Homework Project 4

Given 04/20/2016, Due 05/18/2016

Implement a Bloom filter for 2,000,000 strings with an error rate of less than 3%, using only 2Mbyte of memory. To achieve this, you create eight bit arrays, each of 2,000,000 bits (that is, 250,000 `char`). For each of these, you select a random hash function h_i from a universal family. To insert a string s, you set the $h_i(S)$-th bit to one in the i-th bit array, for $i = 0, \ldots, 7$. To query whether a string q is contained in the set, you check whether $h_i(q)$ is one in the i-th bit array, for all i.

The structure must support the following operations

- `bf_t * create_bf()` creates a Bloom filter with the above specification.
- `void insert_bf(bf_t *b, char *s)` inserts the string `*s` into the Bloom filter `*b`.
- `int is_element(bf_t *b, char *q)` returns 1 if the string `*q` is accepted by the Bloom filter, and 0 else.