

3 ЛАБОРАТОРНАЯ РАБОТА № 3

"ПРОДУКЦИОННАЯ МОДЕЛЬ ПРЕДСТАВЛЕНИЯ ЗНАНИЙ "

Цель работы: сформировать у студентов навыки использования продукционных моделей для представления знаний в информационных аналитических системах.

3.1. Задание к лабораторной работе № 3

1. Изучить методический материал к лабораторной работе, приведенный в п.3.2 и дополнительную литературу.

2. В соответствии с вариантом задания (варианты заданий на лабораторную работу приведены в п.3.3):

- дерево принятия решений с поддеревьями (не менее 2-х поддеревьев);
- дерево принятия решений совместно с поддеревьями, должно содержать не менее 30 ветвей;
- сформировать базу продукционных правил;
- разработать программу, используя любой язык программирования, позволяющую по базе продукционных правил сформировать статические структуры данных:
 - массив переменных условной части;
 - массив переменных логического вывода;
 - массив простых переменных.

3.2. Методические указания к лабораторной работе № 3

Наиболее наглядным и легко реализуемым способом представления поля знаний продукционной СИИ является дерево решений.

Деревья решений (*decision trees*) - иерархические древовидные структуры, состоящие из правил вида «если... то...» и позволяющие выполнять классификацию объектов.

Они являются самым популярным инструментом *data mining*, позволяющим эффективно решать задачи классификации.

В отличие от методов, использующих статистический подход, таких как классификатор Байеса, линейная и логистическая регрессия, деревья решений основаны на машинном обучении и в большинстве случаев не требуют предположений о статистическом распределении значений признаков.

Для построения дерева решений используется обучающая выборка, в которой вся информация об объектах должна быть выражена совокупностью значений признаков и указанием принадлежности к одному из классов. Количество объектов выборки должно быть намного больше, чем количество классов.

В узлах содержатся правила, с помощью которых производится проверка признаков. Множество объектов в узле разбивается на подмножества. *Листья* — это конечные узлы дерева, в которых содержатся подмножества объектов, ассоциированные с классами. Отличием листа от узла является то, что в листе не производится проверка, разбивающая подмножество, и, соответственно, нет ветвления.

В основе работы деревьев решений лежит процесс последовательного разбиения исходного множества объектов на подмножества, ассоциированные с классами. Разбиение производится с помощью решающих правил, в которых осуществляется проверка значений признаков объектов по заданному условию. Поэтому деревья не только классифицируют объекты и наблюдения, но и объясняют, почему объект был отнесен к данному классу. Обладая *высокой объясняющей способностью*, они могут использоваться как инструмент исследования предметной области.

Пример дерева принятия решений приведен на рисунке 3.1.

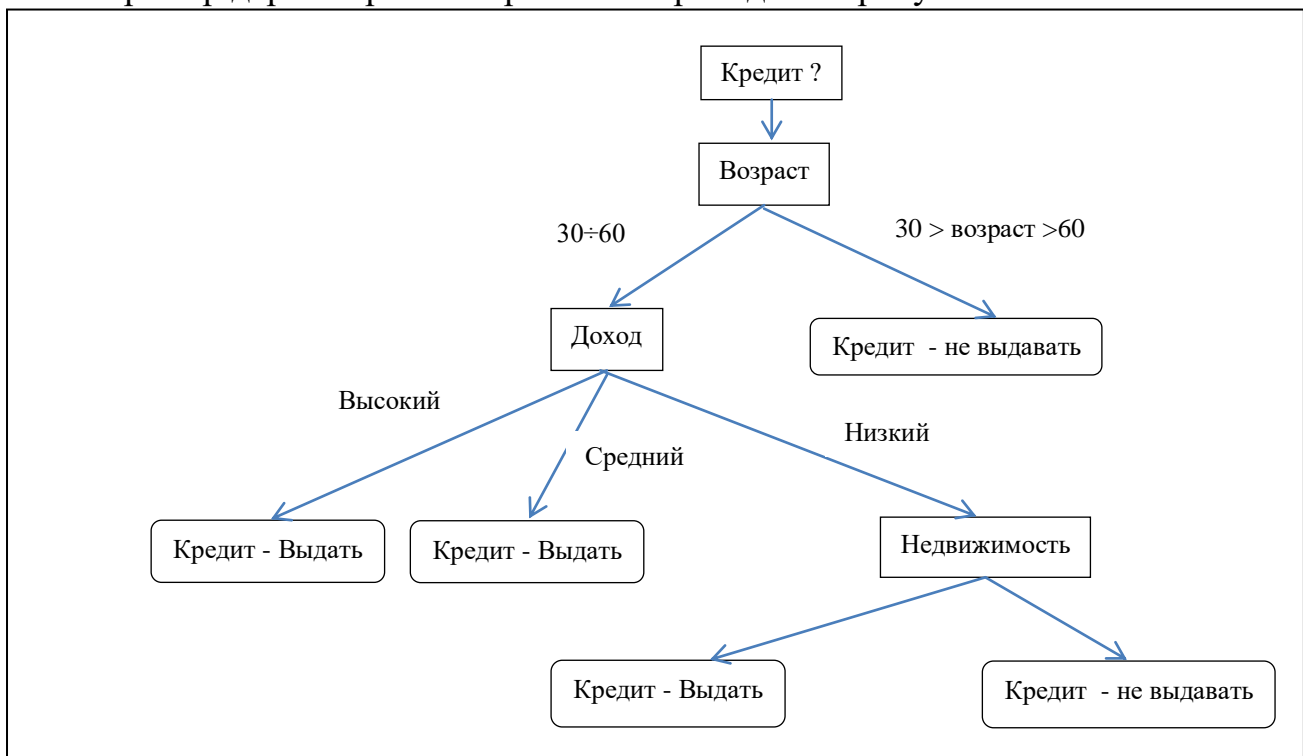


Рисунок 3.1 – Дерево принятия решения по оценке кредитоспособности клиента.

Алгоритм построения дерева решений. Дерево решений последовательно разбивает исходное множество объектов на подмножества. Для ускорения поиска нужного решения при построении деревьев используются оптимизирующие процедуры, основанные на выборе наиболее информативных признаков и исключении незначущих, не предоставляющих существенной информации для разделения объектов на классы.

В общем виде *алгоритм построения дерева решений* следующий:

- выбрать признак, соответствующий максимуму критерия разделимости классов, который вычисляется на основе комбинаторных методов или с помощью энтропии;
- в узле создать ветвь для каждого значения признака;
- для каждой ветви построить подмножества объектов, удовлетворяющие заданному значению признака;
- если подмножество содержит объекты одного класса, то сформировать лист дерева;

- повторить шаги для остальных подмножеств объектов, исключив признак из рассмотрения.

По дереву принятия решений, оценивая кредитоспособность клиентов банка, можно сформулировать следующие правила:

- Если Возраст - $30 \div 60$, Доход - высокий *то* Кредит - Выдать;
- Если Возраст от $30 \div 60$, Доход – низкий, Недвижимость - есть *то* Кредит - Выдать;
- Если Возраст - <30 *то* Кредит - Не выдавать.

Полученные продукционные правила заносятся в базу знаний (БЗ). Фактически БЗ содержит множество продукционных правил или просто продукций (productions).

Каждое продукционное правило состоит из:

- условной части (IF–part) правила - антецедент (antecedent), являющейся шаблоном (образцом), по которому можно определить, в какой момент необходимо использовать (активировать) данное правило;
- части действия (THEN–part) - консеквент (consequent), описывающей соответствующий шаг решения.

Преимущества продукционного подхода:

- естественность подхода, при котором процесс получения выводов аналогичен процессу рассуждения эксперта;
- модульность базы знаний и, как следствие,
- простота модификации;
- однородность базы знаний - одна форма представления знаний;
- простота создания и понимания отдельных правил;
- наличие простого и ясного механизма логического вывода;
- простейший способ объяснения решения - трассировка правил.

Недостатки продукционного подхода:

- неясность взаимных отношений правил;
- сложность описания родовидовых иерархий понятий;
- сложность оценки целостного образа знаний;
- низкая эффективность вывода из-за того, что большинство операций переборного типа.

В связи с этим приходится решать следующие проблемы:

- проверка на непротиворечивость базы знаний;
- проверка на полноту;
- отсеечение избыточности при сопоставлении и разрешении конфликтов;
- выбор критерия разрешения конфликтов так, чтобы вывод был более целенаправленным;
- проблема структурирования знаний.

3.3. Варианты заданий к лабораторной работе № 3

При выполнении лабораторной работы №3 варианты предметных областей соответствуют заданию к лабораторной работе №1 (см.п.1.3).

3.4. Пример выполнения лабораторной работы №3

Используя *ваши поверхностные знания* в предметной области «Музыкальные инструменты» сформировать:

- дерево принятия решений с поддеревьями (не менее 2-х поддеревьев);
- дерево принятия решений совместно с поддеревьями, должно содержать не менее 30 ветвей;
- сформировать базу продукционных правил;
- разработать программу (можно использовать любой язык программирования), позволяющую по базе продукционных правил сформировать статические структуры данных: массив переменных условной части; массив переменных логического вывода; массив простых переменных.

Предполагаем, что при классификации музыкальных инструментов можно подразделять их по типам и подклассам, в зависимости от определённых характеристик.

На рисунке 3.2. приведено поддерево «Тип», позволяющее сделать вывод об отношении музыкального инструмента к одному из типов:

- Струнные;
- Ударные;
- Духовые.

Определить другие типы музыкальных инструментов в приведённом примере – нельзя. Понятие «Тип» будет использовано в поддереве «Подкласс». Тип – подцель, которая может быть достигнута в процессе обработки продукционных правил.

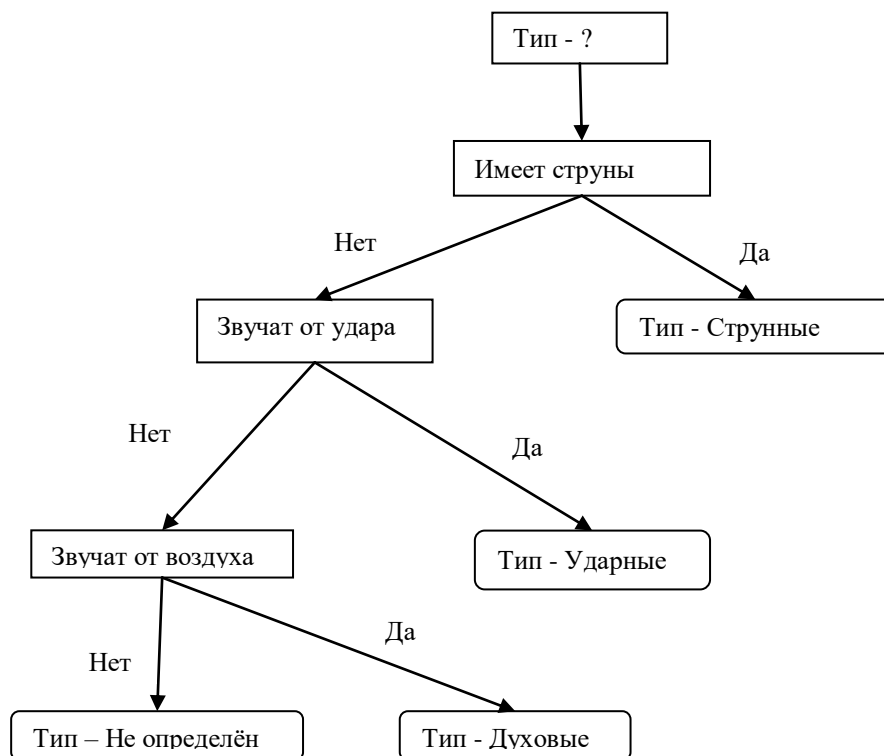


Рисунок 3.2. - Поддерево «Тип»

На рисунке 3.3. приведено поддерево «Подкласс», позволяющее сделать вывод об отношении музыкального инструмента, в зависимости от определённого типа, к одному из подклассов:

- Щипковые;
- Тростевидные;
- Смычковые;
- Лабиальные;
- Ударно-клавишные;
- Амбушюрные.

Определить другие подклассы музыкальных инструментов в приведённом примере – нельзя.

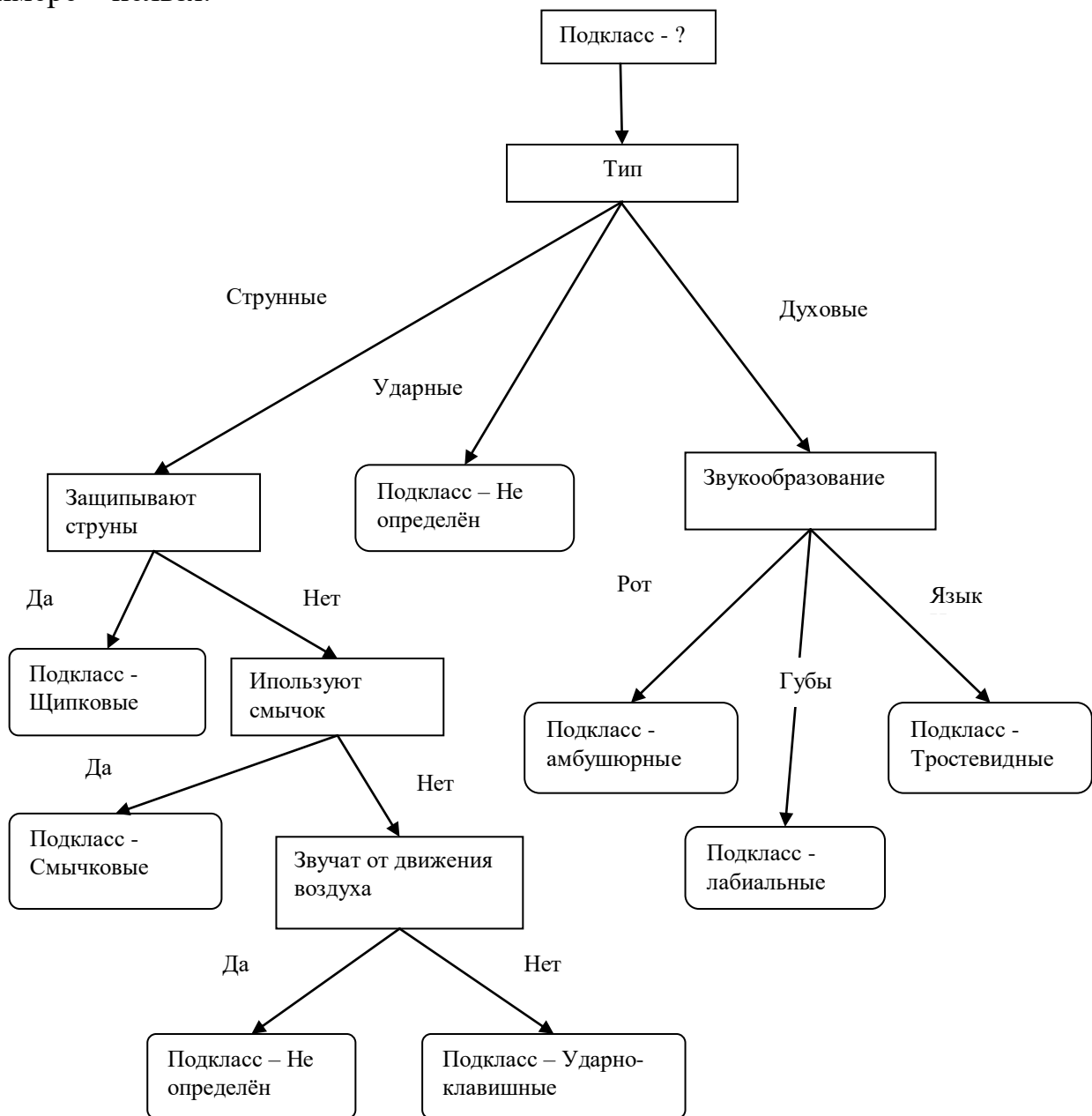


Рисунок 3.3 – Поддерево «Подкласс»

Понятие «Подкласс» будет использовано в дереве «Музыкальные инструменты». Подкласс – подцель, которая может быть достигнута в процессе обработки продукционных правил.

На рисунке 3.4. приведено дерево принятия решений «Музыкальные инструменты», позволяющее определить музыкальный инструмент:

- Кларнет;
- Фагот;
- Гобой;
- Флейта;
- Валторна;
- Труба;
- Туба;
- Колокол;
- Тромбон;
- Арфа;
- Банджо;
- Гитара;
- Мандолина;
- Лютня;
- Домра;
- Балалайка и пр.

Используя дерево принятия решений и поддеревья, формируем базу продукционных правил. Каждое правило в базе соответствует ветке дерева или поддерева принятия решений.

Порядок правил в базе продукционных правил не имеет значения, поэтому можно записывать их в произвольном порядке.

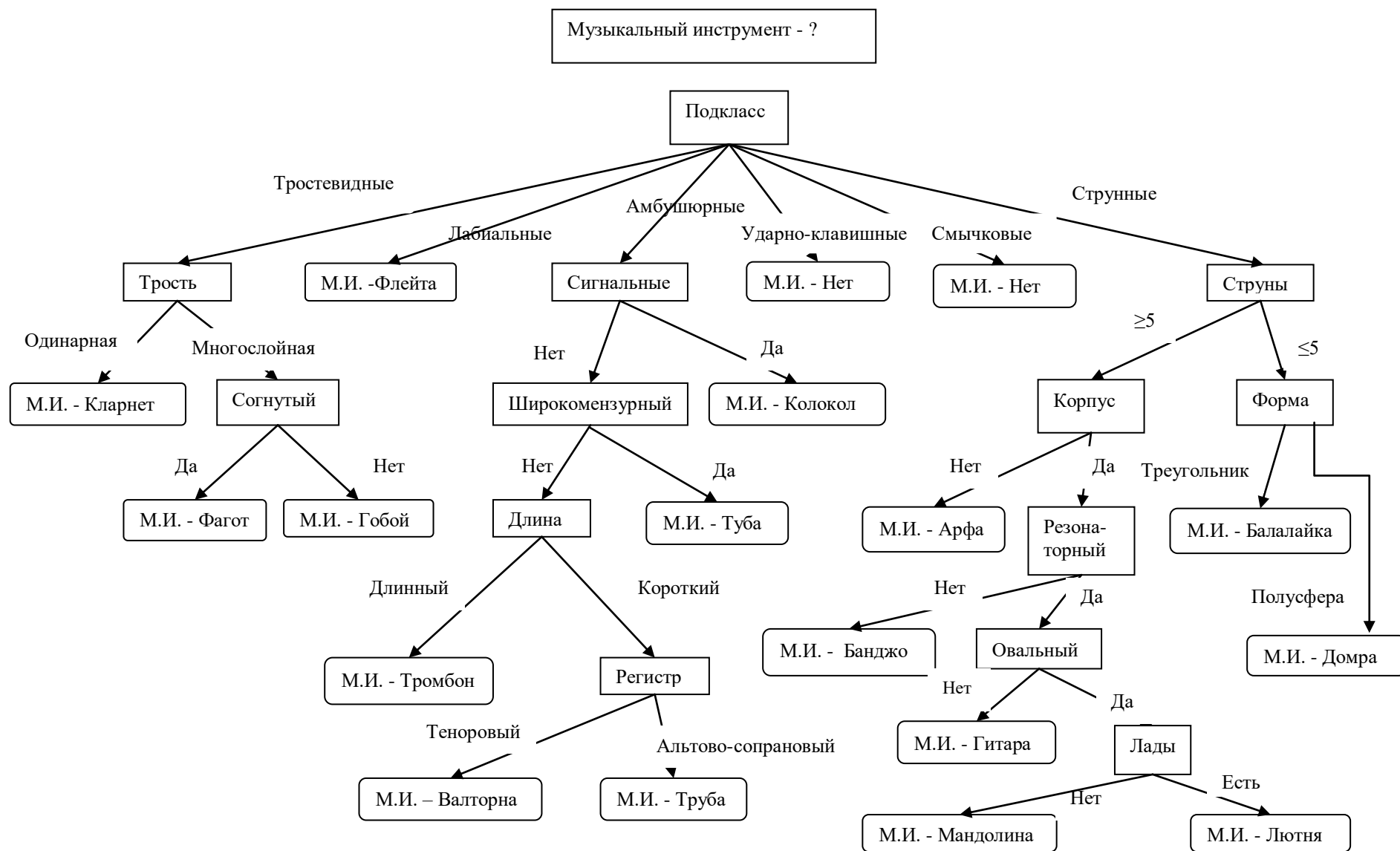


Рисунок 3.4 – Дерево принятия решений «Музыкальные инструменты»

База продукционных правил:

1. **Если** Имеет струны – Да **то** Тип – Струнные
2. **Если** Имеет струны – Нет, Звучат от удара – Да **то** Тип – Ударные
3. **Если** Имеет струны – Нет, Звучат от удара - Нет, Звучат от воздуха – Да **то** Тип – Духовые
4. **Если** Имеет струны – Нет, Звучат от удара - Нет, Звучат от воздуха – Нет **то** Тип – Не определён
5. **Если** Тип – Струнные, Защищают струны – Да **то** Подкласс – Щипковые
6. **Если** Тип – Струнные, Защищают струны – Нет, Используют смычок - Да **то** Подкласс – Смычковые
7. **Если** Тип – Струнные, Защищают струны – Нет, Используют смычок – Нет, Звучат от движения воздуха - Да **то** Подкласс – Не определён
8. **Если** Тип – Струнные, Защищают струны – Нет, Используют смычок – Нет, Звучат от движения воздуха - Нет **то** Подкласс – Ударно-клавишные
9. **Если** Тип – Ударные **то** Подкласс – Не определён
10. **Если** Тип – Духовые, Звукообразование – Рот **то** Подкласс – Амбюшюрные
11. **Если** Тип – Духовые, Звукообразование – Губы **то** Подкласс – Лабиальные
12. **Если** Тип – Духовые, Звукообразование – Язык **то** Подкласс – Тростевидные
13. **Если** Подкласс – Тростевидные, Трость – Одинарная **то** Музыкальный инструмент - Кларнет
14. **Если** Подкласс – Тростевидные, Трость – Многослойная, Согнутый – Да **то** Музыкальный инструмент – Фагот
15. **Если** Подкласс – Тростевидные, Трость – Многослойная, Согнутый – Нет **то** Музыкальный инструмент – Гобой
16. **Если** Подкласс – Лабиальные **то** Музыкальный инструмент – Флейта
17. **Если** Подкласс – Амбюшюрные, Сигнальные – Нет, Широкомензурный – Нет, Длина - Длинный **то** Музыкальный инструмент – Тромбон
18. **Если** Подкласс – Амбюшюрные, Сигнальные – Нет, Широкомензурный – Нет, Длина – Короткий, Регистр - Теноровый **то** Музыкальный инструмент – Валторна
19. **Если** Подкласс – Амбюшюрные, Сигнальные – Нет, Широкомензурный – Нет, Длина – Короткий, Регистр – Альтово-сопрановый **то** Музыкальный инструмент – Труба
20. **Если** Подкласс – Амбюшюрные, Сигнальные – Нет, Широкомензурный – Да **то** Музыкальный инструмент – Туба
21. **Если** Подкласс – Амбюшюрные, Сигнальные – Да **то** Музыкальный инструмент – Колокол

22. **Если** Подкласс – Ударно-клавишные **то** Музыкальный инструмент – Нет
23. **Если** Подкласс – Смычковые **то** Музыкальный инструмент – Нет
24. **Если** Подкласс – Струнные, Струны - ≥ 5 , Корпус – Нет **то** Музыкальный инструмент - Арфа
25. **Если** Подкласс – Струнные, Струны - ≥ 5 , Корпус – Да, Резонаторный - Нет **то** Музыкальный инструмент - Банджо
26. **Если** Подкласс – Струнные, Струны - ≥ 5 , Корпус – Да, Резонаторный – Да, Овальный – Нет **то** Музыкальный инструмент – Гитара
27. **Если** Подкласс – Струнные, Струны - ≥ 5 , Корпус – Да, Резонаторный – Да, Овальный – Да, Лады - Нет **то** Музыкальный инструмент – Мандолина
28. **Если** Подкласс – Струнные, Струны - ≥ 5 , Корпус – Да, Резонаторный – Да, Овальный – Да, Лады - Есть **то** Музыкальный инструмент – Лютя
29. **Если** Подкласс – Струнные, Струны - ≤ 5 , Форма – Треугольник **то** Музыкальный инструмент - Балалайка
30. **Если** Подкласс – Струнные, Струны - ≤ 5 , Форма – Полусфера **то** Музыкальный инструмент – Домра

Полученную базу продукционных правил заносим в файл, который является входной информацией для программы формирующей статические структуры данных. Информация заносится в соответствии с заданными требованиями, что исключает необходимость проведения проверки исходной информации на правильность.

Статические данные будут использованы в программе реализующей прямой и обратный логический вывод.

Статические данные формируются один раз, после создания базы знаний (базы продукционных правил (БПП)). Изменение в эти структуры вносятся только при изменении БПП.

К статическим структурам данных относятся:

- массив переменных условной части;
- массив переменных логического вывода;
- массив простых переменных.

Переменные условной части продукционного правила – переменные, расположенные в условной части правила. Например, для первого правила, переменной условной части является переменная **«Имеет струны»**.

В разработанной программе, формирующей статические структуры данных, массив переменных условной части (МПУЧ) имеет следующий вид:

№ правила	Переменная	Значение

Длина массива может достигать нескольких тысяч строк следовательно поиск информации в нем потребует больших временных затрат, поэтому

данный массив реализуем как хеш-таблицу. Выполнение операции в хеш-таблице начинается с вычисления хеш-функции. В нашем случае используется хеш-функция, позволяющая по номеру правила определяет номер строки МПУЧ, в которой находится первая переменная условной части выбранного правила.

В хеш-функции используются следующие величины:

- N – номер правила;
- dl – максимальное количество переменных в условной части продукционных правил в МПП;
- K – строка МПУЧ, в которой находится первая переменная условной части выбранного правила.

Для разработанной БПП максимальное количество переменных условно части – 6 (28-е правило), значит $dl = 6$.

Для каждого правила в МПУЧ отводим по 6 строк, тогда хеш-функция будет иметь вид:

$$K = (N-1)*dl+1.$$

Массив переменных условной части, полученный после выполнения программы будет иметь вид:

№ правила	Переменная	Значение
1	Имеет струны	Да
2		
3		
4		
5		
6		
7	Имеет струны	Нет
8	Звучат от удара	Да
9		
10		
11		
12		
13	Имеет струны	Нет
14	Звучат от удара	Нет
15	Звучат от воздуха	Да
16		
...		
175	Подкласс	Струнные
176	Струны	≤ 5
177	Форма	Полусфера
178		
179		
180		

Количество строк в массиве переменных условной части – $N*dl=30*6=180$.

Переменные логического вывода продукционного правила – переменные, расположенные в части действия (заключения). Например, для первого правила, переменной логического вывода является переменная «**Тип**».

В разработанной программе, формирующей статические структуры данных, массив переменных логического вывода (МПЛВ) имеет следующий вид:

№ правила	Переменная	Значение

Количество строк в массиве переменных логического вывода соответствует количеству продукционных правил базы продукционных правил. Номер строки в массиве соответствует номеру правила, поэтому хеш-функцию вводить не нужно.

Массив переменных логического вывода, полученный после выполнения программы формирования МПЛВ будет иметь вид:

№ правила	Переменная	Значение
1	Тип	Струнные
2	Тип	Ударные
3	Тип	Духовые
4	Тип	Не определён
5	Подкласс	Щипковые
6	Подкласс	Смычковые
7	Подкласс	Не определён
8	Подкласс	Ударно-клавишные
9	Подкласс	Не определён
10	Подкласс	Амбюшюрные
11	Подкласс	Лабиальные
12	Подкласс	Тростевидные
13	Музыкальный инструмент	Кларнет
14	Музыкальный инструмент	Фагот
15	Музыкальный инструмент	Гобой
16	Музыкальный инструмент	Флейта
17	Музыкальный инструмент	Тромбон
18	Музыкальный инструмент	Валторна
19	Музыкальный инструмент	Труба
20	Музыкальный инструмент	Туба
21	Музыкальный инструмент	Колокол
22	Музыкальный инструмент	Нет
23	Музыкальный инструмент	Нет
24	Музыкальный инструмент	Арфа
25	Музыкальный инструмент	Банджо
26	Музыкальный инструмент	Гитара
27	Музыкальный инструмент	Мандолина
28	Музыкальный инструмент	Лютня
29	Музыкальный инструмент	Балалайка
30	Музыкальный инструмент	Домра

Простые переменные продукционного правила – переменные, расположенные в условной части правила, значения которых определяется в диалоге с пользователем. Например, для первого правила, простой переменной является переменная **«Имеет струны»**.

Массив простых переменных (МПП) может быть сформирован по базе продукционных правил или путем преобразования массива переменных условной части и массива переменных логического вывода.

МПП может быть реализован как обычный массив или как массив динамических структур. В массиве простых переменных должны храниться сами простые переменные и все значения, которые они могут принимать, что позволит при реализации прямого и обратного логических выводов вести диалог с пользователем, предлагая ему выбирать значения из допустимого списка.

3.5. Содержание отчета по лабораторной работе № 3

В отчёте по выполнению лабораторной работы необходимо привести:

- 1) вариант задания;
- 2) описание предметной области;
- 3) дерево принятия решений с поддеревьями;
- 4) базу продукционных правил;
- 5) алгоритм программы формирующей статические массивы (массив переменных условной части, массив переменных логического вывода, массив простых переменных);
- 6) описание структур массивов (МПУЧ, МПЛВ, МПП);
- 7) листинг программы;
- 8) полученные результаты (заполненные информацией массивы);
- 9) выводы по результатам выполнения лабораторной работы с рекомендациями по её модификации.

3.6. Контрольные вопросы к лабораторной работе №3

1. Назначение дерева принятия решений.
2. Описать структуру продукционного правила.
3. Дать определение переменным условной части продукционного правила.
4. Какие переменные продукционного правила называются простыми?
5. Какие статические структуры формируются по базе продукционных правил?
6. Каким образом использование хеш-функций ускоряет обработку массивов?
7. Перечислите основные недостатки продукционных моделей.
8. Перечислите основные преимущества продукционных моделей.
9. Сколько переменных логического вывода может быть в одном продукционном правиле?