Алгоритмы и структуры данных-2

SET 5. Задача A4.

Весна 2024. Клычков М. Д.

Пункт 1. Пусть n — количество хеш-функций обоих фильтров Блума. Пусть хеш-функции $H = h_1, h_2, \dots, h_n \in \mathcal{H}$ для фильтров Блума $F(A)$ и $F(B)$ одинаковы (при различных функциях ответ на поставленный вопрос очевиден). Пусть A, B — битовые массивы, стоящие за фильтрами $F(A), F(B)$ соответственно, $A_j, B_j, j = \overline{1, m}$ — биты, где m — количество битов в фильтрах.

Докажем, что $F(AB), (AB)_j = A_j \& B_j$ будет выдавать положительные ответы о принадлежности объектов из множества $A \cap B$.

Доказательство. Предположим обратное: $\exists x \in A \cap B \colon F(AB) = 0 \Leftrightarrow \exists j \in [1, m] \colon (AB)_j = 0 \Leftrightarrow A_j \& B_j = 0$. Тогда $A_j = 0 \lor B_j = 0 \Leftrightarrow x \notin A \lor x \notin B \Leftrightarrow x \notin A \cap B$. Получили противоречие. $\square$

Пункт 2. Покажем, что $F(AB)$ не обязательно в точности соответствует другому фильтру, который будет получен в результате последовательной вставки объектов из множества $A \cap B$ (обозначим такой фильтр за $F'(AB)$, а его битовый массив за $(AB)'$).

Доказательство. Зафиксируем два множества $A = \{e_A, e_1, e_2, e_3, \dots\}$ и $B = \{e_B, e_1, e_2, e_3, \dots\}$, здесь
$\{e_1, e_2, e_3, \dots\} \in A \cap B$, причем элементы подберем таким образом, чтобы выполнялись дополнительные условия (оставим вне рассмотрения причину, почему такие элементы вообще существуют, очевидно):

$$\forall h \in H \forall e_1, e_2, e_3, \dots \colon (h(e_i))_{j_0} = 0$$
$$\exists h \in H \colon (h(e_A))_{j_0} = 1$$
$$\exists h \in H \colon (h(e_B))_{j_0} = 1$$

Другими словами, элементы, принадлежащие пересечению множеств $A \cap B$ не «вносят вклад» в бит j_0, но находятся элементы $e_A, e_B \notin A \cap B$ такие, что «включают» бит j_0.

Из фильтров Блума $F(A)$ и $F(B)$, получим фильтр $F(AB)$, в нем бит $(AB)_{j_0} = 1$.

Построим новый фильтр Блума $F'(AB)$, последовательно добавляя элементы $e_1, e_2, e_3, \dots \in A \cap B$. Но тогда в таком фильтре Блума $(AB)'_{j_0} = 0$ согласно условию выше, поэтому получаем, что $(AB) \neq (AB)'$ как массивы одинакового размера, то есть $F(AB) \neq F'(AB)$ $\square$