ГОСУДАРСТВЕННЫЙ КОМИТЕТ СССР ПО НАУКЕ И ТЕХНИКЕ
ИНСТИТУТ ПОВЫШЕНИЯ КВАЛИФИКАЦИИ ИНФОРМАЦИОННЫХ РАБОТНИКОВ

Е. В. Падучева

Математические методы в теории научно-технической информации

Часть I

Множества и отношения

ИПКИР

МОСКВА 1979

ГОСУДАРСТВЕННЫЙ КОМИТЕТ СССР ПО НАУКЕ И ТЕХНИКЕ
ИНСТИТУТ ПОВЫШЕНИЯ КВАЛИФИКАЦИИ ИНФОРМАЦИОННЫХ РАБОТНИКОВ

Е.В.Падучева

МАТЕМАТИЧЕСКИЕ МЕТОДЫ В ТЕОРИИ НАУЧНО-ТЕХНИЧЕСКОЙ ИНФОРМАЦИИ

Часть I

МНОЖЕСТВА И ОТНОШЕНИЯ

(Методическое пособие)

Москва
1979

УДК 519.76 + 510.22 (07)

Учебное пособие по курсу "Математические методы в теории научно-технической информации" представляет собой изложение основ математики в применении к задачам теории научно-технической информации.

Пособие состоит из двух частей. Автором первой части является канд.филол.наук Е.В.Падучева, второй - канд.техн.наук В.Б.Борщев.

В части I содержатся сведения об общих понятиях семиотики. Вводятся основные понятия теории множества. Большое внимание уделяется понятию отношения. Рассматриваются операции над отношениями и свойства отношений. Описываются многоместные отношения и их применение к фактографическим информационным поисковым системам. Излагаются применения теории отношений к описанию синтаксической структуры предложения в математической лингвистике.

Пособие предназначено для слушателей системы повышения квалификации информационных работников.

Предисловие

Если 20 лет назад математику преподавали в основном будущим инженерам и физикам (не считая, конечно, будущих математиков), то теперь ее изучают лингвисты и психологи, историки и социологи, биологи и экономисты. Преподают математику и в ИПКИРе. Курс "Математические методы в теории научно-технической информации" входит в программу почти всех специальностей института. Зачем же нужна математика информационным работникам?

Язык математики позволяет четко и ясно говорить об основных проблемах информатики. Без математики трудно понять, как работают информационные системы - как документальные, так и фактографические. Математика помогает увидеть, какие структуры лежат в основе классификаций, каким закономерностям подчиняются информационные потоки, как устроен язык - основной носитель информации.

В программу курса входят далеко не все разделы математики, используемые в информатике. Данное пособие, в свою очередь, не охватывает всей программы. Таким разделам программы, как элементы теории вероятностей, теория графов, теория алгоритмов предполагается посвятить следующие выпуски. Части I и II пособия включают следующие, тесно связанные друг с другом разделы: элементы теории множеств, теории отношений и математической логики. Кроме того в пособие включены разделы, которые формально не входят в программу. Это прежде всего раздел об основных поытиях семиотики. Понятия такого рода, как правило, предполагаются известными слушателями, и в лекциях они либо не обсуждаются, либо вводятся по ходу дела. Было решено посвятить им в пособии особый раздел, который выполняет роль введения. Разделы "Некоторые понятия математической лингвистики" и "Реляционная модель банков данных" (тоже не входящие в программу) включены в качестве подробных примеров того, как аппарат, введенный в основных разделах, "работает" в информатике. Кроме того, эти разделы могут быть полезны сами по себе: лингвистическое обеспечение играет все большую роль при создании автоматизированных информационных систем; банки данных - это тоже бурно развивающаяся область; банк данных является основой всякой современной фактографической информационной системы.

Основу данного пособия составил курс лекций, который авторы в течение ряда лет читали в ИПКИРе. Этот курс читали также Ю.А.Шрейдер, М.М.Новоселов и другие преподаватели. Когда авторы начали читать курс в институте, программа и содержание его уже в основном сложились. Первое пособие по этому курсу принадлежит Ю.А.Шрейдеру.*) Оно, однако, не охватывало некоторую часть программы - элементы теории множеств и основные понятия математической логики. Эти общеизвестные разделы, впрочем, содержатся почти в каждой книге по началам математики. Поэтому данное пособие ни в коей мере не претендует на оригинальность.

Основная задача, которую решали авторы настоящего пособия, - по возможности кратко и популярно изложить соответствующие понятия. При этом авторы старались использовать существующие материалы, заимствуя из них, в частности, структуру изложения, форму определений, примеры. Укажем основные источники, которые

*) Шрейдер Ю.А. Семиотические основы информатики. Лекции. 2-е изд. - М.: ИПКИР, 1975. - 80 с.

были использованы: Столл Р. Множества, логика. Аксиоматические теории. - М., Просвещение, 1968; Шиханович Ю.А. Введение в современную математику. - М.: Наука, 1965; Шрейдер Ю.А. Равенство, сходство, порядок. - М., Наука, 1971; Codd E.F. A Relational model of Data for Large Shared Data Banks.- Comm. of ACM, 1970, v. 13, № 6.

В.Б.Борщев, Е.В.Падучева

Глава I. ОСНОВНЫЕ ПОНЯТИЯ СЕМИОТИКИ

§ I. Знак и знаковая ситуация

Знак - это неопределяемое исходное понятие семиотики - науки о знаковых системах; знак - это объект или событие, который способен что-то о б о з н а ч а т ь , то есть указывать на некоторый другой объект, и (или) что-то о з н а ч а т ь, то есть иметь некоторый смысл. Таким образом, знак - это объект, который несет для его использователей какую-то информацию. Удобнее всего продемонстрировать содержание понятия знака на примерах.

Слово "Луна" обозначает конкретный физический объект и имеет смысл 'естественный спутник Земли'. Знак π обозначает число 3,14... и имеет смысл 'отношение длины окружности к ее диаметру'. Зеленый сигнал светофора означает: 'Движение разрешено'. Имя "Александр Сергеевич Пушкин" обозначает определенного человека, то есть делает возможным упоминание об этом человеке и сообщает определенную информацию о нем, в частности, сообщает, как зовут этого человека и как звали его отца. Сочетание "чемпион мира по шахматам" обозначает (по положению на 1979 г.) Анатолия Карпова и одновременно имеет определенный смысл. Сочетание "автор «Вэверлея»" обозначает Вальтера Скотта и какое-то время было его псевдонимом. Заглавие статьи, пьесы, музыкального произведения является знаком этого произведения. Библиотечный шифр книги является знаком этой книги: он однозначно указывает на книгу и несет определенную, хотя и специфическую информацию о ней.

Приведем примеры не-знаков. Звук [ф] не является сам по себе знаком, поскольку он ничего не означает (хотя звук [и] можно рассматривать как знак - это отдельное слово русского языка). Не каждый объект, сопоставленный каким-то образом другому объекту, является его знаком. Например, у человека обычно есть паспорт; но паспорт не является знаком человека, так как паспорт не используется в ситуации, когда надо всего лишь упомянуть о человеке. То же самое касается других документов - квитанций или гардеробных номерков. 37 копеек не являются знаком бутылки пива, хотя и находятся с ней в определенных соотношениях. Реферат статьи в нормальном случае не является знаком этой статьи, потому что когда говорят о какой-то статье, обычно пользуются ее названием или какой-то описательной характеристикой, но не рефератом (хотя реферат сам по себе является знаком, как и любой языковый текст: он описывает некоторую внеязыковую ситуацию и несет определенный смысл).

С понятием знака непосредственно связаны понятия денотата и концепта знака. Объект, обозначаемый данным знаком, называется его <u>денотатом</u>. Денотат знака вовсе не обязательно является конкретным физическим объектом, имеющим определенные координаты в пространстве и времени. Например, денотат числа π - весьма абстрактная сущность, да и денотат числа 2 - тоже.

Как правило, объект сам по себе не предопределяет однозначно знака, которым он может быть обозначен; иначе говоря, одному объекту может соответствовать много разных знаков. Примеры разных знаков одного объекта:

(I) Пушкин; автор "Евгения Онегина"; первый муж Натальи Николаевны Гончаровой;

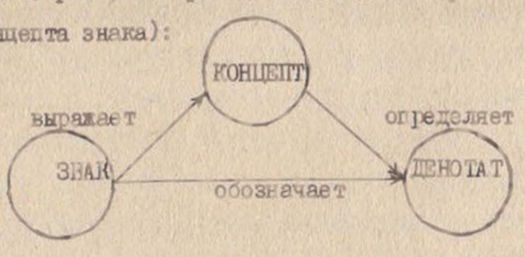
(2) Товарищ Иванов; староста этой группы, человек, сидящий у двери (здесь, разумеется, речь идет о знаках, употребленных в некоторой конкретной ситуации);

(3) Точка пересечения некоторых двух медиан треугольника; точка пересечения другой пары медиан того же треугольника; точке пересечения всех трех медиан этого треугольника;

(4) Один и тот же объект обозначают выражения: $2 + 2$, 2×2, 2^2, $3 + 1$ и $10 - 6$.

Каждый новый знак данного предмета обычно не просто указывает на тот же предмет, а одновременно выявляет какой-то новый аспект предмета, выделяет в нем какой-то иной признак или свойство. Поэтому говорят, что с каждым знаком связан определенный <u>концепт</u>, иначе говоря - смысл знака. Концепт знака - это то, что сохраняется, когда мы по той или иной причине заменяем данный знак предмета на какой-то иной, но возможно более близкий по смыслу. В частности, концепт знака - это то, что можно переводить с одного языка на другой: это то общее, что есть в разных переводах знака. По определению немецкого философа Г.Фреге, концепт знака - это способ, которым знак указывает на свой денотат.

Отношения между знаком, его концептом и денотатом выражает так называемый <u>семиотический треугольник</u>, или <u>треугольник Фреге</u> (по имени Г.Фреге, который доказал логическую необходимость в понятии концепта знака):



Знак <u>выражает</u> свой концепт и <u>обозначает</u> (или называет, указывает, именует) свой денотат. Про концепт говорят, что он <u>опре-</u>

делает свой денотат.*)

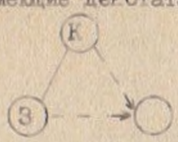
Иногда концепт и денотат, вместе взятые, называют нерасчлененно <u>содержанием</u> (или <u>значением</u>) знака.

Знаки бывают <u>простые</u> (элементарные) и <u>сложные</u>: осмысленное сочетание знаков - это тоже знак, но сложный.

Быть знаком - это для одних объектов постоянное и единственное назначение, а для других - временная функция. Красная лампа, которая служила знаком опасности на дороге, может быть принесена домой и использована при проявлении фотографий, то есть уже не в качестве знака. Когда знак реально вступает во взаимодействие со своим денотатом и концептом, возникает <u>знаковая ситуация</u>.

Наряду с обычными знаками, то есть знаками, имеющими и концепт и денотат, достаточно часто встречаются вырожденные знаки, иначе говоря, возможны знаковые ситуации, в которых одна из вершин семиотического треугольника отсутствует. Рассмотрим основные типы таких ситуаций.

1. Знаки, не имеющие денотата:



Примерами таких знаков являются: "Пегас", "Единорог" (и вообще все имена сказочных и мифологических существ, героев литературных произведений); выражения "нынешний король Франции", "наибольшее натуральное число", целые корни уравнения $2x = 1$, "небесное тело, наиболее удаленное от Земли".

---

*) Заметим, что в лингвистике, начиная с основателя структурализма Ф.де Соссюра, принято считать знаком двустороннюю единицу, состоящую из собственно знака и его концепта; то есть языковой знак - это пара ⟨ знак, концепт ⟩ (в терминах Соссюра, это пара ⟨ означающее, означаемое ⟩).

книг, стен интерпретируются как содержащие нулевой знак родительного падежа множественного числа. Аналогично на фоне отглагольных существительных на -ение, -ание - разложение, восстановление, собирание, убывание - слова приход, вывоз, дрожь, стук воспринимаются как имеющие нулевой суффикс существительного. В реферативном журнале нулевой реферат, то есть реферат, исчерпывающийся библиографическим описанием, может быть воспринят как нулевой знак, смысл которого - статья бессодержательна. Нулевой знак возможен только в составе знаковой системы (см. § 3).

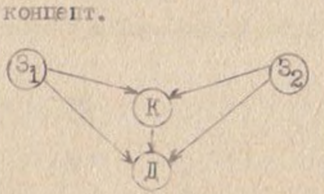
§ 2. Синонимия, омонимия и автонимия знаков

Соответствие между знаками, концептами и денотатами вовсе не всегда является взаимнооднозначным. Поскольку, по мысли Г.Фреге, знак обозначает денотат через посредство своего концепта, мы будем ниже изображать треугольник Фреге "вытянутым" в цепь (З, К, Д - это соответственно знак, концепт, денотат):

При отсутствии взаимнооднозначного соответствия между знаком, концептом и денотатом могут возникать следующие соотношения.

I. Синонимия знаков

Знаки $З_1$ и $З_2$, называются синонимичными, если они выражают один и тот же концепт.



Примеры. Знак & "конъюнкция" в логике синонимичен знаку $\wedge$, а знак $\leftrightarrow$ "эквивалентность" – знаку $\equiv$; знаки $\cdot$, $\times$ и нулевой знак синонимичны в алгебре. Знак "множество всех четных чисел" синонимичен знаку "множество всех чисел, которые делятся на 2". Синонимичны знаки "мой дядя по отцу" и "брат моего отца".

2. Денотативное тождество знаков

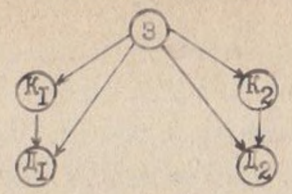
Знаки $З_1$ и $З_2$, каждый со своим концептом, называются <u>денотативно тождественными</u>, если они обозначают один и тот же денотат:

По А.Черчу, такие знаки различаются "способом, которым они указывают на свой денотат". Например, знаки $\{0, 1, 2, 3\}$ и $\{x \mid x$ – натуральное число и $x < 4\}$ денотативно тождественны. Денотативно тождественны знаки "А.С.Пушкин" и "автор «Евгения Онегина»", $\{x \mid$ ЧЕЛОВЕК $(x)\}$ и $\{x \mid$ БЕСПЕРОЕ ДВУНОГОЕ $(x)\}$, "Москва" и "Столица СССР", "жена царя Федора" и "сестра Бориса Годунова", "9" и "число планет солнечной системы".

Денотативное тождество знаков следует отличать от синонимии. Фразы "Лектор стоит спиной к аудитории" и "Лектор стоит лицом к доске" могут быть денотативно тождественны, то есть обозначать одну и ту же ситуацию; но они не синонимичны. Выражения $\frac{2 \cdot 10 - 10}{2}$ и 5 обозначают одно и то же число, но не являются синонимами.

3. Омонимия знака

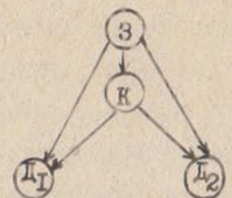
Знак З называется <u>омонимичным</u> (или многозначным), если он может выражать более чем один концепт:



Например, ! - это восклицательный знак и знак факториала. Знак "Москва" обозначает (в Москве) кинотеатр, ресторан, духи, гостиницу и т.д. Слово "соль" - название пищевого продукта и ноты. Слово "симметричный" обозначает свойство (например, фигура может быть симметричной) и отношение (например, точка может быть симметрична другой точке относительно некоторой оси). Омонимичны такие слова, как установка, организация, сопротивление, разряд, простой.

4. Денотативная неоднозначность знака

Знак З называется денотативно неоднозначным, если он, при одном и том же концепте, может обозначать различные денотаты:



Денотативная неоднозначность свойственна всем так называемым указательным знакам - я, ты, здесь, сейчас, этот стол и т.д., - в концепт которых входит отсылка к речевой ситуации или к ее участникам. Например, знак я обозначает в разных случаях разных людей, имея при этом один и тот же смысл: я - 'тот человек, который произносит в данный момент слово я'. Соответственно, денотативно неоднозначными будут все сложные знаки, включающие указательный знак в свой состав, например, мой сосед справа,

отец (= 'мой отец'), дом родной (= 'мой дом родной').

Денотативную неоднозначность знаков не следует смешивать с омонимией.

5. Автонимное употребление знака

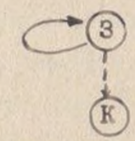
Знаки могут иметь помимо обычного употребления автонимное. Знак З употреблен автонимно, если денотатом знака З в данной знаковой ситуации является сам знак З.

Например, во фразе "Хлеб - существительное мужского рода" слово хлеб обозначает не пищевой продукт, а само слово хлеб. В формальных языках автонимное употребление обычно избегается, и если надо указать, что денотатом слова является оно само, то слово выделяется курсивом или ставится в кавычки.

Автонимно употребленными могут быть целые предложения, ср.: "Волга впадает в Каспийское море - это пример тривиальной истины", "$2 \times 2 = 5$ - ложное высказывание".

Когда знак обозначает сам себя, его нельзя заменить на синоним без изменения смысла текста или предложения. Например, предложение "& похоже на 𝄞" истинно; а предложение "$\wedge$ похоже на 𝄞" ложно, хотя знаки & и $\wedge$ синонимичны. Предложение "Окулист - слово иностранного происхождения" истинно, а предложение "Глазник - слово иностранного происхождения" ложно.

Треугольник Фреге в случае автонимного употребления знака принимает вид:



Здесь концепт перестает быть посредником при отнесении знака к денотату.

§ 3. Знаковая система. Синтактика, семантика, прагматика

Знаковая система - это множество знаков, где имеются определенные регулярные соотношения между знаками, отражающие регулярные же соотношения между их концептами (соответственно, денотатами). Как правило, люди пользуются не изолированными знаками, а именно знаковыми системами.

Один из основателей семиотики, Ч.Моррис, выделил в знаковой системе три аспекта - это синтактика, семантика и прагматика. Каждый из этих аспектов изучается соответствующим разделом семиотики (рис.1).

I. Задачей синтактики является изучение внутреннего устройства знаковой системы - правил построения сложных знаков из простых. Для языковых знаковых систем максимальным (то есть максимально сложным) знаком раньше считалось предложение: сочетание предложений уже не рассматривалось как единый знак. Однако сейчас в лингвистику входит теория связного текста, для которой максимальным знаком является текст.

Если рассматривать естественный язык как знаковую систему, то для нее синтактика - это синтаксис. Задача синтаксиса состоит в том, чтобы определить понятие правильно построенного предложения. Это определение можно дать с помощью порождающей грамматики (см. главу IУ).

В искусственных языках, например, в логике, синтактика определяет, что такое правильно построенная формула. Правильно построенное выражение - это значит потенциально осмысленное.

II. Задача семантики - изучение соотношения, с одной стороны, между знаками и их денотатами и с другой - между знаками и их концептами (смыслами). До последнего времени логическая семантика занималась преимущественно проблемами I-го рода, а проблемами 2-го рода занималась в основном лингвистическая семантика: ее основная задача - описание законов, по которым из смысла отдельных знаков, то есть смысла слов и конструкций, складывается смысл целого, то есть смысл предложения или текста. В настоящее время логическое и лингвистическое понимание семантики расширяются и сближаются одно с другим. При описании смысла в лингвистике исходным является понятие синонимии знаков разной степени сложности - от минимальных знаков до целых текстов. Само понятие смысла определяется как инвариант синонимических преобразований.

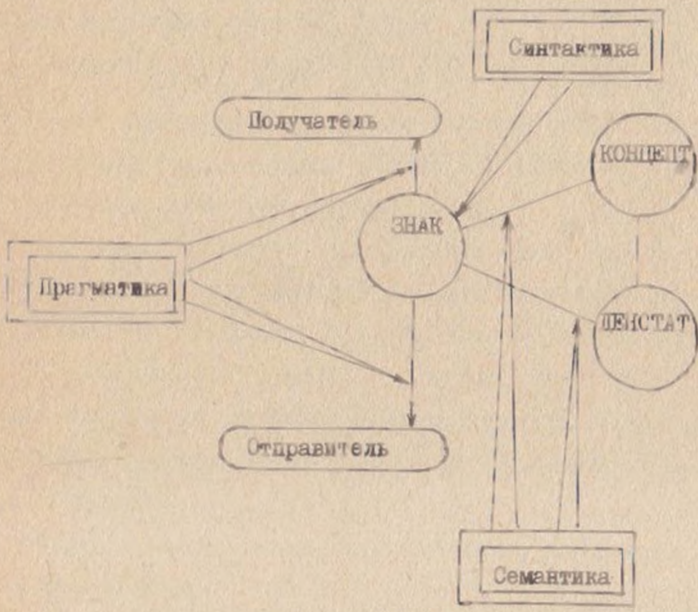


Рис. I

III. Прагматика изучает знаки с точки зрения их отношения к отправителю и адресату сообщений. Слова лицо и рожа соответствуют одному и тому же классу объектов, но различаются с прагматической точки зрения - второе отражает отрицательное отношение говорящего к упоминаемому предмету; аналогичное различие между словами <u>шпионаж</u> и <u>разведка</u>.

Другой прагматический аспект знака - степень его информативности для адресата. Смысл текста не зависит от адресата; он существует объективно. Между тем информативность текста не является его объективной характеристикой: информативность текста для данного адресата зависит от количества знаний, которыми он уже располагает. Информативность текста может быть равна нулю в любом из следующих случаев: 1) если текст известен адресату; 2) если текст непонятен адресату, поскольку тот не имеет необходимых знаний; 3) если текст бессмыслен для данного адресата, поскольку входит в противоречие с его знаниями.

Глава II. ЭЛЕМЕНТЫ ТЕОРИИ МНОЖЕСТВ

§ 1. Основные понятия теории множеств

Множество - это исходное и, тем самым, неопределяемое понятие теории множеств. Его смысл можно пояснить так: <u>множество</u> - это собрание предметов, которые мыслятся как единое целое, как один предмет. Многие объекты, с которыми мы имеем дело, являются множествами. Так, можно говорить о множестве планет Солнечной системы; множестве студентов, обучающихся в данный момент в данном институте; множестве людей в данной комнате; множестве всех целых чисел; множестве всех точек отрезка и т.д.

Само слово "множество" означает примерно то же, что "совокупность". В языке есть много других слов, близких по смыслу к слову множество: "собрание", "класс", "система", "семейство", "комплекс", ансамбль", "коллекция", "группа". Имеются также специальные названия для множеств различной конкретной природы: "стадо" (животных), "отара" (овец), "стая" (птиц), "букет" (цветов), "массив" (документов).

Объект, входящий в некоторое множество, называется <u>элементом</u> этого множества.

Обозначение. $x \in A$ читается: x - <u>элемент</u> множества A, или x <u>входит</u> в A, или x <u>принадлежит</u> A. Знак $\in$ обозначает, таким образом, отношение <u>принадлежности</u> элемента множеству. Выражение $x \notin A$ читается: x не принадлежит A.

Требования, предъявляемые к множествам

1. Требование определенности состава. В математике говорят о множествах только в тех случаях, когда про каждый элемент можно с определенностью сказать, принадлежит он множеству или нет. Так, нельзя говорить, без дальнейших уточнений, о множестве всех слов русского языка, поскольку не ясно, принадлежат ли этому множеству, скажем, такие элементы, как о'кэй, пирла, АЛГОЛ, компьютер, масс медиа, намакаронился, двадцатипятитысячник и т.д.

2. Требование различимости элементов. Про каждые два объекта надо знать, являются ли они одним и тем же элементом множества или разными. Так, в предыдущем примере множества всех слов русского языка надо уточнить, являются ли, скажем, стол и столы одной и той же единицей нашего словаря или двумя разными; аналогично завязать – завязывать, строить – строиться и т.д. Чтобы говорить о множестве всех слов фразы Глухой глухого звал на суд судьи глухого, надо уточнить: 1) считаем ли мы одним и тем же словом разные словоформы и 2) считаем ли мы одним и тем же словом прилагательное глухой и глухой – существительное. Еще пример. Нельзя говорить о множестве всех букв слова молоко, не уточнив, что именно имеется в виду. С одной стороны, можно построить такое множество, в котором будет четыре элемента; с другой – в слове молоко, как известно, шесть букв, так что можно построить и такое множество, в котором будет шесть элементов (в частности, три о – о первое, о второе, о третье).

Специально отметим, что для множества несуществен порядок его элементов. Поэтому, например, слово не является множеством букв: вол и лов – разные слова, хотя буквы в них одни и те же; предложение не является множеством слов: У попа была собака – предложение, а Собака попа была у – нет.

Способы задания множеств

Как указать, о каком множестве идет речь? Это можно сделать двумя способами.

1-й способ задания множества - перечислением элементов.

Обозначение. $A=\{a_1, a_2, \ldots, a_k\}$ читается: множество A состоит из элементов $a_1, a_2, \ldots, a_k$.

Например, $A = \{1, 2, 3, 4, 5\}$ - это множество школьных отметок. Поскольку порядок элементов несущественен, $\{1,2,3,4,5\} = \{5, 4, 3, 2, 1\} = \{1, 5, 2, 4, 3\}$ и т.д.

Другой пример. A = {Иванов, Петров, Сидоров}. Это может быть, например, множество всех отличников 1-го "А" класса школы № 333 г. Москвы. Перечислением может быть задано, скажем, множество всех городов Украинской ССР; множество всех гласных звуков русского языка; множество всех дескрипторов в информационно-поисковом языке (ИПЯ) некоторой поисковой системы.

Во всех этих примерах фигурируют конечные множества. Часто приходится, однако, иметь дело с бесконечными множествами: множество всех положительных чисел; множество всех точек отрезка; множество всех параллелограммов и т.д. Бесконечные множества, естественно, не могут быть заданы перечислением элементов.

2-й способ задания множеств - характеристическим свойством, то есть указанием свойства, которым все элементы данного множества обладают, а все остальные объекты нет.

Характеристическое свойство всех положительных чисел состоит в том, что они больше нуля; характеристическое свойство всех четных чисел - в том, что они делятся на 2 и т.д.

Обозначение. $A=\{x \mid P(x)\}$ читается: A есть множество всех тех и только тех элементов x, которые обладают характеристическим свойством P.

Так, множество N_1 всех положительных чисел может быть задано следующим образом:

$$N_1 = \{x \mid x > 0\}$$

Множество N_2 всех четных чисел задается так:

$$N_2 = \{x \mid x \text{ делится на } 2\}.$$

Разумеется, и конечное множество можно, и иногда удобнее, задавать характеристическим свойством, а не перечислением. Например, $\{1, 2, 3, 4, 5\} = \{x \mid x \text{ целое и } 0 < x < 6\}$.

Два множества называются <u>равными</u> (или <u>совпадающими</u>), если они состоят из одних и тех же элементов. Бывает так, что разные характеристические свойства задают равные множества (то есть одно и то же множество).

Примеры равных множеств.

1) A = $\{x \mid x$ лежит на биссектрисе угла $\varphi\}$; B = $\{x \mid x$ равноудален от сторон угла $\varphi\}$; A = B.

2) A = $\{x \mid x$ – четырехугольник, и у x все углы прямые$\}$; B = $\{x \mid x$ – четырехугольник, и у x три угла прямые$\}$; A = B.

3) $A = \{1,2,3,4,5\}$; $B = \{5,4,3,2,1\}$; A = B.

<u>Виды множеств</u>

В теории множеств рассматриваются такие виды множеств, которые определяются не конкретной природой элементов, а некоторыми общими их свойствами. Различаются следующие виды множеств.

1) <u>Конечное множество</u>.

2) <u>Бесконечное множество</u>.

3) <u>Пустое множество</u>. Это множество, не содержащее ни одного элемента. Необходимость в рассмотрении пустого множества возникает, например, в связи с тем, что можно сформулировать такое характеристическое свойство, которым не обладает ни один объект, например:

- множество всех четырехугольников, у которых три угла прямые, а четвертый - острый;

- множество значимых слов русского языка, не содержащих гласной;

- $\{x \mid x$ - целое число, и $0 < x < 1\}$.

Необходимость в пустом множестве возникает также при определении операций над множествами, см. § 2.

Обозначение для пустого множества - $\varnothing$.

4) Единичное множество. Это множество, состоящее из одного элемента. Единичное множество следует отличать от элемента множества. Например, библиотека, состоящая из одной книги, - это иной объект, чем одноименная книга в составе библиотеки. Другие примеры единичных множеств: багаж пассажира (в общем случае - множество вещей), состоящий из одного чемодана; словарь (в общем случае - множество слов) попугая, состоящий из одного слова и т.д. Полезно обратить внимание на следующее различие. Пусть имеются книги K_1 и K_2; тогда $\{K_1, K_2\}$ - это библиотека, состоящая из двух книг, а $\{\{K_1\}, \{K_2\}\}$ - это множество библиотек.

5) Универсальное множество. Это множество всех объектов, которые в данный момент входят в рассмотрение. В разных ситуациях универсальные множества будут разные. Например, в арифметике универсальное множество - это множество всех чисел; в геометрии - множество всех точек и т.д.

Обозначение для универсального множества - U.

Подмножества

Определение. Множество A называется подмножеством множества B, если всякий элемент A является одновременно элементом B.

Обозначение. $A \subseteq B$ читается: множество A является подмножеством множества B; или иначе: множество A нестро-

2-4

го включено в множество В. Знак $\subseteq$ построен по аналогии со знаком $\leq$ 'меньше или равно' для чисел (ср. $x \leq y$).

Примеры. 1) Пусть С – множество всех ваших сыновей, а Д – множество всех ваших детей; тогда $С \subseteq Д$.

2) Пусть Х – множество всех хищников, П – множество всех позвоночных, М – множество всех млекопитающих; тогда $Х \subseteq П \subseteq М$.

3) Пусть К – множество всех квадратов, ПР – множество всех прямоугольников, ПА – множество всех параллелограммов, Ч – множество всех четырехугольников. Тогда $К \subseteq ПР \subseteq ПА \subseteq Ч$.

Наряду со знаком $\subseteq$ употребляют знак $\supseteq$. Запись $A \supseteq B$ читается: А нестрого включает В. Естественно, если $A \subseteq B$, то $B \supseteq A$.

Как легко видеть, для любого множества А имеет место $A \subseteq A$. Если имеет место одновременно $A \subseteq B$ и $B \subseteq A$, это означает, что $A = B$. Если $A \subseteq B$ и $A \neq B$, то пишут $A \subset B$ (читается: множество А строго включено в множество В). $A \supset B$ читается: А строго включает В.

Следует обратить внимание на различие между знаком $\in$ принадлежности элемента множеству и знаком $\subseteq$ нестрого включения множества в множество. Так, если $A = \{1, 2, 3, 4, 5\}$, то $1 \in A$, но неверно, что $1 \subseteq A$. Отметим следующее различие знаков $\in$ и $\subseteq$. Если $A \subseteq B$ и $B \subseteq C$, то $A \subseteq C$, ср. выше пример 2. Но если $a \in B$, а $B \in C$, то отсюда не следует, что $a \in C$. Например, представим себе ситуацию в мебельном магазине, где шкаф a входит в состав гарнитура (множества объектов) В, т.е. является элементом множества В, а В входит в множество объектов С, поступивших в продажу. В этой ситуации шкаф a не принадлежит множеству С (то есть не может быть сам по себе куплен).

§ 2. Операции над множествами

Операции над множествами позволяют по двум множествам строить однозначным образом третье множество, являющееся результатом данной операции (аналогично тому, как операция над числами, например, сложение, состоит в том, что по двум числам находится третье, которое является их суммой).

Следует различать название операции и название результата применения этой операции к заданным множествам (ср. в арифметике: сложение – сумма, умножение – произведение и т.д.), хотя иногда они будут называться одним и тем же словом.

Пусть множества A и B являются подмножествами универсального множества $\mathcal{U}$ (т.е. $A, B \subseteq \mathcal{U}$).

I. Операция <u>объединения</u>. Результат операции – <u>объединение</u>, или <u>сумма</u> множеств. О б о з н а ч е н и е. $A \cup B$ читается: объединение множеств A и B.

О п р е д е л е н и е. $A \cup B = \{x \mid x \in A$ или $x \in B$ (или $x \in A$ и $x \in B$ одновременно)$\}$ (читается: <u>объединением</u> множеств A и B является множество всех тех и только тех элементов x, которые принадлежат или множеству A, или множеству B, или обоим).

<u>Примеры</u>. 1) $\{1, 2, 3\} \cup \{2, 3, 4\} = \{1, 2, 3, 4\}$

2) МАЛЬЧИКИ $\cup$ ДЕВОЧКИ = ДЕТИ.

3) $\{x \mid x$ делится на 2$\} \cup \{x \mid x$ делится на 5$\} = \{x \mid x$ делится на 2 или на 5 или на оба числа сразу$\}$.

4) $\{x \mid x > 2\} \cup \{x \mid x > 5\} = \{x \mid x > 2\}$.

5) ПОЭТЫ $\cup$ АРТИСТЫ КИНО = $\{x \mid x$ поэт или артист кино$\}$.

II. Операция <u>пересечения</u>. Результат операции – <u>пересечение</u> множеств. О б о з н а ч е н и е. $A \cap B$ читается: пересечение множеств A и B.

О п р е д е л е н и е. $A \cap B = \{x \mid x \in A$ и $x \in B\}$ (читается: пересечением множеств A и B является множество всех тех и только тех элементов x, которые принадлежат одновременно множеству A и множеству B).

Примеры. I) $\{1, 2, 3\} \cap \{2, 3, 4\} = \{2, 3\}$.

2) МАЛЬЧИКИ $\cap$ ДЕВОЧКИ $= \varnothing$.

3) $\{x \mid x$ делится на $2\} \cap \{x \mid x$ делится на $5\} =$ $= \{x \mid x$ делится на $10\}$.

4) $\{x \mid x > 2\} \cap \{x \mid x > 5\} = \{x \mid x > 5\}$.

5) ПОЭТЫ $\cap$ АРТИСТЫ КИНО $= \{$ Евтушенко, Высоцкий, ...$\}$.

Ш. Операция вычитания. Результат операции – разность множеств. Обозначение. $A \setminus B$ читается: разность множеств A и B.

Определение. $A \setminus B = \{x \mid x \in A$ и $x \notin B\}$ (читается: разностью множеств A и B является множество, состоящее из тех и только тех элементов x, которые принадлежат A и не принадлежат B).

Примеры. I) $\{1, 2, 3\} \setminus \{2, 3, 4\} = \{1\}$.

2) ДЕТИ $\setminus$ МАЛЬЧИКИ = ДЕВОЧКИ

Операция вычитания, в отличие от объединения и пересечения, не коммутативна: $A \setminus B \neq B \setminus A$; так, $\{2, 3, 4\} \setminus \{1, 2, 3\} = \{4\}$.

IУ. Операция перехода к дополнению. Результат операции – дополнение множества. Обозначение. $\overline{A}$ читается: дополнение множества A.

Определение. $\overline{A} = \{x \mid x \in U$ и $x \notin A\}$ (читается: дополнением множества A является множество всех тех элементов x, которые принадлежат универсальному множеству и не принадлежат A).

Примеры. I) Пусть A – множество всех четных чисел. Тогда $\overline{A}$ – множество нечетных чисел (если U – множество всех целых чисел).

2) $\overline{\text{МОСКВИЧИ}}$ = НЕМОСКВИЧИ (если U – множество людей; если U – множество всех объектов вообще, то $\overline{\text{МОСКВИЧИ}}$ = НЕМОСКВИЧИ $\cup$ НЕ ЛЮДИ).

3) $\overline{\{0\}} = \{x \mid x > 0$ или $x < 0\}$ (если U – числа).

§ 3. Взаимное расположение множеств

Пусть A и B – подмножества некоторого универсального множества U, т.е. $A, B \subseteq U$. Имеется всего пять в принципе возможных соотношений множеств A и B друг с другом:

1) A строго включено в B ($A \subset B$); т.е. если $x \in A$, то $x \in B$, и существует $y \in B$ такой, что $y \notin A$. Например, $A \subset B$ имеет место, если A – множество четных чисел, а B – множество всех целых чисел.

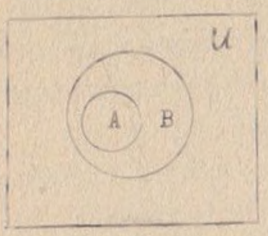
2) B строго включено в A ($B \subset A$); т.е. если $x \in B$, то $x \in A$ и существует $y \in A$ такой, что $y \notin B$. Например, $B \subset A$, если B – множество всех квадратов, а A – множество всех четырехугольников.

3) A и B частично пересекаются; то есть существует элемент x такой, что $x \in A$ и $x \in B$, существует $y \in A$ такой, что $y \notin B$, и существует $z \in B$ такой, что $z \notin A$. Например, пусть A – множество всех поэтов, а B – множество всех переводчиков; тогда, очевидно, A частично пересекается с B.

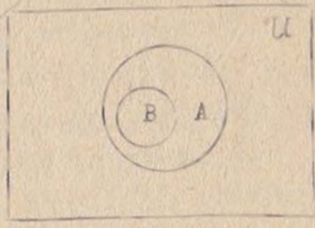
4) A и B не пересекаются ($A \cap B = \varnothing$); то есть не существует элемента x такого, что $x \in A$ и $x \in B$. Непересекающимися являются, например, множество школьников и множество студентов; множество стран Азии и множество стран Южной Америки.

5) A и B равны ($A = B$); то есть если $x \in A$, то $x \in B$, и если $x \in B$, то $x \in A$. Например, если $A = \{1, 2, 3, 4, 5\}$, а $B = \{5, 4, 3, 2, 1\}$, то A и B равны.

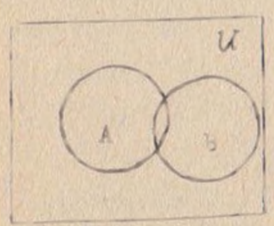
Взаимное расположение множеств удобно представить с помощью так называемых <u>диаграмм Венна</u>: универсальное множество изображается в виде прямоугольника, а множества A и B – в виде кругов; предполагается, что элементам множеств соответствуют точки плоскости. Диаграммы Венна для пяти случаев взаимного расположения множеств представлены на рис. 2.



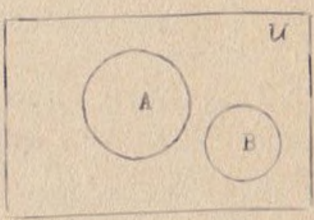
I) $A \subset B$



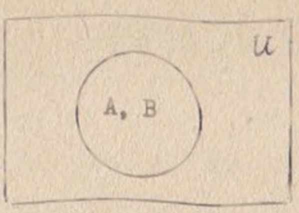
2) $B \subset A$



3) A частично пересекается с Б



4) $A \cap B = \varnothing$



5) A = B

Рис. 2

§ 4. Разбиения и покрытия

Определение. Разбиением множества A называется система (то есть множество) $\mathcal{P}$ непустых подмножеств $\{A_I, A_2, \ldots\}$ множества A такая, что:

I) $A_I \cup A_2 \cup \ldots = A$ (иначе говоря, любой элемент $x \in A$ входит по крайней мере в одно из подмножеств $A_i \in \mathcal{P}$);

2) для любых двух различных подмножеств A_i и $A_j \in \mathcal{P}$ $A_i \cap A_j = \emptyset$ (иначе говоря, ни один элемент $x \in A$ не входит одновременно в два подмножества).

Например, разбиением множества A славянских языков будет система $\mathcal{P}$, состоящая из подмножеств A_I - восточнослаянские языки, A_2 - западнославянские языки и A_3 - южнославянские языки.

Подмножества A_I, A_2, ... из разбиения $\mathcal{P}$ называются обычно классами разбиения.

Если от системы $\mathcal{P}$ требуется, чтобы она обладала свойством I, но не требуется, чтобы она обладала свойством 2, то она называется покрытием множества A.

Например, если выделить в множестве A всех целых чисел подмножество A_I четных чисел, A_2 нечетных чисел и A_3 - положительных чисел, то эта система подмножеств будет покрытием множества A; она не будет разбиением, так как, например, число 2 входит одновременно в A_I и A_3.

§ 5. Пример использования операций над множествами при формулировке поискового предписания в ИПС дескрипторного типа

В информационно-поисковой системе (ИПС) дескрипторного типа применяется координатное индексирование документов, составляющих информационный массив системы. А именно, каждому документу $t \in T$ (где T - множество документов, составляющих информационный массив) приписывается набор (неупорядоченный) дескрипторов из множества Д = $\{d_1, \dots, d_n\}$ дескрипторов ИПЯ, используемого в данной системе. Этот набор дескрипторов является <u>индексом</u>, который характеризует содержание документа. Например, если документу $t \in T$ приписан набор дескрипторов

$$\{\text{БИБЛИОТЕКА, АВТОМАТИЗАЦИЯ}\} ,$$

это значит, что в t идет речь об автоматизации библиотек; набор

$$\{\text{ПРОИЗВОДСТВО, САМОЛЕТЫ, ДВИГАТЕЛИ}\} ,$$

приписанный документу t, означает, что в t идет речь о производстве самолетных двигателей, и т.д.

Выбрав некоторый порядок (произвольный, но фиксированный) для множества Т всех документов информационного массива и для множества Д дескрипторов данной ИПС, можно построить следующую таблицу (Т = $\{t_1, \dots, t_m\}$; Д = $\{d_1, \dots, d_n\}$):

	d_1	d_2	d_3		...	d_n
t_1		×	×		...	×
t_2			×		...	
t_3	×	×			...	
⋮	⋮	⋮	⋮			⋮
t_m			×			

В таблице каждая строка задает индекс соответствующего документа. Например, документ t_1 имеет индекс $\{d_2, d_3 \dots d_n\}$. Каждый столбец задает <u>инверсное множество</u> соответствующего дескриптора, то есть множество всех документов информационного массива, которые имеют в составе своего индекса данный дескриптор. Если обозначить инверсное множество дескриптора d через

$[d]$, то, например, $[d_3] = \{t_1, t_2, \dots, t_m\}$. Для любого $d_i \in Д$ $[d_i] \subseteq Т$, и $[d_1] \cup [d_2] \cup \dots \cup [d_n] = Т$.

Легко видеть, что любой содержательный запрос, то есть требование о выдаче документов информационного массива, имеющих то или иное содержание, можно перевести в поисковое предписание, которое будет представлять собой задание некоторого множества документов через теоретико-множественные операции над инверсными множествами исходного информационного массива.

Примеры.

Запрос "Выдать документы по автоматизации научно-технических библиотек" переводится в предписание: "Найти множество

[АВТОМАТИЗАЦИЯ] $\cap$ (БИБЛИОТЕКИ] $\cap$ [НАУКА И ТЕХНИКА]",

где [АВТОМАТИЗАЦИЯ] - инверсное множество дескриптора АВТОМАТИЗАЦИЯ и т.д.

Запрос "Выдать документы о перфокартах ручного обращения" переводится в предписание: "Найти множество

[КРАЕВЫЕ ПЕРФОКАРТЫ) $\cup$ [ЩЕЛЕВЫЕ ПЕРФОКАРТЫ] [СУПЕРПОЗИЦИОННЫЕ]".

Запрос "Выдать документы, касающиеся иностранных авиалиний в Азии" переводится в предписание: "Найти множество

([АВИАЦИОННЫЕ ЛИНИИ] $\cap$ [АЗИЯ]) $\setminus$ [СССР]".

Формулировка поискового предписания в виде теоретико-множественных операций над инверсными множествами дескрипторов одновременно а) отражает смысловое содержание вопроса; б) является точной инструкцией для отыскания множества документов, которые соответствуют запросу.

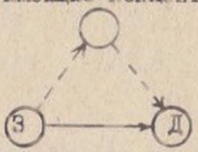
Упражнения

1. Представить с помощью диаграмм Венна операции объединения, пересечения и вычитания для каждого из пяти случаев взаимного расположения двух множеств (результирующее множество заштриховать).

Знаком без денотата является переменная в математике. Действительно, смысл переменной (свободной) – например, числовой – состоит в том, что вместо нее можно подставлять в качестве ее значений различные числа, причем для одной и той же переменной данной формулы – одно и то же число (но для разных переменных – не обязательно разные!). Но эти числа не являются денотатами переменной: у этих чисел есть свои обозначающие их знаки.

К числу знаков без денотата относятся также так называемые <u>неполные</u> знаки; это знаки, которые не имеют денотата сами по себе, но способны входить в состав сложных знаков, которые уже имеют денотат. Так, союз "и" не имеет собственного денотата, но с его помощью образуются знаки, имеющие денотат, например, "Платон и Сократ". Слово "стол" не имеет денотата само по себе; но с его помощью образуются знаки, имеющие денотат в каждой конкретной ситуации, например, "мой стол", "этот стол".

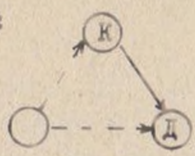
2. Знаки, не имеющие концепта:



Согласно некоторым из существующих теорий (в частности, согласно теории английского философа С.Милля) к знакам без концепта относятся собственные имена, а также чисто указательные знаки: " это", "то", "вот". Более убедительной, однако, является точка зрения, что у собственных имен и указательных знаков есть концепты. Вот что пишет об этом английский логик А.Черч: "В противоположность Миллю мы не будем допускать существование имен, имеющих точное значение ⟨денотат⟩, но не имеющих сопутствующего значения ⟨концепта⟩, и будем считать, что имя всегда должно указывать свое точное значение ⟨денотат⟩ <u>каким-то образом</u>, то есть посредством своего смысла. Это не исключает того, что в некоторых специальных случаях смысл имени может сводиться к тому, что денотат зовется так-то и так-то (например, в случае личных имен), или к тому, что денотат есть то, что находится в данном месте в данное время (как в некоторых случаях указательного "это")".*) Несомненно, однако, что у собственного имени концепт является менее определенным, чем у других знаков. Какой, например, концепт у имени "Наполеон"? Различные определения, которые могут быть даны Наполеону (например, "победитель битвы при Аустерлице", "император, проигравший войну 12-го года", "узник Святой Елены" и проч.), - это сведения о Наполеоне, которые мы знаем не из смысла его имени, а из истории.

К числу знаков без собственного концепта относятся так называемые несобственные, или синкатегорематические знаки; это знаки, которые не имеют смысла сами по себе, и только в сочетании с другими собственными и несобственными знаками образуют выражения, уже имеющие смысл. Таковы, например, скобки в составе математических формул.

3. Нулевые знаки:



Примером нулевого знака является знак операции умножения в алгебре; например, выражение $a\,b$ понимается как $a \times b$. Большое количество нулевых знаков имеется в естественном языке. Так, на фоне того, что существительное в русском языке обычно имеет окончание (книг-а, книг-и, стол-ы, стол-ов), словоформы

*) Черч А. Введение в математическую логику. - М.: Изд-во иностр. лит-ры, 1960, т.I, с. 343.

2. Обозначим через U множество всех документов информационного массива; через P множество документов, релевантных по отношению к некоторому запросу Z ("содержательная выдача"). Предположим, что в ответ на запрос Z система выдала множество документов B. Охарактеризовать с помощью операций над множествами B и P:

- множество M_1 релевантных выданных документов;
- множество M_2 релевантных невыданных документов;
- множество M_3 выданных нерелевантных документов.

3. Построить диаграммы Венна для всех возможных случаев взаимного расположения множеств B и P в упражнении 2. В каких случаях полнота выдачи равна 100%? В каких случаях точность выдачи равна 100%?[ж)]

4. Показать, с помощью диаграмм Венна, что

а) $A \setminus B = A \cap \bar{B}$;

б) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$;

в) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.

Литература

1. Шиханович Ю.А. Введение в современную математику. – М.: Наука, 1965, гл. II.

2. Столл Р.Р. Множества. Логика. Аксиоматические теории. – М.: Просвещение, 1968. 232 с, гл. I.

3. Черный А.И. Введение в теорию информационного поиска. – М.: Наука, 1975. 238 с. гл. 5.

ж) Полнота выдачи считается равной 100%, если все релевантные документы выданы; точность выдачи считается равной 100%, если все выданные документы релевантны.

ГЛАВА III. ОТНОШЕНИЯ

§ I. Понятие отношения

В нашей практике мы часто встречаемся с отношениями. Приведем примеры отношений. Отношения между числами: МЕНЬШЕ, ДЕЛИТСЯ НА (= КРАТНО), ЯВЛЯЕТСЯ ДЕЛИТЕЛЕМ. Отношения между геометрическими фигурами: ПОДОБЕН (например, один треугольник другому), ПАРАЛЛЕЛЕН (например, отрезок отрезку), ПРИНАДЛЕЖИТ (точка прямой), КОНГРУЭНТЕН (угол углу). Между множествами: СТРОГО ВКЛЮЧЕНО (одно множество в другое), ЧАСТИЧНО ПЕРЕСЕКАЕТСЯ (множество с множеством). Между другими объектами: СЕВЕРНЕЕ (один город другого), СТАРШЕ (один человек другого), БРАТ (один человек другому), СИНОНИМИЧНО (одно слово другому), ПРОТИВОРЕЧИТ (одно высказывание другому) и т.д.

Можно заметить, что каждое из этих отношений выделяет пары объектов, принадлежащих некоторому множеству, и устанавливает связь между объектами этих пар. Определение понятия отношения строится на основе понятия пары.

Введем сначала некоторые вспомогательные понятия.

<u>Парой</u> (или <u>упорядоченной парой</u>) будем называть два объекта, взятые в определенном порядке.

О б о з н а ч е н и е. $\langle x, y \rangle$ - это пара.

Например, $\langle 1, 2 \rangle$ - пара чисел.

В паре различается 1-я и 2-я компонента. В частном случае, в паре 1-й и 2-й компонентой может быть один и тот же объект. Например, $\langle 2, 2 \rangle$ - пара.

Чем пара отличается от двухэлементного множества?

Во-первых, тем, что в паре порядок элементов значим; так, $\langle 1, 2 \rangle$ и $\langle 2, 1 \rangle$ - это разные пары, т.е. $\langle 1, 2 \rangle \neq \langle 2, 1 \rangle$ (хотя $\{1, 2\} = \{2, 1\}$).

Во-вторых, тем, что в паре 1-й и 2-й компонентой может быть один и тот же объект (а в множество каждый объект входит "один раз").

Прямым произведением множеств А и В называется множество всех пар $\langle a, b \rangle$ таких, что $a \in A$ и $b \in B$.

Обозначение. $A \times B$ - прямое произведение множества А на множество В.

На рис. 3 точки соответствуют элементам множеств А и В, а стрелки - парам, т.е. элементам множества $A \times B$.

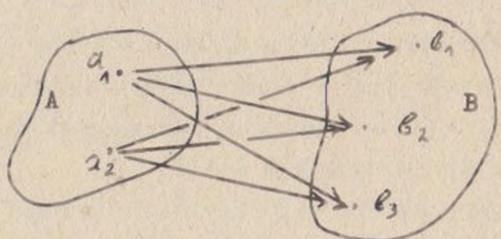


Рис. 3

Например, если А = {Иван, Петр} , В = {Иванович, Петрович} , то $A \times B$ = {⟨Иван, Иванович⟩ , ⟨Иван, Петрович⟩ , ⟨Петр, Иванович⟩ , ⟨Петр, Петрович⟩} .

В частности, можно рассмотреть прямое произведение множества А с самим собой, $A \times A = A^2$. В этом случае говорят о квадрате множества А. Так, для множества А в примере на рис. 3 $A^2 = \{\langle a_1, a_1 \rangle, \langle a_1, a_2 \rangle, \langle a_2, a_1 \rangle, \langle a_2, a_2 \rangle\}$.

Теперь можно определить само понятие отношения.

Определение. <u>Отношением</u> (бинарным) на множестве A называется любое множество R такое, что $R \subseteq A^2$. Иначе говоря, отношение на множестве A – это некоторое множество упорядоченных пар элементов из A.

Обозначение. xRy читается: x находится в отношении R к y. То же самое можно записать так: $R(x,y)$, или так: $\langle x,y\rangle \in R$; читается: пара $\langle x,y\rangle$ принадлежит отношению R.

<u>Примеры</u>. 1) Рассмотрим семью, состоящую из семи человек. Николай – муж Марии. У Николая есть брат Егор, у Марии – сестра Дарья. Петя, Ваня и Таня – дети Николая и Марии. Тогда на множестве A = {Николай, Мария, Егор, Дарья, Петя, Ваня, Таня} отношение СЕСТРА состоит из четырех пар: ⟨Мария, Дарья⟩, ⟨Дарья, Мария⟩, ⟨Таня, Петя⟩, ⟨Таня, Ваня⟩. Отношение МУЖ на этом множестве состоит из одной пары ⟨Николай, Мария⟩. А отношению ДЕД не принадлежит ни одной пары, так как нет ни одной пары людей (в множестве A), находящихся в этом отношении. Заметим, что на другом множестве названные отношения состояли бы из других пар.

2) Отношение МЕНЬШЕ на множестве $A = \{1,2,3\}$ задается множеством пар $R = \{\langle 1, 2\rangle, \langle 2, 2\rangle, \langle 1, 3\rangle\}$. Множеству R принадлежат все те и только те пары элементов из A, которые находятся в отношении МЕНЬШЕ. Таким образом, множество задает отношение МЕНЬШЕ на A, то есть является отношением МЕНЬШЕ.

3) Отношение МЕНЬШЕ на множестве N всех натуральных чисел задается так:

$$R = \{\langle x,y\rangle \mid \text{существует } z \in N \text{ такое, что } x + z = y\}.$$

Характеристическое свойство позволяет из всех пар множества N^2 выбрать только такие, которые принадлежат отношению МЕНЬШЕ, то есть такие, в которых 1-я компонента меньше 2-й.

4) Отношение предпочтения на множестве школьных отметок $\{2, 3, 4, 5\}$ задается множеством пар $\{\langle 5, 4\rangle, \langle 5, 3\rangle, \langle 5, 2\rangle, \langle 4, 3\rangle, \langle 4. 2\rangle, \langle 3, 2\rangle\}$.

5) Отношение СЕВЕРНЕЕ на множестве городов {Рига, Минск, Киев, Одесса} задается парами: {⟨Рига, Минск⟩, ⟨Рига, Киев⟩, ⟨Рига, Одесса⟩, ⟨Минск, Киев⟩, ⟨Минск, Одесса⟩, ⟨Киев, Одесса⟩}.

6) Отношение принадлежности точки прямой на рис. 4 задается парами $\langle A, n\rangle, \langle B, n\rangle, \langle B, l\rangle, \langle C, l\rangle, \langle C, m\rangle$.

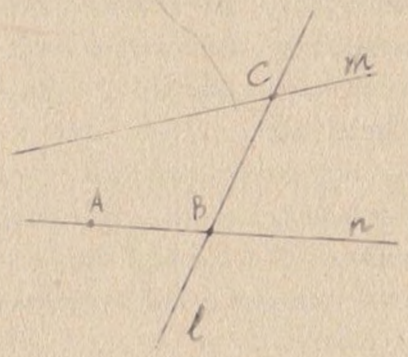


Рис. 4

Рассмотренные примеры показывают, что всякое отношение является множеством пар. Разумеется, не всякое множество пар $R \subseteq A^2$ задает какое-то "разумное" отношение – аналогично тому, как из множеств объектов одни являются "разумными", то есть соответствуют каким-то содержательным свойствам, а другие вводятся "для общности".

Заметим, что к понятию "отношения" можно подходить двояко. С одной стороны, имеется идея, или концепт отношения (см. § 1 главы 1). Концепт отношения – это смысл слова (или словосочетания), которое используется в естественном языке в качестве имени этого отношения (ср. слова СЕСТРА, МЕНЬШЕ и т.д.). С другой стороны, имеется денотат отношения – множество пар объектов, находящихся в этом отношении. Теоретико-множественное понятие отношения отражает второй из этих подходов.

Все рассмотренные до сих пор отношения были <u>бинарными</u>, то есть это были отношения между двумя объектами. Возможны и небинарные отношения, например, трехместные: ЛЕЖИТ МЕЖДУ (точка А между точкой В и точкой С), ОДИНАКОВО УДАЛЕН (луч от двух сторон угла); четырехместные и т.д. Трехместное отношение задается не множеством пар, а множеством упорядоченных троек; четырехместное – множеством упорядоченных четверок и т.д. Число мест отношения называется <u>арностью</u> отношения. Иногда n-арное отношение обозначают R^n. Упорядоченная последовательность произвольной (но конечной) длины называется <u>кортежем</u>. Таким образом, в общем случае, n-арное отношение – это множество кортежей длины n. В дальнейшем будут в основном рассматриваться бинарные отношения (и под отношением будет пониматься бинарное отношение), а переход к небинарным отношениям специально оговаривается.

<u>Способы задания отношения</u>

I. Задание отношения перечислением пар. Естественно, этот способ применим только к отношениям на конечном множестве. См. выше примеры 1, 2, 4, 5, 6.

II. Задание отношения характеристическим свойством пар.

<u>Примеры</u>. 1) Отношение $>$, это множество

$\{\langle x, y\rangle \mid$ существует положительное z такое, что $x = y + z\}$,

иначе говоря, отношение $>$ задается множеством всех тех пар, у которых ко 2-му элементу надо прибавить некоторое число (положительное), чтобы получить I-й.

2) Отношение ОБРАТНОЕ ЧИСЛО – это множество

$$\left\{ \langle x, y\rangle \;\middle|\; x = \frac{1}{y} \right\}.$$

Иначе говоря, число x является обратным для y (то есть пара $\langle x, y\rangle$ принадлежит отношению), если $x = \frac{1}{y}$.

Поскольку отношение – это разновидность множества, два отношения <u>равны</u>, если они задаются одним и тем же множеством пар.

В частности, два разных характеристических свойства могут задавать одно и то же отношение; ср. другое характеристическое свойство для отношения ОБРАТНОЕ ЧИСЛО:

ОБРАТНОЕ ЧИСЛО $= \{ \langle x, y\rangle \mid xy = 1 \}$.

<u>Графические способы изображения отношения</u> (на конечном множестве)

I. Матрица отношения. Пусть отношение R задано на конечном множестве A, содержащем n элементов. Тогда <u>матрицей</u> отношения R является таблица из n строк и n столбцов; каждая строка и каждый столбец соответствуют некоторому элементу множества A, а каждая клетка, расположенная на пересечении i-й строки и j-го столбца, соответствует некоторой паре элементов из A, то есть некоторому элементу множества A^2. Крестики ставятся в клетки, соответствующие тем элементам множества A^2, которые принадлежат отношению R.

<u>Пример</u>. Матрицу отношения $<$ на множестве $\{I, 2, 3\}$ см. на рис. 5: $< \; = \{ \langle I, 2\rangle, \langle I, 3\rangle, \langle 2, 3\rangle \}$.

	I	2	3
I	.	x	x
2	.	.	x
3	.	.	.

Рис. 5

II. Отношение R на множестве A может быть задано **графом**; при этом элементы множества A изображаются точками, а каждой паре $\langle x, y \rangle$, входящей в R, соответствует стрелка, ведущая от x к y.

Примеры. I) Граф отношения $<$ на том же множестве $A = \{1, 2, 3\}$ см. на рис. 6.

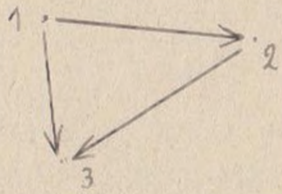


Рис. 6

2) Граф отношения непосредственного алфавитного предшествования на множестве слов $\{$арба, арбуз, баран, вол, галка$\}$ см. на рис. 7.

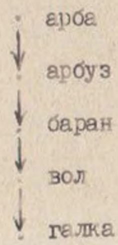


Рис. 7

III. Отношение может быть задано с помощью **таблицы**. Таблицу, задающую отношение между человеком и местом его рождения (для некоторого конечного множества людей), см. на рис. 8.

МЕСТО РОЖДЕНИЯ	
I	2
Иванов И.И.	Москве
Петров П.П.	Москва
Сидоров С.С.	Тула
Колесов К.К.	Норильск
Шагал М.З.	Витебск

Рис. 8

С помощью таблицы может быть задано не только бинарное отношение, но и отношение любой другой арности. См. таблицу трехместного отношения между человеком, местом рождения и годом рождения на рис. 9.

МЕСТО И ГОД РОЖДЕНИЯ		
I	2	3
Иванов И.И.	Москва	1930
Петров П.П.	Москва	1898
Сидоров С.С.	Тула	1935
Колесов К.К.	Норильск	1937
Шагал М.З.	Витебск	1889

Рис. 9

Виды отношений

Среди отношений, которые могут быть заданы на множестве A, выделяется <u>пустое</u> отношение $\varnothing$ (ср. пустое множество), то есть отношение $\mathcal{R}$, задаваемое пустым множеством пар; <u>полное</u> отношение U (ср. универсальное множество), задаваемое множеством $\mathcal{R} = A^2$; <u>диагональное</u> отношение Δ, задаваемое множеством всех пар вида $\langle x, x\rangle$ для всех $x \in A$.

Эти отношения, будучи заданы на множестве $A = \{1, 2, 3\}$, изображаются следующими матрицами:

<u>Пустое</u> отношение

	1	2	3
1	.	.	.
2	.	.	.
3	.	.	.

<u>Полное</u> отношение

	1	2	3
1	x	x	x
2	x	x	x
3	x	x	x

<u>Диагональное</u> отношение

	1	2	3
1	x	.	.
2	.	x	.
3	.	.	x

Рис. 10

§ 2. Операции над отношениями

Поскольку отношение есть, по определению, множество пар, то есть м н о ж е с т в о, над отношениями можно производить все те же операции, что и над множествами. Таким образом, первую группу операций над отношениями составляют теоретико-множественные операции. Ко второй группе относятся операции, применимые специально к отношениям.

<u>Теоретико-множественные операции</u>

Пусть имеются отношения R и S, заданные на множестве A. Таким образом, $R, S \subseteq A^2$.

1. <u>Объединением</u> отношений R и S называется отношение

$$R \cup S = \{ \langle x,y \rangle \mid \langle x,y \rangle \in R \text{ или } \langle x,y \rangle \in S \}.$$

Иначе говоря, объединением отношений R и S является отношение, задаваемое множеством всех пар $\langle x, y \rangle$ таких, что $\langle x, y \rangle \in R$ или $\langle x, y \rangle \in S$ (или обоим).

<u>Примеры</u>. 1) $< \cup = \; = \; \leqslant$

2) $< \cup > \; = \; \neq$

3) МАТЬ $\cup$ ОТЕЦ = РОДИТЕЛЬ ;

4) БРАТ $\cup$ СЕСТРА = GESCHWISTER (т.е. дети общих родителей).

2. Пересечением отношений R и S называется отношение

$$R\cap S=\{\langle x,y\rangle \mid \langle x,y\rangle\in R \text{ и } \langle x,y\rangle\in S\}.$$

То есть пересечением отношений R и S является множество всех пар $\langle x, y\rangle$ таких, что $\langle x, y\rangle\in R$ и одновременно $\langle x, y\rangle\in S$.

Примеры. 1) $\leqslant \cap < \; = \; <$;

2) БРАТ $\cap$ СЕСТРА = $\varnothing$.

3. Разностью отношений R и S называется отношение

$$R\setminus S=\{\langle x, y\rangle \mid \langle x, y\rangle\in R \text{ и } \langle x, y\rangle\notin S\}.$$

Иначе, разностью отношений R и S называется множество всех пар $\langle x, y\rangle$ таких, что $\langle x, y\rangle\in R$ и $\langle x, y\rangle\notin S$.

Примеры. 1) $\leqslant \setminus < \; = \; =$;

2) СОСЛУЖИВЕЦ $\setminus$ НАЧАЛЬНИК = ПОДЧИНЕННЫЙ $\cup$ СОТРУДНИК.

4. Дополнением отношения R на множестве A^2 называется отношение

$$\overline{R}=\{\langle x, y\rangle \mid \langle x, y\rangle\in A^2 \text{ и } \langle x, y\rangle\notin R\}.$$

Примеры. 1) $\overline{=} \; = \; \neq$;

2) ЮЖНЕЕ = СЕВЕРНЕЕ $\cup$ НА ОДНОЙ ШИРОТЕ.

Прочие операции над отношениями

5. Отношение называется инверсией отношения R (или обратным для R), если оно состоит из всех пар $\langle x, y\rangle$ таких, что пара $\langle y, x\rangle\in R$.

Обозначение. Инверсия отношения R обозначается R^{-1}. Таким образом, $R^{-1}=\{\langle x, y\rangle \mid \langle y, x\rangle\in R\}$.

Примеры. 1) МУЖ$^{-1}$ = ЖЕНА;

2) РОДИТЕЛЬ$^{-1}$ = РЕБЕНОК;

3) Инверсией отношения $>$ является отношение $<$;

4) Инверсией отношения $=$ является само отношение $=$;

5) СИНОНИМ$^{-1}$ = СИНОНИМ.

Если отношения R и R^{-1} заданы матрицами, то матрица R^{-1} симметрична матрице R относительно главной диагонали.

6. Отношение называется композицией отношений R и S, если оно состоит из всех пар $\langle x, y\rangle$, для которых существует $z\in A$ такое, что $\langle x, z\rangle\in R$ и $\langle z, y\rangle\in S$.

Обозначение. Композиция отношений R и S обозначается $R\circ S$. Таким образом,

$$R\circ S=\{\langle x,y\rangle \mid \text{существует } z\in A \text{ такое, что } \langle x,z\rangle\in R \text{ и } \langle z,y\rangle\in S\}.$$

Примеры. 1) БРАТ $\circ$ РОДИТЕЛЬ = ДЯДЯ.

Действительно, если я прихожусь дядей Тимоше, это значит, что существует человек такой, что я являюсь братом этого человека, а он является родителем (то есть отцом или матерью) Тимоши. (Здесь не принимается во внимание другой случай, когда дядя является не братом родителя, а мужем сестры родителя.)

2) МАТЬ $\circ$ МУЖ = СВЕКРОВЬ.

Отметим, что операция композиции не коммутативна: $R\circ S\neq S\circ R$; действительно, скажем, мать мужа - это не то же, что муж матери.

7. Композиция отношения R с самим собой называется квадратом отношения R.

Пример. ОТЕЦ^2 = ДЕД.

Вообще, можно возводить отношение в любую степень: ОТЕЦ^3 = ПРАДЕД, и т.д.

8. Транзитивным замыканием отношения R называется отношение, обозначаемое $\widehat{R}$ и состоящее из всех пар $\langle x, y\rangle$ таких, что $\langle x, y\rangle\in R$, или $\langle x, y\rangle\in R^2$, или $\langle x, y\rangle\in R^3$.

и т.д. Таким образом,

$$\hat{R} = \{ \langle x,y \rangle \mid \langle x,y \rangle \in R \text{ или } \langle x,y \rangle \in R^2 \text{ или } \langle x,y \rangle \in R^3 \text{ и т.д.} \}.$$

Иначе можно сказать, что

$$\hat{R} = R \cup R^2 \cup R^3 \cup \ldots .$$

Примеры. 1) Транзитивное замыкание отношения ОТЕЦ - это отношение ПРЕДОК ПО МУЖСКОЙ ЛИНИИ.

2) Если R - отношение непосредственного алфавитного предшествования, то $\hat{R}$ - отношение просто предшествования. Так, в примере на рис. 7 пара $\langle$арба, баран$\rangle$ не принадлежит отношению R, но она принадлежит отношению $\hat{R}$.

3) Для отношения $>$ транзитивное замыкание есть то же самое отношение $>$.

Две последние операции, 9 и 10, отличаются от операций 5 - 8 тем, что применимы не только к бинарным отношениям: они применимы к отношениям произвольной степени.

9. Сцеплением отношений R и S, имеющих соответственно арности m и n, называется отношение, состоящее из всех кортежей, которые получаются в результате того, что находится кортеж отношения R, у которого последняя компонента совпадает с первой компонентой некоторого кортежа из S, и каждая такая пара кортежей "склеивается" в один по этой общей компоненте.

Обозначение. Сцепление отношений R и S обозначается $R * S$. Таким образом, $R*S = \{ \langle r_1,\ldots,r_m,s_2,\ldots,s_n \rangle \mid \langle r_1,\ldots,r_m \rangle \in R$ и $\langle s_1,\ldots,s_n \rangle \in S$ и $r_m = s_1 \}$.

Если отношения R и S заданы таблицами, то таблица отношения $R*S$ состоит из всех строк вида $\langle r_1,\ldots,r_m,s_2,\ldots,s_n \rangle$, таких, что строка вида $\langle r_1,\ldots,r_m \rangle$ входит в таблицу отношения R, а строка вида $\langle s_1,\ldots,s_n \rangle$ входит в таблицу отношения S и $r_m = s_1$ (т.е. отношение $R * S$ состоит из всех строк, получаемых склеиванием строк отношений R и S по общей компоненте, которая является последней в строке для R и первой в строке для S).

См. пример сцепления отношений на рис. II.

R	
1	2
a	b
c	b
c	d
e	a

S	
I	2
b	e
d	e
e	d

$R * S$		
I	2	3
a	b	e
c	b	e
c	d	a

Р и с. II

Если R имеет арность m, а S – арность n, то $R * S$ будет иметь арность $m + n - 1$.

Аналогичным образом можно определить сцепление отношения не с другим отношением, а с множеством: $R * A$ – это отношение, задаваемое множеством всех тех кортежей отношения R, в которых последняя компонента совпадает с одним из элементов множества A; $A * R$ – это отношение, задаваемое множеством всех тех кортежей отношения R, в которых первая компонента является одновременно элементом множества A.

I0. Проекцией отношения R (арности n) на его компоненту K_i $(i \leq n)$ называется множество, которому принадлежат все те и только те элементы, которые являются i-ми компонентами кортежей из R.

О б о з н а ч е н и е. Проекция отношения R на его компоненту K_i обозначается $R[K_i]$.

Таким образом, $R[K_i]$ состоит из всех различных элементов, входящих в i-й столбец таблицы отношения R.

Например, проекция отношения R на рис.II на его компоненту K_2 – это множество $\{b, d, a\}$.

Проекция отношения R арности n на пару компонент $\langle K_i, K_j \rangle$ $(i,j \leq n)$ – это отношение $R[K_i, K_j]$, состоящее из множества всех пар $\langle r_i, r_j \rangle$ таких, что r_i, r_j являются соответственно i-й и j-й компонентой некоторого кортежа из отношения R.

Например, для отношения $R*S$ на рис. II проекция на пару компонент K_2, K_3 – это отношение $R*S$ $[K_2, K_3] = \{\langle b, e \rangle, \langle d, a \rangle\}$

Аналогично определяется проекция отношения на любую упорядоченную последовательность его компонент. Например, проекция отношения $R*S$ на рис. II на тройку компонент $[K_1, K_3, K_2]$ – это то же отношение $R*S$, в таблице которого, однако, столбцы 2 и 3 поменялись местами.

§ 3. Свойства отношений

При всем разнообразии конкретных отношений между ними можно обнаружить некоторые сходства. Например, отношения РАВЕН, СИНОНИМИЧЕН, РОДСТВЕННИК сходны в том отношении, что если осуществить над ними операцию инверсии, то мы получим то же самое отношение – в отличие, скажем от отношений МЕНЬШЕ, ПРЕДШЕСТВУЕТ, ПРЕДОК, для которых инверсия даст в каком-то смысле противоположное отношение. Другой пример. Отношение ПАРАЛЛЕЛЕН и РОДСТВЕННИК похожи друг на друга в том смысле, что если подвергнуть их транзитивному замыканию, то получится то же отношение; тогда как отношение НЕПОСРЕДСТВЕННО ПРЕДШЕСТВУЕТ и ОТЕЦ принадлежат к другой группе – их транзитивное замыкание является новым отношением.

Рассмотрим некоторые общие свойства отношений, которые позволят объединять отношения в такого рода классы.

I. Отношение R на множестве A называется рефлексивным если для всех $x \in A$ $\langle x, x\rangle \in R$.

Примеры рефлексивных отношений: параллельность, равенство, синонимия.

Другое, эквивалентное определение рефлексивности. R рефлексивно, если и только если $\Delta \subseteq R$.

Отсюда следует простой способ распознавания рефлексивных отношений по их матрице: отношение рефлексивно, если оно включает диагональ. Например, на рис. I2 отношение $\geqslant$ рефлексивно, а $>$ не рефлексивно:

$\geqslant$	I	2	3
I	x	.	.
2	x	x	.
3	x	x	x

$>$	I	2	3
I	.	.	.
2	x	.	.
3	x	x	.

Рис. I2

Отношение R на множестве A называется антирефлексивным, если для всех $x \in A$ $\langle x, x\rangle \notin R$.

Например, отношение $<$ антирефлексивно.

Другое, эквивалентное определение: R антирефлексивно, если и только если

$$R \cap \Delta = \varnothing.$$

Отношение может быть одновременно не рефлексивным и не антирефлексивным. Например, это отношение, задаваемое матрицей на рис. I3.

	I	2	3
I	.	.	.
2	.	x	x
3	.	x	x

Рис. I3

Оно не рефлексивно, так как не включает пару $\langle 1, 1\rangle$, и не антирефлексивно, так как включает пары $\langle 2, 2\rangle$ и $\langle 3, 3\rangle$. Примером реального отношения такого типа может быть, скажем, отношение УВАЖАЕТ, заданное на каком-то множестве людей: одни люди с уважением относятся к самим себе, а другие, быть может, его утратили.

2. Отношение R на множестве А называется <u>симметричным</u>, если для любой пары $\langle x, y\rangle \in R$ пара $\langle y, x\rangle$ тоже принадлежит R.

Примеры симметричных отношений: равенство, параллельность, синонимия, родство.

Другое определение. Отношение R симметрично, если и только если

$$R^{-1} = R.$$

Симметричное отношение легко распознается по своей матрице: матрица симметричного отношения симметрична относительно диагонали. Примеры отношений с симметричными матрицами см. на рис. I4.

	I	2	3
I	.	x	x
2	x	.	x
3	x	x	.

	I	2	3
I	x	x	x
2	x	x	x
3	x	x	x

Р и с. I4

Отношение R называется <u>антисимметричным</u>, если из того, что $\langle x, y\rangle \in R$ и $\langle y, x\rangle \in R$, следует, что $x=y$. Иначе говоря, R антисимметрично, если и только если

$$R \cap R^{-1} \subseteq \Delta.$$

Примеры антисимметричных отношений: $\geqslant$, $\leq$, $\subseteq$, СТАРШЕ, ЯВЛЯЕТСЯ ЧАСТЬЮ.

Замечание. Есть еще одно свойство, тоже, так сказать, противоположное симметричности, но более сильное, чем антисимметричность – это асимметричность. Отношение R называется асимметричным, если из того, что $\langle x, y\rangle \in R$, следует, что $\langle y, x\rangle \notin R$. Асимметричность, в отличие от антисимметричности, не допускает наличия в составе R пар вида $\langle x, x\rangle$, т.е. пар, состоящих из одинаковых элементов. Эквивалентная формулировка асимметричности: R асимметрично, если и только если

$$R \cap R^{-1} = \varnothing.$$

Легко видеть, что R асимметрично, если и только если R антисимметрично и антирефлексивно.

3. Отношение R на множестве A называется транзитивным, если для любых трех элементов $x, y, z \in A$ из того, что $\langle x, z\rangle \in R$ и $\langle z, y\rangle \in R$ следует, что $\langle x, y\rangle \in R$.

Примеры транзитивных отношений: параллельность, равенство, строгое и нестрогое включение, синонимия, родство. Отношения $\perp$ и $\in$ не транзитивны. Не является транзитивным отношение знакомства.

Другое определение транзитивности: отношение R транзитивно, если и только если

$$\hat{R} = R.$$

4. Отношение R на множестве A называется связным, если для любых двух различных элементов $x, y \in A$ пара $\langle x, y\rangle$ или пара $\langle y, x\rangle$ (или обе) принадлежит R.

Например, отношение $<$ связно, а отношение $=$ не связно.

Элементы $x, y \in A$ такие, что ни $\langle x, y\rangle$, ни $\langle y, x\rangle$ не принадлежит R, называются несравнимыми относительно R. Иначе можно сказать, что R на A связно, если любые два различных элемента из A сравнимы относительно R.

§ 4. Типы отношений

I. Эквивалентность и толерантность

Определение I. Отношение R на множестве А называется отношением типа <u>эквивалентности</u> (или просто <u>эквивалентностью</u>), если R рефлексивно, симметрично и транзитивно, т.е. если

I) для всех $x \in A$ $\langle x,x\rangle \in R$,

2) для всех $x \in A$, если $\langle x,y\rangle \in R$, то $\langle y,x\rangle \in R$,

3) для любых x, y, $z \in A$, если $\langle x,z\rangle \in R$ и $\langle z,y\rangle \in R$, то $\langle x,y\rangle \in R$.

Примерами эквивалентностей являются: отношение подобия, равенства, родства; отношение ЖИВЕТ В ОДНОМ ДОМЕ, СОСЛУЖИВЕЦ, ОДНОКУРСНИК, ОДНОФАМИЛЕЦ.

Не является эквивалентностью, скажем, отношение $\perp$, поскольку оно не транзитивно: если $a \perp b$ и $b \perp c$, то неверно, что $a \perp c$. Не является эквивалентностью отношение "иметь общий делитель"; так, числа 2 и I0 имеют общий делитель 2, и числа I0 и 5 имеют общий делитель 5; а числа 2 и 5 взаимно простые.

Подчеркнем, что эквивалентность – это название типа отношения, а не конкретного отношения. Так, на множестве $A = \{1,2,3\}$ можно задать несколько различных эквивалентностей, например, отношение равенства, полное отношение, отношение "иметь одинаковую четность".

Эквивалентности обладают еще одним замечательным свойством: если отношение R на множестве A является эквивалентностью, то R задает разбиение (см. § 4 главы II) множества А на классы $\{A_1, A_2, \ldots\}$ такое, что для любых двух элементов $x, y \in A$, принадлежащих одному и тому же классу A_i этого разбиения, $\langle x,y\rangle \in R$, а для любых двух элементов $z, u \in A$, принадлежащих разным классам, $\langle z,u\rangle \notin R$.

Так, отношение "начинаться на одну и ту же букву" на множестве слов некоторого словаря - это отношение типа эквивалентности; действительно, оно рефлексивно, симметрично и транзитивно. Отсюда следует то обстоятельство (известное из практики), что множество слов можно разбить на непересекающиеся классы так, что в одном классе будут слова, начинающиеся на одну и ту же букву, а в разных классах - на разные.

Аналогично, из того, что отношение подобия на множестве многоугольников является эквивалентностью, следует, что это множество можно разбить на классы так, что в одном классе все многоугольники будут подобны один другому и никакие два подобных многоугольника не окажутся в разных классах.

Верно и обратное: любое разбиение $\mathcal{P}$ множества A на классы задает на A отношение эквивалентности R такое, что $\langle x,y\rangle \in R$, если и только если x и y принадлежат некоторому одному и тому же классу разбиения.

Это дополнительное свойство отношений эквивалентности можно использовать для более простого графического представления этих отношений: чтобы указать на множестве A некоторое отношение эквивалентности, достаточно указать, на какие классы это отношение разбивает множество A.

Например, чтобы задать на множестве A отношение родства, достаточно задать разбиение на классы родственников.

О п р е д е л е н и е 2. Отношение, которое является рефлексивным и симметричным, но не обязательно транзитивным, называется <u>толерантностью.</u>

Например, отношение знакомства, в общем случае, не является транзитивным: человек X может быть знаком с человеком z,

z с y, а x и y могут быть и незнакомы друг с другом. Поскольку отношение знакомства рефлексивно и симметрично, оно является толерантностью.

Другой пример толерантности – отношение соавторства: оно тоже рефлексивно и симметрично, но, вообще говоря, не транзитивно.

Если отношение R на множестве A является толерантностью, то ему соответствует система $\mathcal{P} = \{A_1, A_2 \ldots\}$ подмножеств множества A такая, что $x, y \in A_i$ из $\mathcal{P}$, если и только если $\langle x,y \rangle \in R$, и такая, что $\mathcal{P}$ является покрытием (см. § 4 главы I) множества A. Таким образом, отношению толерантности на множестве A соответствует система подмножеств множества A, вообще говоря, пересекающихся.

Например, пусть граф на рис. 15 а задает отношение соавторства на множестве $A = \{Г, И, М, П, Ч\}$. Система подмножеств $\{A_i\}$ множества A таких, что для любых элементов $x, y \in A$ $x, y \in A_i$, если и только если x и y – соавторы, изображена на рис. 15 б. Ясно, что эта система является покрытием множества A, но не является разбиением – поскольку подмножества пересекаются.

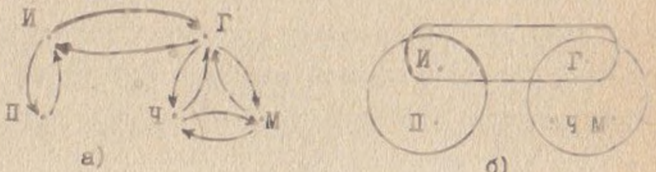


Рис. 15

2. Отношения порядка

Строгие порядки

О п р е д е л е н и е. Отношение R на множестве A называется отношением строгого порядка, если:

1) R антирефлексивно, то есть для всех $x \in A$ $\langle x,x \rangle \notin R$.

2) R транзитивно, то есть для любых $x, y, z \in A$, если $\langle x,z \rangle \in R$ и $\langle z,y \rangle \in R$, то $\langle x,y \rangle \in R$

Например, отношения $<$, $\subset$, ПРЕДОК, ПРЕДШЕСТВУЕТ – все являются отношениями строгого порядка. Отношения $\in$, ОТЕЦ не являются строгими порядками, так как они не транзитивны. Отношение непосредственного алфавитного предшествования тоже не относится к строгим порядкам.

Легко доказать, что если отношение R на множестве A антирефлексивно и транзитивно, то R антисимметрично. Действительно, предположим, что R не антисимметрично, то есть что существуют $x, y \in A$ такие, что $\langle x, y\rangle \in R$ и $\langle y, x\rangle \in R$. Но тогда, в силу транзитивности, $\langle x, x\rangle \in R$, что противоречит условию антирефлексивности R . Следовательно, R антисимметрично.

Таким образом, отношение строгого порядка обладает тремя свойствами: оно антирефлексивно, антисимметрично и транзитивно.

На одном множестве можно задать различные отношения строгого порядка: например, на множестве $A = \{1,2,3,4\}$ можно задать отношения $<$ и $>$, которые оба являются строгими порядками.

Другие примеры отношений строгого порядка: отношение СТАРШЕ на множестве всех людей, отношение НАЧАЛЬНИК на множестве сотрудников некоторого учреждения; отношение СЕВЕРНЕЕ на множестве населенных пунктов; отношение БЛИЖЕ К НАЧАЛУ КООРДИНАТ на множестве точек координатной плоскости.

Если R – отношение строгого порядка, то его граф не содержит петель (то есть замкнутых путей, которые ведут от элемента к нему самому).

Поскольку отношения строгого порядка транзитивны, естественно пользоваться, при их задании с помощью графа, <u>сокращенными графами</u>, то есть такими, в которых опущены стрелки, вытекающие из транзитивности.

П р и м е р . На рис. 16 изображен полный(а) и сокращенный (б) граф отношения $>$ на множестве $\{1,2,3,4\}$

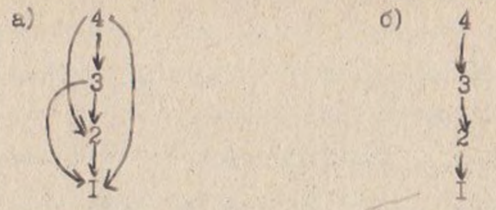


Рис. I6

Нестрогие порядки

Определение. R называется отношением нестрогого порядка, если R I) рефлексивно, 2) антисимметрично и 3) транзитивно.

Примеры. I) Отношение $\subseteq$ на системе подмножеств множества А – это отношение нестрогого порядка.

2) Отношение ДЕЛИТСЯ НА (имеется в виду, делится без остатка) на множестве натуральных чисел – это тоже нестрогий порядок.

3) На рис. I7 приводится сокращенный граф отношения включения одним словом другого на множестве слов $\{$ параграф, графин, пара, граф, пар, ара, па, ар, ра$\}$ (" слово a включает слово b " – значит 'b целиком вкладывается в a'). Это отношение является нестрогим порядком; в частности, оно рефлексивно: каждое слово включает само себя. Поскольку отношение нестрогого порядка всегда рефлексивно, в сокращенном графе этого отношения естественно опускать стрелки, вытекающие из рефлексивности.

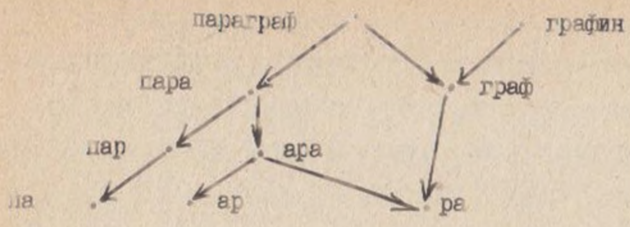


Рис. I7

Строгие и нестрогие порядки связаны следующими простыми соотношениями: если R - нестрогий порядок на множестве A, то $R \setminus \Delta$ - строгий порядок на том же A; если R - строгий порядок на множестве A, то $R \cup \Delta$ - нестрогий порядок на том же множестве.

Совершенные и древесные порядки

Среди порядков, как строгих, так и нестрогих, выделяется частный случай - совершенные порядки.

Определение. Отношение порядка R (строгого или нестрогого) на множестве A называется отношением совершенного порядка, если оно связно, т.е. если любые два различных элемента $x, y \in A$ сравнимы по R (либо $\langle x, y\rangle \in R$, либо $\langle y, x\rangle \in R$).

Например, отношение $<$ - это отношение совершенного порядка (строгого), поскольку из двух различных чисел всегда одно меньше другого. А отношение $\subset$ уже не является совершенным порядком, поскольку два подмножества A_I, A_2 множества A могут быть различными, и тем не менее не будет верно ни $A_I \subset A_2$ ни $A_2 \subset A_I$; например, A_I и A_2 могут не пересекаться, как в а) на рис. I8, или частично пересекаться, как в б).

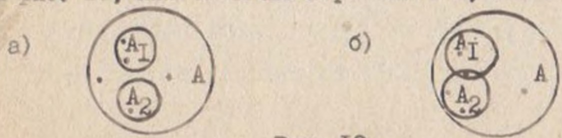


Рис. I8

Отношение СТАРШЕ не является совершенным порядком, поскольку два человека могут быть одинакового возраста; отношение СЕВЕРНЕЕ не является отношением совершенного порядка, поскольку два населенных пункта могут лежать на одной параллели, так что ни один не будет севернее другого.

Из отношений нестрогого порядка отношение $\leqslant$ является совершенным порядком, а отношение $\subseteq$ не является.

Граф отношения совершенного строгого порядка (так же, как сокращенный граф отношения совершенного нестрогого порядка) является цепью.

Другим частным случаем порядка (как строгого, так и нестрогого) является древесный порядок.

Чтобы определить, что такое древесный порядок, нужно сначала ввести понятия максимального и наибольшего элемента.

На одном и том же множестве можно задать много разных отношений порядка. Договоримся содержательно понимать любое отношение порядка R как "старше".

Элемент x множества A с отношением порядка R называется <u>максимальным</u>, если не существует $y \in$ A такого, что $\langle y,x\rangle \in R$ (т.е. если не существует элемента, который "старше" x).

Элемент x множества A с отношением порядка R называется <u>наибольшим</u>, если для любого $y \in$ A $\langle x,y\rangle \in R$ (т.е. если x "старше" всех остальных элементов A).

Если множество A конечно и R – совершенный порядок (строгий или нестрогий), то в A существует единственный максимальный элемент, и он же наибольший. В противном случае наибольшего элемента может не быть. Например, во множестве натуральных чисел нет наибольшего числа. Отношение включения на множестве слов на рис. 17 – это порядок (нестрогий) без наибольшего элемента.

О п р е д е л е н и е. Порядок R на конечном множестве A называется древесным, если

1) для любых $x, y, z \in A$ если $\langle x,z\rangle \in R$ и $\langle y,z\rangle \in R$, то x, y сравнимы (т.е. если имеются два элемента, которые старше z, то один из них старше другого);

2) в A существует наибольший элемент.

Иначе. Порядок R является древесным, если его сокращенный граф является деревом.

Примеры древесных порядков.

1) Отношение служебного подчинения на множестве сотрудников учреждения – это древесный порядок (строгий).

2) Классификация животных по классам, отрядам, семействам, родам и видам задает на множестве этих рубрик отношение древесного порядка; фрагмент этой классификации представлен графом на рис. 19.

Рис. 19

3) На множестве рубрик УДК, начинающихся с общей цифры, задается отношение древесного порядка. См. пример на рис. 20.

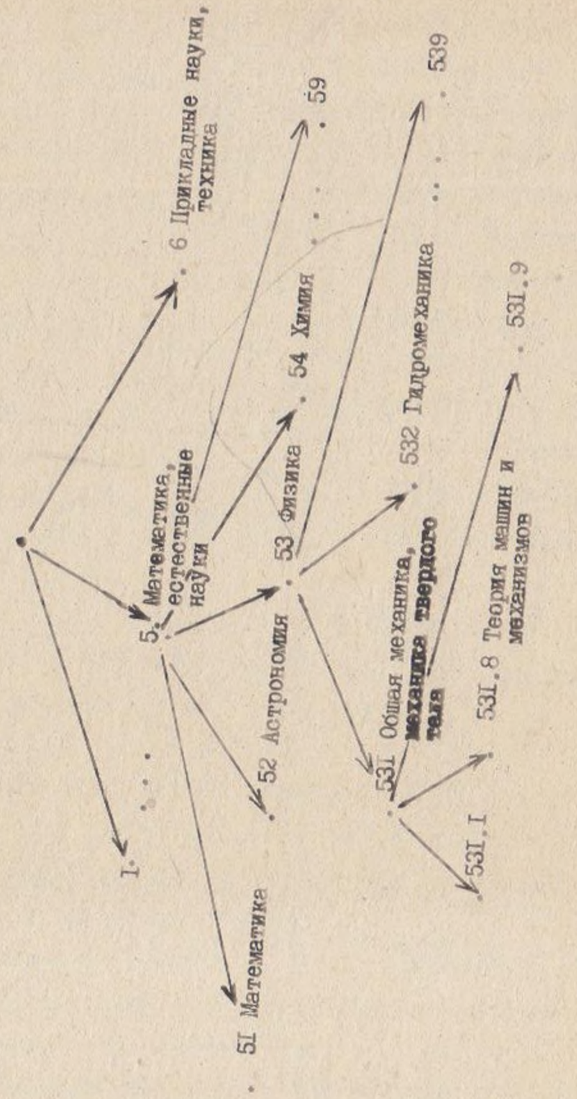


Рис. 20

Табл. I позволяет сравнить друг с другом разные типы отношений с точки зрения их свойств. Скобки в таблице означают, что данное значение данного признака вытекает из значений, принимаемых другими признаками; ноль означает, что значение этого признака несущественно для данного типа отношений.

Таблица I

Свойства / Отношения	Рефлексивность	Антирефлексивность	Симметричность	Антисимметричность	Транзитивность	Связность
Эквивалентность	+	(-)	+	(-)	+	0
Толерантность	+	(-)	+	(-)	0	0
Строгий порядок	(-)	+	(-)	(+)	+	0
Совершенный строгий порядок	(-)	+	(-)	(+)	+	+
Нестрогий порядок	+	(-)	(-)	+	+	0
Совершенный нестрогий порядок	+	(-)	(-)	+	+	+

§ 5. Пример использования теории отношений в информатике : банки данных

Основой фактографической ИПС (то есть ИПС, предназначенной, в отличие от документальных ИПС, для выдачи не библиографических, а фактических сведений) является банк данных, то есть совокупность сведений, подлежащих различным видам переработки в зависимости от нужд пользователя и характера информационного запроса. Способ представления этой исходной совокупности сведений в памяти ЭВМ должен обеспечивать а) простоту и универсальность процедур поиска информации в ответ на разнообразные запросы, а также б) удобство процедур добавления новой информации и устранения устаревшей. Таким образом, от структуры банка данных существенным образом зависит эффективность фактографического поиска.

В числе предлагавшихся и реализованных на практике структур банка данных одной из наиболее перспективных является так называемая р е л я ц и о н н а я структура, (подробнее см. ч. П ж) настоящего пособия), прямым образом опирающаяся на теоретико-множественное понятие отношения.

Рассмотрим в качестве примера банк данных Б, содержащий всевозможные сведения о необходимых заказчику изделиях. Изделия могут быть различных наименований, изготовляться из разных материалов и иметь разные размеры; они, далее, могут поставляться различными предприятиями в различных количествах; предприятия могут находиться в различных городах, каждое имеет свое название и свою категорию, причем все эти различия могут, вообще говоря, быть существенны для заказчика.

ИПС с банком данных Б должна быть в состоянии выдавать ответы, например, на следующие запросы:

(1) Какие детали поставляет такое-то предприятие?

(2) Какие предприятия поставляют болты такого-то размера?

(3) Есть ли предприятия, поставляющие одновременно болты, кулачки и муфты?

(4) Есть ли предприятия, поставляющие винты и гайки одинакового размера?

(5) К какой категории относятся предприятия, поставляющие такую-то деталь и находящиеся в таком-то городе, и т.д.

Существенно, что не только конкретное содержание, но и самый тип вопроса не обязан быть жестко закрепленным заранее.

ж) Борщев В.Б. Математические методы в теории научно-технической информации. Ч. П. Математическая логика. Базы данных. - М.: ИПКИР, 1979.

Банк данных реляционной структуры представляет собой совокупность отношений – как бинарных, так и трех-, четырех- и более местных. В практически возникающих задачах множества объектов, представляющих интерес, конечные (хотя, быть может, и очень большие), и отношения на таких множествах могут быть заданы прямым перечислением всех упорядоченных пар (или соответственно троек, четверок и т.д. – в зависимости от арности отношения), принадлежащих отношению. Этот перечень удобно задавать для каждого отношения соответствующей ему таблицей.

Например, рассматриваемый банк данных состоит из трех отношений – отношение ИЗДЕЛИЕ (И), отношение ПРЕДПРИЯТИЕ (П) и отношение ПОСТАВЩИК-ИЗДЕЛИЕ (ПИ).

ИЗДЕЛИЕ – это четырехместное отношение, указывающее а) наименование изделия; б) материал; в) размер (в каких-то единицах измерения, в данном случае для нас несущественных); кроме того, у изделия должен быть код, который позволял бы различать изделия, не совпадающие хотя бы в одном из существенных параметров – по наименованию, материалу или размеру. Разумеется, число параметров изделия можно было бы и увеличить.

Отношение ПРЕДПРИЯТИЕ в нашем примере будет тоже четырехместным, оно определяет: а) код предприятия; б) название; в) категорию; г) местонахождение (город).

ПОСТАВЩИК-ИЗДЕЛИЕ – это трехместное отношение, указывающее, какие изделия (то есть коды) поставляет каждое предприятие (код) и в каком количестве (количества задаются в каких-то единицах измерения, здесь не уточняемых).

Предположим, что на какой-то отрезок времени на интересующем нас участке отношения И, П и ПИ задаются следующими таблицами (для каждой компоненты отношения указан ее номер и название); И # – код изделия, П #-код предприятия):

	ИЗДЕЛИЕ (И)		
И#	НАИМЕНО-ВАНИЕ	МАТЕРИАЛ	РАЗМЕР
I	2	3	4
И1	муфты	чугун	19
И2	болты	латунь	17
И3	болты	чугун	25
И4	винты	сталь	15
И5	винты	сталь	17
И6	гайки	сталь	17

ПРЕДПРИЯТИЕ (П)

П#	НАЗВА-НИЕ	КАТЕГОРИЯ	ГОРОД
I	2	3	4
П1	"Салют"	I	Тула
П2	"Заря"	I	Тула
П3	"Рассвет"	2	Серпухов
П4	"Восход"	3	Серпухов
П5	"Рассвет"	3	Чехов

ПОСТАВЩИК-ИЗДЕЛИЕ (ПИ)		
П#	И#	КОЛИЧЕ-СТВО
I	2	3
П1	И1	3
П1	И2	2
П1	И5	4
П2	И1	5
П2	И2	5
П3	И3	3
П3	И4	4
П3	И6	4
П4	И2	I,5
П5	И1	2

Ответ на любой из вопросов типа (I) – (5) может быть получен применением, быть может многократным, к исходным отношениям банка данных Б двух операций над отношениями – это операция сцепления отношений и операция проекции отношения на его компоненту или последовательность компонент (см.§ 2). Тем самым перевод вопроса к реляционному банку данных на язык операций над отношениями становится поисковым предписанием, то есть точной инструкцией для отыскания ответа.

Ниже приведены примеры запросов на язык операций над отношениями; для наглядности при задании операций проекции компоненты отношений обозначаются своими названиями, а не порядковыми номерами, как в определении из § 2.

(а) Вопрос: "Какие предприятия (коды) поставляют изделие И1 ?" Перевод: (ПИ [П#, И#] * {И1}) [П#] .

Вычисление ответа:

ПИ [П#, И#]	
П#	И#
П1	И1
П1	И2
П1	И5
П2	И1
П2	И2
П3	И3
П3	И4
П3	И6
П4	И2
П5	И1

ПИ [П#, И#] * {И1}	
П#	И#
П1	И1
П2	И1
П5	И1

Ответ: {П1, П2, П5} .

(б) Вопрос: "Какие изделия (наименования) поставляет предприятие П2?"

Перевод: ((ПИ * {П2}) [П#, И#] * И) [НАИМЕН]

Вычисление ответа:

ПИ * {П2}		
П#	И#	КОЛИЧ
П2	И1	5
П2	И2	5

(ПИ * {П2}) [П#, И#]	
П#	И#
П2	И1
П2	И2

(ПИ * {П2}) [П#, И#] * И				
П#	И#	НАИМЕН	МАТЕРИАЛ	РАЗМЕР
П2	И1	муфты	чугун	19
П2	И2	болты	латунь	17

Ответ: {муфты, болты}

(в) Вопрос: "Какие предприятия (коды) поставляют изделия, сделанные из стали?"

Перевод: ((ПИ [П#, И#] * И) * {сталь}) [П#]

Вычисление ответа:

ПИ [П#, И#]
см.вопрос(а)

ПИ [П#, И#] * И				
П#	И#	НАИМЕН	МАТЕРИАЛ	РАЗМЕР
П1	И1	муфты	чугун	19
П1	И2	болты	латунь	17
П1	И5	винты	сталь	17
П2	И1	муфты	чугун	19
П2	И2	болты	латунь	17
П3	И3	болты	чугун	2[illegible]
П3	И4	винты	сталь	15

П3	И6	гайки	сталь	I7
П4	И2	болты	латунь	I7
П5	ИI	муфты	чугун	I9

(ПИ [П#, И#] * И) * {сталь}				
П#	И#	НАИМЕН	МАТЕРИАЛ	КОЛИЧ
ПI	И5	винты	сталь	I7
П3	И4	винты	сталь	I5
П3	И6	гайки	сталь	I7

Ответ: {ПI, П3}.

(г) Вопрос: "Для каждого изделия (кода) указать все города, из которых оно может быть получено".

Перевод: (ПИ [И#, П#] * П [П#, ГОРОД]) [И#, ГОРОД]

Вычисление ответа:

ПИ [И#, П#]	
И#	П#
ИI	ПI
ИI	П2
ИI	П5
И2	ПI
И2	П2
И2	П4
И3	П3
И4	П3
И5	ПI
И6	П3

П [П#, ГОРОД]	
П#	ГОРОД
ПI	Тула
П2	Тула
П3	Серпухов
П4	Серпухов
П5	Чехов

П [И#, П#] * П [П#, ГОРОД]		
И#	П#	ГОРОД
ИI	ПI	Тула
ИI	П2	Тула
ИI	П5	Чехов
И2	ПI	Тула
И2	П2	Тула
И2	П4	Серпухов
И3	П3	Серпухов
И4	П3	Серпухов
И5	ПI	Тула
И6	П3	Серпухов

Ответ:

И#	ГОРОД
И1	Тула
И1	Чехов
И2	Тула
И2	Серпухов
И3	Серпухов
И4	Серпухов
И5	Тула
И6	Серпухов

Упражнения

1. Определить, являются ли рефлексивными, симметричными и транзитивными следующие отношения:

R_1 – отношение непосредственного алфавитного предшествования (на некотором множестве слов);

R_2 – отношение "состоять из одинакового числа букв" (на произвольном множестве слов);

R_3 – отношение антонимии (на произвольном множестве слов);

R_4 – отношение соавторства на множестве авторов {Иванов, Петров, Сидоров, Чернов, Белов, Рыжов, Михайлов}, выступивших в печати в следующих авторских коллективах:

а) Иванов, Петров, Сидоров;

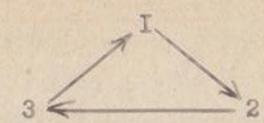
б) Сидоров, Чернов;

в) Белов, Рыжов;

г) Белов, Чернов;

д) Иванов, Михайлов, Чернов.

2. Построить транзитивное замыкание отношения, задаваемого следующим графом:



К какому типу принадлежит это отношение?

3. Перевести на язык операций над отношениями следующие вопросы:

(1) Какие детали (коды) поставляют предприятия города Тулы?

(2) Из каких городов могут быть получены болты?

Литература

1. Шиханович Ю.А. Введение в современную математику. – М.: Наука, 1965.

2. Столл Р.Р. Множества. Логика. Аксиоматические теории. – М.: Просвещение, 1968. Гл. 1, § 6, 7, 10.

3. Шрейдер Ю.А. Равенство, сходство, порядок. – М., Наука, 1971.

4. Date C. J. An introduction to database systems. – Reading (Mass.) : Addison Wesley, 1975.

Г л а в а IУ. НЕКОТОРЫЕ ПОНЯТИЯ МАТЕМАТИЧЕСКОЙ ЛИНГВИСТИКИ

Главная задача лингвиста, изучающего естественный язык, - описать ф о р м у языковых выражений, то есть описать, как они построены, и охарактеризовать их с м ы с л, то есть указать, что они означают. Форма языкового выражения - это то, что передается отправителем сообщения и воспринимается адресатом, то есть последовательности букв, звуков и т.д. Смысл - это то, что имеется в виду отправителем и понимается адресатом.

Соответствие между формой и смыслом в языке не является ни простым, ни однозначным. Возможны разные по форме предложения, которые имеют один и тот же смысл, см. пример (I):

(I) а. Иван уехал потому, что у него заболела мать;

б. Причиной отъезда Ивана была болезнь матери;

в. У Ивана заболела мать, и он уехал;

г. Поскольку у Ивана заболела мать, он уехал;

д. Болезнь матери заставила Ивана уехать;

е. Иван уехал из-за болезни матери.

В то же время может быть так, что одно и то же предложение допускает разные понимания, то есть имеет более чем один смысл, см.пример (2):

(2) а. Он начал печь (то ли 'строить печь', то ли 'печь нечто', например, пироги);

б. Сухопарая торговка вяленой воблой торчала среди ящиков (то ли женщина торговала воблой, то ли она торчала, как вобла);

в. Врач подошел к мужичку на костылях (то ли мужичок был на костылях, то ли врач);

г. Лошадь Пьера отставала от адъютанта и равномерно встряхивала его (его - то ли Пьера, то ли адъютанта).

Примеры (I) и (2) показывают, что общий смысл предложения зависит не только от смысла его слов, но и от синтаксической структуры предложения. Синтаксическая структура предложения - это совокупность сведений о связях между его словами и словосочетаниями. Ниже в §I-3 излагаются способы описания синтаксической структуры предложения, разработанные в математической лингвистике; в § 4 вводится понятие порождающей грамматики.

§ I. Представление структуры предложения в виде дерева подчинения

Слова в предложении связаны друг с другом по смыслу и по форме, то есть находятся в определенном отношении друг к другу; причем отношение это, вообще говоря, несимметричное: одни слова как бы подчиняют себе другие. Ф о р м а л ь н о е подчинение состоит в том, что одно слово определяет грамматическую форму другого; например, существительное определяет род, число и падеж согласованного с ним прилагательного и, вообще, определения, ср. дом большой - река большая - село большое - реки большие и т.д.; предлог определяет падеж управляемого им существительного, ср. у школы (* у школе), но при школе (* при школы); глагол определяет выбор предлога, ср. войти в дом, но отойти от дома и т.д. С м ы с л о в о е подчинение состоит в том, что слову, как правило, соответствует некоторая ситуация с определенным набором участников, и слово, обозначающее такую ситуацию, подчиняет себе слова, обозначающие участников этой ситуации и ее сопутствующие обстоятельства; например, в предложении Художник нарисовал этот портрет углем слово нарисовал обозначает ситуацию с тремя участниками – кто?, что? и чем? и, соответственно, подчиняет себе слова художник, портрет и углем; в предложении Книга лежит на столе слово лежит подчиняет себе слова книга и на (столе); в сочетании книга и тетрадь соединительный союз и подчиняет себе слова, которые он соединяет.

Предложение состоит из слов, точнее – из словоформ. Синтаксическая структура предложения может быть представлена деревом синтаксического подчинения, или просто деревом подчинения, заданным на множестве словоформ предложения.

Введем определения.

Определение 1. Бинарное отношение $\mathcal{D}$ на множестве M называется отношением синтаксического подчинения, если оно обладает следующими свойствами:

1) У каждого элемента $x \in M$, кроме одного, есть хотя бы одно подчиняющее. Иначе говоря, для всех $x \in M$, кроме одного, существует $y \in M$ такой, что $\langle y, x\rangle \in \mathcal{D}$. Элемент, который не имеет подчиняющего, называется вершиной M.

2) Ни один $x \in M$ не может иметь двух подчиняющих. Иначе говоря, не существует $y, z \in M$ ($y \neq z$) таких, что $\langle y, x\rangle \in \mathcal{D}$ и $\langle z, x\rangle \in \mathcal{D}$.

3) Если x – вершина, то для любого $y \in M$ ($y \neq x$) $\langle x, y\rangle \in \hat{\mathcal{D}}$ (где $\hat{\mathcal{D}}$ – транзитивное замыкание отношения $\mathcal{D}$, см. § 2 главы III).

Иначе можно сказать, что $\mathcal{D}$ есть синтаксическое подчинение, если его транзитивное замыкание $\hat{\mathcal{D}}$ есть отношение древесного порядка (см. § 4.2 главы III). Таким образом, если $\mathcal{D}$ есть синтаксическое подчинение, то его граф (полный) является деревом.

Если $\langle x,y \rangle \in \mathcal{D}$, говорят, что x синтаксически подчиняет y, или просто, что x подчиняет y.

Определение 2. Объект $\langle M, \mathcal{D} \rangle$, где M – множество словоформ предложения $\mathcal{S}$, а $\mathcal{D}$ – отношение синтаксического подчинения на M, называется деревом подчинения предложения $\mathcal{S}$.

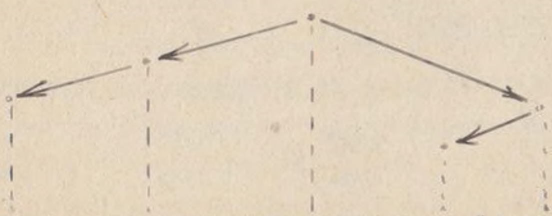
Пример. Дерево подчинения предложения (1) Тамбовские колхозники собрали богатый урожай см. на рис. 21.

Рис. 21.

Определение 3. Объект $\langle M, \mathcal{D}, \mathcal{L} \rangle$, где $\mathcal{D}$ – синтаксическое подчинение на множестве M, а $\mathcal{L}$ – отношение строгого порядка на M, называется расположенным деревом подчинения.

Содержательно, отношение $\mathcal{L}$ – это линейное расположение слов в предложении.

Пример. На рис. 22 указаны два способа графического изображения расположенного дерева подчинения предложения (1).



а) Тамбовские колхозники собрали богатый урожай

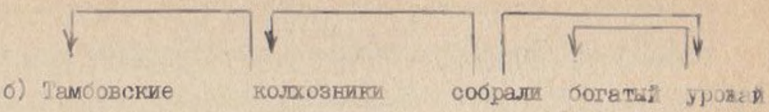


Рис. 22.

Возможность сопоставить предложению два различных дерева подчинения обычно свидетельствует о его неоднозначности. Примеры см. на рис. 23.

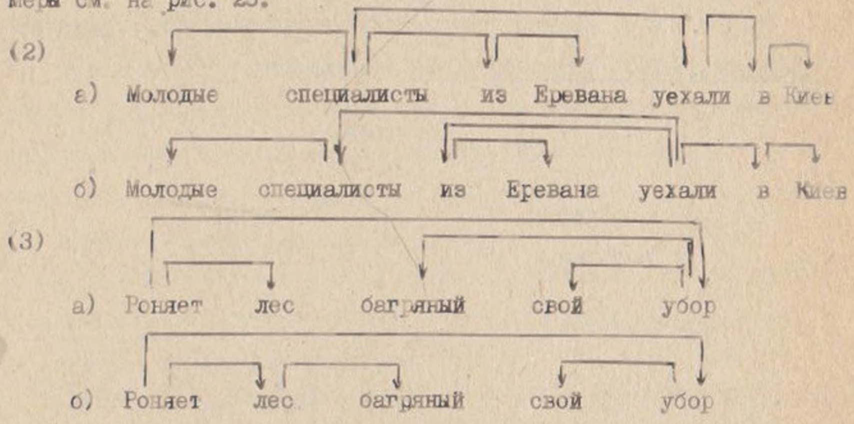
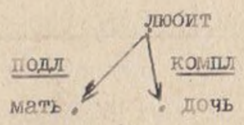


Рис. 23

Иногда различие пониманий предложения может быть выражено деревом подчинения только при условии, что различаются <u>типы</u> подчинения. Дерево подчинения с различением типов подчинения называется <u>размеченным деревом подчинения</u>. На графе тип подчинения указывается пометой при стрелке. Примеры см.на рис. 24.

(4) <u>Мать любит дочь</u>:

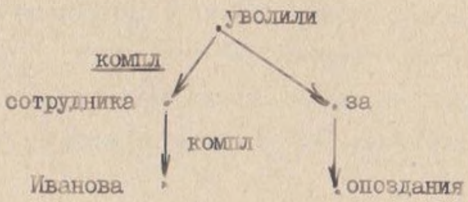
а)



б)

(5) Сотрудника Иванова уволили за опоздания:

а)



б)

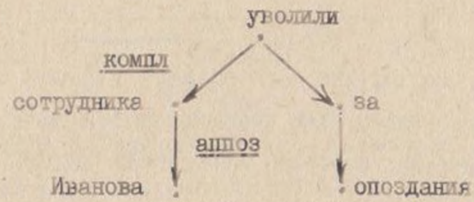


Рис. 24

О б о з н а ч е н и я : подл (от слова подлежащее) - подчинение сказуемому подлежащего; компл (от слова комплетив) - подчинение глаголу или отглагольному имени дополнения; аппоз (от слова аппозитив) - подчинение слову приложения.

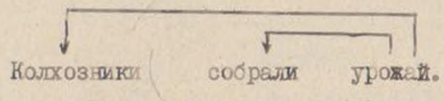
Задачи, связанные с деревом подчинения, могут быть следующие.

З а д а ч а I. Определить, является ли отношение R , заданное на множестве словоформ М, синтаксическим подчинением.

З а д а ч а 2. Понять, соответствует ли данное синтаксическое подчинение $\mathcal{D}$ на множестве словоформ предложения S' самому предложению S ; т.е. является ли данное синтаксическое подчинение адекватным отражением синтаксических связей в данном предложении.

З а д а ч а 3. Построить алгоритмическую процедуру, приписывающую предложению его дерево подчинения.

Задача I решается элементарно: достаточно проверить, удовлетворяет ли предъявляемое отношение условиям $I^0 - 3^0$ из определения I. Задача 2 более сложная. В простых случаях соответствие и несоответствие отношения $\mathcal{D}$ предложению очевидно; например, отношение $\mathcal{D}$, задаваемое графом примера (6), является синтаксическим подчинением, но очевидным образом не соответствует предложению (6):

(6) 

Однако, скажем, для предложения (7) способ представления его синтаксических связей в виде дерева подчинения не вполне ясен:

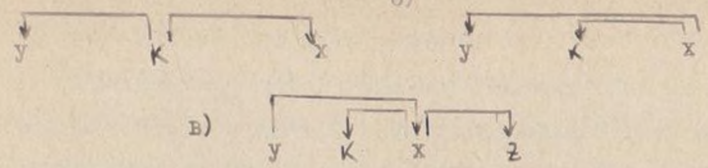
(7) Солнце всходит и заходит.

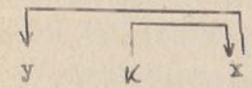
Задача 3 решается применительно к тому или иному решению проблемы 2, в частности, в системах автоматического анализа текста при машинном переводе.

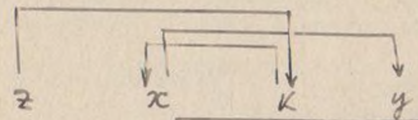
§ 2. Понятие проективности

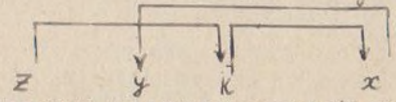
О п р е д е л е н и е. В расположенном дереве подчинения $\langle M, \mathcal{D}, \mathcal{L} \rangle$ отношение синтаксического подчинения $\mathcal{D}$ называется проективным относительно линейного порядка $\mathcal{L}$ (или иначе: для отношений $\mathcal{D}$ и $\mathcal{L}$ на множестве М соблюдается условие проективности), если для любых $x, y, k \in M$ таких, что $\langle x, y \rangle \in \mathcal{D}$ а k лежит между х и у, $\langle x, k \rangle \in \hat{\mathcal{D}}$.

Так, в примере (I) условие проективности выполнено, а в примерах (2) - (4) отношение $\mathcal{D}$ непроективно относительно $\mathcal{L}$:

(I) а) 

(2) 

(3) 

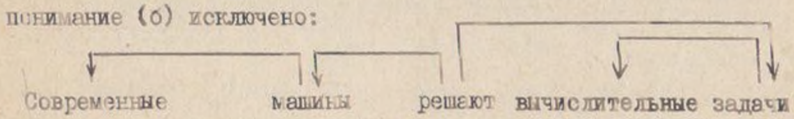
(4) 

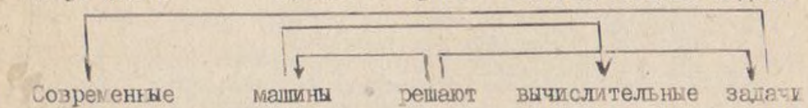
Применительно к принятому способу графического изображения расположенного дерева подчинения (см. рис. 22б), критерий проективности состоит в том, что стрелки, выражающие подчинение: а) не должны пересекаться (как в (3) и (4)); б) не должны покрывать вершину (как в (2)).

Как правило, при наличии разных возможностей человек понимает предложение таким образом, чтобы отношение подчинения в нем было проективным. Так для предложения

(5) Современные машины решают вычислительные задачи

естественно понимание (а) с проективным подчинением, а непроективное понимание (б) исключено:

а) 

б) 

Есть предположение, что у строчек из "Евгения Онегина"

Он из Германии туманной

Привез учености плоды

(не вполне понятых, поскольку эпитет <u>туманная</u> не очень естествен по отношению к Германии) имеется понимание, которое в смысловом отношении более естественно - Он из Германии привез плоды туманной учености - но которое обычно не приходит в голову, поскольку оно соответствует непроективному подчинению:

Он из Германии туманной привез учености плоды.

Проективность дает такое расположение слов, при котором синтаксически связанные слова находятся, в целом, в максимальной близости друг к другу.

Некоторые типы структур, встречающиеся в русском просторечии, воспринимаются как стилистические дефектные именно из-за нарушения проективности:

(6) Прибывающий электропоезд на 2-й путь следует до Мытищ

(7) Купленные товары в другом магазине предъявляйте продавцу

Не следует, однако, думать, что непроективность во фразах естественного языка полностью исключена. Так, фраза (8), хотя она и непроективна, вполне удовлетворительна:

(8) <И> волчью вашу я давно натуру знаю

§ 3. Система составляющих

В общем случае синтаксические (формальные и смысловые) связи в предложении не сводятся естественным образом к связям между отдельными словами. Поэтому возможен другой способ представления синтаксической структуры предложения, при котором в предложении выделяются группировки слов (как правило, это цельные "отрезки" предложения), связанные друг с другом. Определенным образом устроенное множество отрезков предложения (см. ниже) называется его системой составляющих, а каждый отрезок в этой системе – составляющей предложения. Пример системы составляющих (все выделяемые отрезки – кроме однословных – ограничены скобками):

(1) [[Тамбовские колхозники] [собрали [богатый урожай]]].

О п р е д е л е н и е. <u>Системой составляющих</u> называется множество $\sum$ отрезков предложения (состоящих из одного или нескольких слов), такое, что:

1) В $\sum$ входят все отрезки, состоящие из одного слова, и отрезок, равный всему предложению.

2) Отношение строгого включения $\supset$ на множестве отрезков $\sum$ является древесным порядком, т.е.

а) если для $x, y, z \in \sum$ верно, что $x \supset z$ и $y \supset z$, то или $x \supset y$ или $y \supset x$ (т.е. x и y сравнимы по $\supset$);

б) существует наибольший отрезок, т.е. такой отрезок $x \in \sum$, что для любого $y \in \sum$ $(y \neq x)$ верно, что $x \supset y$.

Отрезки из $\sum$ называются <u>составляющими</u>.

Из условия (2а) следует, что составляющие не могут частично пересекаться (см. § 3 главы II); т.е. если $x \cap y \neq \emptyset$, то либо $x \supset y$, либо $y \supset x$. Содержательно это означает, что при обозначении границ отрезков скобками не может быть "неправильных" расстановок скобок – например, таких: (Брат [его) пьет]. Два отрезка, ограниченных скобками, могут иметь общий элемент только в том случае, если один из них целиком входит в другой; если отрезки не входят целиком один в другой, то они вообще не пересекаются.

Система составляющих предложения может быть представлена графом (сокращенным) отношения $\supset$ на множестве составляющих Σ. Этот граф называется <u>деревом составляющих</u>. Граф, который задает одновремеьно отношение $\supset$ и отношение $\mathcal{L}$ линейного расположения составляющих в предложении называется <u>расположенным деревом составляющих</u>, см. рис. 25.

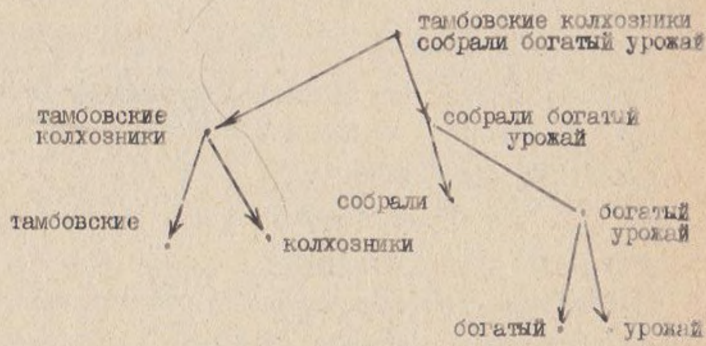


Рис. 25.

Дерево составляющих, как правило, изображается сокращенным образом: выписываются полностью только те отрезки предложения, которые соответствуют концевым узлам дерева, а неконцевым узлам соответствует тот отрезок предложения, который получается, если прочесть подряд все слова, до которых можно "дойти" по стрелкам от данного узла (рис. 26).

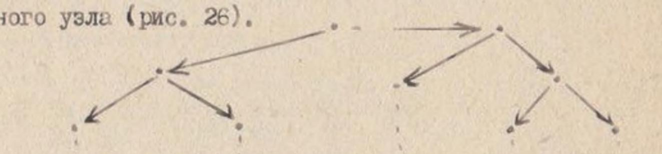


Рис. 26.

Система составляющих называется <u>размеченной</u>, если каждой составляющей сопоставлен ее грамматический тип. Пример. Размеченная система составляющих представлена с помощью дерева на рис. 27.

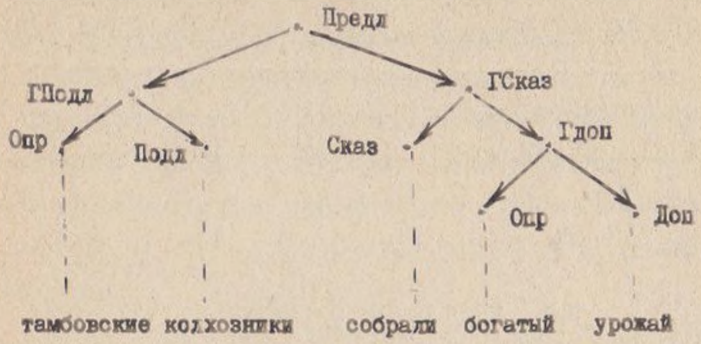


Рис. 27.

Сокращения: <u>Подл</u> - подлежащее, <u>Сказ</u> - сказуемое, <u>Опр</u> - определение; <u>ГПодл</u>, <u>Гдоп</u> и <u>ГСказ</u> - группа, соответственно, подлежащего, дополнения и сказуемого; <u>Предл</u> - предложение.

Размеченная система составляющих может быть задана также с помощью размеченных скобок:

$$\Big\{\Big[\big(\text{Тамбовские}\big)_{\text{Опр}}\ \big(\text{колхозники}\big)_{\text{Подл}}\Big]_{\text{ГПодл}}\ \Big[\big(\text{собрали}\big)_{\text{Сказ}}\ \Big[\big(\text{богатый}\big)_{\text{Опр}}\ \big(\text{урожай}\big)_{\text{Доп}}\Big]_{\text{ГДоп}}\Big]_{\text{ГСказ}}\Big\}_{\text{Предл}}$$

Система составляющих, как и дерево подчинения, может быть использована для различения разных пониманий неоднозначного предложения:

(1) а) ([Молодые специалисты] (из Еревана)] [уехали [в Киев]];

б) [Молодые специалисты] [[из Еревана] уехали (в Киев]].

Пример предложения, в котором для различения омонимии необходима размеченная система составляющих:

(2) а) $\left\{ (\text{Победители})_{\text{Подл}} \; - \; (\text{бакинцы})_{\text{Сказ}} \right\}_{\text{Предл}}$

б) $\left\{ (\text{Бакинцы})_{\text{Подл}} \; - \; (\text{победители})_{\text{Сказ}} \right\}_{\text{Предл}}$

Система составляющих называется иерархированной, если в каждом наборе непосредственных составляющих выделена главная. Иерархированная система составляющих однозначно определяет для предложения его дерево подчинения. Таким образом, между двумя способами представления структуры предложения имеется естественная связь.

§ 4. Порождающие грамматики

Одна из главных задач описания синтаксиса естественного языка – определить, какие сочетания слов воспринимаются как правильные предложения (см. общие задачи синтактики при описании знаковой системы в § 3 главы I). Не менее важно определить источники неправильности реально встречаемых предложений. Какие, например, закономерности нарушены в предложениях (I) – (4) ?

(I) *Подъезжая к станции, у меня упала шляпа;

(2) *Твоя моя не понимай;

(3) *Я взял мой пиджак;

(4) *Это зависит.

Порождающая грамматика позволяет решать задачу исчерпывающего описания множества правильных предложений языка, одновременно эксплицитно указав действующие в нем синтаксические закономерности.

Естественно, при решении этой задачи нельзя рассматривать каждое предложение изолированно. Чтобы решить эту задачу, надо признать, что некоторые предложения, разные по лексическому составу, имеют одну и ту же синтаксическую структуру; что какие-то предложения, разные по сложности, имеют сходную синтаксическую структуру. Например, одну и ту же синтаксическую структуру имеют предложения:

(5) а. Тамбовские колхозники собрали богатый урожай;

б. Молодые специалисты решили сложную задачу;

в. Молодые колхозники увидели большой город.

Предложения примера (6) близки по структуре к предложениям примера (5), но структура (6а) проще, а (6б) сложнее, чем (5):

(6) а. Колхозники собрали урожай;

б. Молодые тамбовские специалисты решили сложную вычислительную задачу.

В основе порождающей грамматики лежит идея структурного сходства между лексически различными предложениями и идея структурной близости более сложных предложений к более простым.

Формально, порождающая грамматика - это объект вида $\langle W, V, \mathcal{I}, R \rangle$, где W - словарь терминальных символов (т.е. слов языка), V - словарь вспомогательных символов (т.е. грамматических типов составляющих), $\mathcal{I}$ - начальный символ, R - множество правил подстановки вида $\mathcal{I} \to \varphi$, где $\mathcal{I} \in V$, а φ - цепочка элементов из $W \cup V$. Вывод в порождающей грамматике - это последовательность строк таких, что I-я строка - это $\mathcal{I}$, последняя состоит из одних только терминальных символов и каждая последующая получается из предыдущей применением какого-нибудь правила подстановки. Цепочка порождается данной грамматикой, если она является последней строкой некоторого вывода (то есть если она может быть выведена правилами подстановки из начального символа и не содержит вспомогательных символов).

Пример I. Порождающая грамматика G_I, задающая один простой фрагмент русского языка:

R = { Предл ⟶ Подл Сказ;

Подл ⟶ колхозники, специалисты;

Сказ ⟶ трудятся, работают, отдыхают } ;

W = колхозники, специалисты, трудятся, работают, отдыхают;

V = { Предл, Подл, Сказ }

I = Предл.

Один из возможных выводов в данной порождающей грамматике:

```
              Предл
             /     \
         Подл       Сказ
          |           |
     колхозники     Сказ
          |           |
     колхозники    трудятся
```

Эта грамматика порождает, следовательно, предложение (1) <u>Колхозники трудятся</u>, а также (2) <u>Колхозники отдыхают</u>, (3) <u>Специалисты трудятся</u> и т.д.

П р и м е р 2. Порождающая грамматика G_2 , задающая чуть более сложный фрагмент русского языка.

R = { Предл ⟶ ГПодл ГСказ,

ГПодл ⟶ Опр ГПодл,

ГПодл ⟶ Подл,

ГСказ ⟶ Сказ ГДоп,

ГДоп ⟶ Опр ГДоп,

ГДоп ⟶ Доп,

Подл ⟶ колхозники, специалисты, машины,

Сказ ⟶ собирают, решают, посещают,

Доп ⟶ урожай, задачи, города,

Опр ⟶ тамбовские, молодые, большие, сложные, вычислительные } ;

W = { колхозники, ... , сложные, вычислительные } ;
V = { Предл, ГПодл, ГСказ, Опр, Подл, Сказ, ГДоп, Доп } ;
S = Предл.

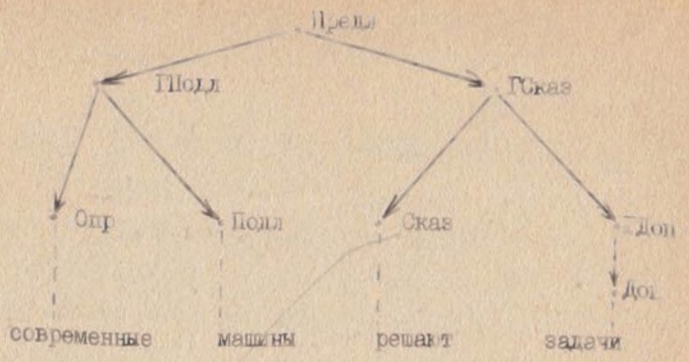
<u>Пример</u> вывода в грамматике G_2 :

Предл
ГПодл ГСказ
Опр Подл ГСказ
Опр Подл Сказ ГДоп
Опр Подл Сказ Доп
современные Подл Сказ Доп
. . . .
современные машины решают задачи

Эта грамматика порождает, помимо предложения (4) Современные машины решают задачи, другие вполне нормальные предложения - (5) Современные машины решают сложные вычислительные задачи; (6) Молодые тамбовские колхозники собирают большие урожаи; (7) Современные специалисты посещают большие города и т.д. Наряду с этим, она порождает большое количество предложений в том или ином отношении аномальных - например, предложение (8) *Вычислительные специалисты посещают тамбовские урожаи, которое, однако, с синтаксической точки зрения вполне правильно.

Вывод предложения в грамматике автоматически сопоставляет ему размеченное дерево составляющих. Например, вывод предложения (4) сопоставляет ему следующее дерево (чтобы получить дерево по выводу, достаточно в выводе соединить каждый символ с теми символами, в которые он перешел при применении правила подстановки, и устранить повторения):

6-4



Литература

1. Гладкий А.В. Формальные грамматики и языки. – М.: Наука, 1973, с. 282–304 (прил. I. Системы составляющих и деревья синтаксического подчинения).

2. Падучева Е.В. О способах представления синтаксической структуры предложения. – Вопросы языкознания, 1964, № 2, с. 99–113.

3. Иорданская Л.Н. Свойства правильной синтаксической структуры и алгоритм ее обнаружения. – Проблемы кибернетики, 1964, вып. 11, с. 215–244.

Указатель терминов

СОДЕРЖАНИЕ

Формат бумаги 60x84 I/I6
В печать от 4/XII-1979 г. Т-I98II Тираж 4000 экз.
Печ.л. 5,75 Уч.-изд.л. 3,78 Заказ 9373

Производственно-издательский комбинат ВИНИТИ
Люберцы, Октябрьский проспект, 403