

SET-5

Группа БПИ243, Тупицин Тимофей Романович

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 $F(AB)$.

Система оценки

Ответьте на следующие вопросы:

1. 3 балла Верно ли, что $F(AB)$ будет выдавать положительные ответы о принадлежности объектов из множества $A \cap B$? Почему (нет)?
2. 3 балла 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.*

1. Ответ: да.

Если элемент содержится в множестве A , то в фильтре блума по индексу хэша будет стоять 1. Аналогично с множеством B . Тогда при побитовом И, если в фильтре блума(A) была 1 и фильтре блума(B) была 1, то в фильтре блума(AB) будет стоять единица. Фильтр выдаст положительный ответ для всех элементов из $A \cap B$.

2. Ответ: нет.

В фильтре блума(AB) могут появиться лишние единицы, которых не было бы в фильтре, созданном только по элементам $A \cap B$. Например, какой-то бит стал равен 1 в фильтре блума(A) из-за элемента, которого нет в B . И этот же самый бит стал равен 1 в фильтре блума(B) из-за другого элемента, которого нет в A . При побитовом И этот бит сохранится в фильтре блума(AB), хотя ни один элемент из пересечения его не активировал. Из-за таких единиц фильтр блума будет чаще выдавать ложно-положительные ответы, поэтому он не будет в точности соответствовать фильтру по $A \cap B$.