

## Задача А4. Совместимость фильтров Блума

Предположим, что по двум множествам объектов  $A$  и  $B$  создано два фильтра Блума одинакового размера —  $F(A)$  и  $F(B)$  соответственно. Путем выполнения поэлементного побитового И над фильтрами  $F(A)$  и  $F(B)$  был получен третий фильтр, который обозначен  $F(AB)$ . В рамках задания фильтр Блума можно рассматривать в виде одного битового вектора.

**1. Верно ли, что  $F(AB)$  будет выдавать положительные ответы о принадлежности объектов из множества  $A \cap B$ ? Почему (нет)?**

Да

При добавлении элемента  $x$  в Bloom фильтр устанавливаются биты по позициям  $h_1(x), h_2(x), \dots, h_k(x)$  (где  $k$  - число используемых хеш-функций). Если  $x \in A \cap B$ , то при построении фильтров  $F(A)$  и  $F(B)$  в соответствующих позициях устанавливаются единицы:

$$F(A)[h_i(x)] = 1 \text{ для всех } i = 1, \dots, k$$

$$F(B)[h_i(x)] = 1 \text{ для всех } i = 1, \dots, k$$

При выполнении побитового И фильтров:

$$F(AB)[h_i(x)] = F(A)[h_i(x)] \wedge F(B)[h_i(x)] = 1 \wedge 1 = 1.$$

Таким образом, при проверке элемента  $x$  в  $F(AB)$  все требуемые биты окажутся равными единице, и фильтр вернёт положительный ответ. Это соответствует фундаментальному свойству Bloom фильтров: отсутствие ложных отрицательных срабатываний для уже вставленных элементов.

**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)? Хеш-функции не меняются.**

Нет

При последовательной вставке объектов из  $A \cap B$  в пустой фильтр Bloom для каждого элемента вычисляются его хэш функции, и соответствующие биты устанавливаются в 1. То есть, итоговый фильтр содержит единицы только в тех битах, которые были затронуты при хэшировании элементов из  $A \cap B$ .

Однако при построении  $F(A)$  и  $F(B)$  устанавливаются биты не только от объектов из  $A \cap B$ , но и от объектов, принадлежащих  $A \setminus B$  или  $B \setminus A$ . Вполне возможно, что в обоих фильтрах некоторые биты окажутся равными 1 по разным причинам (то есть из-за разных объектов, не принадлежащих пересечению). Тогда при выполнении операции И получим:

$$F(AB)[j] = 1 \quad \text{если} \quad F(A)[j] = 1 \text{ и } F(B)[j] = 1$$

что может происходить и для битов, которые не были бы установлены при вставке только элементов из  $A \cap B$