



ИПМ им.М.В.Келдыша РАН • Электронная библиотека

Препринты ИПМ • Препринт № 140 за 1986 г.



ISSN 2071-2898 (Print)
ISSN 2071-2901 (Online)

Гаврилов С.В., [Кирильченко А.А.](#),
[Кугушев Е.И.](#), [Платонов А.К.](#)

Программный комплекс для
создания и анализа
семантических сетей,
описывающих системы
ПОНЯТИЙ

Рекомендуемая форма библиографической ссылки: Программный комплекс для создания и анализа семантических сетей, описывающих системы понятий / С.В.Гаврилов [и др.] // Препринты ИПМ им. М.В.Келдыша. 1986. № 140. 24 с.
URL: <http://library.keldysh.ru/preprint.asp?id=1986-140>

О р д е н а Л е н и н а
ИНСТИТУТ ПРИКЛАДНОЙ МАТЕМАТИКИ
имени М.В. Келдыша
А к а д е м и и н а у к С С С Р

С.В. Гаврилов, А.А. Кирильченко, Е.И. Кугушев, А.К. Платонов

ПРОГРАММНЫЙ КОМПЛЕКС
ДЛЯ СОЗДАНИЯ И АНАЛИЗА СЕМАНТИЧЕСКИХ СЕТЕЙ,
ОПИСЫВАЮЩИХ СИСТЕМЫ ПОНЯТИЙ

Препринт № 140 за 1986 г.

Москва 1986 г.

АННОТАЦИЯ

Речь идет о диалоговой системе для создания, редактирования и анализа описания предметной области в виде семантической сети (графа). Для сетей больших размеров требуется удобный язык редактирования и быстрый доступ. Создана система, удовлетворяющая таким требованиям. В работе содержится общее руководство по использованию системы, описаны результаты конкретных применений.

Ключевые слова и фразы: граф, семантическая сеть, тезаурус, система понятий.

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ	3
ОПИСАНИЕ СИСТЕМЫ "ТЕЗАУРУС"	4
Структура данных.....	4
Структура "Словаря".....	5
Структура "Графа",	6
Язык редактирования	7
<i>Редактор "Словаря"</i>	7
<i>Редактор "Графа"</i>	9
Процедуры для исследования графа.....	12
<i>Процедуры ТСУ, ТW</i>	12
<i>Алгоритмы перебора на графе</i>	13
ВОЗМОЖНЫЕ ПРИМЕНЕНИЯ СИСТЕМЫ "ТЕЗАУРУС"	14
ИССЛЕДОВАНИЕ ЗАМКНУТОЙ СИСТЕМЫ ПОНЯТИЙ.....	14
Схема тезауруса понятий.....	14
Цели исследования и общая характеристика результата	16
"Отладка" графа понятий.	17
<i>Поиск противоречий типа "цикл"</i>	17
<i>Взаимоотношение связей разных типов</i>	18
<i>Анализ взаимных связей</i>	18
ХАРАКТЕРИСТИКИ ЗАМКНУТОЙ СИСТЕМЫ ПОНЯТИЙ	21
Анализ иерархической структуры	21
Частотный анализ	22
ЗАКЛЮЧЕНИЕ	24
Литература	24

ВВЕДЕНИЕ

При работе со структурами типа семантических сетей возникает необходимость хранения и анализа сложных графов на ЭВМ. Граф, как представление знаний посредством описания связей объектов, удобен для широкого диапазона проблем. Количество задач, возникающих при этом, достаточно велико. Однако анализ этих задач позволяет выделить ряд алгоритмов, к решению которых обычно сводятся обработка и исследование (см. табл. I).

Таблица1. Анализ задач, возникающих при работе с графами

Области применения	Требования "отладки" графа	Задачи исследования
1. Словарь: <i>вершины</i> - слова, <i>рёбра</i> - ссылки	Отсутствие циклов	Анализ цепочек иерархического подчинения
2. План (карта): <i>вершины</i> - тупики и перекрёстки. <i>рёбра</i> - дорога	Перекрёсток - не менее трёх связей. тупик - одна связь	Нахождение пути
3. Описание изображения: <i>вершины</i> – фигуры, <i>рёбра</i> - взаимное расположение (слева - справа, дальше - ближе и др.)	Транзитивность одноимённых отношений взаимного расположения	Сравнение участков изображений
4. Блок-схемы программ: <i>вершины</i> – блоки, <i>рёбра</i> - передачи управления	Наличие входа и выхода для любого блока, кроме "Финиш" и "Старт"	Эффективность выполнения программ
5. Комплекс программ: <i>вершины</i> - операторы и переменные. <i>рёбра</i> - "вход" и "выход" оператора	Чередование операторов с переменными	Нахождение последовательности операторов по указанным переменным "входа" и "выхода"

Прежде всего, задачи могут быть разделены в соответствии с этапами работы

1. Формирование и заведение структуры данных.
2. "Отладка" графа.
3. Исследование полученной системы, выявление закономерностей.

В связи с этим разработаны соответствующие программные комплексы. Построенная для этих целей система "Тезаурус" включает возможности редактирования и исследования сетей достаточно больших размеров и сложности. Язык редактирования обеспечивает чтение и запись данных, обновление или исключение отдельных элементов. Программы

обработки предназначены для анализа сетей, выявления закономерностей или противоречий.

Созданный комплекс программ обработки открыт для расширения и включает стандартные процедуры для работы с графами: сравнение элементов, нахождение кратчайшего пути, поиск циклов и др.

В работе приводятся результаты примера процесса создания и исследования сетей, построенных на базе системы "Тезаурус", — исследование системы психологических понятий. Система психологических понятий построена на базе словаря К.К. Платонова [1]. Словарь содержит около 900 слов. В определениях словаря используются связи трех типов. Количество слов и данный набор связей оптимальны для использования словаря в качестве теста системы.

В связи с проблемой анализа и синтеза схем расчета для определения характеристик движения роботов и космических аппаратов построены алгоритмы процедур перебора на графе типа И-ИЛИ. На основе выполненного тестирования системы "Тезаурус" сделан вывод о возможности её использования для автоматизированного поиска решений в системе проектирования орбит космического аппарата или систем управления мобильных роботов.

ОПИСАНИЕ СИСТЕМЫ "ТЕЗАУРУС"

Комплекс программ "Тезаурус" служит для обработки семантических сетей смысловых связей понятий (слов или словарных статей). Хранение данных на внешних носителях в сжатой форме даёт возможность работать с большим количеством слов.

Система (как комплекс программ, так и структура данных) разделяется на две части, которые могут использоваться независимо:

- "Словарь" — для хранения словарных статей и их редактирования: выборки в алфавитном порядке, поиска по контексту, замены, исключения, добавления и др.;
- "Граф" — для работы с графом.

Система "Тезаурус" позволяет строить графы, вершинами которых являются словарные статьи "Словаря". Направленные рёбра графа поименованы также статьями "Словаря". Внутри "Словаря" различий между ребрами и вершинами "Графа" не существует.

Структура данных

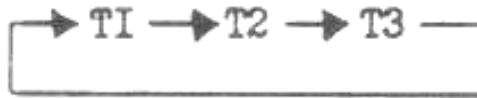
Весь объем данных состоит из 5 файлов, длины которых задаются пользователем при создании словаря:

T1 — файл словарных статей,

- T2 — линейный каталог,
- T3 — алфавитный каталог,
- T4 — каталог узлов,
- T5 — файл описания связей.

Структура "Словаря"

"Словарь" занимает 3 файла: T1, T2, T3, составляющих в целом зацикленную структуру со ссылками:



Полная схема взаимных ссылок показана на рис. 1. Такая структура устанавливает взаимно-однозначное соответствие между самими словами, их линейными номерами (номер слова в порядке поступления) и их алфавитными номерами. Линейный номер закрепляется постоянно за словом и используется для формирования связей в "Графе". Алфавитный номер применяется при двоичном поиске и при выборке в алфавитном порядке.

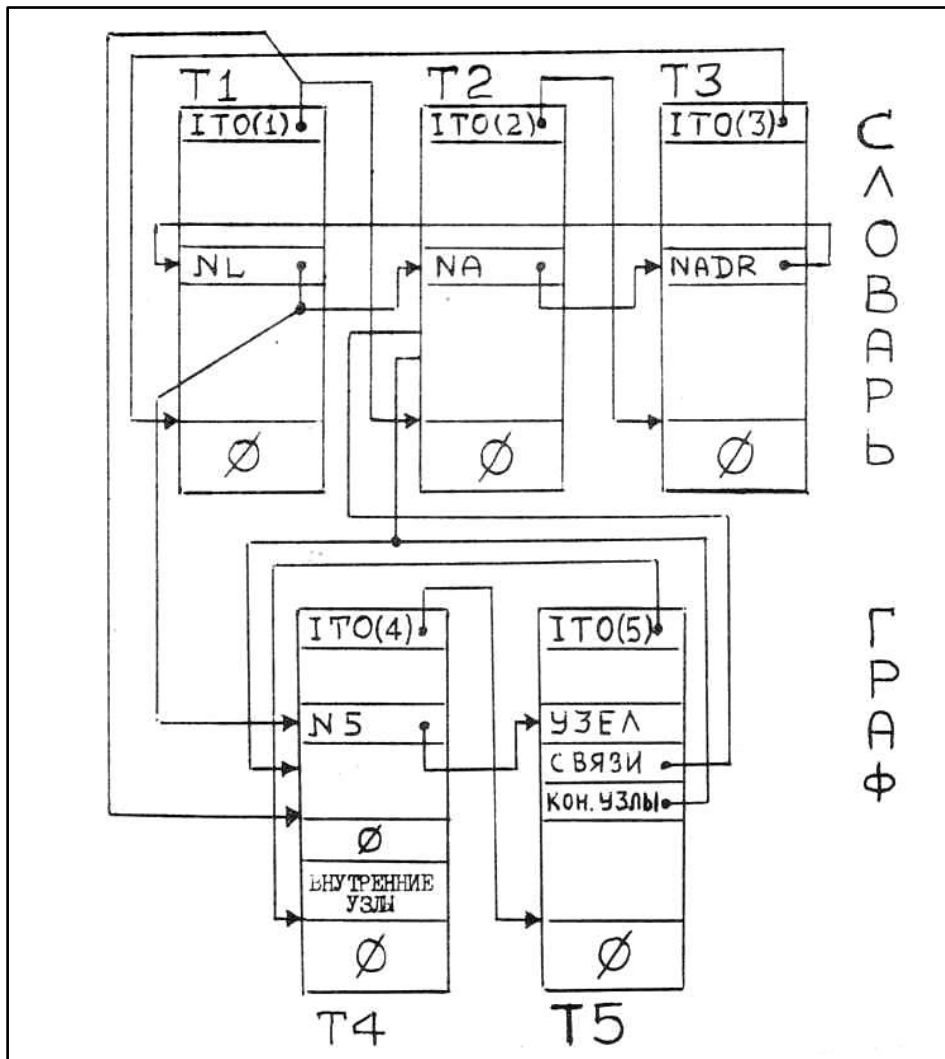


Рис. 1. Структура данных и схема взаимных ссылок файлов T1-T5.

T1 — файл словарных статей. Элемент отдельной записи включает:

1. Линейный номер NL – номер слова в порядке поступления → ссылка на T2.
2. Число букв $L \leq 80$ и Слово словарной статьи.

T2 — линейный каталог. Этот файл состоит из алфавитных номеров NA.

Алфавитные номера (ссылки на T3) расположены в порядке возрастания NL.

T3 — алфавитный каталог. Элемент файла — адрес словарной статьи в T1 — NADR.

Пример:

Параметры словарной статьи "АЗАРТ": NA=46, NADR=32, NL=11, L =5

NADR – Индекс массива файла T1 в памяти БЭСМ-6: T1(32):

T1(33):

NL → Место в массиве файла T2 = $11/3 = 3+2/3 \Rightarrow T2(3)$:

NA → Место в массиве файла T3 = $46/3 = 15+1/3 \Rightarrow T3(15)$:

И	И	5	А	З	А
Р	Т				
.....			46	
.....		32		

Структура “Графа”

Семантическая сеть описывается в файлах T4, T5 с использованием линейных номеров, каждому из которых в "Словаре" поставлена в соответствие словарная статья. Вершины графа, не имеющие словесных имён, будем называть "внутренними" узлами.

T4 — каталог узлов. Элемент файла содержит:

1. N5 — адрес описания соответствующего узла, ссылка на T5.
2. Признак узла IP1 :

$$IP1 = \begin{cases} 0 & \text{— узел "внутренний"} \\ 1 & \text{— узел "текстовый"} \end{cases}$$

3. Рабочий признак IP5. Признак IP5 может быть использован для хранения меток при узлах. Предусмотрено изменение IP5 по заданию пользователя.

T5 — файл описания связей. Описание узла содержит связи двух направлений: вход и выход. При добавлении новой связи формируются два описания со связями противоположного направления в разных узлах. Такое дублирование информации даёт быстрый доступ к данным в алгоритмах движения по графу. Ниже перечислены элементы описания для одного узла.

T5(N5) — заголовок узла:

NL — линейный номер начального узла,

LY — количество связей в узле,

LYY — количество конечных узлов (суммарное по всем связям).

$T5(N5+1) - T5(N5+LY)$ — заголовки связей.

Заголовок связи:

LC — линейный номер связи,

ID — направление связи:

$$ID = \begin{cases} 0 & \text{— "выход"} \\ 1 & \text{— "вход"} \end{cases}$$

LC — длина связи (количество конечных узлов),

NB — начало описания — первый узел в списке конечных узлов.

Узлы, относящиеся к данной связи, расположены в списке конечных узлов подряд, начиная с № NB.

$T5(N5+LY+1) - T5(N5+LY+(LYY+2)(3))$ - список линейных номеров (NL) конечных узлов.

Язык редактирования

Язык редактирования предназначен для внесения изменений в структуру данных, чтения элементов. Как и структура данных, язык разделяется на две части. Действия редактора "Словаря" ограничено рамками файлов T1, T2, T3. Редактор же "Графа" кроме изменений в T4-T5 может выполнять чтение элементов "Словаря" и добавление новых статей. Считывание инструкций редактирования организуется вызовом соответствующего монитора:

call слвр — для редактора "Словаря",

call ГРАФ — для редактора "Графа".

Перечень карт редактирования дан в таблицах 2,3.

Редактор "Словаря"

Редактор "Словаря" организует чтение, запись, замену или исключение словарных статей. Различаются карты редактирования двух типов:

1. \$... — управляющая карта (УК).
2. Словарная статья (СС). Первый символ отличен от \$, Длина — не более 72-х символов.

Первая карта — произвольная УК. Закрывается сегмент редактирования признаком конца \$E. Карты, следующие за \$E, не воспринимаются.

\$I — заведение словарной статьи в "Словаре". Одной карты достаточно для оформления приказа **на** заведение нескольких СС.

\$E — конец редактирования. По этой команде происходит возврат из редактора "Словаря" в монитор. Карта \$E необходима во всех случаях, когда требуется запись новых данных.

Виды управляющих карт даны в таблице 2.

Таблица 2 • Управляющие карты (УК) редактора "СЛОВАРЯ"

УК	Название	Синтаксис	Процедура
\$I	Заведение словарной статьи	\$I {CC}	IT
\$WC	Поиск по контексту	\$WC_[NL1[_NL2]] {CC}	WC
\$WA	Выборка по алфавитным номерам	\$WA_[NA1[_NA2]]	WR
\$WL	Выборка по линейным номерам	\$WL_[NL1[_NL2]]	WR
\$DA	Вычеркивание по алфавитным номерам	\$DA_[NA1[_NA2]]	DN
\$DL	Вычеркивание по линейным номерам	\$DL_[NA1[_NA2]]	DN
\$CA	Исключение вычеркнутых слов	\$CA_[NA1[_NA2]]	CL
\$RC	Замена слова	\$RC_[NLØ] CC	RT
\$E	Конец редактирования	\$E	IOEND (после return)
\$ØØ	Подготовка МД (МЛ)	\$ØØ_Nz1_Nz2	DØØØ

При выполнении \$E выводится на печать (экран):

"КОНЕЦ ЗАДАНИЯ"

и длины файлов:

ITO(1) — максимальный линейный номер,

ITO(2) — максимальный алфавитный номер,

ITO(3) — длина файла T1 в словах ЕЭСМ-6,

ITO(4) — длина файла T5 в словах БЭСМ-6,

ITO(5) — максимальный линейный номер узла.

\$WC[NL1 [NL2]] — поиск по контексту в интервале линейных номеров $NL1 \leq NL \leq NL2$. Контекст — карты типа CC, указанные после УК. Если интервал не задан, то выполняется **поиск по** всему "Словарю", т.е.: $NL1=1, NL2=ITO(1)$

\$WL[NL1 [NL2]] — печать словарных статей с линейными номерами в интервале $NL1 \leq NL \leq NL2$

\$WL NL1 = \$WL NL1 NL1

\$WL = \$WL 1 ITO(1)

\$WA [NA1 [NA2]] — выборка в алфавитном порядке.

Задаётся интервал алфавитных номеров.

\$WA NA1 = \$WA NA1 NA1

\$WA=\$WA 1 ITO(2)

\$DL[NL1 [NL2]] -вычеркивание слов по линейным номерам.

\$DA[NA1 [NA2]] -вычеркивание слов по алфавитным номерам.

Исключение слов требует изменения ссылок по всему "Словарю", поэтому введено временное "вычеркивание" слов* Впоследствии "вычеркнутые" слова можно исключить по приказу са или же восстановить на прежнем месте.

\$CA [NA1 [NA2]] — чистка "Словаря" (исключение вычеркнутых, слов).

Сжимается файл T1 и изменяются алфавитные номера.

Соответствие "линейный номер - слово" сохраняется.

\$00 N1 N2 — подготовка МД(МЛ). N1 - N2 — интервал зон, подлежащих чистке. Нумерация зон десятичная. Эта карта используется при формировании нового "Словаря".

\$RC NLO — замена слова. Новая словарная статья получает при заведении заданный линейный номер NLO. Прежнее слово с линейным номером NLO исключается.

ПРИМЕР

***CALL СЛВР**

\$I

АБУЛИЯ - заведение словарных статей

АБСТРАКЦИЯ

АВАРИЙНАЯ СИТУАЦИЯ

\$WA 1 — чтение слова с NA=1

\$WL 1 3 — чтение слов с NL = 1, 2, 3

\$WC — выборка по контексту

АБ — контекст

ЦИЯ — контекст

\$E — конец редактирования

*END FILE

Редактор "Графа"

Редактор "Графа" предоставляет возможность заводить, исключать, наменять или читать элементы семантической сети. Предусмотрена возможность задавать данные в словесной и численной форме. При добавлении новых связей, заданных словесным образом, предварительно заводятся отсутствующие в "Словаре" имена узлов, т.е. выполняются функции редактора "Словаря". Предусмотрено заведение общего узла для двух и более словарных статей, например, для слов-синонимов. Виды управляющих карт даны в табл. 3.

При словесном описании в УК перечисляется набор словарных статей, причем для управляющих карт \$1 и \$D первая СС воспринимается как начальный узел, вторая — как имя связи, остальные — как конечные узлы:

\$Д где Д=1 или Д=D.

УЗЕЛ

ИМЯ СВЯЗИ

КОН. УЗЕЛ 1

КОН. УЗЕЛ 2

... ..

ТАБЛИЦА 3. Управляющие карты (УК) редактора "ГРАФА"
<H>::= —>(<—

УК	Название	Синтаксис	Процедура
\$I	Добавление Связи	$\$I[_{NL}[(IP5)][_ {NC}[\langle H \rangle\{NL1[-NL2]\}]]]$ { CC } { CC }	ICY IY YY
\$D	Исключение Связи	$\$D[_{NL}_ {NC}[\langle H \rangle\{NL1[-NL2]\}]]]$ { CC } { CC }	DCY DC
\$W	Текстовая Печать	$\$W[_{NL1}[-NL2]\}]]]$ { CC }	WMM
\$L	Цифровая Печать	$\$L[_{NL1}[-NL2]\}]]]$ { CC }	WY
\$A	Атрибуты Узла	$\$A[_{NL1}[-NL2]\}]]]$ { CC }	WY WMM
\$C	Уплотнение Графа	$\$C[_{NL1}[-NL2]\}]]]$ { CC }	CL1
\$E	Конец редактирования	\$E	IOFND (после return)

$\$I [NL [(IP5)] [NC [\langle H \rangle\{NL1 [-NL2]\}]]]$ — добавление связи. $\langle H \rangle::= -\rangle\langle-$

а) $\$I NL(IP5) NC \rightarrow NL1$ — формируется узел с линейным номером NL, рабочим признаком (меткой) IP5. Если IP5 не задано, то сохраняется прежнее значение метки. Добавляются описания связей:

NL NC -> NL1

NL1 NC <- NL

б) $\$I NL (IP5)$ — заведение узла или изменение метки.

в) $\$I NL, N12$ — формируется общий узел для двух слов.

г) Случай словесного описания — эквивалентен случаям а), б) или в) в зависимости от количества указанных словарных статей.

$\$D [NL NC [\langle H \rangle\{NL1 [-NL2]\}]]]$ — исключение связи.

$\$D NL NC$ — связь исключается полностью со всеми конечными узлами.

Эквивалентная текстовая форма — две словарные статьи после УК \$D.

$\$W \{NL [-NL2]\}$ — печать узла в текстовом виде. Начальные узлы задаются перечислением словарных статей или линейных номеров и интервалов.

$\$L \{NL1 [-NL2]\}$ — печать информации о содержимом узла в цифровом виде.

Печатаются (выводятся на экран) линейные номера связей и конечных узлов.

\$A {NL1 [-NL2]} — печать атрибутов узла: N5, LY, LYY, IP5, IP1 (см. 1.1.2).

\$C {NL1 [-NL2]} — чистка (уплотнение) "Графа". Сжимается файл T5 за счет исключения пустых узлов.

\$E — конец редактирования. Выполняются те же функции, что и при работе УК HE в редакторе "Словаря" (см. 1.2.1).

Пример.

* **CALL ГРАФ**

\$1

УЗЕЛ 0

СВЯЗЬ А

КОН. УЗЕЛ 1 — добавление связей

КОН. УЗЕЛ 2

КОН. УЗЕЛ 3

\$D

УЗЕЛ 0 - исключение связей

СВЯЗЬ А

КОН. УЗЕЛ 2

КОН. УЗЕЛ 3

\$W

УЗЕЛ 0 - печать

КОН. УЗЕЛ 1

\$1 3(15) 2 → 7-8 12 17 — добавление связей

\$1 3 7 — общий узел

\$D 7 2←3 — исключение связи

\$E — конец редактирования

***END FILE**

Процедуры для исследования графа

Описанные выше средства редактирования позволяют формировать и обновлять структуры данных типа семантических сетей. Данный раздел посвящен описанию средств для анализа и исследования полученных структур.

Типичные задачи, связанные с сетями или графами — это сравнение отдельных элементов, нахождение пути, выделение подструктур и ряд других. Описываемый набор программ служит для решения части задач подобного типа и может быть расширен.

Задание пользователя оформляется как последовательность операторов языка типа ФОРТРАН. Для входа в систему нужен предварительный вызов:

CALL IOBEG

Процедуры TSY, TW

Базовыми элементами процедур для исследования являются программы раскрытия узла на один шаг, т.е. программы для нахождения непосредственно следующих или предыдущих вершин графа: TSY и TW.

Обращение к процедуре TSY:

COMMON (CY)LY, NN, LC, NB, N2(500)

CALL TY(NL, N5, IP5, IP1)

CALL TSY(N5, NC, ID, IP)

Вход: NL — линейный номер исходного узла,

NC — линейный номер связи,

ID — направление связи (ID=0 — "выход", ID=I — "вход").

Выход: IP5 — рабочий признак (метка),

IP1 — признак узла: IP1=0 — узел
"внутренний"

IP1=1 — узел "текстовый"

IP — признак результата поиска.

При IP=I будет получено:

LY — количество связей в узле,

LC — количество найденных конечных узлов,

N2 — массив конечных узлов.

Если требуется отобразить все вершины, независимо от имени связи, но в данном направлении, используется процедура TW:

CALL TW(NL, ID, IP5, M), COMMON (CP1) NK(500), NG(500)

Вход: NL — линейный номер исходного узла,

ID — направление связи.

Выход: IP5 — метка узла,

M — количество найденных узлов.

При $M > 0$:

NK — массив конечных узлов,

$NG(I)$ — цена пути от NL до $NK(I)$.

Алгоритмы перебора на графе

Для нахождения пути или дерева вывода (для графа типа И-ИЛИ) применяются следующие процедуры (определение графа И-ИЛИ — см. [3]):

- 1) .PO — для полного перебора по методу волн,
- 2) .PI — для нахождения кратчайшего пути по методу равных цен,
- 3) P2, P6 — для получения дерева-вывода в графе типа И-ИЛИ

Полный перебор.

SUBROUTINE PO(NL, NC, ID, NFIN, JG, IP).

NL — линейный номер начальной вершины,

NFIN — линейный номер вершины-цели,

NC — линейный номер связи,

ID — направление связи,

IP — результат, признак достижения цели.

Метод равных цен.

SUBROUTINE PI(NL, ID, NFIN, JG, IP).

JG — результат, стоимость полученного пути,

NL, ID, NFIN, IP — см. PO.

Процедура позволяет получить путь от начальной вершины до целевой, стоимость, которого минимальна. В качестве характеристики стоимости берётся IP5 связи.

Граф И-ИЛИ.

SUBROUTINE P6(NL, ID, NGO) NGO = max(IP5).

Методы перебора на графе И-ИЛИ общего типа описаны в [3].

Рассматриваемая процедура охватывает частный случай:

- 1) Вершины типа И и ИЛИ чередуются между собой.
- 2) Начальные и конечные узлы поиска — вершины типа ИЛИ.

Примером, удовлетворяющим этим условиям, является граф связи переменных через операторы-процедуры.

В этом случае

- вершины типа И — операторы,
- вершины типа ИЛИ — переменные.

ВОЗМОЖНЫЕ ПРИМЕНЕНИЯ СИСТЕМЫ "ТЕЗАУРУС"

Система "Тезаурус" рассчитана на широкий круг задач, решаемых с использованием представления в виде графа. Следует отметить ряд реализованных задач из перечисленных в таблице I.

Первой из них была задача исследования словаря понятий, подробно рассмотренная в следующем разделе на примере обработки Словаря психологических понятий. Другая задача о поиске трассы движения шагающего робота на местности с малым числом возможных мест для постановки ног сводится к применению алгоритмов перебора, описанных выше в разделе "Процедуры для исследования графа".

Построения нужной последовательности операторов для решения задач прикладной небесной механики [4,5] также решаются методами поиска пути на графе семантики этих операторов. В работе [4] дано описание системы проектирования орбит КА, где приведена семантическая сеть операторов-переменных задачи двух тел для расчета характеристик движения и положения космического аппарата. Граф этой сети принадлежит к рассматриваемому типу графов И-ИЛИ. Это позволяет средствами системы "Тезаурус" определять нужную последовательность обращения к операторам системы проектирования орбит для определения характеристик движения по группе известных данных.

В рамках исследования возможностей системы "Тезаурус" для подобного рода применений была проведена работа по анализу применения системы в качестве базы данных при автоматизированном поиске решения. Речь идет о таком режиме диалога с ЭВМ, когда пользователь, не заботясь о требуемой последовательности операций, может формировать задание, описывая входные данные и цель решения. Задачу нахождения цепочки вывода, которая должна привести от данного входа к требуемому выходу, после этого выполняет программа управления поиском пути на графе.

ИССЛЕДОВАНИЕ ЗАМКНУТОЙ СИСТЕМЫ ПОНЯТИЙ

Рассмотрим пример применения системы "Тезаурус" для построения и исследования замкнутой системы понятий в статьях их толкового словаря. Для анализа был взят Словарь психологических понятий [1]. Ставилась цель: проверить на примере данного Словаря работу редактора и процедур обработки системы "Тезаурус".

Схема тезауруса понятий

В системе понятий исследуемого Словаря определены взаимные ссылки трёх типов:

1. Каждое понятие подведено под другое, более широкое или родовое понятие, определяя тем самым подчинение или субординацию

2. Указываются специфические различия (*Differentia specifica*) для данного понятия.
3. Подчинённые понятия, наиболее тесно связанные с данным, выделены как его координация.

В соответствии с этим был построен граф с тремя видами связей: "Субординация", "*Differentia specifica*", "Координация". По каждой связи может быть несколько ссылок. В качестве примера приведены данные для понятий "Жаргон" и "Желание":

ЖАРГОН:

Субординация → ЯЗЫК

Координация → ГРУППА (ЛЮДЕЙ), ПОНИМАНИЕ, ПОНЯТИЯ
КОРРЕКТНЫЕ И НЕКОРРЕКТНЫЕ, СЛОВО

ЖЕЛАНИЕ:

Субординация → НАПРАВЛЕННОСТЬ ЛИЧНОСТИ

Differentia specifica → ЭМОЦИИ

Координация → ВЛЕЧЕНИЕ, ИНТЕРЕС, ОСОЗНАННОЕ, СТРЕМЛЕНИЕ,
ЦЕЛЬ

Координация ← ДЕМОКРАТИЧЕСКИЙ СТИЛЬ РУКОВОДСТВА,
ЗАКОН ЭФФЕКТА

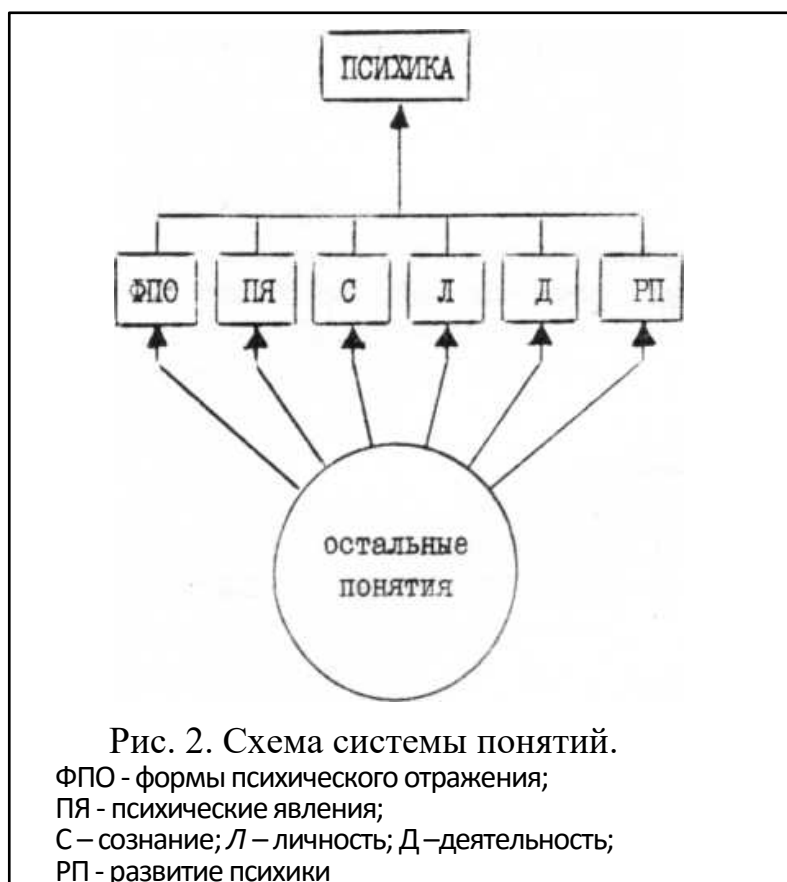
Связи обратного направления — указание слов, в определении которых есть ссылки соответствующего типа на данное слово.

Система понятий построена таким образом, что каждое понятие через связь "Субординация" определяется через более общее понятие, и так — до одной из общепсихологических категорий, имеющей в свою очередь непосредственный выход на понятие "Психика". Таким образом, графом понятий, казалось бы, должен быть граф, незначительно отличающийся от дерева, вершина которого — "Психика", а на следующем уровне имеется 6 категорий Словаря: "Формы психического отражения", "Психические явления", "Сознание", "Личность", "Деятельность", "Развитие психики". Все остальные понятия должны быть подвешены к этим категориям (см. рис. 2).

Анализ показал, однако, что это не совсем так. Наряду с понятиями, непосредственно раскрывающими и конкретизирующими явление "Психика", в словарь вошли:

- Слова языка психологической науки — средство раскрытия предмета психологии: "Дерево психологической науки", "Метод", "Понятие" и им подчиненные.
- Понятия и термины, заимствованные из других наук (философии, физиологии и др.). Такие термины получают в психологической

системе свой специфический аспект и поэтому тоже приводятся в Словаре, однако полная иерархия, показанная на рис. 2, для них не выполняется.



Цели исследования и общая характеристика результата

Поставленная цель исследования Словаря — нахождение и исправление его слабых мест с точки зрения формализма связей понятий и выявление закономерностей полученной системы. Проведенная работа привела к выводу, что Словарь в данном виде не может рассматриваться как формализованная система со строго выдержанным иерархическим подчинением. Для этой цели сам стиль изложения должен быть строго формализован, что усложнило бы чтение Словаря. Однако, оказалось возможным, внося ряд изменений, выделить подмножество Словаря — систему понятий, для которой достаточно точно выдержаны формальные требования:

1. Каждое понятие определено через более общее, стоящее на ступень выше в иерархической структуре.
2. Цепочка последовательных связей с именем "Субординация" выходит на одну из общепсихологических категорий.
3. Выполнены требуемые взаимоотношения разных связей (рис.3 - ниже).

Для получения системы понятий, удовлетворявшей таким требованиям, из неё были исключены некоторые слова смежных наук и ряд терминов пси-

хологической науки, т.е. рассматривалась система в рамках схемы, изображенной на рис. 2.

"Отладка" графа понятий.

Для получения формализованной системы была проведена работа по "отладке" графа связи понятий. С этой целью решались следующие задачи по обнаружению недостатков:

1. Поиск противоречий типа "цикл" в системе связей "Субординация" и "Differentia specifica".
2. Анализ взаимоотношения связей "Субординация" и "Координация".
3. Анализ положения каждого из понятий по отношению к соседним в графе

Поиск противоречий типа "цикл"

"Субординация" и "Differentia specifica" — это ссылки на другие понятия, через которые определяется данное понятие. Если проследить цепочку связей "Субординация", переходя с каждым шагом к более широкому понятию, то в этом процессе должен быть достигнут один из таких узлов:

1. Понятие одной из 6-ти обще-психологических категорий,
2. Понятие психологической науки или её отдела,
3. Понятие из языка смежной науки.

Аналогичная картина должна иметь место для цепочек связи "Differentia specifica".

Естественно, что возвращение в цепочке назад, к одному из пройденных понятий, недопустимо внутри рассматриваемой системы. Это — случай "цикла" при движении в графе связей в одном направлении.

Программа просмотра цепочки связей находит также места достижения одного понятия двумя различными путями (это тоже "цикл", если считать связь ненаправленной). В ряде случаев это не является противоречием, но иногда возникает и как ненужное следствие лишних ссылок (см. рис. 5).

В системе связей "Субординация" Словаря был найден единственный замкнутый цикл, указывающий на необходимость различать слова системы психологических понятий и термины языка психологической науки:

ПОНЯТИЕ → МЫШЛЕНИЕ → ФОРМЫ ПСИХИЧЕСКОГО ОТРАЖЕНИЯ →
КАТЕГОРИИ ОБЩЕПСИХОЛОГИЧЕСКИЕ → ПОНЯТИЕ.

В этом неудачном цикле "Понятие", как один из "Элементов мышления", в системе терминов, описывающих психические явления, находится на одном из нижних уровней иерархии. С другой стороны, "Понятие" в системе слов языка психологической науки иерархически находится наверху этих уровней и имеет в качестве подчиненных "Общепсихологические категории". Фактически здесь должны быть применены два разных слова, например — "Мыслимое понятие" и "Понятие". Подобные различия подразумеваемого в сказанном обычны во

фразах естественного языка, но они должны быть исключены из терминов формальной системы понятий любой науки.

Взаимоотношение связей разных типов

Речь идет об уточнении отношения "Координация" использованием данных по связи "Субординация". "Субординация" — это ссылка на более общее, родовое понятие, которое в иерархической структуре находится на ступень выше данного. "Координация" же строится из подчиненных понятий. Т.е., по смыслу, это две взаимно обратные связи (рис. 3). Однако при создании Словаря не ставилось прямой цели соблюдать такое взаимоотношение связей, и в перечень связей "Координация" кроме подчинённых понятий вошел ряд других понятий, имеющих отношение к определяемому понятию.

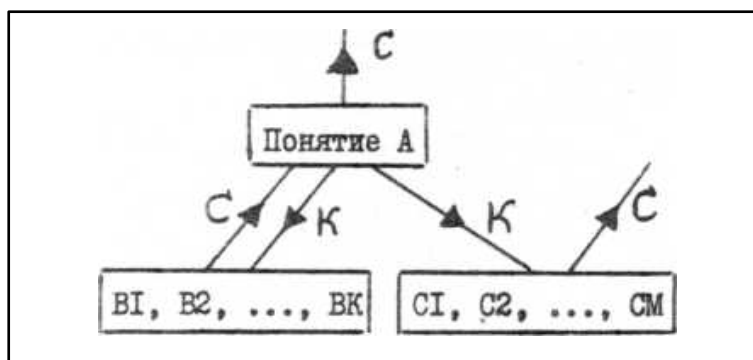


Рис. 3. Взаимоотношение связей "Субординация" (С) и "Координация" (К).

Задача уточнения "Координации" — это нахождение тех узлов, которые ссылаются на данное понятие, как на более общее, но не упоминаются среди его "Координации".

Анализ взаимных связей

Каждое понятие словаря находится во взаимоотношении с другими. Для анализа взаимных связей, определяющих положение конкретного понятия, можно получить цепочки "Субординации" вверх (каждое следующее есть родовое по отношению к предыдущему) вплоть до общепсихологических категорий, а также цепочки подчиненных понятий до слов нижнего уровня, через которые ни какие другие не определяются (по связи "Субординация" в обратном направлении). В результате для каждого понятия возникают два дерева (за исключением простых случаев двух путей между парой понятий) — рис. 4 и 5. Полученные деревья характеризуют место данного понятия в системе и, следовательно, место явления, им обозначаемого, в структуре психических явлений. Наглядность полученной информации позволяет выявлять неточности определений словаря.

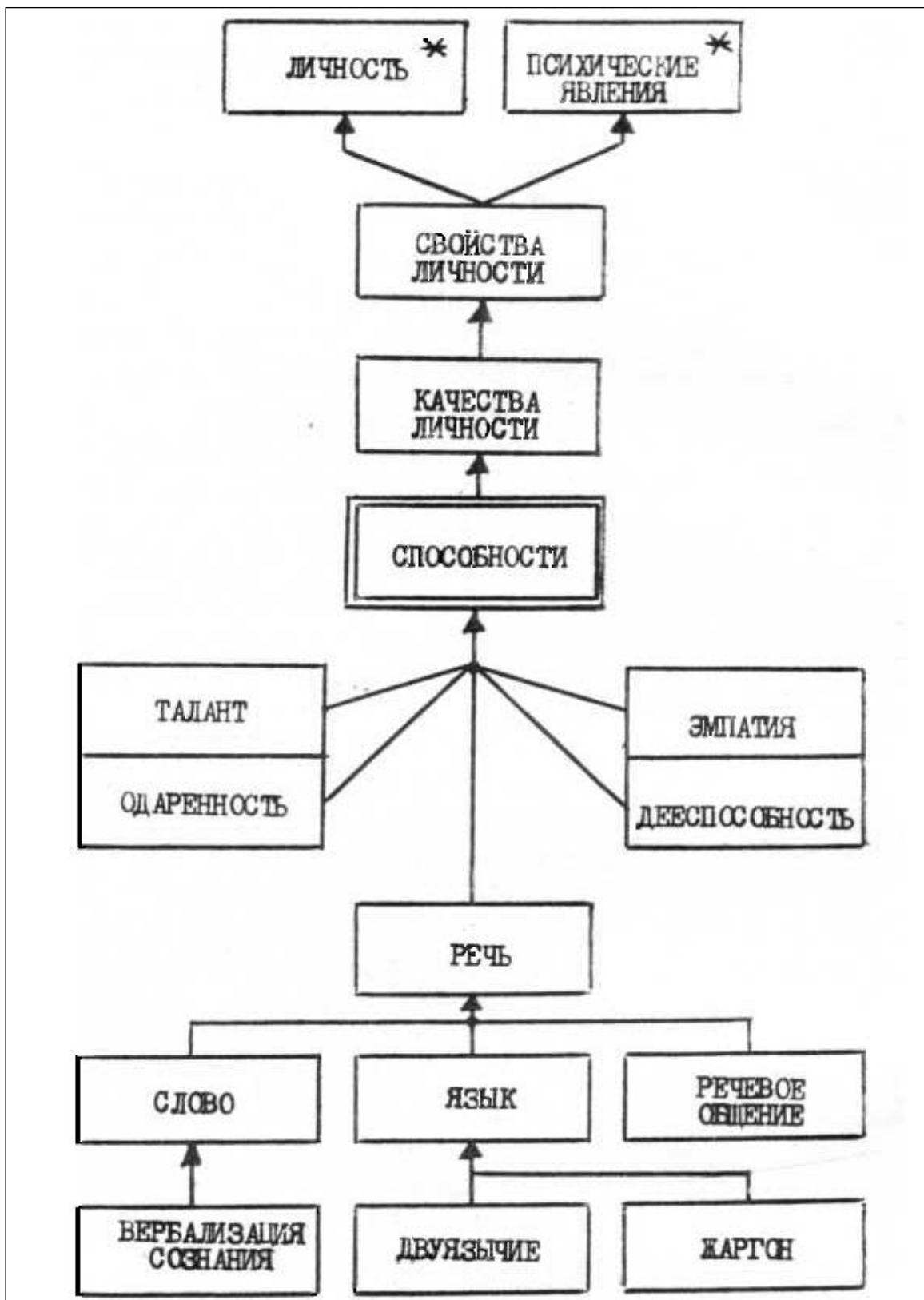


Рис 4 Деревья от вершины "СПОСОБНОСТИ"

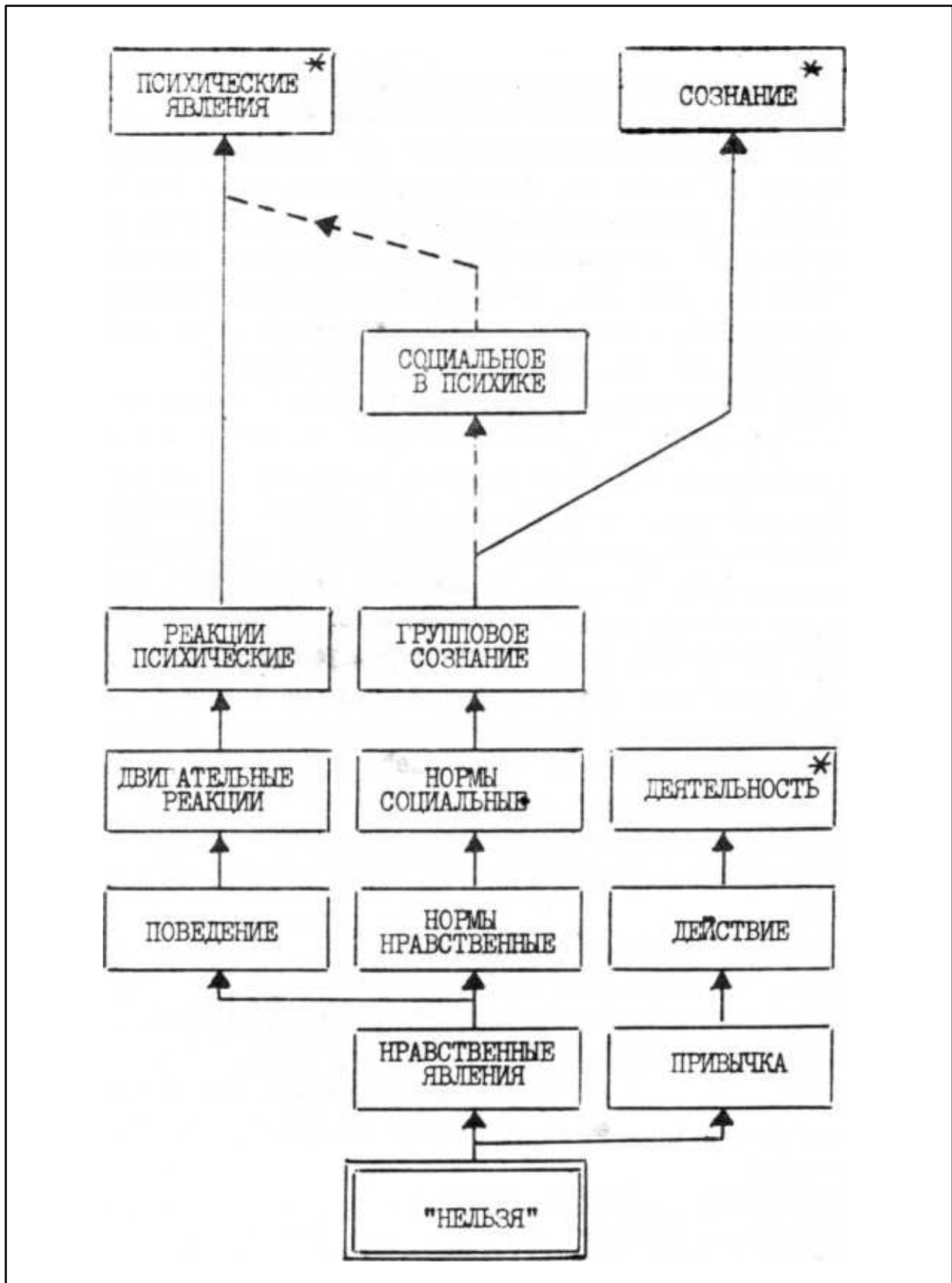


Рис. 5. Пример двух путей между парой понятий.
(Нарушение структуры "дерева")

ХАРАКТЕРИСТИКИ ЗАМКНУТОЙ СИСТЕМЫ ПОНЯТИЙ

В этом разделе приводятся результаты исследования словаря на основе получения ряда статистических характеристик. Полученные данные позволяют делать выводы о степени полноты исследуемой системы понятий и о возможности предсказания новых понятий, которые следует включить в Словарь.

Анализ иерархической структуры

Цель анализа — создание формализованной системы с понятием "Психика" на нулевой и подчиненными ему общепсихологическими категориями на первой ступени иерархии так, чтобы полученная система могла уточняться и пополняться только за счет дополнительных понятий, включаемых в нижние и следующие за ними ступени после просмотренных. Была поставлена задача определения глубины иерархии и получения распределения по ступеням для слов, подчиненных общепсихологическим категориям.

Анализ подтвердил лингвистический вывод о возможном количестве ступеней иерархии: 7 ± 2 (число Миллера – объем внимания человека). Таблица 4 составлена после добавления ряда слов из языка психологической науки, что нарушило нормальное распределение по ступеням.

Таблица 4. Иерархическая структура системы психологических понятий

Категория	Количество ступеней	Количество слов в ступенях 0-8									Сумма
		0	1	2	3	4	5	6	7	8	
ДЕЯТЕЛЬНОСТЬ	4 (3)	1	23	31	24	9					88
ЛИЧНОСТЬ	6	1	22	50	20	38	11	4			146
ПСИХИЧ. ЯВЛЕНИЯ	8 (6)	1	45	90	39	69	28	18	3	1	294
РАЗВИТИЕ ПСИХИКИ	4	1	11	21	12	2					47
СОЗНАНИЕ	5	1	15	10	19	6	2				53
ФОРМЫ ПСИХИЧ. ОТРАЖ.	5	1	9	98	48	17	4				179

Лишним звеном были понятия: "Категории психологии труда", "Категории социальной психологии", "Категории психологии управления". Если это учесть, то получим более точное число ступеней, указанное в скобках. Из таблицы видно, что функция распределения имеет максимум в районе 2-ой ступени. Дополнительные максимумы (4-ая ступень для категорий "Личность" и "Психические явления" или 3-я ступень для категории "Сознание") обусловлены, главным образом, включением слов, заимствованных из других

наук или отраслей психологической науки. Так, например, категории "Психические явления" подчинены ряд следующих слов социальной психологии (см. рис. 6):

- ступень 0: ПСИХИЧЕСКИЕ ЯВЛЕНИЯ ←
- ступень 1: СОЦИАЛЬНОЕ В ПСИХИКЕ ←
- ступень 2: КАТЕГОРИИ СОЦИАЛЬНОЙ ПСИХОЛОГИИ ←
- ступень 3: ГРУППА(ЛЮДЕЙ) ←

Принимая во внимание уточнения в скобках, видим, что максимальное число ступеней (шагов между уровнями) равно 6-ти. Учитывая подчинение:

ПСИХИЧЕСКИЕ ЯВЛЕНИЯ → ПСИХИКА

ЛИЧНОСТЬ → ПСИХИКА,

получим максимум 7-9 ступеней иерархии системы понятий.

Частотный анализ

Частота определяется как количество использований данного понятия среди других определений.

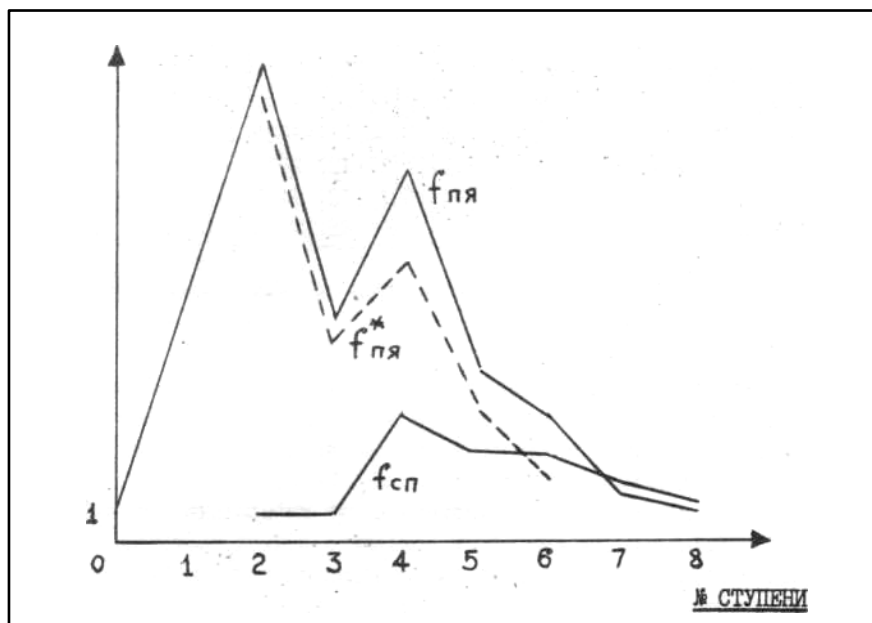


Рис. 6. Распределение слов по ступеням иерархического подчинения (для понятий, подчиненных категории "Психические явления").

$f_{пя}$ — распределение, полученное в эксперименте,

$f_{сп}$ — распределение для понятий социальной психологии,

$f_{пя}$ — распределение без учета понятий, связанных с социальными явлениями.

Частотный анализ позволяет выявить наиболее употребительные понятия и группы понятий. Распределение частот дано в таблице 5. Здесь под частотой подразумевается суммарное количество ссылок на данное понятие по связям

"Субординация" и "Differentia specifica".

Первые две графы дают статистическую информацию о связи ранга (порядкового номера группы понятий с данной частотой) и частоты. Например, по 2 раза зарегистрированы частоты 51, 40, 30, 14 (порядковые номера 2, 3, 7, 17). "Накопленная частота" представляет собой сумму частот группы понятий, в которую входят все понятия от самого частотного (ранг 1) до данного ранга. Так для группы с порядковым номером 7 эта величина равна 395. По этой сумме можно определить, какую долю в определениях составляет данная группа понятий, число которых указано в последней графе.

Таблица 5. Частотный анализ слов понятий.

Ранг	Частота	Число компонентов	Накопленная частота	Накопленное число компонентов
1	52	1	52	1
2	51	2	154	3
3	40	2	234	5
4	35	1	269	6
5	34	1	303	7
6	32	1	335	8
7	30	2	395	10
8	29	1	424	11
9	27	1	451	12
10	26	1	477	13
11	25	1	502	14
12	22	1	524	15
13	19	4	600	19
14	17	1	617	20
15	16	1	633	21
16	15	5	708	26
17	14	2	736	28
18	13	3	775	31
19	12	3	811	34
20	11	6	877	40
21	10	10	977	50
22	9	3	1004	53
23	8	9	1076	62
24	7	9	1139	71
25	6	11	1205	82
26	5	21	1310	103
27	4	29	1426	132
28	3	48	1570	180
29	2	69	1708	249
30	1	199	1907	448

Как видим из таблицы, лишь меньшая часть понятий используется в определениях в качестве ссылок "Субординация" и "Differentia specifica" — 448 из 932, т.е. 48% от общего количества слов словаря. Будем называть эту

часть понятий дескрипторами.

Заметим, что 50 самых частотных терминов (5.4% от общего числа понятий или 11.2% дескрипторов) покрывают более половины дефиниций (977 из 1907, т.е. 51%). Можно также видеть, что 10 самых активных понятий (1.1% от общего числа или 2.3% дескрипторов) покрывают свыше 20% дефиниций (395 из 1907).

Отношение общего числа дефиниций к числу понятий характеризует среднюю частоту или степень "активности" группы понятий — K . Так для дескрипторов:

$$K = 1907/448 \approx 4.3$$

Среднее значение по словарю:

$$K = 1907/932 \approx 2.1$$

Данные распределения по ступеням иерархии (таблица 4, рис.6) позволяют сделать вывод, что наиболее активными являются категории и понятия, непосредственно им подчиненные (наибольший рост функции распределения — участки 0 - 2). Однако не следует буквально это воспринимать, поскольку таблица иерархического распределения не учитывает связи "Differentia specifica", кроме того возможны две и более ссылки по связи "Субординация".

ЗАКЛЮЧЕНИЕ

Опыт применения комплекса "Тезаурус" подтвердил возможность его использования для работы с семантическими сетями большого объёма. Удобный язык редактирования и быстрый доступ пригодны для эксплуатации в диалоговом режиме. Разработанные методы анализа графов могут быть использованы в любой области, где требуется выполнение операций, аналогичных описанным в работе.

Литература

1. Платонов К.К. Краткий словарь системы психологических понятий. М., БШ, 1981.
2. Холл П. Вычислительные структуры. М., "Мир", 1978.
3. Нильсон Н. Искусственный интеллект. М., "Мир", 1973.
4. Платонов А.К., Казакова Р.К. Система проектирования орбит в прикладных задачах небесной механики. Препринт ИПМ АН СССР. 1976. № 106, 39 с.
5. Платонов А.К., Казакова Р.К. Язык для расчета характеристик движения в прикладных задачах небесной механики. Препринт ИПМ АН СССР. 1074. № 78, 49 с.