

## Problem 1

### 'Cos

The word 'cos is a shortened form of the word because, the normalized form should be **because**.

### Shi'ite

The word Shi'ite should be normalized to **Shiite** based on the rationale that most people will not add the apostrophe while searching for it.

### cont'd

The word cont'd should be normalized to **continue**. The word is a shortened form of the word "continued", but we should normalize all forms of the word to present tense to catch as many results as possible during searching.

### Hawai'i

The word Hawai'i should be normalized to **Hawaii** under the same rationale that was used to normalize Shi'ite: most people will not type the apostrophe.

### O'Rourke

This should be normalized to **Orourke**. This follows from the same justification used for Hawai'i and Shi'ite: users are lazy.

### ain't

This word is a contraction of the phrase is not. Since that is the case, we should normalize the word into **is not**.

## Problem 2

## Problem 3

The query *"fools rush in"* returns documents 2, 4, and 7.

The query *"fools rush in" AND "angels fear to tread"* returns only document 4.

Problem 4

Problem 5

		p		a		r		i		s	
	0	1		2		3		4		5	
a	1	$\frac{1}{1}$	$\frac{1}{1}$	$\frac{1}{1}$	$\frac{1}{1}$	$\frac{1}{1}$	$\frac{1}{1}$	$\frac{1}{1}$	$\frac{1}{1}$	$\frac{1}{1}$	$\frac{1}{1}$
l	2	$\frac{1}{1}$	$\frac{1}{1}$	$\frac{1}{1}$	$\frac{1}{1}$	$\frac{1}{1}$	$\frac{1}{1}$	$\frac{1}{1}$	$\frac{1}{1}$	$\frac{1}{1}$	$\frac{1}{1}$
i	3	$\frac{1}{1}$	$\frac{1}{1}$	$\frac{1}{1}$	$\frac{1}{1}$	$\frac{1}{1}$	$\frac{1}{1}$	$\frac{1}{1}$	$\frac{1}{1}$	$\frac{1}{1}$	$\frac{1}{1}$
c	4	$\frac{1}{1}$	$\frac{1}{1}$	$\frac{1}{1}$	$\frac{1}{1}$	$\frac{1}{1}$	$\frac{1}{1}$	$\frac{1}{1}$	$\frac{1}{1}$	$\frac{1}{1}$	$\frac{1}{1}$
e	5	$\frac{1}{1}$	$\frac{1}{1}$	$\frac{1}{1}$	$\frac{1}{1}$	$\frac{1}{1}$	$\frac{1}{1}$	$\frac{1}{1}$	$\frac{1}{1}$	$\frac{1}{1}$	$\frac{1}{1}$

Problem 7