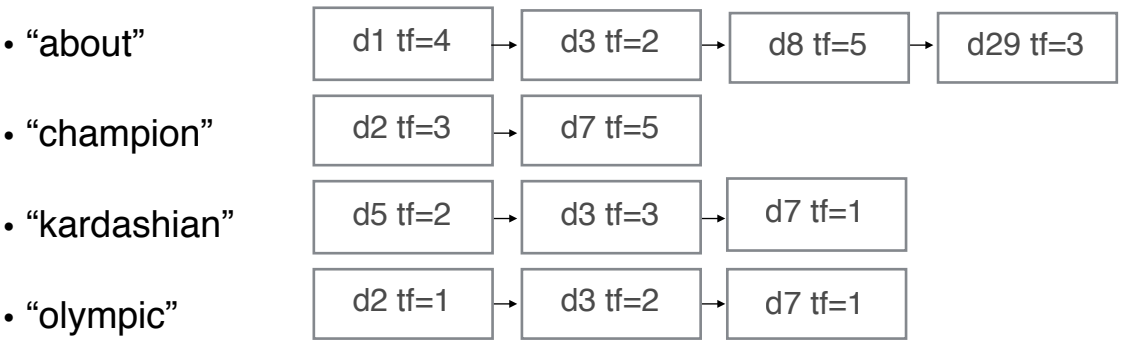
## Algorithm sketch

**Iterate** through all query terms $q_i$:
    for each term, **iterate** through its postings $d_j$:
        **update** the respective **accumulators** by $w_{iq} * w_{ij}$

**return** sorted documents by their final **accumulator** scores
(after normalizing each accumulator by the document norm)

Use: IDF(w) = N / DF(w), N = 300, DF(w) = n. docs that contain w
$w_{ij} = tf_{ij} * idf_j$
$w_{iq} = 1$
do not normalize by the document norm

## Inverted index




## IDF values

| Key | Value |
|---|---|
| "about" | |
| "champion" | |
| "kardashian" | |
| "olympic" | |

## Accumulators

| Key | Value |
|---|---|

## Final ranking