CIS 3362 Reference Sheet

Letter	A	B	C	D	E	F	G	H	I	J	K	L	M
Number	0	1	2	3	4	5	6	7	8	9	10	11	12
Letter	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Number	13	14	15	16	17	18	19	20	21	22	23	24	25

Value	1	3	5	7	11	17	25
Inverse Mod 26	1	9	21	15	19	23	25

Frequency Information of each letter in regular English text

Let	A	B	C	D	E	F	G	H	I	J	K	L	M
Frq	.082	.015	.028	.043	.127	.022	.020	.061	.070	.002	.008	.040	.024
Let	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Frq	.067	.075	.019	.001	.060	.063	.091	.028	.010	.023	.001	.020	.001

Common Digrams and Trigrams, in order of frequency of appearance

Digrams: TH, HE, IN, ER, AN, RE, ED, ON, ES, ST, EN, AT, TO, NT, HA, ND, OU, EA, NG, AS, OR, TI, IS, ET, IT, AR, TE, SE, HI, and OF

Trigrams: THE, ING, AND, HER, ERE, ENT, THA, NTH, WAS, ETH, FOR and DTH

$$IC = \sum_{i=1}^{k} \frac{f_i(f_i - 1)}{n(n-1)}$$, (where f_i is the frequency of the ith letter, n is the total number of letters, and k is the alphabet size)

$$MIC = \sum_{i=1}^{k} \frac{f_i g_i}{nm}$$ (where f_i and g_i are the frequencies of the ith letter in both samples respectively, n and m are the total number of letters in both samples respectively, and k is the alphabet size)