

Поиск и сортировка информации в базах данных

Что нужно знать:

- при составлении условия отбора можно использовать знаки отношений $<$, \leq (меньше или равно), $>$, \geq (больше или равно), $=$ (равно), \neq (не равно)
- последовательность выполнения логических операций в сложных запросах: сначала выполняются отношения, затем – «И», потом – «ИЛИ»
- для изменения порядка выполнения операции используют скобки
- реляционные базы данных обычно хранятся в памяти компьютера в виде нескольких связанных таблиц
- столбцы таблицы называются **полями**, а строки – **записями**

- каждая таблица содержит описание одного типа объектов (человека, бригады, самолета) или одного типа связей между объектами (например, связь между автомобилем и его владельцем)
- в каждой таблице есть **ключ** – некоторое значение (это может быть одно поле или комбинация полей), которое отличает одну запись от другой; в таблице не может быть двух записей с одинаковыми значениями ключа
- на практике часто используют суррогатные ключи – искусственно введенное числовое поле (обычно оно называется **идентификатор, ID**)
- таблицы связываются с помощью ключей; чаще всего используется связь 1:N (или 1:∞), когда одной записи в первой таблице может соответствовать много записей во второй таблице, но не наоборот; например:

Компании

ID	Название	Телефон
14	Альфа	271-34-98
23	Бета	275-12-34
24	Гамма	220-45-32

ID	Товар	ID_компании
18	Бумага	23
45	Бензин	14
28	Канцелярия	23
64	Корм для кошек	24

Пример задания:

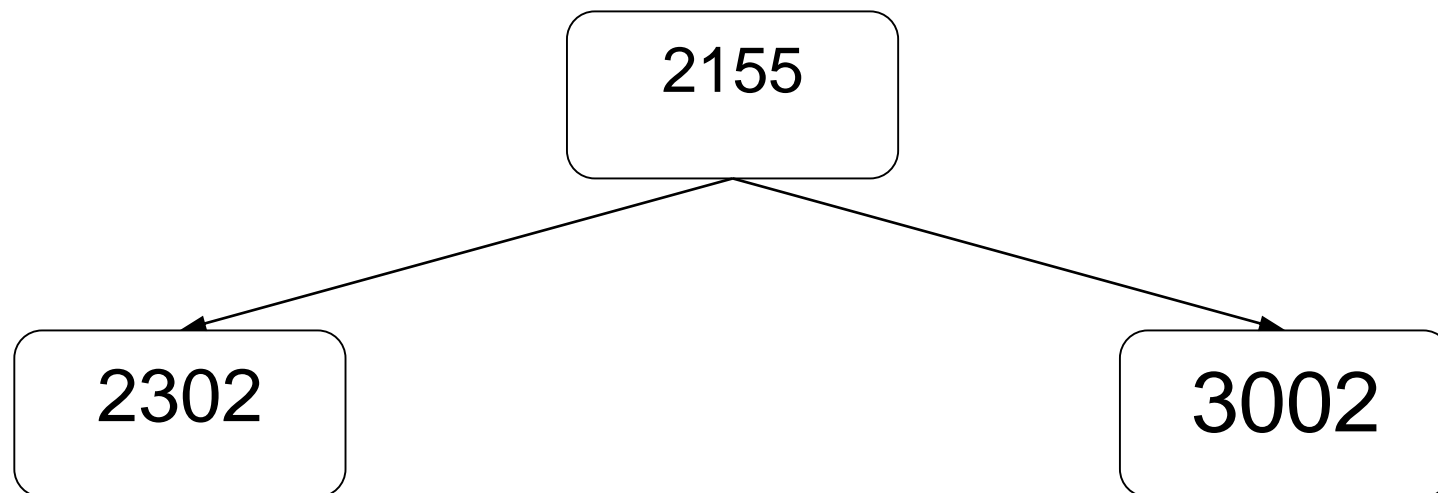
- **P-06.** В фрагменте базы данных представлены сведения о родственных отношениях. На основании приведённых данных определите, сколько прямых потомков (т.е. детей и внуков) Павленко А.К. упомянуты в таблице 1.

ID	Фамилия_И.О.	Пол
2146	Кривич Л.П.	Ж
2155	Павленко А.К.	М
2431	Хитрук П.А.	М
2480	Кривич А.А.	М
2302	Павленко Е.А.	Ж
2500	Сокол Н.А.	Ж
3002	Павленко И.А.	М
2523	Павленко Т.Х.	Ж
2529	Хитрук А.П.	М
2570	Павленко П.И.	М
2586	Павленко Т.И.	Ж
2933	Симонян А.А.	Ж
2511	Сокол В.А.	Ж
3193	Биба С.А.	Ж

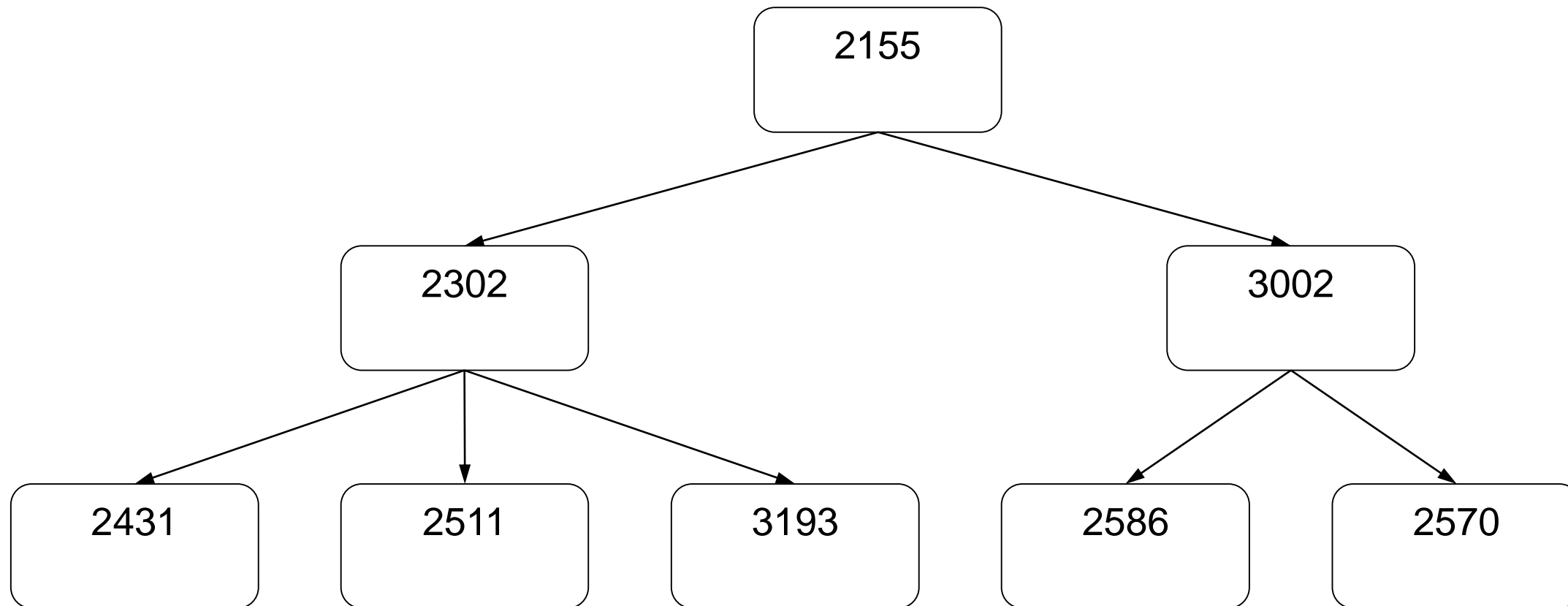
ID_Родителя	ID_Ребенка
2146	2302
2146	3002
2155	2302
2155	3002
2302	2431
2302	2511
2302	3193
3002	2586
3002	2570
2523	2586
2523	2570
2529	2431
2529	2511

Решение:

- сначала находим в таблице 1 Павленко А.К. (ID = 2155)
- теперь по таблице 2 ищем его детей – их идентификаторы 2302 и 3002; можно строить генеалогическое дерево:



далее так же определяем внуков 2155, то есть, детей 2302 и 3002



как следует из таблицы, данных о правнуках 2155 в таблице нет
всего прямых потомков 7 – двое детей и 5 внуков.

Ответ: 7.

пример

- **P-05.** В фрагменте базы данных представлены сведения о родственных отношениях. Определите на основании приведенных данных фамилию и инициалы дяди Леоненко В.С. Пояснение: дядей считается брат отца или матери.

ID	Фамилия_И.О.	Пол	ID_Родителя	ID_Ребенка
14	Леоненко Н.А.	Ж	23	24
23	Геладзе И.П.	М	44	24
24	Геладзе П.И.	М	24	25
25	Геладзе П.П.	М	64	25
34	Леоненко А.И.	Ж	23	34
35	Леоненко В.С.	Ж	44	34
33	Леоненко С.С.	М	34	35
42	Вильямс О.С.	Ж	33	35
44	Гнейс А.С.	Ж	14	33
45	Гнейс В.А.	М	34	42
47	Вильямс П.О.	М	33	42
57	Паоло А.П.	Ж	24	57
64	Мээр П.А.	Ж	64	57

Решение:

- лицо мужского пола не может быть бабушкой, поэтому ответы 3 и 4 неверны
- ищем в первой таблице *Леоненко В.С.*, определяем, что ее код 35
- чтобы найти родителей *Леоненко В.С.*, ищем во второй таблице записи, где код ребенка равен 35: таким образом, её родители имеют коды 33 и 34
- ищем бабушек и дедушек, то есть, записи во второй таблице, где код ребенка равен 33 или 34: соответствующие коды бабушек и дедушки *Леоненко В.С.* – это 14, 44 и 23
- ищем детей персон с кодами 14, 44 и 23 – это братья и сестры родителей *Леоненко В.С.*, то есть, её дяди и тётки; находим, что это человек с кодом 24, *Геладзе П.И.*
- Ответ: 2

пример

- **P-04.** В фрагменте базы данных представлены сведения о родственных отношениях. Определите на основании приведенных данных фамилию и инициалы бабушки Ивановой А.И.

ID	Фамилия_И.О.	Пол
71	Иванов Т.М.	М
85	Петренко И.Т.	М
13	Черных И.А.	Ж
42	Петренко А.И.	Ж
23	Иванова А.И.	Ж
96	Петренко Н.Н.	Ж
82	Черных А.Н.	М
95	Цейс Т.Н.	Ж
10	Цейс Н.А.	М
	...	

ID_Родителя	ID_Ребенка
23	71
13	23
85	23
82	13
95	13
85	42
82	10
95	10
...	...

Решение:

- лицо мужского пола не может быть бабушкой, поэтому ответ 1 неверен
- ищем в первой таблице Иванову А.И., определяем, что ее код 23
- чтобы найти родителей Ивановой А.И., ищем во второй таблице записи, где код ребенка равен 23: таким образом, её родители имеют коды 85 и 13
- теперь ищем бабушек и дедушек, то есть, записи во второй таблице, где код ребенка равен 85 или 13: соответствующие коды бабушки и дедушки нашей Ивановой – это 82 и 95
- в таблице 1 смотрим, кто имеет коды 82 и 95: это Черных А.Н. (мужского пола) и Цейс Т.Н (женского пола); последняя явно является бабушкой

- ответ: Цейс Т.Н.

пример

- **P-03.** Результаты тестирования представлены в таблице:
- Сколько записей в ней удовлетворяют условию «**Пол = 'ж' ИЛИ Химия > Биология**»?
- 1) 5 2) 2 3) 3 4) 4

Фамилия	Пол	Математика	Русский язык	Химия	Информатика	Биология
Аганян	ж	82	56	46	32	70
Воронин	м	43	62	45	74	23
Григорчук	м	54	74	68	75	83
Роднина	ж	71	63	56	82	79
Сергеенко	ж	33	25	74	38	46
Черепанова	ж	18	92	83	28	61

Решение:

- заданное сложное условие отбора состоит из двух простых
- **У1: Пол = 'ж'**
- **У2: Химия > Биология**
- которые связаны с помощью логической операции «ИЛИ»
- заметим, что столбцы «Фамилия», «Математика», «Русский язык» и «Информатика» никак не влияют на результат; уберем их из таблицы и добавим два новых столбца, в которых будем отмечать, выполняются ли условия У1 и У2 для каждой строчки

Пол	Химия	Биология	Пол = 'ж'	Химия > Биология
ж	46	70	+	
м	45	23		+
м	68	83		
ж	56	79	+	
ж	74	46	+	+
ж	83	61	+	+

- логическая операция «ИЛИ» означает выполнение хотя бы одного из двух условия (или обоих одновременно), поэтому заданному сложному условию удовлетворяют все строки, где есть хотя бы один плюс; таких строк пять, они выделены зеленым фоном:

Пол	Химия	Биология	Пол = 'ж'	Химия > Биология
ж	46	70	+	
м	45	23		+
м	68	83		
ж	56	79	+	
ж	74	46	+	+
ж	83	61	+	+

таким образом, правильный ответ – 1.

Файловая система

Что нужно знать:

- данные на дисках хранятся в виде файлов (наборов данных, имеющих имя)
- чтобы было удобнее разбираться с множеством файлов, их объединяют в каталоги (в *Windows* каталоги называются «папками»)
- каталог можно воспринимать как контейнер, в котором размещаются файлы и другие каталоги, которые называются подкаталогами или *вложенными* каталогами (они находятся внутри другого каталога, вложены в него)
- каталоги организованы в многоуровневую (иерархическую) структуру, которая называется «деревом каталогов»
- главный каталог диска (который пользователь видит, «открыв» диск, например, в Проводнике *Windows* или аналогичной программе) называется *корневым* каталогом или «корнем» диска, он обозначается буквой логического диска, за которой следует двоеточие и знак «\» (обратный слэш); например, **A:** – это обозначение корневого каталога диска A
- Для разделения имен каталогов в адресе в разных операционных системах применяют прямой слэш «/» или обратный слэш «\». В системе *Windows*, которая наиболее распространена в России, стандартным разделителем считается «\», именно такой знак чаще всего используется в задачах ЕГЭ.

- каждый каталог (кроме корневого) имеет (один единственный!) «родительский» каталог – этот тот каталог, внутри которого находится данный каталог
- полный адрес каталога – это перечисление каталогов, в которые нужно войти, чтобы попасть в этот каталог (начиная с корневого каталога диска); например
C:\USER\BIN\SCHOOL
- полный адрес файла состоит из адреса каталога, в котором он находится, символа «\» и имени файла, например
C:\USER\BIN\SCHOOL\Вася.txt
- маска служит для обозначения (выделения) группы файлов, имена которых имеют общие свойства, например, общее расширение
- в масках, кроме «обычных» символов (допустимых в именах файлов) используются два специальных символа: звездочка «*» и знак вопроса «?»;
- звездочка «*» обозначает любое количество любых символов, в том числе, может обозначать пустую последовательность;
- знак вопроса «?» обозначает ровно один любой символ
- при выводе списка имен файлов они могут быть отсортированы по имени, типу (расширению), дате последнего изменения, размеру; это не меняет их размещения на диске;
- если установлена сортировка по имени или типу, сравнение идет по кодам символов, входящих в имя или в расширение

Пример

В каталоге находится 6 файлов:

- **maveric.map**
- **maveric.mp3**
- **taverna.mp4**
- **revolver.mp4**
- **vera.mp3**
- **zveri.mp3**

Ниже представлено восемь масок. Сколько из них таких, которым соответствуют ровно четыре файла из данного каталога?

ver*.mp

***?ver?.mp?**

?*ver*.mp?*

v*r*?.m?p

???*???.mp*

???*???.m*

a*.*a

a*.*p

Решение:

- фактически нужно проверить каждое имя файла на соответствие каждой маске
- сформулируем по-русски правила, которые задаёт каждая маска:

<code>*ver*.mp*</code>	в имени содержится «ver», расширение начинается с «.mp»
<code>*?ver?.mp?</code>	в имени содержится «ver», слева и справа от «ver» есть, по крайней мере, по одному символу; расширение начинается с «.mp» и состоит из 3 символов
<code>?*ver*.mp?*</code>	в имени содержится «ver», слева от «ver» есть, по крайней мере, один символ; расширение начинается с «.mp» и состоит не менее, чем из 3 символов
<code>*v*r*?.m?p*</code>	в имени содержится буква «v» и после неё – «r», справа от «r» есть, по крайней мере, один символ; расширение начинается с «m», потом идёт ещё один какой-то символ, за ним – «p»
<code>???*???.mp*</code>	имя состоит не менее, чем из 6 символов, расширение начинается с «.mp»
<code>???*???.m*</code>	имя состоит не менее, чем из 6 символов, расширение начинается с «.m»
<code>*a*.a*</code>	в имени есть буква «a», в расширении тоже есть буква «a»
<code>*a*.p*</code>	в имени есть буква «a», в расширении – буква «p»

- составляем таблицу проверки, выделяем красным фоном символы, не совпадающие с маской; если все совпало, ячейку выделяем зеленым фоном:

ver.mp*	maveric.ma p	maveric.mp3	taverna.mp4	revolver.mp4	vera.mp3	zveri.mp3
*?ver?.mp?	maveric.ma p	maveric.mp3	taverna.mp4	rev o lver.mp4	v era.mp3	zveri.mp3
?*ver*.mp?*	maveric.ma p	maveric.mp3	taverna.mp4	revolver.mp4	v era.mp3	zveri.mp3
*v*r*.r						mp p
???*???						mp3
???*???						mp3
a.a*	maveric.map	maveric. mp3	taverna. mp4	revolver. mp4	vera. mp3	zveri. mp3
a.p*	maveric.map	maveric.mp3	taverna.mp4	revolver .mp4	vera.mp3	zveri .mp3
a.p*	maveric.map	maveric.mp3	taverna.mp4	revolver .mp4	vera.mp3	zveri .mp3

видим, что трём маскам: **?*ver*.mp?***, **???*???.m*** и ***a*.p***, соответствует ровно по 4 файла.

Ответ: 3.

пример

- *Перемещаясь из одного каталога в другой, пользователь последовательно посетил каталоги **DOC, USER, SCHOOL, A:\, LETTER, INBOX**. При каждом перемещении пользователь либо спускался в каталог на уровень ниже, либо поднимался на уровень выше. Каково полное имя каталога, из которого начал перемещение пользователь?*

пример

- *Каталог содержит файлы с именами*
 - а) **q.c**
 - б) **qq.cpp**
 - в) **qq.c**
 - г) **q1.c1**
 - д) **qaa.cmd**
 - е) **q12.cpp**
- *Определите, в каком порядке будут показаны файлы, если выбрана сортировка по типу (по возрастанию).*

Решение

- при сортировке по типу сравниваются расширения имен файлов
- при сравнении используют коды символов
- отсутствие символа (когда расширение закончилось) считается «меньше» любого символа, то есть, файл с расширением `.c` будет находиться в списке выше, чем файлы с расширениями `.c1` и `.cmd`
- коды цифр размещаются в таблице символов раньше, чем коды букв, то есть, файл с расширением `.c1` будет находиться в списке выше, чем файл с расширением `.cmd`

- теперь можно распределить имена файлов по расширениям
- а) **q.c**
- в) **qq.c**
- г) **q1.c1**
- д) **qaa.cmd**
- б) **qq.cpp**
- е) **q12.cpp**
- осталась еще одна проблема – решить, что делать, если расширения совпадают; в этом случае в большинстве программ для определенности используется дополнительная сортировка по имени, поэтому файл с именем **q12.cpp** будет стоять в списке выше, чем файл **qq.cpp** (код цифры '1' меньше, чем код буквы 'q')
- в итоге получаем
- а) **q.c**
- в) **qq.c**
- г) **q1.c1**
- д) **qaa.cmd**
- е) **q12.cpp**
- б) **qq.cpp**

таким образом, мы доказали, что правильный ответ – **авгдеб**

пример

- *Для групповых операций с файлами используются маски имен файлов. Маска представляет собой последовательность букв, цифр и прочих допустимых в именах файлов символов, в которых также могут встречаться следующие символы: Символ «?» (вопросительный знак) означает ровно один произвольный символ. Символ «*» (звездочка) означает любую последовательность символов произвольной длины, в том числе «*» может задавать и пустую последовательность. Определите, по какой из масок будет выбрана указанная группа файлов:*

- **1234.xls**
- **23.xml**
- **234.xls**
- **23.xml**
- **1) *23*.*x***
- **2) ?23?.x??**
- **3) ?23?.x***
- **4) *23*.*??**