SYMBOLS & NOTATION

$D_i = \{w_{i1}, w_{i2}, ..., w_{it}\}$ represents a document (or query) vector and w_{ik} is the weight of some feature T_k in document D_i. A weight of zero is used for features that are absent from a particular document, and positive weights characterize features actually assigned. t is the number of unique terms in the document.

f_{ik}	Occurrence frequency of T_k in D_i.	u_i	Number of unique terms in document D_i.
N	Collection size.	$avgu$	Average unique-term-length of documents.
n_k	Number of documents with feature T_k assigned.	b_i	Byte length of document D_i.
$\max(f_{ik})$	Maximum frequency of all T_k in D_i.	$avgb$	Average byte-length of documents.
s	*slope* in the context of pivoted document length normalization.	G	Global collection statistics.

SMART'S TERM-WEIGHTING TRIPLE NOTATION

$tf(f_{ik})$				$df(N, n_k)$				$g(G, D_i)$			
	b	1	Binary weight	**x**	**n**	1	Multiplier of 1, disregards the collections frequency	**x**	**n**	1	1, disregards length normalization factor
t	**n**	f_{ik}	raw term frequency	**f**		$\log\left(\frac{N}{n_k}\right)$	inverse collection frequency		**c**	$\sqrt{\sum_{k=1}^{t} w_{ik}^2}$	cosine normalization
	a	$0.5 + 0.5 \cdot \frac{f_{ik}}{\max(f_{ik})}$	augmented normalized term frequency (normalized to be in [0.5, 1])		**t**	$\log\left(\frac{N+1}{n_k}\right)$	inverse collection frequency		**u**	$1 - s + s \cdot \frac{u_i}{avgu}$	pivoted unique normalization
	l	$1 + \log(f_{ik})$	log	**p**		$\log\left(\frac{N - n_k}{n_k}\right)$	probabilistic inverse collection frequency		**b**	$1 - s + s \cdot \frac{b_i}{avgb}$	pivoted byte size normalization
	L	$\frac{1 + \log(f_{ik})}{1 + \log(avg(f_{ik}))}$	average term frequency based normalization								
	d	$1 + \log(1 + \log(f_{ik}))$	double logarithm								