

ПРОДУКЦИОННАЯ МОДЕЛЬ

- Впервые метод продукций был предложен Эмилем Постом в 1936 г. в контексте уточнения понятия «алгоритм». Данный вычислительный формализм был основан на правилах замены строк символов. В 1963 г. этот метод был использован Ноамом Хомским для определения языков и формальных грамматик. А с начала семидесятых годов метод продукций находит широкое применение для представления знаний в экспертных системах и в системах, основанных на знаниях.

- Продукционное правило, или продукция, – это оператор преобразования, представляющий собой выражение следующего вида:
- <ситуация> → <заключение>.
- Левая часть связана с распознаваемой ситуацией и содержит список признаков, которые должны быть выполнены, чтобы было выполнено заключение, содержащееся в правой части. Возможна и другая трактовка понятия продукции:
- <ситуация> → <действие>.
- Здесь правая часть содержит список действий, которые должны быть выполнены, если выполнен список признаков, описывающих ситуацию.

- Продукционная система – это машина, которая воспринимает совокупность известных фактов и строит новые заключения, работая в режиме распознавание – заключение (действие).
Продукционная система состоит из трех основных компонентов:
- Первый из них – это набор правил, используемый как база знаний, иногда его еще называют базой правил. Второй компонент – это база фактов или рабочая память – память для временного хранения, в которой хранятся предпосылки, касающиеся конкретных задач предметной области, и результаты выводов, получаемые на их основании. Третий компонент реализует механизм логического вывода, обрабатывающий правила в соответствии с содержанием рабочей памяти; другое название этого компонента – машина логического вывода.

- Способ получения вывода представляет собой либо правило модус поненс:
 - посылка p , правило $p \rightarrow q$, заключение q ;
 - либо модус толленс:
 - посылка $\sim q$ правило $p \rightarrow q$, заключение $\sim p$.
- Рабочая память (РП) в начале работы продукционной системы содержит исходные данные для поставленной задачи и пополняется фактами, установленными в процессе работы системы. Правила продукций применяются к содержимому рабочей памяти.
- правило состоит из двух частей: условной и заключительной.
- Условная часть может состоять из одного или нескольких условий. Если условная часть правила удовлетворяет содержимому рабочей памяти, то правило может быть применено, что приведет к изменению рабочей памяти.

- В общем виде продукция может быть представлена выражением следующего вида:

- (i); $Q;P;A \Rightarrow B;N$

- i - имя продукции, в его качестве может выступать некоторая лексема, отражающая суть данной продукции (например, "покупка книги").
- Элемент **Q** характеризует сферу применения продукции.
- Разделение знаний на отдельные сферы позволяет экономить время при поиске решения задачи. Например, часть продукций описывает процесс приготовления пищи, а вторая – выбора маршрута путешествия и т. п.
- Основным элементом продукции является ее **ядро**: $A \Rightarrow B$. ("**ЕСЛИ A, ТО B**")

- Элемент **P** есть условие применимости ядра продукции: когда **P** принимает значение "истина", ядро продукции активизируется.
- Обычно **P** представляет собой логическое выражение. Когда **P** принимает значение «истина», ядро продукции активизируется. Если **P** ложно, то ядро продукции не может быть использовано. Например, если в продукции «НАЛИЧИЕ ДЕНЕГ; ЕСЛИ ХОЧЕШЬ КУПИТЬ ВЕЩЬ X, ТО ЗАПЛАТИ В КАССУ ЕЕ СТОИМОСТЬ И ОТДАЙ ЧЕК ПРОДАВЦУ» условие применимости ядра продукции ложно, т.е. денег нет, то применить ядро продукции невозможно.

- Элемент **N** описывает **постусловия продукции** (*актуализируются только в том случае, если ядро продукции реализовалось*).
- Постусловия продукции описывают действия и процедуры, которые необходимо выполнить после реализации **B**. Например, после покупки некоторой вещи в магазине необходимо в описи товаров, имеющихся в этом магазине, уменьшить количество вещей такого типа на единицу. Выполнение **N** может происходить не сразу после реализации ядра продукции.
- Продукционная модель обладает тем недостатком, что при накоплении достаточно большого числа (порядка нескольких сотен) продукций они начинают противоречить друг другу.

- Любое продукционное правило, содержащееся в базе знаний, состоит из двух частей: **антецедента** и **консеквента**. Антецедент представляет собой посылку правила (условную часть) и состоит из элементарных предложений, соединенных логическими связками «и», «или». Консеквент (заключение) включает одно или несколько предложений, которые выражают либо некоторый факт, либо указание на определенное действие, подлежащее исполнению. Продукционные правила принято записывать в виде антецедент-консеквент.
- **Все ядра делятся на два больших типа:**
- **детерминированные** (при актуализации ядра и при выполнимости A правая часть ядра выполняется обязательно)
- **недетерминированные** (B может выполняться или не выполняться: "ЕСЛИ A, ТО ВОЗМОЖНО B")

- **Детерминированные продукции могут быть однозначными и альтернативными.**

Во втором случае в правой части ядра указываются альтернативные возможности выбора.

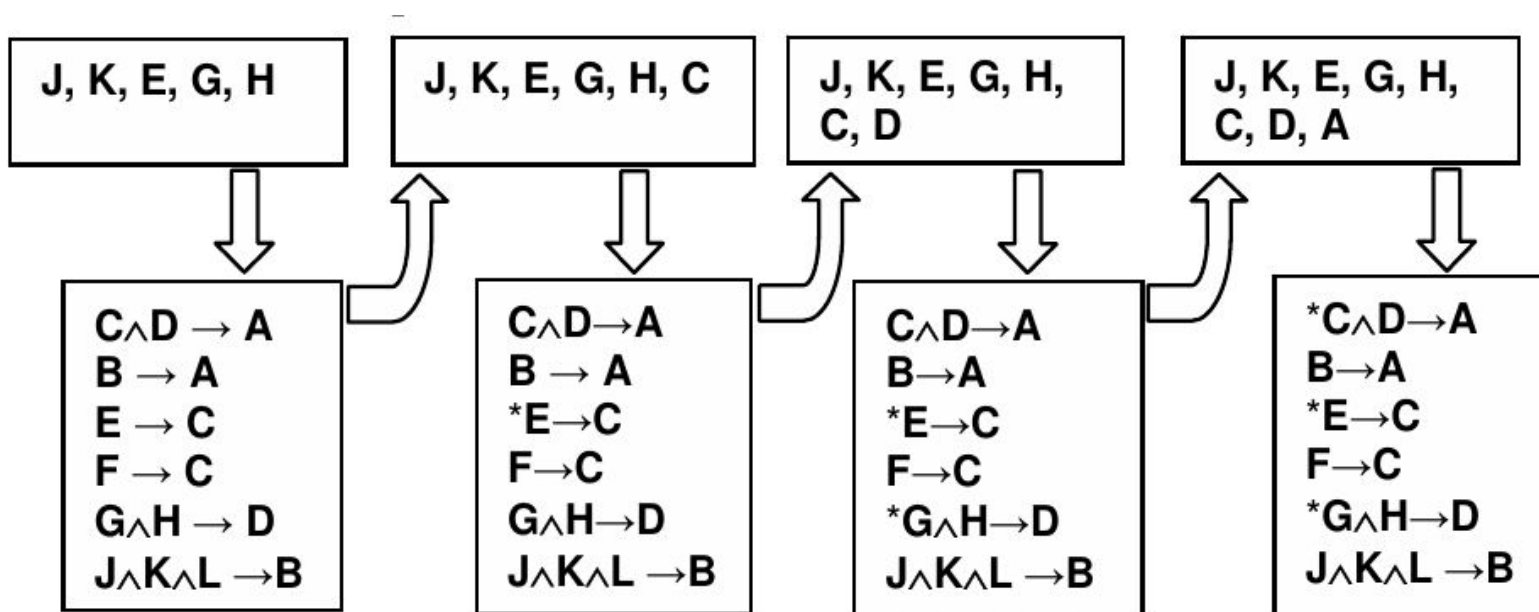
Особым типом являются *прогнозирующие продукции*, **в которых описываются последствия, ожидаемые при актуализации А.**

- Продукционная модель часто дополняется определённым порядком, вводимым на множестве продукций, что упрощает механизм логического вывода. Порядок может выражаться в том, что отдельная следующая по порядку продукция может применяться только после попыток применения предшествующих ей продукций. Примерно похожее влияние на продукционную модель может оказать использование приоритетов продукций, означающее, что в первую очередь должна применяться продукция, имеющая наивысший приоритет.
- Рост противоречивости продукционной модели может быть ограничен путём введения механизмов исключений и возвратов. **Механизм исключений** означает, что вводятся специальные правила-исключения. Их отличает большая конкретность в сравнении с обобщёнными правилами. При наличии исключения основное правило не применяется. **Механизм возвратов** же означает, что логический вывод может продолжаться в том случае, если на каком-то этапе вывод привёл к противоречию. Просто необходимо отказаться от одного из принятых ранее утверждений и осуществить возврат к предыдущему состоянию.

- Существуют два типа выполнения систем продукций: прямой и обратный
- **Прямой вывод** называется также выводом, управляемым данными, или нисходящим. В таких системах поиск идет от исходных данных (фактов) к заключениям. Т.е. проверяются условия **A**, включающие известные факты, и активизируются те продукции, для которых **A** истинно. После этого в рабочую память заносятся промежуточные заключения **B'**, которые в дальнейшем выступают как дополнительные факты для **A'** и так до тех пор, пока не будет получено итоговое заключение **B**.
- **Обратный вывод** называется также выводом, управляемым целями, или восходящим. В таких системах выдвигается некоторая гипотеза **B**, а затем идет поиск промежуточных фактов **A'**, подтверждающих эту гипотезу. После этого в рабочую память заносятся промежуточные факты **A'**, которые в дальнейшем выступают как промежуточные гипотезы (заключения) **B'**. Если принятая гипотеза **B** приводит к известным фактам **A**, то она считается итоговым заключением.
- Существуют также системы с двунаправленными выводами.

- **Прямой вывод** рекомендуется использовать в следующих случаях:
 - - все или большинство исходных данных заданы в постановке задачи (например, в экспертной системе PROSPECTOR);
 - - существует большое количество потенциальных целей, но мало способов использования фактов (например, в экспертной системе DENDRAL эффективно использовались факт, что для любого органического соединения существует чрезвычайно большое число возможных структур, однако данные масс-спектрографа позволяют оставить лишь небольшое количество таких комбинаций);
 - - сформировать цель или гипотезу очень трудно (например, в экспертной системе DENDRAL изначально мало информации о возможной структуре соединения).
- **Обратный вывод** рекомендуется использовать в следующих случаях:
 - - цель поиска или гипотеза явно присутствует в постановке задачи или может быть легко сформулирована (например, при доказательстве математических теорем);
 - - имеется большое количество правил, которые на основе полученных фактов продуцируют всевозрастающее число заключений и целей. Своевременный отбор целей позволяет отсеять множество возможных ветвей, что делает процесс поиска более эффективным;
 - - исходные данные не приводятся в задаче, но подразумевается, что они должны быть известны решателю. Например, выбирается предварительный медицинский диагноз (гипотеза), а потом под него подбираются симптомы (факты).

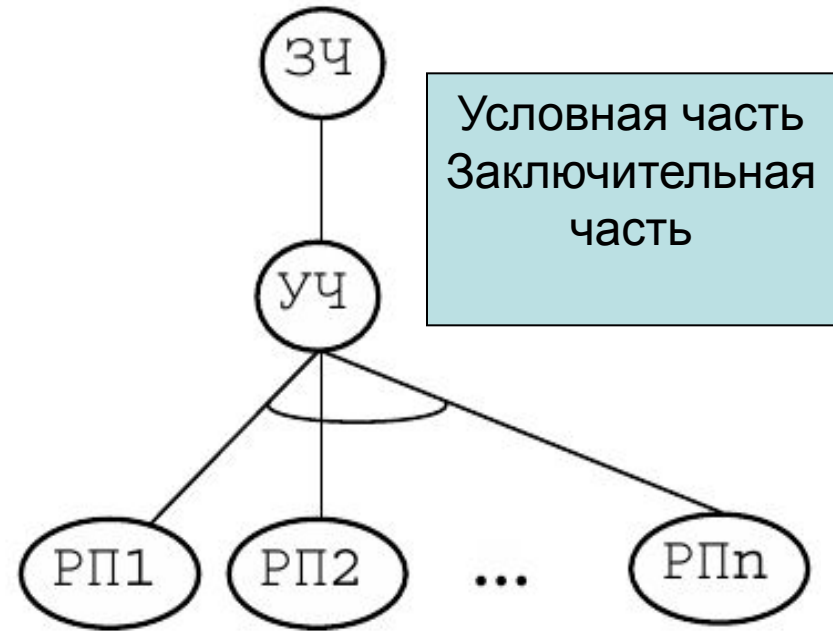
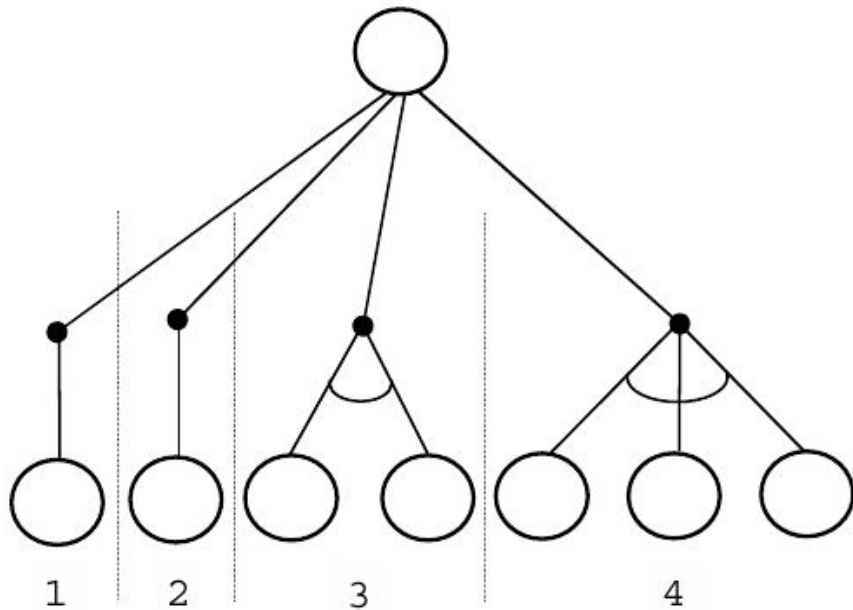
- Рассмотрим пример продукционной системы с простым набором правил:
- $C \wedge D \rightarrow A$, $B \rightarrow A$, $E \rightarrow C$, $F \rightarrow C$, $G \wedge H \rightarrow D$, $J \wedge K \wedge L \rightarrow B$.



- процедура обратного вывода
- Шаг 1. В систему поступает указание о необходимости вывода A . Сначала осуществляется проверка наличия A в базе фактов и после отрицательного ответа поиск правил, заканчивающихся на A , то есть имеющих A справа от стрелки. Система находит два правила $C \wedge D \rightarrow A$, $B \rightarrow A$ и принимает решение о том, что для подтверждения A необходимо установить наличие или C и D , или B .
- Шаг 2. Система пытается установить C в результате проверки сначала базы фактов и последующего обнаружения правил, приводящих к C . Обнаружено два таких правила: $E \rightarrow C$ и $F \rightarrow C$. Система приходит к решению о том, что для получения вывода о присутствии C ей необходимо установить наличие или E , или F .
- Шаг 3. Система обнаруживает E в базе фактов, подтвердив возможность вывода C .
- Шаг 4. Система пытается установить D сначала в базе фактов, потом в результате поиска правил, приводящих к D . Обнаружено одно такое правило $G \wedge H \rightarrow D$. Система приходит к решению о том, что для получения вывода о присутствии D ей необходимо установить наличие G и H .
- Шаг 5. Система обнаруживает G и H в базе фактов, подтвердив возможность вывода D .
- Шаг 6. Система выполняет правило $E \rightarrow C$ и устанавливает факт C .
- Шаг 7. Система выполняет правило $G \wedge H \rightarrow D$ и устанавливает факт D .
- Шаг 8. Система выполняет правило $C \wedge D \rightarrow A$ и устанавливает факт A .

- На шаге 1 необходимо определить, какое правило сначала подтвердить: $C \wedge D \rightarrow A$ или $B \rightarrow A$, а на шаге 2 необходимо выбрать между $E \rightarrow C$ или $F \rightarrow C$.
- Если на каждом этапе логического вывода существует множество применяемых правил, то это множество носит название конфликтного набора, а выбор одного из них называется разрешением конфликта.
- Чтобы повысить эффективность продукционной системы, необходимо решить проблему управления последовательностью применения правил или управления выводом.

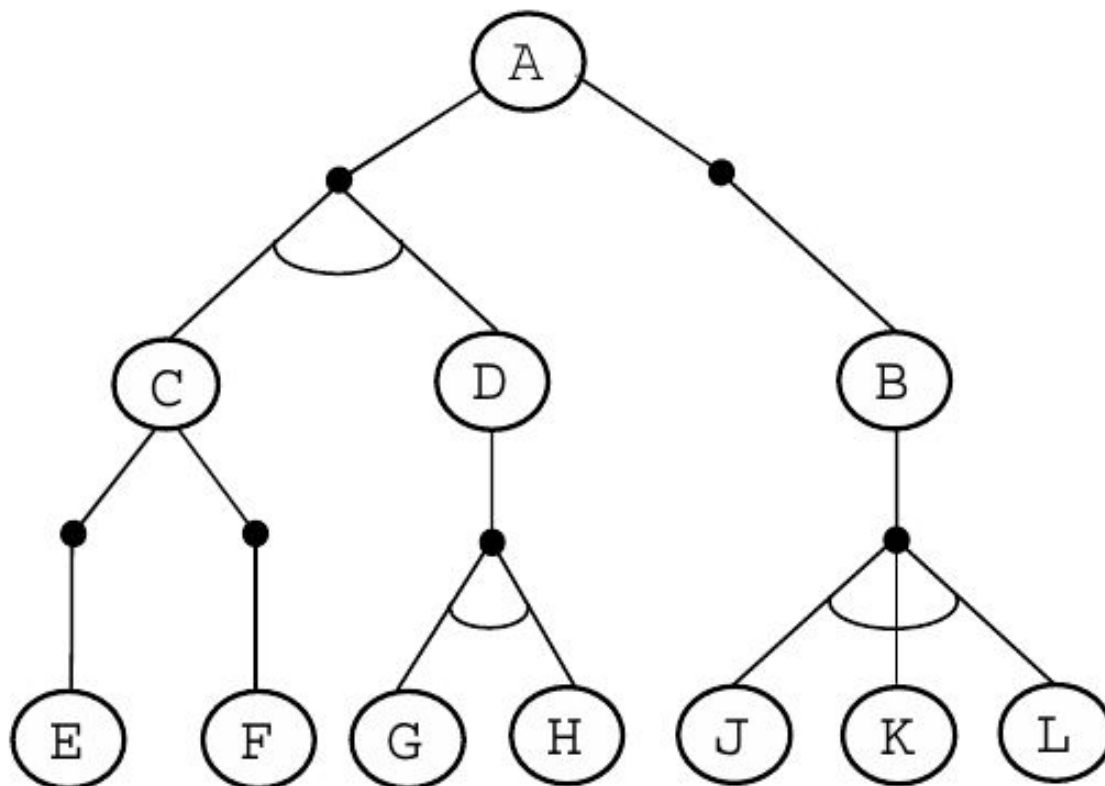
- Рассмотренные выше правила представляют собой отношение вывода, установленное между содержимым рабочей памяти, ссылкой на которое осуществляется из условной части (УЧ), и содержимым, указываемым в заключительной части (ЗЧ). Визуально такое отношение можно представить в виде графа с древовидной структурой



Здесь РП1, РП2, ..., РПn – данные, хранящиеся в рабочей памяти, к которым обращается данное правило.

Если существует множество правил, из которых выводится одно и то же заключение, то, выполнив операцию ИЛИ над всеми заключениями, получаемыми с помощью этих правил, можно показать отношение между результатами отдельного вывода и данными, на основании которых делается вывод

- Следовательно, если в такой форме представить отношение между всеми правилами продукционной системы и содержимым РП, то систему можно представить в виде одного графа типа И/ИЛИ. В самых нижних узлах этого графа будут располагаться основные системные данные, а в самом верхнем узле – заключения, выводимые системой. В итоге вывод, получаемый с помощью продукционной системы, можно представить как совокупность серии правил, поддерживающих отдельное заключение, и данных, на основании которых делается вывод.



- С помощью такого графа обратный вывод можно представить как проблему поиска пути на данном графе. Другими словами, для подтверждения одной цели из всех связей ИЛИ, определенных по отношению к узлам графа, соответствующим этой цели, выбирается одна, и делается попытка подтвердить все узлы, являющиеся предусловиями этой цели. В случае неудачи выбирается следующая связка ИЛИ и повторяется аналогичная процедура. Если обнаружится, что хотя бы одна из связок ИЛИ позволяет вывести последнюю цель, то доказательство этой цели считается успешным, в противном случае – неуспешным. Выбор одной из связок ИЛИ есть не что иное, как выбор одного из правил, таким образом, мы возвращаемся к проблеме разрешения конфликтов. Кроме того, определение последовательности оценки связок И, раскрывающих связки ИЛИ, соответствует последовательности оценки УЧ-правил.

- При выполнении условия применимости одновременно для нескольких продукции возникает дилемма выбора продукции или их группы (в случае возможности параллельной обработки), которая в данной ситуации будет активизирована в целях наискорейшего достижения поставленной цели. Решение этой задачи возлагается на систему активизации продукции

- **Принцип «стопки книг».** Основан на идее, что наиболее часто используемая продукция является наиболее полезной. Готовые продукции как бы образуют «стопку», в которой порядок определяется накопленной частотой использования продукции в прошлом. На самом верху «стопки» находится продукция, которая использовалась чаще всех. Из некоторого набора продукции с истинными условиями для исполнения выбирается та продукция или продукции, у которых частота использования максимальна. Подобный принцип управления особенно хорош, когда частота исполнения подсчитывается с учетом некоторой ситуации, в которой ранее применялась продукция, и это имело положительную оценку. При такой обратной связи метод стопки книг может превратиться в обучающуюся процедуру, адаптирующуюся к тем задачам, которые возникают во внешней среде. Управление по принципу стопки книг целесообразно применять, если продукции относительно независимы друг от друга.

- **Поиск в глубину.** При поиске решения в качестве очередной подцели выбирается та, которая соответствует следующему, более детальному уровню описания задачи. Например, диагностирующая система, сделав на основе известных симптомов предположение о наличии определенного заболевания, будет продолжать запрашивать уточняющие признаки и симптомы этой болезни до тех пор, пока полностью докажет или опровергнет выдвинутую гипотезу.
- **3. Поиск в ширину.** При поиске в ширину, напротив, система вначале анализирует условия (классификационные признаки) одного уровня детализации, а затем переходит к следующему (более детальному) уровню. Например, диагностирующая система вначале проанализирует все симптомы, находящиеся на одном уровне детализации, даже если они относятся к разным заболеваниям, и лишь затем перейдет к симптомам следующего уровня.

- **4. Принцип наиболее длинного условия** (разновидность поиска в глубину). Заключается в выборе из набора готовых продукций той, у которой стало истинным наиболее «длинное» условие выполнимости ядра. Этот принцип опирается на соображение «здравого смысла», что частные правила, относящиеся к узкому классу ситуаций, важнее общих правил, относящихся к широкому классу ситуаций, так как первые учитывают больше информации о ситуации, чем вторые. Трудность использования данного принципа состоит в том, что надо заранее упорядочить условия по вхождению друг в друга по отношению «частное — общее». Управление по этому принципу наиболее подходит для систем, в которых база знаний (набор продукции) образует иерархическое дерево со связями типа «частное — общее».

- **Принцип метапродукций.** Основан на идее ввода в систему продукций специальных метапродукций, задачей которых является организация управления при возможности неоднозначного выбора из набора готовых продукций. Приведем пример использования метапродукций:
 - ЕСЛИ «погода теплая» И «идет снег»
 - ТО продукции, у которых в А имеется «отдых на улице», следует активизировать раньше, чем продукции, содержащие в А «отдых в помещении»
 - В качестве условия, записанного в метапродукции, может выступать и некоторое утверждение об исключении определенных продукций из набора активизируемых продукций.
- **6. Принцип декомпозиции** (разбиения задачи на подзадачи). Подразумевает разбиение набора продукций на сферы применения.

- **Принцип приоритетного выбора.** Связан с введением статических или динамических приоритетов на продукции. Статические приоритеты могут формироваться на основании сведений о важности производственных правил в данной проблемной области. Эти сведения, как правило, представляют собой информацию, предоставляемую экспертами. Динамические приоритеты вырабатываются в процессе функционирования системы производств и могут отражать, например, такой параметр, как время нахождения продукции в наборе активных производств.

- **Управление по именам.** Основано на задании для имен продукций, входящих в некоторую систему, некоторой формальной грамматики или другой процедуры, обеспечивающей сужение фронта готовых продукций и выбор из него очередной продукции для выполнения.

- Например, пусть система продукций представлена четырьмя простейшими productions:
- (a) $A \rightarrow B$;
- (б) B и $D \rightarrow A$;
- (в) A или $B \rightarrow C$;
- (г) A и $D \rightarrow C$.
- Если выполняется **A**, то в набор активируемых продукций включает продукты с именами (a) и (в), а если выполняются **B** и **D**, то продукты с именами (б) и (в). Для устранения подобной недетерминированности может быть введена некоторая грамматика для имен продукций: (a) \square (в); (б) \square (в); (б) \square (г).