From Amit Singhal, "Modern Information Retrieval: A Brief Overview"
IEEE Data Engineering Bulletin, 2001

tf	is the term's frequency in document
qtf	is the term's frequency in query
N	is the total number of documents in the collection
df	is the number of documents that contain the term
dl	is the document length (in bytes), and
avdl	is the average document length

Okapi weighting based document score: [23] = Robertson, Walker, Beaulieu, OKAPI at TREC-7 (1999)

$$\sum_{t\in Q,D} ln\frac{N-df+0.5}{df+0.5}\cdot\frac{(k_1+1)tf}{(k_1(1-b)+b\frac{dl}{avdl})+tf}\cdot\frac{(k_3+1)qtf}{k_3+qtf}$$

this paren should be after k_1

k_1 (between 1.0–2.0), b (usually 0.75), and k_3 (between 0–1000) are constants.

Pivoted normalization weighting based document score: [30] = Singhal, Choi, Hindle, Lewis, Pereira, AT&T at TREC-7 (1999)

$$\sum_{t\in Q,D}\frac{1+ln(1+ln(tf))}{(1-s)+s\frac{dl}{avdl}}\cdot qtf\cdot ln\frac{N+1}{df}$$

s is a constant (usually 0.20).

Table 1: Modern Document Scoring Schemes

under the vector space model are often based on researchers' experience with systems and large scale experimentation. [26] In both models, three main factors come into play in the final term weight formulation. a) Term Frequency (or tf): Words that repeat multiple times in a document are considered salient. Term weights based on *tf* have been used in the vector space model since the 1960s. b) Document Frequency: Words that appear in many documents are considered common and are not very indicative of document content. A weighting method based on this, called inverse document frequency (or *idf*) weighting, was proposed by Sparck-Jones early 1970s. [15] And c) Document Length: When collections have documents of varying lengths, longer documents tend to score higher since they contain more words and word repetitions. This effect is usually compensated by normalizing for document lengths in the term weighting method. Before TREC, both the vector space model and the probabilistic models developed term weighting schemes which were shown to be effective on the small test collections available then. Inception of TREC provided IR researchers with very large and varied test collections allowing rapid development of effective weighting schemes.

Soon after first TREC, researchers at Cornell University realized that using raw *tf* of terms is non-optimal, and a dampened frequency (e.g., a logarithmic *tf* function) is a better weighting metric. [4] In subsequent years, an effective term weighting scheme was developed under a probabilistic model by Steve Robertson and his team at City University, London. [22] Motivated in part by Robertson's work, researchers at Cornell University developed better models of how document length should be factored into term weights. [29] At the end of this rapid advancement in term weighting, the field had two widely used weighting methods, one (often called *Okapi weighting*) from Robertson's work, and the second (often called *pivoted normalization weighting*) from the work done at Cornell University. Most research groups at TREC currently use some variant of these two weightings. Many studies have used the phrase *tf-idf weighting* to refer to any term weighting method that uses *tf* and *idf*, and do not differentiate between using a simple document scoring method (like $\sum_{t\in Q,D} tf\cdot ln\frac{N}{df}$) and a state-of-the-art scoring method (like the ones shown in Table 1). Many such studies claim that their proposed methods are far superior than *tf-idf weighting*, often a wrong conclusion based on the poor weighting formulation used.