# Лекция 2.

Критерий допустимости информационных графов.
Критерий полноты базового множества
для типа задач информационного поиска.

## 1 Критерий допустимости информационных графов

Пусть нам дана ЗИП $I = \langle X, V, \rho \rangle$.

Скажем, что ИГ $U$ *решает* ЗИП $I = \langle X, V, \rho \rangle$, если для любого запроса $x \in X$ ответ на этот запрос содержит все те и только те записи из $V$, которые удовлетворяют запросу $x$, то есть

$$\mathcal{J}_U(x) = \{y \in V : x\rho y\}.$$

ИГ $U$, решающий ЗИП $I$, будем также называть *допустимым* для задачи $I$.

Пусть $U$ — некоторый ИГ, $y$ — запись из $Y$. Через $L_U(y)$ обозначим множество листьев ИГ $U$, которым соответствует запись $y$.

Справедлива следующая теорема.

**Теорема 1.** *ИГ $U$ решает ЗИП $I = \langle X, V, \rho \rangle$ тогда и только тогда, когда для любой записи $y \in V$, такой, что $O(y, \rho) \neq \emptyset$, справедливо $L_U(y) \neq \emptyset$ и $\bigvee\limits_{\alpha \in L_U(y)} \varphi_\alpha = \chi_{y,\rho}$, а для любой записи $y \in V$, такой, что $O(y, \rho) = \emptyset$, справедливо либо $L_U(y) = \emptyset$, либо $\bigvee\limits_{\alpha \in L_U(y)} \varphi_\alpha \equiv 0$.*

*Доказательство. Достаточность.*

Возьмем произвольный запрос $x \in X$.

Возьмем произвольную запись $y \in V$.

Если $O(y, \rho) = \emptyset$ и, следовательно, $x \notin O(y, \rho)$, то по предположению либо $L_U(y) = \emptyset$, либо $\bigvee\limits_{\alpha \in L_U(y)} \varphi_\alpha \equiv 0$. Откуда следует, что $y \notin \mathcal{J}_U(x)$.

Теперь рассмотрим случай, когда $O(y, \rho) \neq \emptyset$.

Если $x\rho y$, то $\chi_{y,\rho}(x) = 1$, и согласно предположению

$$\bigvee_{\alpha \in L_U(y)} \varphi_\alpha(x) = 1.$$

Откуда следует, что существует лист $\alpha \in \mathcal{L}(U)$ такой, что $\varphi_\alpha(x) = 1$. Следовательно, $y \in \mathcal{J}_U(x)$.

Если $x \notin O(y, \rho)$, то $\chi_{y,\rho}(x) = 0$, и согласно предположению

$$\bigvee_{\alpha \in L_U(y)} \varphi_\alpha(x) = 0.$$

Откуда следует, что $y \notin \mathcal{J}_U(x)$.

Таким образом, мы показали, что

$$\mathcal{J}_U(x) = \{y \in V : x\rho y\}$$

и тем самым доказали достаточность.

*Необходимость.*

Возьмем произвольную запись $y \in V$.

Если $O(y, \rho) = \emptyset$ и, следовательно, для любого запроса $x \in X$ $x \notin O(y, \rho)$, значит, для любого $x \in X$ $y \notin \mathcal{J}_U(x)$. Откуда следует, что либо $L_U(y) = \emptyset$, либо $\bigvee\limits_{\alpha \in L_U(y)} \varphi_\alpha \equiv 0$.

Теперь рассмотрим случай, когда $O(y, \rho) \neq \emptyset$.

Предположим, что для данной записи $y$ не выполняются предположения теоремы, то есть существует такой запрос $x$, что

$$\bigvee_{\alpha \in L_U(y)} \varphi_\alpha(x) \neq \chi_{y,\rho}(x).$$

Но это означает, что $y$ принадлежит в точности одному из множеств $\mathcal{J}_U(x)$ или $\{y \in V : x\rho y\}$.

Следовательно, при этом $x$

$$\mathcal{J}_U(x) \neq \{y \in V : x\rho y\},$$

и, значит, ИГ $U$ не решает ЗИП $I$.

Тем самым теорема доказана. □

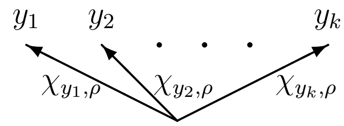


Рис. 1: Информационный граф переборного алгоритма

По сути теорема 1 говорит, что если нам дана ЗИП $I = \langle X, V, \rho \rangle$, и мы хотим построить ИГ, решающий эту ЗИП, мы должны построить многополюсник, который между корнем и остальными полюсами реализует как функции проводимости все функции $\chi_{y,\rho}$, где $y \in V$.

# 2 Критерий полноты базового множества для типа задач информационного поиска

Если нам дана ЗИП $I = \langle X, V, \rho \rangle$ и базовое множество $\mathcal{F} = \langle F, G \rangle$, то возникает вопрос, а можно ли построить информационный граф над базовым множеством $\mathcal{F}$, решающий эту задачу $I$? Если для любой записи $y_i \in V = \{y_1, \dots, y_k\}$ $\chi_{y_i,\rho} \in F$, то ответ на этот вопрос положительный, и граф, изображенный на рисунке 1, решает задачу $I$.

В данном разделе дается более полный ответ на этот вопрос.

Пусть нам дан тип $S = \langle X, Y, \rho \rangle$, где $X$ — множество запросов, $Y$ — множество записей, $\rho$ — отношение поиска, заданное на $X \times Y$, и базовое множество $\mathcal{F} = \langle F, G \rangle$.

Скажем, что базовое множество $\mathcal{F}$ *полно* для типа $S = \langle X, Y, \rho \rangle$, если для любой ЗИП $I = \langle X, V, \rho \rangle$ типа $S$ существует ИГ $U$ над базовым множеством $\mathcal{F}$, решающий ЗИП $I$.

Справедлив следующий результат, относящийся к проблеме полноты для ИГ.

**Теорема 2.** *Пусть заданы множества запросов $X$, записей $Y$, отношение поиска $\rho$ на $X \times Y$ и базовое множество $\mathcal{F} = \langle F, G \rangle$, такое, что предикат тождественная 1 принадлежит множеству $F$. Тогда $\mathcal{F}$ будет полным для типа $S = \langle X, Y, \rho \rangle$ тогда и только тогда, когда для любой записи $y \in Y$ такой, что $O(y, \rho) \neq \emptyset$, функцию $\chi_{y,\rho}(x)$ можно представить формулой вида*

$$\chi_{y,\rho}(x) = \bigvee_{i=1}^{n} \bigwedge_{j=1}^{m_i} f_{ij}(x),$$

*где* $f_{ij} \in F \cup \widehat{G}$.

*Доказательство. Достаточность.*

Пусть нам дана произвольная ЗИП $I = \langle X, V, \rho \rangle$, где $V = \{y_1, \ldots, y_k\} \subseteq Y$. По предположению каждую из функций $\chi_{y_l,\rho}(x) \quad (l = \overline{1,k})$ можно представить формулой

$$\chi_{y_l,\rho}(x) = \bigvee_{i=1}^{n_l} \bigwedge_{j=1}^{m_{li}} f_{lij}(x),$$

где $f_{lij} \in F \cup \widehat{G},\ l = \overline{1,k}$.

Без ограничения общности можно считать, что каждая конъюнкция $\bigwedge_{j=1}^{m_{li}} f_{lij}(x)$ содержит предикат из $F$, причем этот предикат стоит первым в конъюнкции. В противном случае мы всегда можем добавить предикат тождественная 1.

ИГ, решающий ЗИП $I$, будем строить следующим образом.

Сначала возьмем $k+1$ вершину и объявим одну из них корнем, а остальные объявим листьями и мысленно перенумеруем, начиная с 1 до $k$. Затем для каждого $l \in \{\overline{1,k}\}$ проделаем следующее.

Припишем $l$-му листу (обозначим его $\alpha_l$) запись $y_l$.

Если $O(y_l, \rho) \neq \emptyset$, то проведем из корня в лист $\alpha_l$ $n_l$ ориентированных цепей, причем $i$-я цепь $(i = \overline{1, n_l})$ будет состоять из $m_{li}$ ребер. Теперь для каждого $i$ $(i \in \{\overline{1, n_l}\})$ проделаем следующее. Если $f_{lij} \in F$, то $j$-е ребро $i$-й цепи объявим предикатным и припишем ему предикат $f_{lij}$. Если $f_{lij} \in \widehat{G}$ (то есть $f_{lij} = \xi_g^n$, где $g \in G$, а $n \in \mathbf{N}$), то вершину $\beta$, из которой исходит $j$-е ребро $i$-й цепи, объявим переключательной припишем ей переключатель $g$, $j$-му ребру $i$-й цепи припишем число $n$, из вершины $\beta$ выпустим еще $r-1$ ребро, где $r$ — мощность области значений переключателя $g$, и сопоставим им взаимно однозначно числа из множества $\{\overline{1,r}\} \setminus \{n\}$.

Нетрудно видеть, что в этом случае

$$\varphi_{\alpha_l}(x) = \bigvee_{i=1}^{n_l} \bigwedge_{j=1}^{m_{li}} f_{lij}(x) = \chi_{y_l,\rho}(x).$$

Поскольку это условие выполняется для всех листьев, то согласно теореме 1 построенный ИГ решает ЗИП $I$.

*Необходимость.*

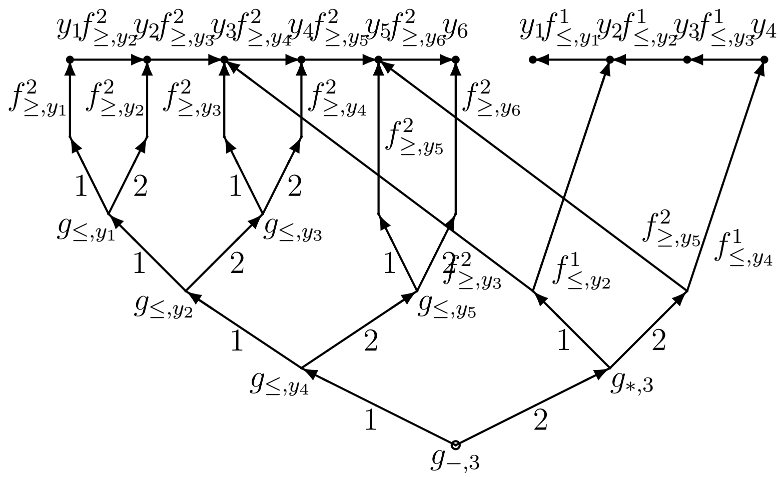


Рис. 2:

Пусть $\mathcal{F}$ полно для отношения $\rho$. Возьмем произвольную запись $y \in Y$ такую, что $O(y,\rho) \neq \emptyset$. Пусть $V$ — любая библиотека, содержащая запись $y$.

Так как $\mathcal{F}$ полно, то существует ИГ $U$, решающий ЗИП $I = \langle X, V, \rho \rangle$. Рассмотрим множество $L_U(y)$ листьев ИГ $U$, которым соответствует запись $y$. Так как $U$ решает задачу $I$, то согласно теореме 1

$$\bigvee_{\alpha \in L_U(y)} \varphi_\alpha(x) \equiv \chi_{y,\rho}(x).$$

Поскольку каждая из функций $\varphi_\alpha(x)$ есть функция проводимости от корня к листу $\alpha$, а функция проводимости по определению есть дизъюнкция конъюнкций некоторых предикатов из $F \cup \widehat{G}$, то необходимость доказана, что и доказывает теорему. $\square$

# Упражнения

**1.** Пусть $S = \langle X, X, = \rangle$ — тип поиска идентичных объектов, множество предикатов $F$ задается соотношением (1),

$$F = \{f_{=,a}(x) = \begin{cases} 0, & \text{если } x \neq a \\ 1, & \text{если } x = a \end{cases} : a \in X\}, \tag{1}$$

базовое множество имеет вид $\mathcal{F} = \langle F, \emptyset \rangle$, $V = \{y_1, y_2, \ldots, y_k\} \subseteq X$. Приведите пример информационного графа над базовым множеством $\mathcal{F}$, решающего ЗИП

$I = \langle X, V, = \rangle$.

**2.** Пусть $S = \langle X, X, = \rangle$ — тип поиска идентичных объектов, множество переключателей имеет вид

$$G = \{g_{=,a}(x) = \begin{cases} 1, & \text{если } x = a \\ 2, & \text{если } x \neq a \end{cases} : a \in X\}, \tag{2}$$

базовое множество имеет вид $\mathcal{F} = \langle \emptyset, G \rangle$, $V = \{y_1, y_2, \ldots, y_k\} \subseteq X$. Приведите пример информационного графа над базовым множеством $\mathcal{F}$, решающего ЗИП $I = \langle X, V, = \rangle$.

**3.** Пусть $X = \{1, 2, ..., N\}$, $S = \langle X, X, = \rangle$ — тип поиска идентичных объектов, множество переключателей имеет вид

$$G = \{g_a(x) = \begin{cases} 1, & \text{если } x < a \\ 2, & \text{если } x = a \\ 3, & \text{если } x > a \end{cases} : a \in X\}, \tag{3}$$

базовое множество имеет вид $\mathcal{F} = \langle \emptyset, G \rangle$, $V = \{3, 5, 7, 11, 13, 17, 19\}$. Постройте информационный граф над базовым множеством $\mathcal{F}$, решающий ЗИП $I = \langle X, V, = \rangle$.

**4.** Пусть $X = \{1, 2, ..., N\}$, $S = \langle X, X, = \rangle$ — тип поиска идентичных объектов, $V = \{y_1, y_2, \ldots, y_k\} \subseteq X$. Предположим, что $y_1 < y_2 < \cdots < y_k$. Метод блочного поиска с размером блока $m$, решающий задачу $I = \langle X, V, = \rangle$, состоит в следующем. Если на вход алгоритма поиска подается запрос $x \in X$, то, начиная с $i = 1$ до $i = k/m$, просматриваем записи $y_{i \cdot m}$. Если $x > y_{i \cdot m}$, то увеличиваем $i$ на 1, иначе по очереди просматриваем записи $y_{(i-1)m+1}, y_{(i-1)m+2}, \ldots, y_{i \cdot m}$ и сравниваем их с запросом $x$. При равенстве мы нашли нужную запись, если же ни для какой записи равенства не наблюдается, то ответ на запрос $x$ пуст. Опишите базовое множество и постройте информационный граф над этим базовым множеством, который бы решал ЗИП $I = \langle X, V, = \rangle$ методом блочного поиска.

**5.** Пусть $X = \{1, 2, ..., N\}$, $V \subseteq X$, $\rho_c$ — отношение поиска, задаваемое на $X \times V$ и определяемое соотношением

$$x \rho_c y \Longleftrightarrow (y \in V) \& (x \leq y) \& (\neg (\exists y')((y' \in V) \& (x \leq y') \& (y' < y))), \tag{4}$$

т.е. $x \rho_c y$, если $y \in V$, ближайшее справа к $x$. При выполнении этих условий ЗИП $I = \langle X, V, \rho_c \rangle$ называется задачей о близости. Пусть базовое множество имеет вид $\mathcal{F} = \langle \emptyset, G \rangle$, где множество переключателей $G$ задается соотношением (5)

$$G = \{g_{\leq,a}(x) = \begin{cases} 1, & \text{если } x \leq a \\ 2, & \text{если } x > a \end{cases} : a \in X\}. \tag{5}$$

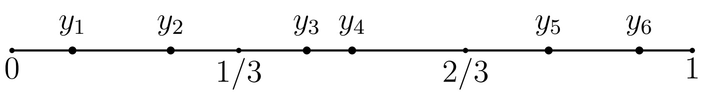


Рис. 3:

Постройте информационный граф над базовым множеством $\mathcal{F}$, решающий ЗИП $I = \langle X, V, \rho_c \rangle$, если $V = \{3, 5, 7, 11, 13, 17, 19\}$.

**6.** Одномерная задача о доминировании задается типом $S_{dom1} = \langle [0,1], [0,1], \geq \rangle$. Пусть $V = \{y_1, y_2, \dots, y_k\} \subseteq [0,1]$. Опишите некоторое базовое множество и постройте какой-либо информационный граф над этим базовым множеством, который бы решал ЗИП $I = \langle [0,1], V, \geq \rangle$.

**7.** Пусть $S_{int} = \langle X_{int}, Y_{int}, \rho_{int} \rangle$ — тип одномерного интервального поиска, где $X_{int} = \{(u,v) : 0 \leq u \leq v \leq 1\}$, $Y_{int} = [0,1]$, отношение $\rho_{int}$ определяется соотношением (6)

$$(u,v)\rho_{int} y \Longleftrightarrow u \leq y \leq v, \tag{6}$$

где $(u,v) \in X_{int},\ y \in Y_{int}$, $V = \{y_1, y_2, \dots, y_6\}$, где $y_1 = 1/6$, $y_2 = 1/4$, $y_3 = 3/8$, $y_4 = 2/5$, $y_5 = 3/4$, $y_6 = 7/8$. Пусть

$$f^1_{\leq,a}(u,v) = \begin{cases} 1, & \text{если } u \leq a \\ 0, & \text{если } u > a \end{cases}, \tag{7}$$

$$f^2_{\geq,a}(u,v) = \begin{cases} 1, & \text{если } v \geq a \\ 0, & \text{если } v < a \end{cases}, \tag{8}$$

$$g_{\cdot,m}(u,v) = \max(1, ]u \cdot m[), \tag{9}$$

$$g_{-,m}(u,v) = \begin{cases} 1, & \text{если } v - u < 1/m \\ 2, & \text{если } v - u \geq 1/m \end{cases}, \tag{10}$$

$$g_{\leq,a}(u,v) = \begin{cases} 1, & \text{если } u \leq a \\ 2, & \text{если } u > a \end{cases}. \tag{11}$$

Решает ли информационный граф, изображенный на рисунке 2, где функции определяются соотношениями (7)–(11), задачу информационного поиска $I = \langle X_{int}, V, \rho_{int} \rangle$? Обоснуйте ответ.

**8.** Докажите, что информационный граф, изображенный на рисунке 4, решает одномерную задачу интервального поиска $I = \langle X_{int}, V, \rho_{int} \rangle$, где $V = \{y_1, y_2, y_3, y_4, y_5, y_6\}$ — библиотека, изображенная на рисунке 3.

**9.** Пусть $S_{int} = \langle X_{int}, Y_{int}, \rho_{int} \rangle$ — тип одномерного интервального поиска, где отношение $\rho_{int}$ определяется соотношением (6), $V = \{1/8, 1/7, 1/5, 3/7,$

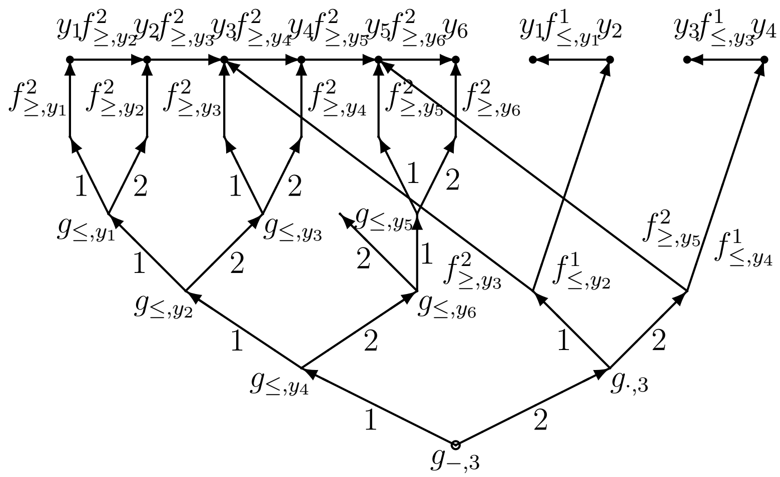


Рис. 4: Решение одномерной задачи интервального поиска

$3/5, 4/5, 7/8\}$. Опишите некоторое базовое множество и постройте какой-либо информационный граф над этим базовым множеством, который бы решал ЗИП $I = \langle X_{int}, V, \rho_{int} \rangle$.

**10.** Пусть $S = \langle X, X, = \rangle$ — тип поиска идентичных объектов, базовое множество имеет вид $\mathcal{F} = \langle \emptyset, G \rangle$, где множество переключателей $G$ задается соотношением (5). Будет ли полно базовое множество $\mathcal{F}$ для типа $S$?

**11.** Включающий поиск описывается типом $S_{bool} = \langle B^n, B^n, \overset{b}{\succeq} \rangle$, где $B^n$ — $n$-мерный булев куб, $\overset{b}{\succeq}$ — отношение поиска на $B^n \times B^n$, определяемое следующим соотношением

$$(x_1, \ldots, x_n) \overset{b}{\succeq} (y_1, \ldots, y_n) \Longleftrightarrow x_i \geq y_i, \ i = \overline{1, n}. \tag{12}$$

Приведите пример базового множества, полного для типа $S_{bool}$. Приведите пример минимального по мощности базового множества, полного для типа $S_{bool}$.