Пусть F - данный фильтр Блума

Definition: Область определения фильтра Блума

Через D_F обозначим область определения фильтра Блума F, т.е. множество значений, которые можно добавить в фильтр

Claim Принадлежность фильтру Блума

Пусть задан некоторый фильтр Блума $F(A)$ на множестве $A \subseteq D_F$
Введём обозначение: $\forall x : x \in F(A) \iff$ фильтр выдал ответ о принадлежности объекта x фильтру

Note

По построению фильтра Блума F:
$\forall A \subseteq D_F \, \forall x \, (x \in A \implies x \in F(A))$

Т.к. в условии задачи не сказано обратное, будем считать, что для всех фильтров Блума $F(A), A \subseteq D_F$ выбрано одинаковое количество хэш функций так, что в одном фильтре функции могут быть не равны, но для любых двух фильтров последовательность их хэш функций совпадает.

(если у каждого фильтра свои хэш функции, то в общем случае ответ на оба вопроса, очевидно, нет)

Иначе говоря, нам дан один фильтр Блума, реализации которого для различных множеств A будут различаться, но правила построения (т.е. длина массива и последовательность хэш функций) всегда одинаковые.

Claim Дополнительные обозначения

Для всех рассматриваемых фильтров Блума через n обозначим длину их битового массива, а через m обозначим количество хэш функций. Последовательность значений хэш функций по модулю n обозначим $(h_1, h_2, ..., h_m)$, то есть $\forall i \in \{1, 2, ..., m\} : (h_i : D_F \longrightarrow \{0, 1, ..., n-1\})$

Note

$\forall x \in D_F$ значения $h_1(x), h_2(x), ..., h_m(x)$ не обязательно попарно различны

Question 1

Верно ли, что $F(AB)$ будет выдавать положительные ответы о принадлежности объектов из множества $A \cap B$? Почему (нет)?
Докажем, что утверждение верно, т.е. ответ - да.

Proof:

1. По условию даны фильтр Блума $F(A)$, построенный на множестве $A \subseteq D_F$,
и фильтр Блума $F(B)$, построенный на множестве $B \subseteq D_F$.
Через $F(AB)$ обозначен фильтр с битовым массивом, полученным путём побитового И
над битовыми массивами фильтров $F(A)$ и $F(B)$
2. Пусть $x \in A \cap B$, тогда, т.к. в фильтрах $F(A)$ и $F(B)$ используются одинаковые хэш-функции,
то и в битовом массиве фильтра $F(A)$, и в битовом массиве фильтра $F(B)$ на позициях
$h_1(x), h_2(x), ..., h_m(x)$ будет стоять 1, тогда после применения побитового И к массивам
в получившемся битовом массиве на этих позициях будет стоять 1 $\implies x \in F(AB)$

$\blacksquare$

Question 2

Верно ли, что $F(AB)$ будет в точности соответствовать другому фильтру, который будет получен в результате последовательной вставки объектов из множества $A \cap B$? Почему (нет)?
Докажем, что утверждение неверно, т.е. ответ - нет, приведя контрпример:

1. Рассмотрим фильтр Блума, в котором $n = 3, m = 2$

Рассмотри 3 попарно различных объекта x, y, z, положим $D_F = \{x, y, z\}, A = \{x\}, B\{y\}$

тогда $A \cap B = \varnothing$

2. Определим 2 хэш функции (по модулю n):

$h_1 : D_F \longrightarrow \{0, 1, 2\}$

$h_1(x) = 0$

$h_1(y) = 1$

$h_1(z) = 1$

$h_2 : D_F \longrightarrow \{0, 1, 2\}$

$h_2(x) = 1$

$h_2(y) = 2$

$h_2(z) = 1$

3. Тогда битовый массив в $F(A)$ - это кортеж из 3 битов $(1, 1, 0)$

Битовый массив в $F(B)$ - это кортеж из 3 битов $(0, 1, 1)$

Битовый массив в $F(AB)$ - это кортеж из 3 битов $(1, 1, 0)\&(0, 1, 1) = (0, 1, 0)$

Битовый массив в $F(A \cap B) = F(\varnothing)$ - это кортеж из 3 битов $(0, 0, 0)$

Получилось, что в $F(AB)$ и $F(A \cap B)$ битовые массивы не совпадают.

В частности, $z \in F(AB) \land z \notin F(A \cap B)$