CS 1656 – Introduction to Data Science – Fall 2016

Prof. Alexandros Labrinidis – Department of Computer Science – University of Pittsburgh

02 – Information Retrieval – Aug 31, 2016

Assume the 6 documents:
- Document #1: The University of Pittsburgh is located in Pittsburgh.
- Document #2: Carnegie Mellon University is located in Pittsburgh.
- Document #3 Pittsburgh was voted most livable city. Steelers. Steelers!
- Document #4: The Steelers won over the Cleveland Browns.
- Document #5: The Pittsburgh Steelers have won 6 Super Bowls.
- Document #6: Cleveland is located in Ohio.

(Q1) Compute the **frequency F** (=num of occurrences) and the **term frequency TF** (=1 + $\log_2(F)$) of the keyword Steelers against each of the above documents.

	Doc #1	Doc #2	Doc #3	Doc #4	Doc #5	Doc #6
F(Steelers,j)						
TF(Steelers,j)						

(Q2) Compute the **inverse document frequency** (IDF) of the keyword Steelers against all of the above documents.

(Q3) Compute the **weight W** of the keyword Steelers against each of the above documents.

	Doc #1	Doc #2	Doc #3	Doc #4	Doc #5	Doc #6
w(Steelers,j)						

Cheat sheet:

$\log_2(1) = 0$, $\log_2(2) = 1$, $\log_2(4) = 2$