РОССИЙСКАЯ АКАДЕМИЯ НАУК

ВСЕРОССИЙСКИЙ ИНСТИТУТ НАУЧНОЙ И ТЕХНИЧЕСКОЙ ИНФОРМАЦИИ
(ВИНИТИ)

НАУЧНО • ТЕХНИЧЕСКАЯ ИНФОРМАЦИЯ

Серия 2. ИНФОРМАЦИОННЫЕ ПРОЦЕССЫ И СИСТЕМЫ

ЕЖЕМЕСЯЧНЫЙ НАУЧНО-ТЕХНИЧЕСКИЙ СБОРНИК

Издается с 1961 г. **№ 11** **Москва 2012**

60 ЛЕТ ВСЕРОССИЙСКОМУ ИНСТИТУТУ НАУЧНОЙ И ТЕХНИЧЕСКОЙ ИНФОРМАЦИИ РОССИЙСКОЙ АКАДЕМИИ НАУК

Шестьдесят лет назад, в 1952 г. произошло крупное событие в российской науке: Президиум АН СССР по инициативе и при активном содействии ее президента А.Н. Несмеянова принял решение о создании в системе Академии наук Института научной информации, основной задачей которого было информационное обеспечение «большой» науки.

В настоящее время это Всероссийский институт научной и технической информации Российской академии наук. Когда в западном мире прошел шок от запуска в СССР первого искусственного спутника Земли, американские и европейские специалисты писали, что они были поражены осведомленностью советских ученых о положении дел в мировой науке. По оценкам западной прессы, создание мощнейшего информационного центра по своей значимости превосходит даже первый полет человека в космос. Сегодня ВИНИТИ РАН – один из крупнейших информационных центров в мире – здесь сосредоточено ядро национальных ресурсов научно-технической информации России.

Институт формирует уникальный информационный фонд в области точных, естественных, технических наук, экономики, медицины и охраны окружающей среды. На его основе ученым и специалистам предлагаются различные современные информационные продукты и услуги, в том числе Реферативный журнал в печатной и электронной формах, бюллетени, сборники и другие информационные издания. В Институте создан банк данных, доступный через Интернет, выдаются цифровые и печатные копии первоисточников.

Информационные продукты и услуги Института составляют основу для информационного сопровождения различных направлений исследовательской, образовательной и инновационной деятельности страны. В настоящее время Институт является не только базой информационного обеспечения науки в стране, он координирует эту деятельность в Содружестве независимых государств. Решением Совета глав правительств СНГ Всероссийскому институту научной и технической информации РАН придан статус базовой организации государств - участников СНГ по межгосударственному обмену научно-технической информацией.

Наши пользователи – практически все регионы России, ведущие НИИ, библиотеки, государственные структуры и бизнес, 10 государств - участников СНГ, зарубежные университеты и ведущие библиотеки мира. Помимо выпуска традиционных реферативных и обзорных изданий, научных журналов и баз дан-

ных, распространяемых в России, СНГ и в дальнем зарубежье, Институт на договорной основе сотрудничает с информационными организациями и университетами Болгарии, Великобритании, Германии, США.

Многие научные работники считают, что реферативный журнал, возникший ещё в XVIII в. и получивший нынешний вид в начале прошлого века, безнадежно устарел. В эпоху Интернета, многочисленных баз и банков данных, электронных коммуникаций это дорогостоящее и громоздкое средство оповещения о новых научных достижениях якобы не пользуется спросом, во всяком случае, в его нынешнем виде, и его пора сдать в архив. Заблуждение относительно того, что «в Интернете всё есть» не имеет под собой оснований. Интернет это сеть компьютеров, и она может привести только к сайтам, которые кем-то и чем-то наполнены. Кроме того, в Интернете пользователь в ответ на свой запрос получает огромное число непроверенных и часто ошибочных сведений, относящихся к разному (и не всегда указанному) времени.

Современное информационное обслуживание ученых гораздо больше диверсифицировано по сравнению со временем появления реферативных журналов. Оно включает поиск по многочисленным базам и банкам данных таким, например, как *Web of Science* фирмы *Thomson-Reuter* или *Scopus* издательства *Elsevir*, которые помимо самой информации выдают и метаинформацию в виде модного теперь числа библиографических ссылок, неправомерно называемого по-русски цитированием. В Институте имеются разнообразные средства навигации не только по текстовым, но и по цифровым базам данных, интеллектуальные информационные системы, в частности такие, которые позволяют автоматически генерировать гипотезы о причинно-следственных зависимостях, и многое другое. Но это не умаляет значения Реферативного журнала, который имеет собственные, никакими другими средствами не повторяемые свойства.

До сих пор три четверти всей новой значимой информации учёный получает из журнальных статей. По закону рассеяния Бредфорда из журналов, которые в каждом номере публикуют большинство статей по определенной теме, можно получить лишь треть статей по этой теме. Остальные рассеяны по смежным и многоотраслевым журналам. Реферативные издания в печатном и электронном виде собирают эти рассеянные статьи по узким темам-рубрикам по интересам конкретных пользователей. Они избавляют учёного, который хочет сам искать нужную информацию, от потери времени на поиск в Интернете и библиотеках, поскольку их генераторы берут на себя ответственность за отбор релевантных статей из фиксированного списка журналов. В мире выходит около 300 тыс. научных журналов, из которых не менее четверти содержат информацию по естественным наукам и технике.

Реферативные журналы и базы данных публикуют рефераты статей на языке пользователя. Разумеется, большинство высокоцитируемых в мире журналов выходит на английском языке, а приличный учёный должен им владеть. Но думает он на своём родном языке, на котором мир особым образом поделён на понятия для всех думающих на этом языке. Если мы хотим иметь российскую науку, то в ней должна быть своя терминология и своё видение любой актуальной проблемы. Это и дают рефераты на русском языке и классификаторы - рубрикаторы Реферативного журнала, составляемые квалифицированными отечественными специалистами.

В рамках программы Президиума РАН «Прогноз потенциала индустриализации России» в Институте уделяется внимание подготовке аналитических обзоров по отдельным областям знания и направлениям развития науки, техники и технологий. При подготовке обзоров используются банк данных ВИНИТИ (более 30 млн документов), фонд научно-технической литературы, насчитывающий около 2,5 млн единиц хранения, полнотекстовые зарубежные электронные ресурсы (более 12 тыс. зарубежных журналов через Интернет).

Одним из важнейших направлений деятельности ВИНИТИ РАН является навигация по мировым информационным ресурсам. Для этого в Институте разработаны специальные методические и программно-технологические средства, позволяющие проводить поиск по разнородным информационным ресурсам.

Эта деятельность возможна благодаря проводимым на протяжении всей истории Института научным исследованиям. Под руководством его директора профессора А.И. Михайлова (1906-1987) в нашей стране была создана научная дисциплина *информатика*; нынешний его директор академик РАН Ю.М. Арский является выдающимся представителем *геоинформатики.* Президиум РАН ежегодно отмечает достижения в области создания интеллектуальных информационных систем, развития ДСМ-метода автоматического порождения гипотез на основе исследований, ведущихся научным коллективом под руководством профессора В.К. Финна. Результаты этих исследований защищаются в докторских и кандидатских диссертациях, которые

подготавливаются в аспирантуре и на базовых кафедрах Института совместно с ведущими университетами страны. Инновационный фонд «Сколково» проявил интерес к разработкам Института, касающимся извлечения знаний из больших массивов информации, которые базируются на результатах фундаментальных исследований в области искусственного интеллекта и интеллектуальных систем.

Сегодня Институт по-прежнему является главным научно-методическим центром страны в области научно-технической информации, осуществляя координацию работ по созданию и развитию общесистемной нормативно-методической базы. В настоящее время в нем ведется и развивается эталонный массив Государственного рубрикатора научно-технической информации, который издается и распространяется в печатной и электронной формах. По этому Рубрикатору индексируются все выполняемые в стране научно-исследовательские и опытно-конструкторские работы, а также базы данных Научной электронной библиотеки (*eLibrary*).

ВИНИТИ РАН является членом международного Консорциума Универсальной десятичной классификации, ему переданы авторские права на полное издание таблиц УДК на русском языке. На данный момент подготовлено 12 томов и 5 выпусков изменений и дополнений, которыми обеспечиваются библиотеки и информационные центры страны.

На базе Института работает Технический комитет по стандартизации «НТИ, библиотечное, издательское и архивное дело». ВИНИТИ РАН непосредственно участвует в разработке стандартов, осуществляет координацию соответствующей деятельности, экспертизу стандартов, рубрикаторов многих организаций, проводит конференции, семинары, а также оказывает методическую помощь издательствам и научным учреждениям по включению издаваемых ими журналов в базы данных *Web of Science* и *Scopus* для увеличения показателей цитируемости российских ученых и повышения импакт-факторов изданий. С этой целью для издающих организаций Институт проводит семинары, на которых подробно обсуждаются проблемы подготовки журналов для названных указателей цитирования. Активное участие в этих семинарах принимают разработчики Российского индекса научного цитирования (РИНЦ).

Президиум РАН два года назад обязал академические институты предоставлять обширную статистику, включая библиометрические показатели: количество публикаций сотрудников института в различных БД, цитируемость и средневзвешенный импакт-фактор (ИФ) организации. На наших глазах библиометрия превратилась в самостоятельную отрасль информационной деятельности. Многие зарубежные финансирующие организации, поддерживающие фундаментальные исследования, усилили внимание к библиометрии как источнику качества научной продукции. В соответствии с этим в Институте проводятся библиометрические исследования по сопоставительному анализу библиометрических индикаторов отечественной и мировой науки. Особое внимание уделяется тенденциям научной продуктивности и цитируемости ученых РАН и сопоставлению этих показателей с зарубежными.

В Институте ведется разработка методики библиометрической оценки эффективности научной деятельности с использованием информационных ресурсов компании *Thomson-Reuters*, информационной системы *Web of Science* и аналитических БД: Указатель цитируемости научных журналов (*Journal Citation Reports*) и Основные показатели науки (*Essential Science Indicators*). В настоящее время проводится исследование по оценке влияния конкурсного финансирования на публикационную активность и показатели цитируемости отечественных исследователей и по оценке эффективности библиометрической активности университетов.

Подготовка научно-информационных продуктов и проводимая в Институте научно-исследовательская работа отражаются в публикациях сотрудников, в которых описываются новые методы информационного анализа, что позволяет ученым выйти на наиболее перспективные направления научного поиска, способствует заметному сокращению разрыва в уровне информационного обеспечения между отечественными и зарубежными исследовательскими центрами.

Институт тесно сотрудничает с Библиотечным советом РАН, ведущими библиотеками страны, такими как Библиотека Российской академии наук (БАН), Библиотека РАН по естественным наукам (БЕН), Государственная научно-техническая библиотека (ГПНТБ) и другими.

В работе ВИНИТИ РАН есть серьезные трудности. Это недостаточный уровень финансирования, ограничивающий полноту входного потока литературы, определяющий низкий уровень оплаты труда сотрудников, невозможность привлечения к работе молодых специалистов. Эти трудности присущи всем академическим институтам, да и всей бюджетной сфере. Мы стараемся находить возможные пути их преодоления.

В Институте создан научный образовательный центр, в котором ведется подготовка студентов различных университетов по проблемам современных информационных технологий. Среди двух десятков этих университетов такие крупные, как Московский государственный университет им. М.В. Ломоносова, Российский государственный гуманитарный университет, имеющий в ВИНИТИ базовую кафедру и лабораторию, Московский университет инженеров транспорта, Московский автодорожный университет и многие другие. Студенты этих вузов становятся нашими аспирантами и составляют возможный резерв институтских кадров.

Другой путь преодоления трудностей лежит в модернизации технологии подготовки информационных продуктов и услуг, в создании системы подготовки аналитических и прогнозных материалов по ключевым проблемам науки и технологии. Он предполагает отказ от малотиражных изданий на бумажных носителях, перевод этих и других изданий на электронные носители, создание локальных специализированных баз данных, таких как «Химические формулы и соединения», «Экономия энергии». Предусматривается перевод части фонда научно-технической литературы в машиночитаемый вид, создание электронной библиотеки депонированных научных работ (свыше 200 тыс. ед.).

Вступая в седьмое десятилетие, Институт считает важной свою роль в информатизации российского общества. Для этого предполагается:

– расширение доступа российским ученым и специалистам к мировой научно-технической литературе;

– совершенствование информационного обеспечения отечественной науки;

– развитие современной информационной инфраструктуры для перехода к информационному обществу;

– обеспечение организационных, методических, правовых условий для создания единого информационного пространства России и других стран СНГ;

– содействие в подготовке обзоров и прогнозов научно-технического развития России на базе накопленного массива информации и интеллектуальных систем ее поиска и обработки.

У ВИНИТИ РАН есть потенциал для ответа на вызовы современности.

От редакции

ОБЩИЙ РАЗДЕЛ

УДК 002.63ВИНИТИ

Памяти Мириам Львовны Аврух

В. Б. Борщев

Лаборатория электромоделирования ВИНИТИ: 1958-1959 гг.*

В МОСКВУ...

Летом 1958 г. я приехал в Москву делать диплом. Не один. Часть нашей группы, человек 10-12, послали в ЛЭМ - Лабораторию электромоделирования…

Я учился тогда на Радиотехническом факультете Казанского авиационного института (КАИ). Когда мы были на пятом курсе, в Казани решили строить завод вычислительных машин[1]. И на нашем факультете выделили группу «для подготовки кадров» для этого завода. По сути дела, ввели новую специальность, хотя в чем-то и близкую тому, чему нас учили раньше. Но на факультете квалифицированных преподавателей, чтобы учить нас, в общем-то не было. Нам читали какие-то спецкурсы, не помню уже какие, видимо что-то было скроено на скорую руку. На какие-то занятия нас посылали на создаваемый завод. Завод только начинали строить. А пока там стоял небольшой барак и в нем ютилось несколько инженеров. Один из них нам рассказывал, как устроены вычислительные машины.

Летом 1958 г. нам объявили, что нас посылают в Москву, в Академию наук СССР, в Лабораторию электромоделирования на преддипломную практику и диплом. Про Лабораторию мы ничего не слышали, а слова *Академия наук* производили впечатление. Провинциальное воображение рисовало «здание полное света и воздуха» (расхожий газетный штамп того времени) и мраморные лестницы, покрытые коврами... Действительность, как водится, оказалась несколько иной.

Лаборатория электромоделирования

ЛЭМ была частью Всесоюзного института научной и технической информации – ВИНИТИ, но какой-то автономной частью – с собственным бюджетом, бухгалтерией, отделом кадров и другими признаками автономного существования. Это было, видимо, связано с ее историей. ЛЭМ вошла в состав ВИНИТИ сравнительно недавно, в 1957 г., а до этого она была частью ИТМ и ВТ[2], а еще раньше – частью Энергетического института АН СССР.

Располагалась Лаборатория далеко от ВИНИТИ, во 2-м Бабьегородском переулке, между Москва-рекой и Якиманкой. От этого переулка остался сейчас только маленький кусочек, большая его часть снесена, там теперь Парк искусств, рядом с ЦДХ (Центральным Домом художников). Помещалась Лаборатория в бараке, в котором когда-то держали пленных немцев[3]. И уж, конечно, никакого мрамора и никаких ковров. Ходила легенда, что когда в этом бараке захотели по каким-то причинам сменить трубы, приехавшие водопроводчики сказали, что этого делать нельзя: если удалить старые трубы, барак может рухнуть...

Но внутри барака было очень интересно! Интересными были люди, которые там обитали. Нас, практикантов из Казани, распределили по отделам или каким-то другим подразделениям. Мы с Феликсом Рохлиным, моим другом еще со школьных лет, попали под крыло Мириам Львовны Аврух – это была очень симпатичная молодая женщина, всего лет на 10 старше нас, хотя она была одним из ветеранов Лаборатории и помнила историю ее «кочевья» по разным институтам Академии наук.

Работали у Мириам Львовны, в основном, молодые люди: Натан Бирман, Лева Магидсон и Галя Са-

---

* Предварительный сокращенный вариант этого материала опубликован в сборнике «Петербургская библиотечная школа». -2012. - № 2.

[1] Слова *компьютер* в русском языке тогда еще не было, говорили о *вычислительных машинах,* в ходу была аббревиатура ЭВМ – электронная вычислительная машина.

[2] Институт точной механики и вычислительной техники АН СССР, там была создана БЭСМ, одна из первых в СССР вычислительных машин.

[3] Я описываю картинку, которая сохранилась в моей памяти. Натан Бирман [1] пишет об этом точнее и подробнее: «Лаборатория до 1960 года размещалась на территории бывшего концлагеря для немецких военнопленных, который остался бесхозным после отъезда его обитателей на родину, в ГДР. Расположен был этот концлагерь прямо напротив входа в ЦПКиО им. Горького, там, где теперь стоит выставочный комплекс художественных музеев. А в те годы здесь были три полуразрушенных барака, обнесенных забором с колючей проволокой, вышки без пулеметчиков и прочие атрибуты концлагерного прошлого».

енко, недавние выпускники Института связи. Частыми гостями были молодые математики, закончившие МГУ и работавшие в соседнем отделе: Наташа Рикко, Гриша Клейнерман, Лева Нисневич и Гена Стецюра (он, впрочем, был, кажется, физиком). Заходили логики и лингвисты, тоже недавно закончившие МГУ: Виктор Финн и Делир Лахути, Лена Падучева и Мара Ланглебен.

Лаборатория тогда расширялась, на работу брали и привлекали к сотрудничеству не только выпускников МГУ и других вузов, но и уже известных людей самых разных специальностей. Там работали математик, ученик А.Н. Колмогорова, В.А. Успенский, знаменитый уже тогда лингвист Вяч.Вс. Иванов, химик Г.Э. Влэдуц. Как-то был связан с Лабораторией и химик Д.А. Бочвар, уникальный человек, «по совместительству», один из самых известных советских логиков.

Чуть позже в Лаборатории появились еще два логика – А.С. Есенин-Вольпин и И.Х. Шмаин. Есенин-Вольпин тогда только что перевел знаменитую книгу «Введение в метаматематику» знаменитого американского логика С. Клини. Шмаин заканчивал мехмат МГУ и буквально не расставался с этой книгой Клини. Оба они были уже тогда известны не только и не столько своей профессиональной деятельностью.[4]

Лев Израилевич Гутенмахер

Создал эту Лабораторию и заведовал ею Лев Израилевич Гутенмахер, очень колоритная фигура. О нем подробнее. Приведу с некоторыми сокращениями статью из Википедии (http://ru.wikipedia.org/wiki/Гутенмахер,_Лев_Израилевич)[5].

Лев Изра́илевич Гутенма́хер (1908, Тарутино Аккерманского уезда Бессарабской губернии — 1981, Одесса) — советский математик, специалист в области электрического моделирования, один из пионеров развития электронно-вычислительной (компьютерной) технологии в СССР. Доктор технических наук (1940), профессор (1943). Лауреат Сталинской (1946) и Государственной (1962) премий.

Биография

Лев Израилевич Гутенмахер родился в бессарабской колонии Тарутино (ныне райцентр Тарутинского района Одесской области Украины) в 1908 году. Окончил Донской политехнический институт в 1931 году и был оставлен в аспирантуре там же. В 1934 году защитил кандидатскую диссертацию, продолжил преподавать в институте (теперь Новочеркасском политехническом институте) до 1938 года.

С 1938 года — в Энергетическом институте Академии наук СССР, где в 1939 году основал и возглавил лабораторию электромоделирования (ЛЭМ). Одновременно преподавал в Московском инженерно-физическом институте (с 1943 года — профессор). С 1948 по 1956 год лаборатория электромоделирования Гутенмахера входила в Институт точной механики и вычислительной техники АН СССР, а с 21 июня 1957 года вошла в состав Всесоюзного института научной и технической информации АН СССР и с 1962 года — в состав Всесоюзного научно-исследовательского института природных газов (ныне Всероссийский нефтегазовый институт им. академика А. П. Крылова). Лабораторию составляли 8 отделов, в том числе логико-математический, аналогового электрического моделирования, математической лингвистики, математических методов в химии и другие. В последние годы жизни был профессором кафедры автоматизации вычислительных процессов электротехнического факультета Одесского политехнического института и жил в Одессе на углу Дерибасовской и Ришельевской улиц.

Основные научные результаты — в области приближённых и численных методов вычислительной техники. Л.И. Гутенмахер с самого начала научной карьеры в 1930-х годах занимался использованием электрических сетей для моделирования сложных информационных систем и для решения уравнений в частных производных. В 1950 году им был представлен проект безламповой электронной вычислительной машины (ЭВМ) с использованием электромагнитных бесконтактных реле на феррит-диодных элементах, разработанных в его лаборатории на основе магнитных усилителей трансформаторного типа; сама машина ЛЭМ-1 была создана к 1954 году. В 1956 году он опубликовал научное сообщение об успешном использовании матрикса накопителей для хранения информации.

Л. И. Гутенмахеру принадлежит ряд теоретических работ в области кибернетики и одна из первых монографий в этой области в СССР (*Системы извлечения информации*), уже в 1950-е годы он занимался компьютерным моделированием когнитивных процессов (таких как логическое мышление, чтение) и математической лингвистикой. В публикациях этих лет им были детально рассмотрены различные аспекты хранения и извлечения информации, программное обеспечение, компьютерная техника, связь по телефонным каналам, включая электронный доступ к библиотечным фондам. Среди монографий учёного: «Электрическое моделирование: электроинтегратор» (1943), «Электрические модели и их применение в технике и физике» (1946), «Электрические модели» (1949), «Электронные информационно-логические машины» (1960, 1962), «Ассоциативные запоминающие устройства» (1967) и «Импульсное электромоделирование» (1983), часть из которых переведена на английский, немецкий, французский и испанский языки.

Эта статья в Википедии вызвала у меня смешанную реакцию. Я надеюсь, что там правильно перечислены основные факты жизни Л.И. Гутенмахера. В основном они не противоречат тому, что я помню про этот период. С оценочными и просто общими суждениями дело обстоит сложнее. Но об этом позже.

[4] А.С. Есенин-Вольпин – сын С. Есенина, логик, поэт, философ и правозащитник. Он уже успел к тому времени посидеть в тюрьме и психушке. Илья Шмаин был арестован на втором курсе мехмата за участие в кружке (кампании друзей), где обсуждались самые разные вещи. В 1954 г. он был реабилитирован и восстановился на мехмате (см. о А.С. Есенине-Вольпине текст в Википедии (http://ru.wikipedia.org/wiki/Есенин-Вольпин_Александр_Сергеевич), а об И.Х. Шмаине тексты на сайте «Библиотека Якова Кротова» (http://krotov.info/spravki/persons/20person/shmain.htm), в частности, мои тексты о нем, которые частично были опубликованы в сборнике НТИ. Сер. 2. – 2007. - № 12. – С. 1-9).

[5] Пометка на версии статьи: «Это версия страницы, ожидающая проверки. Последняя подтверждённая версия датируется *5 ноября 2010*».

А пока некоторые фактические уточнения и дополнения. Сталинская премия была присуждена Гутенмахеру в 1948 г.[6] (а не в 1946 г., как сказано в цитированной выше статье). С Государственной премией какая-то нестыковка. Эти премии присуждались в СССР, начиная с 1967 г. Поэтому Гутенмахер не мог получить эту премию в 1962 г. В списке лауреатов этих премий по науке и технике, приведенном в той же Википедии, фамилия Гутенмахера не значится.

ЛЭМ И ПЕРВЫЕ СОВЕТСКИЕ ЭВМ[7]

Нас (студентов Казанского авиационного института) послали в ЛЭМ, так как здесь занимались вычислительными машинами и задачами, для которых эти машины создаются. Работы, проводимые в ЛЭМ в конце 40-х и в 50-е гг., естественно рассматривать в историческом контексте. Тогда в СССР проектировали и создавали первые вычислительные машины. История их создания не лишена интриги. ЛЭМ при этом играла скорее маргинальную роль. Но и эта роль по-своему интересна.

Исходным толчком послужили сообщения о создании в США электронной вычислительной машины ENIAC (Electronic Numerical Integrator And Computer), которая начала работать в 1946 г. ENIAC был монстром: 17 с половиной тысяч электронных ламп, больше 7 тысяч кристаллических диодов, около 5 миллионов контактных паек. Машина весила около 30 тонн и занимала площадь 63 кв. метра.

Как пишет в своей книге Б.Н. Малиновский [2], публикации, доходившие до СССР, были скудными, но они, видимо, взбудоражили если не начальство, то некоторых проницательных специалистов. Первый детальный советский проект вычислительной машины был разработан в 1948 г. в Энергетическом институте АН И.С.Бруком и Б.И.Рамеевым. Проект не был реализован – это требовало больших организационных усилий, которые тогда явно превышали возможности авторов. В том же году Брук и Рамеев получили авторское свидетельство на изобретение "Автоматическая цифровая электронная машина" – первое изобретение в СССР, зарегистрированное в области цифровой электронной вычислительной техники.

В конце 40-х гг. в Киеве С.А. Лебедевым была начата разработка МЭСМ (малой электронносчетной машины), которая начала работать в конце 1951 г. Примерно в то же время Бруком и его новыми сотрудниками (Н.Я.Матюхиным, М.А.Карцевым и др.) была разработана машина М-1 с хранимой программой (малая ЭВМ – в ней было раз в десять меньше ламп, чем в МЭСМ, но работала она значительно быстрее). Машина начала работать в начале 1952 г.

В 1948 г. зашевелилось и начальство. Были созданы две организации, задачей которых была разработка вычислительной техники – Институт точной механики и вычислительной техники (ИТМ и ВТ) в АН СССР и СКБ-245 в Министерстве машиностроения и приборостроения СССР. В СКБ-245 впоследствии была разработана машина «Стрела», а в ИТМ и ВТ – машина БЭСМ. И вот тут в этой истории впервые возникает ЛЭМ – она вошла в состав ИТМ и ВТ.

ЛЭМ в Институте точной механики и вычислительной техники

О том, как ЛЭМ попала в ИТМ и ВТ, можно прочитать в воспоминаниях П.П. Головистикова [3]. Он пишет, что в 1948 г. ИТМ и ВТ был

«образован на базе трех институтов АН СССР: Института машиноведения, Энергетического и Математического институтов. Из Института машиноведения выделен Отдел точной механики во главе с академиком Н.Г. Бруевичем. Из Энергетического института Лаборатория электромоделирования во главе с профессором Л.И. Гутенмахером. Из сотрудников Математического института им. В.А. Стеклова образованы Отдел приближенных вычислений с начальником отдела чл.-корр. Л.А. Люстерником и Экспериментально-счетная лаборатория во главе с И.Я. Акушским». Директором института был назначен академик Бруевич.

«Институт точной механики и вычислительной техники был создан для разработки новых средств вычислительной техники. Его название (сохранившееся до сих пор) отражало состояние вычислительной техники того времени. Тогда весьма распространенными были еще механические вычислительные системы: например, прецизионные дисковые дифференциальные анализаторы, которыми занимался Отдел точной механики...

Из Энергетического института с Лабораторией электромоделирования в ИТМ и ВТ было переведено 19 человек. В их числе: проф. Л.И. Гутенмахер (зав. лабораторией), ст. инженер В.Ф. Артюхов (зам. зав. лабораторией), мл. научн. сотр. Н.В. Корольков, В.В. Бардиж, Г.К. Кузьминок, М.Л. Аврух, И.А. Виссонова, ст. инженер Н.П. Зубрилин и др. Позднее в лабораторию была принята И.В. Логинова (Виппер).

После перехода в ИТМ и ВТ сотрудники Лаборатории продолжали заниматься электрическими моделями, предназначенными для решения уравнений с частными производными (уравнений Лапласа, Пуассона), а также созданием ламповых интеграторов для решения линейных дифференциальных уравнений, внедрением этих устройств в производство…».

В общем, ИТМ и ВТ организовали по принципу «я тебя слепила из того, что было»: отдел Бруевича занимался механическими вычислительными системами и моделями, ЛЭМ – электрическим моделированием, чем-то вроде аналоговых вычислительных систем, а «вычислительные математики» из института Стеклова – организацией расчетов на тогдашних калькуляторах типа «рейнметалл» (что-то вроде электрических арифмометров). По сути это было ме-

[6] Это была премия третьей степени за выдающиеся изобретения и коренные усовершенствования методов производственной работы. Премия была присуждена «за создание нового счётного аппарата - электроинтегратора». Вместе с Гутенмахером этой премии были удостоены Н.В. Корольков, Б.А. Волынский и В.П. Лебедев.

[7] Этот раздел является, в основном, компиляцией из материалов, опубликованных в Интернете: Википедии и воспоминаний участников событий. Я их цитирую или пересказываю. В первую очередь, хочу отметить очень интересную книгу Б.Н. Малиновского [2]. Остальные источники указываются по ходу изложения.

ханическим объединением разнородных организмов, страдающим от «тканевой несовместимости». А современными вычислительными машинами занимались тогда немногочисленные энтузиасты в других местах – в Киеве у Лебедева и в Москве у Брука.

Но вскоре в ИТМ и ВТ сменилось начальство и произошла реорганизация.

Малиновский пишет в своей книге:

«Через год после образования института его работу проверяла комиссия Президиума АН СССР под председательством В.М. Келдыша. Весьма возможно, что причиной этого явилось письмо Лаврентьева Сталину. Комиссия пришла к неутешительному выводу: цифровой электронной вычислительной технике, быстро развивающейся на Западе, уделяется очень мало внимания...

Был подготовлен проект постановления правительства о ...разработке цифровой электронной вычислительной машины... Однако при рассмотрении подготовленного проекта постановления правительства случилось непредвиденное. Присутствующий Л.И. Гутенмахер, руководитель одной из лабораторий ИТМ и ВТ АН СССР, выступил с предложением выполнить машину не на электронных лампах, а на разработанных в его лаборатории безламповых элементах - электромагнитных бесконтактных реле (на основе магнитных усилителей трансформаторного типа). Его предложение вызвало живой интерес у министра П.И. Паршина... Результатом совещания стал проект постановления правительства о создании двух вычислительных машин - электронной в Академии наук СССР и на элементах Гутенмахера - в министерстве.

... Гутенмахер, ободренный поддержкой министра, упорно работал. В начале 1950 г. он представил в СКБ-245 эскизный проект вычислительной машины на феррит-диодных элементах... К этому времени ситуация в министерстве, на его беду, резко изменилась. В СКБ-245 появился Б.И. Рамеев, разработавший еще в 1948 г. в соавторстве с И.С. Бруком проект цифровой ЭВМ с программным управлением (это был первый в нашей стране проект электронной ЭВМ!).

Рамеев сразу подключился к работам. И очень быстро подготовил аванпроект ЭВМ на электронных лампах. Далее события развивались весьма своеобразно. Технический совет СКБ-245 в отсутствии Рамеева рассмотрел проект Гутенмахера. Затем заслушали Рамеева (при отсутствии Гутенмахера). Итогом стало решение - создавать ЭВМ на электронных лампах, а не на элементах Гутенмахера. У БЭСМ появилась серьезная соперница - ЭВМ "Стрела".

Остается добавить, чем завершилась работа по феррит-диодной ЭВМ. Л.И. Гутенмахер, лишившись поддержки СКБ-245, продолжал работу собственными силами. В его лаборатории в ИТМ и ВТ АН СССР была спроектирована и создавалась параллельно БЭСМ вычислительная машина на феррит-диодных элементах. Позднее, где-то году в 1954-м, мне удалось ознакомиться с ней, когда она уже работала. Ее производительность была невысокой. Вследствие низкого качества элементов надежность работы также оставляла желать лучшего. Импульсный источник питания был громоздок и неэкономичен. Под предлогом секретности вход в лабораторию был практически запрещен. В начале 60-х годов она была закрыта. Строгая секретность, которую вносил Гутенмахер в свои исследования, привела к тому, что о его машине мало кто знает. Тем не менее - это определенная веха в истории вычислительной техники».

Машина на феррит-диодных элементах, о которой пишет Малиновский – это ЛЭМ-1, упоминаемая в цитированной выше статье из Википедии.

В 1950 г. директором ИТМ и ВТ был назначен академик М.А. Лаврентьев, в институте была создана лаборатория № 1, заведующим которой был назначен С.А. Лебедев, который, как пишет в уже цитированном тексте П.П. Головистиков, «был ответственным за предъявление Государственной комиссии в I кв. 1951 г. эскизного проекта Большой электронной счетной машины (БЭСМ)».

В воспоминаниях Головистикова ничего не говорится о взаимодействии ЛЭМ с основными подразделениями ИТМ и ВТ, где началась разработка БЭСМ. По-видимому Лаборатория была достаточно автономна. Она даже не переехала в новое здание, построенное для ИТМ и ВТ на Ленинском проспекте. Но какое-то взаимодействие было. В частности, при создании ферритовой оперативной памяти вычислительных машин, так называемого МОЗУ (магнитное оперативное запоминающее устройство)[8]. Такая память пришла на смену предыдущим системам – акустической памяти (ртутные линии задержки) и памяти на электронно-лучевых трубках (потенциометрах).

Первое в СССР МОЗУ было разработано в ИТМ и ВТ для машины БЭСМ. И идея той модификации МОЗУ, которая использовалась в БЭСМ и некоторых других советских машинах, принадлежала Гутенмахеру, хотя отлаживалось это МОЗУ в лаборатории № 1, см. выдержки из воспоминаний В.В.Бардижа[9] [4]:

«В Институте точной механики и вычислительной техники им. С.А.Лебедева АН СССР в течение многих лет велись успешные работы по магнитным элементам для электронных вычислительных машин.

Основными из этих элементов были кольцевые ферритовые сердечники с прямоугольной петлей гистерезиса (ППГ). Такие сердечники являлись основой для создания оперативных запоминающих устройств (ОЗУ) электронных вычислительных машин, создаваемых в Советском Союзе с середины 50-х до конца 70-х годов.

Положительными качествами ферритовых сердечников с ППГ являются: возможность хранения информации без затраты энергии, длительный, практически неограниченный срок службы, простота технологического процесса изготовления, радиационная стойкость...

Первые выпуски ферритовых сердечников с ППГ в СССР были начаты в 1954 г. в Институте ТМ и ВТ АН СССР в Лаборатории электромоделирования под руководством Л.И.Гутенмахера. При этом использовалась технология, разработанная А.А.Косаревым на основе принципов порошковой металлургии.

[8] Ферритовый сердечник – это небольшое колечко диаметром в несколько миллиметров. При намагничивании он может быть в двух состояниях, представляющих то или иное двоичное значение (1 или 0). Это значение может быть «считано» с сердечника и любое значение может быть туда записано. Т.е. сердечник может служить двоичным элементом оперативной памяти.

[9] Замечу, что Бардиж перешел в 1950 г. из ЛЭМ в лабораторию № 1, руководимую Лебедевым.

...Проверка осуществлялась тогда ручным способом, воздействием на каждый сердечник серии импульсов тока, имитирующих работу сердечника в ОЗУ.

На основе ферритовых сердечников с ППГ в Институте было создано магнитное ОЗУ (МОЗУ), которое в 1956 г. вошло в состав электронной вычислительной машины БЭСМ АН СССР.

Созданное МОЗУ было первым в Советском Союзе. В нем использовалась система 2Д с двумя сердечниками на бит, предложенная Л.И. Гутенмахером и доведенная до практического использования в основном А.С. Федоровым и М.Л. Сычевой. В отладке МОЗУ активное участие принимал М.Д. Великовский»[10].

ЛЭМ в ВИНИТИ

В 1957 г. ЛЭМ была включена в состав ВИНИТИ. А в 1958 г., когда я туда попал, она процветала и бурно расширялась. Чем там только не занимались: читающими устройствами и ассоциативной памятью, лингвистикой и информационными задачами в химии, а также многими другими проблемами. Все это как-то должно было служить провозглашенной цели – построению информационной машины. А пока я коснусь только некоторых работ, проводившихся в Лаборатории раньше и в описываемый тут период.

Логические элементы на ферритовых сердечниках (магнитные элементы). В конце сороковых и в пятидесятые годы вычислительные машины строились, в основном, на электронных лампах, а позже – на полупроводниках. В Лаборатории «пошли другим путем» и предложили строить машины из логических элементов на ферритовых сердечниках (см. выше воспоминания Малиновского). Если идея оперативной памяти на ферритовых сердечниках была тогда стандартной, то использование таких сердечников в логических элементах ЭВМ было оригинальным решением. Причем, идея эта возникла еще до разработки МОЗУ в СССР.

Каждый такой элемент представлял собой небольшую коробочку (размером со спичечный коробок или чуть меньше), внутри которой была некоторая схема из ферритовых сердечников, кристаллических диодов и, может быть, чего-то еще. Элементы были нескольких типов, каждый тип соответствовал одной из булевых операций – дизъюнкции, конъюнкции, отрицанию и некоторым другим. На входы такого элемента в каждом такте работы машины подавались двоичные сигналы 1 или 0, а на выходе элемента (через такт) появлялся результат соответствующей логической операции. Такты задавались управляющими импульсами, подаваемыми одновременно на все элементы вычислительного устройства. Т.е. каждый элемент был, фигурально говоря, небольшим «черным ящиком», реализующим ту или иную логическую операцию.

То, что такие элемены в точности соответствовали логическим операциям, было их несомненным достоинством. Это упрощало проектирование из них логических схем, так как схема строилась как формальный объект. Такое проектирование было в чем-то аналогично программированию или составлению схем конечных автоматов.

Предполагалось также, что элементы на ферритовых сердечниках будут надежнее электронных ламп. Лампы действительно были не очень надежны. А в вычислительных машинах использовались тысячи, а иногда и десятки тысяч ламп. Скептики считали, что такого рода системы в принципе не могут работать. Но разработанные методы ежедневной профилактики как-то решали эту проблему.

Б.Н. Малиновский, ссылаясь на С.А. Лебедева, конструктора БЭСМ, пишет, что одним из инструментов профилактики была кувалда (!):

«Посетив нашу лабораторию и дотошно оглядев ЦЭМ-1, Сергей Алексеевич удивил нас вопросом: "А кувалдочкой вы по ней не стучите?". Оказалось, что на БЭСМ кувалда - это штатный инструмент, а удары ею по железному каркасу машины - один из элементов профилактики! Столь же удивительным теперь показался бы приказ не допускать решения задачи дольше 15 минут без повторного пересчета с тем, чтобы не расходовать машинное время впустую».

Однако разработанные в ЛЭМ магнитные элементы и создаваемые на них устройства были не слишком надежными. Но в пятидесятые годы в Лаборатории проектировали и строили вычислительные машины на этих элементах.

ЛЭМ-1, ЛЭМ-2 и ЛЭМ-4 – вычислительные машины на магнитных элементах[11]. Википедия пишет, что ЛЭМ-1 была создана в 1954 г. Это подтверждается приведенным выше рассказом Малиновского. Когда я попал в Лабораторию, ЛЭМ-1 была уже историей, никаких материальных следов от нее не осталось. Я помню только разговоры о ней, в частности, кто-то говорил Мириам Львовне Аврух, которая была руководителем моего дипломного проекта: «Надо было Вам защищаться на ЛЭМ-1». Видимо, она имела к этой машине самое непосредственное отношение.

А в 1958 г. в Лаборатории шла наладка машины ЛЭМ-2. Этим занимался, кажется, целый отдел – инженеры, отвечающие за те или иные блоки, и программисты. Главным в этом деле был аспирант Гутенмахера Юнис Махмудов. ЛЭМ-2 была темой его диссертации. Юнис был из Азербайджана. Очень симпатичный и, видимо, очень способный. Наверное, он успешно защитился.

Но машина надежно работать так и не стала. Я помню, что Гриша Клейнерман, математик, который, отлаживал на ЛЭМ-2 какие-то программы, говорил, что, приходя на работу, он обычно спрашивал у инженеров, обслуживающих машину: «Как больная?». По-видимому, «больная» так и не встала на ноги.

Уже позже в Лаборатории стали создавать новую машину, ее первоначальное название – ЛЭМ-4**.** Ав-

[10] В.В. Бардиж [4] пишет, что в американских машинах, где МОЗУ появились раньше, использовалась другая идея (3D, D – это dimensions, идеи различались «геометрией» выбора нужной ячейки памяти), там было в два раза меньше сердечников, но к их качеству предъявлялись гораздо более высокие требования, которым советские сердечники в те годы не удовлетворяли. Модификация, предложенная Гутенмахером, позволяла, в частности, работать с не очень качественными сердечниками.

[11] Я именую (нумерую) эти машины, доверяясь собственной памяти. Некоторые люди нумеруют их по-другому.

торами проекта были Г. Клейнерман, Л. Нисневич и Г. Стецюра – очень способные люди. В проекте было много интересных идей. Но машина тоже так и не заработала.

Н.П. Брусенцов и его усовершенствования магнитных элементов. «Сетунь» – самая оригинальная советская ЭВМ с не очень счастливой судьбой.

Прерву свой рассказ о работах ЛЭМ и добавлю еще немного исторического контекста. Про Брусенцова и «Сетунь» пишет тот же Малиновский. Оказывается это тоже связано с ЛЭМ и Гутенмахером.

Про «Сетунь» я слышал давно, еще в начале 60-х гг. – это чуть ли не единственная в мире машина в троичной системе счисления. Но я не знал, что она была построена на магнитных (ферритовых) элементах.

«Сетунь» проектировали в ВЦ МГУ по инициативе С.Л. Соболева. Решено было строить ее на магнитных элементах. Малиновский пишет:

«Соболев договорился с Л.И. Гутенмахером, в лаборатории которого в ИТМ и ВТ АН СССР к этому времени была создана двоичная ЭВМ на магнитных элементах, о стажировке Брусенцова в его лаборатории.

Авторитет Соболева "открыл двери" закрытой для всех лаборатории. "Мне показали машину и дали почитать отчеты, которые в электротехническом отношении, на мой взгляд, оказались весьма слабыми, - вспоминает Н.П. Брусенцов. - ...Разобравшись в этих заблуждениях, я легко нашел схему, в которой работают все сердечники, но не одновременно, что и требовалось для реализации троичного кода".

Именно тогда у него возникла мысль использовать троичную систему счисления. Она позволяла создать очень простые и надежные элементы, уменьшала их число в машине в семь раз по сравнению с элементами, используемыми Л.И.Гутенмахером. Существенно сокращались требования к мощности источника питания, к отбраковке сердечников и диодов...

После стажировки он разработал и собрал схему троичного сумматора, который сразу же и надежно заработал. С.Л. Соболев, узнав о его намерении создать ЭВМ с использованием троичной системы счисления, горячо поддержал замысел... Изобрести сумматоры, счетчики и прочие типовые узлы не составило особого труда для Брусенцова: "Летом 1957 г. на пляже в Новом Афоне все детали были прорисованы в тетрадке, которую я захватил с собой." ...

В 1958 г. сотрудники лаборатории своими руками изготовили первый образец машины...

На десятый день комплексной наладки ЭВМ заработала! Такого в практике наладчиков разрабатываемых в те годы машин еще не было! Машину назвали "Сетунь" - по имени речки неподалеку от Московского университета».

Итак, «брусенцовские» магнитные элементы были несравненно надежнее и гутенмахеровских элементов, и ламп. Построенная на этих элементах «Сетунь» работала безотказно. Мне, к сожалению, не удалось выяснить, знали ли в ЛЭМ об этих «брусенцовских» усовершествованиях магнитных элементов и вообще о работах Брусенцова.

«Сетунь» была наверное самой оригинальной советской машиной, может быть, единственной оригинальной советской разработкой в области вычислительной техники. Другие советские машины того времени, хотя и были самостоятельными разработками, все-таки, так сказать, «догоняли Америку».

История «Сетуни» совершенно замечательна, правда, она не самая счастливая. Но я не буду здесь ее пересказывать, подробнее о ней – в книге Малиновского.

Подведу итог разработкам ЛЭМ того времени. Гутенмахером и его сотрудниками в эти годы были выдвинуты три оригинальные технические идеи:

- модификация МОЗУ (которая в ЛЭМ называлась «система Z», а Бардиж называет ее 2D);
- ДЕЗУ – постоянная память, о ней ниже;
- магнитые (феррит-диодные) элементы.

Все эти идеи требовали серьезной инженерной проработки и, видимо, в ЛЭМ это не всегда получалось.

Как пишет Бардиж [4], МОЗУ для БЭСМ делали по идее Гутенмахера, но довели его до практического использования в лебедевской Лаборатории № 1.

ДЕЗУ не удалось довести до рабочего состояния.

Феррит-диодные элементы работали, по крайней мере вначале, не очень надежно. И создаваемые на них в ЛЭМ машины ЛЭМ-1 и ЛЭМ-2 так и не удалось по-настоящему отладить.

Брусенцов усовершенствовал эти элементы и на них была сделана «Сетунь», удивительно надежная и уникальная по своей конструкции машина. Но если бы не было гутенмахеровских элементов, нечего было бы усовершенствовать.

Долговременное емкостное запоминающее устройство – ДЕЗУ

Видимо, в середине 1950-х гг. Гутенмахер изобрел ДЕЗУ, постоянную память на конденсаторах, емкостное запоминающее устройство. Грубо говоря, память вычислительных машин бывает двух типов – оперативная и постоянная. Содержимое оперативной памяти можно оперативно менять. В постоянную память можно записать что-нибудь раз и навсегда, а потом можно только считывать эту информацию, изменить ее уже невозможно.

Тогда оперативная память стоила очень дорого. ДЕЗУ должно было быть несравненно дешевле. На листах бумаги, не помню уже какой, особой или обычной, напылялись с двух сторон, друг под другом, небольшие металлические кружочки, кажется, по нескольку десятков на листе. Каждая пара кружочков образовывала небольшой конденсатор, который можно было зарядить, подав на него небольшое напряжение. А если этот конденсатор проткнуть (проткнуть бумагу там, где напылена эта пара кружочков), то зарядить его невозможно. Таким образом, кружочек мог быть элементом памяти, представлять 1 или 0. Конечно, эти кружочки надо было как-то коммутировать, для этого на бумагу напылялись полосочки-соединения, а листы бумаги «прошивались» насквозь металлическими стержнями.

Память должна была выглядеть как стопки бумаги и стоить ненамного дороже, чем бумага. Идея, вроде бы, была простой до гениальности. У ДЕЗУ был один недостаток – в реальных условиях оно не работало. Но этот недостаток надеялись со временем испра-

вить. А пока надо было придумать, где эту память можно было использовать.

Гутенмахер выдвинул идею создания информационной машины, основой которой будет ДЕЗУ. Такая машина должна была решать широкий спектр информационно-логических задач. И ЛЭМ перекочевала в 1957 г. в ВИНИТИ – где же, как не в Институте научной информации заниматься информационно-логическими задачами и строить специальные информационно-логические машины.

ИНФОРМАЦИОННАЯ МАШИНА: ФАНТАЗИИ И РЕАЛЬНОСТЬ

Первые вычислительные машины использовались, в основном, для вычислений, так сказать, в соответствии со своим тогдашним названием. Их вообще тогда было очень мало. Было даже мнение, что большого количества вычислительных машин и не нужно, что самих вычислительных задач не так уж и много, можно их быстро все перерешать (в цитируемой выше книге Б.Н. Малиновского говорится, что так, например, считал Базилевский, главный конструктор «Стрелы», одной из первых советских вычислительных машин).

Конечно, быстро стало понятно, что вычислительная машина – это универсальный инструмент, на ней можно решать самые разные задачи, если ... вы умеете это делать. Так, уже в 1950-е гг. появились первые (довольно примитивные) программы машинного перевода текстов и первые шахматные программы и т.п. Тогда же активно обсуждалась целесообразность создания специализированных машин для тех или иных задач.

Как уже говорилось, Л.И. Гутенмахер считал, что нужны специализированные информационные машины. В 1957 г. он опубликовал в Вестнике АН СССР статью «Электрическое моделирование некоторых процессов умственного труда» [5]. Очень любопытная статья. Как по жанру, так и по содержанию. Для жанра, наверное, самое подходящее слово – прожектерство, если освободить его от отрицательных коннотаций. Близкие жанры – научная фантастика или сон Веры Павловны из знаменитого романа Чернышевского.

Гутенмахер писал: «Лабораторией (имеется в виду ЛЭМ – *В.Б.*) проведена подготовительная работа к созданию опытного образца информационной машины с быстродействующей “памятью” на миллиард двоичных знаков. По замыслу машина должна будет выполнять с большой скоростью многие сложные процессы умственного труда».

Быстродействующая память – это ДЕЗУ, опытные образцы которого не удавалось (и не удалось) отладить. Не очень понятно, откуда он взял это число – миллиард. Это сейчас 10^9 бит – мелочь, меньше, чем память обычной флешки. А тогда это число завораживало. В статье довольно много места занимают подсчеты: сколько книг можно записать в такую память, сколько места она будет занимать (всего-то (!) 100 м^3), сколько электричества понадобится (10 квт – «меньше мощности двигателя легкового автомобиля»).

Гутенмахеру казалось, что с такой памятью можно творить чудеса. Говорят, что на каком-то высоком собрании он сказал: «В память такой машины можно будет записать Большую Советскую Энциклопедию. И машина будет отвечать на различные вопросы голосом любимого артиста»[12].

Приводимое в статье [5] описание возможностей проектируемой информационной машины ненамного подробнее и четче этого высказывания: «Информация может поступать в машину... от разнообразных источников», в частности от «фотоэлектронного устройства, позволяющего считывать печатные буквы и цифры с книг и журналов» (реально работающие читающие устройства появятся через десятки лет). «Вопросы могут быть поставлены по существу содержания информации и ответы на них требуют машинного логического анализа и синтеза материала».

Зато активно используются метафоры. Память машины сравнивается с книгохранилищем, а сама машина – с динамической библиотекой, к которой может быть устроен удаленный доступ для тысяч абонентов. Абоненты, пользуясь чем-то вроде каталога, получают нужную им информацию. Детали не уточняются.

Приведенный выше пересказ – это первая часть цитируемой статьи, тут, по крайней мере можно представить, какие технические задачи он хочет решать и чего он хочет добиться. В 1957 г. все это казалось (и было) фантазией. ДЕЗУ, как уже говорилось, так и не удалось отладить, а весь проект информационной машины, основанный на дешевой памяти большого объема, оказался мыльным пузырем.

Но любопытно взглянуть на эти, так сказать, технические фантазии Гутенмахера, рисуемые в первой части его статьи, из сегодняшнего времени. Сейчас объем памяти даже персональных компьютеров составляет сотни гигабайт. Причем, это оперативная память. Постоянная память, которая, как надеялся Гутенмахер, обеспечит информационный прорыв, играет в современных технологиях маргинальную роль. Да и само по себе многократное увеличение объема памяти не вызывает особых революций, важны системы, которые используют эту память. Подлинную информационную революцию принесли базы данных, персональные компьютеры, а потом Интернет, системы типа Google и Яндекс, а также отраслевые информационные системы типа LinguistList и т.п.

Перейду ко второй, гораздо более фантастичной части статьи Гутенмахера, относящейся к «электрическому моделированию процессов умственного труда». Он пишет, что в память информационной машины можно «записать содержание учебников..., результаты научных исследований, ... опыт практической деятельности того или иного работника умственного труда». Дальше воображаемая картина становится совсем мутной. Говорится о стереотипных ассоциациях к заданному машине вопросу, об операциях с этими ассоциациями, смысловыми связями и системами понятий, «в результате которых машина могла бы дать ответ на вопрос, требующий умозаключения, т.е. определения неизвестного отношения

[12] Это речение засело в памяти многих знакомых мне бывших сотрудников ЛЭМ. В частности, об этом вспоминает В.А. Успенский (2002) [6].

между двумя понятиями на основании известного отношения их к третьему». И т.д., и т.п.

Видимо, работавшие в то время в ЛЭМ логики рассказывали ему что-то про теории, аксиомы и правила вывода. Он пишет, что все это можно будет делать после предварительной формализации информации на языке логики предикатов, записав в памяти посылки и получая из них все возможные следствия. Для этого потребуется дальнейшее развитие математической логики. И машина на заданные ей вопросы должна давать ответы, которые не записаны в ее памяти. Более того, «новая ценная информация будет вырабатываться самой машиной в процессе ответов на вопросы, это уже будет не пассивное обучение, а, так сказать, активное накопление практического опыта – активное "самообучение"...». Это было время, когда искусственный интеллект только начинал зарождаться. И обещал много чудес.

Как уже говорилось, в 1957 г. ЛЭМ вошла в состав ВИНИТИ. И создание информационной машины стало главной задачей Лаборатории.

Шутка. После гражданской войны, во время разрухи, Ленин выдвинул лозунг «Коммунизм есть советская власть плюс электрификация всей страны». Герберт Уэллс назвал Ленина кремлевским мечтателем. А местные остряки преобразовали лозунг: «Советская власть – это коммунизм минус электрификация».

По аналогии цитируемую выше статью можно отреферировать как тезис Гутенмахера: «Информационная машина – это ДЕЗУ плюс электрическое моделирование умственного труда». А так как ДЕЗУ никогда не работало и, поэтому, не требовало электричества, то тезис можно трансформировать: «Информационная машина без ДЕЗУ и электричества – это моделирование умственного труда». Для этого в ЛЭМ были созданы теоретические отделы.

ИНФОРМАЦИОННЫЕ ЗАДАЧИ И ТЕОРЕТИЧЕСКИЕ ПОДРАЗДЕЛЕНИЯ ЛЭМ

В. А. Успенский: ЛЭМ, информационная машина и информационный язык

Об информационной машине, предложенной Гутенмахером и о теоретических отделах ЛЭМ пишет В.А. Успенский см. [6 (2002) т. 2, с. 945-946]. Приведу цитату (с сокращениями) из этой работы.

«Совещание на улице Грицевец[13], да и вся деятельность теоретических подразделений Лаборатории электромоделирования служат прекрасной иллюстрацией к высказанной уже мысли, что и неправильные идеи могут порой быть полезными. В данном случае речь идет о принадлежащей Л.И. Гутенмахеру идее («под которую» и была создана его Лаборатория) построения "информационной машины с большой долговременной памятью". Идея носила чисто технический характер и касалась способов записи информации – способов не семиотических, а электротехнических (с помощью ферритовых сердечников прежде всего[14]). .. Кажется, идея оказалась порочной прежде всего с электротехнической точки зрения (первым мне сказал об этом имевший электротехническое образование В.М. Глушков – причем сказал прямо на тротуаре улицы Грицевец, во время перерыва)... Как бы то ни было, задуманная Гутенмахером информационная машина так никогда и не была создана. Однако именно эта, оказавшаяся бесплодной, идея Л.И. Гутенмахера стимулировала теоретические разработки в области прикладной семиотики, относящиеся к способам записи информации на логических языках и информационному поиску. Дело в том, что сам Л.И. Гутенмахер и его ближайшие сотрудники претендовали лишь на изобретение некоего технического способа... хранения большого массива информации и технического же способа поиска в таком массиве. Каков должен был быть язык представления информации, было совершенно неясным. Был поставлен вопрос о создании специального языка для записи информации – так называемого информационного языка, имеющего более отчетливую, чем естественные языки, логическую структуру. Одновременно вставал вопрос о логике информационного поиска. Теоретический аспект всего этого комплекса семиотических проблем располагался на границе между логикой и лингвистикой. Именно в лаборатории электромоделирования Е.В. Падучевой были начаты первые в СССР систематические исследования по логическому анализу естественного языка.

В организационном отношении указанная деятельность привела к созданию внутри Лаборатории электромоделирования специального теоретического отдела, состоявшего из двух групп: группы математической логики и группы математической лингвистики. Отдел так и назывался: математической логики и математической лингвистики. Должность руководителя отдела в Лаборатории электромоделирования называлась «начальник отдела». Начальником отдела математической логики и математической лингвистики Гутенмахер назначил меня. Одновременно я числился руководителем группы математической логики. Группой математической лингвистики руководил В.В. Ива́нов – разумеется, совершенно самостоятельно, так что я не считал себя начальствующим над ним. Ядро отдела составляли Н.М. Ермолаева, А.В. Кузнецов, Д.Г. Лахути, Е.В. Падучева, В.К. Финн, И.Н. Шелимова, Ю.А. Шиханович, А.Л. Шумилина.

Впоследствии, после поглощения в 1960 г. Лаборатории Институтом научной информации АН СССР (он же – Всесоюзный институт научной и технической информации, короче – ВИНИТИ), этот отдел математической логики и математической лингвистики Лаборатории электромоделирования составил ту основу, из которой образовалось существующее и поныне теоретическое подразделение ВИНИТИ...

Л.И. Гутенмахер был противоречивой (как сейчас модно говорить, неоднозначой фигурой). Проекты его были безумны – во всяком случае при том уровне техники, какой был в СССР в 1957 г. Предполагалось, что информационная машина, которую он собирался построить, будет сообщать информацию «голосом любимого артиста» (выражение Гутенмахера, стремившегося, полагаю, таким способом потрафить вкусам членов Политбюро). По-видимому, он искренне верил в свои проекты... Тем не менее именно Льву Израилевичу Гутенмахеру во многом обязана советская семиотика – она начала развиваться под его крылом».

[13] Речь идет о «Совещании по комплексу вопросов, связанных с разработкой и построением машин с большой долговременной памятью», проводившемся в 1957 г. Лабораторией электромоделирования.

[14] Тут В.А. Успенский не совсем точен – долговременная память (ДЕЗУ, основа машины) предполагалась не на ферритовых сердечниках, см. мой текст выше (*В.Б.*)

Теоретические исследования в ЛЭМ

Из цитированного выше текста В.А. Успенского [6] видно, что он не очень верил в реальность создания информационной машины. Но было ясно другое. Для того чтобы решать на вычислительных машинах информационно-логические задачи, нужно не только подходящее «железо» (hardware), нужно думать и о самих задачах. Видимо, ему удалось убедить в этом Гутенмахера. И в ЛЭМ были созданы теоретические отделы, в которых работали математики, логики, лингвисты и химики (о химии и химиках дальше).

Основной целью, поставленной перед этими подразделениями, было создание «специального языка для записи информации – так называемого информационного языка[15]». Эта задача обсуждалась уже в докладах на организованном ЛЭМ Совещании в 1957 г. (см.сноску 13) и в нескольких статьях опубликованного вскоре после этого сборника «Сообщения Лаборатории электромоделирования, выпуск 1». Приведу оглавление этого сборника, в дальнейшем для краткости буду называть его «Сообщения ЛЭМ».

Оглавление сборника «Сообщения Лаборатории электромоделирования»[16]

Многие статьи этого сборника являются изложением докладов на упомянутом Совещании, в том числе статья самого Успенского, которой открывался сборник[17]. В этой статье Успенский писал: «Чтобы лучше изучить наш предмет, отвлечемся от всех посторонних вопросов, в том числе и от техники. Будем руководствоваться следующим принципом: “то, что может быть формализовано, может быть и автоматизировано”. Итак, наша общая задача – формализовать поиск нужной информации, а на первых порах – создать язык, пригодный для формализации поиска, сравнения и отождествления информации».

Опубликованные в «Сообщениях ЛЭМ» статьи, прежде всего работы по созданию информационного языка, – это первые попытки осмыслить, какие формальные средства подходят для описания разного рода информационно-логических задач и методов их решения, в частности задач информационного поиска. Надо сказать, что само понятие информационного языка было не очень четким, как и цели, для которых его собирались создавать. Очевидно было только, что для этого предлагается использовать арсенал средств математической логики – понятия теории, метатеории, логического вывода и т.п.

Предполагалось, что информационный язык может быть создан для каждой предметной области. И в работах, опубликованных в «Сообщениях ЛЭМ», говорится о информационных языках для нескольких областей – органической химии, биологии и геометрии. Однако более или менее подробно, насколько мне известно, был разработан только информационный язык для элементарной (школьной) геометрии. Первые наметки этого языка содержатся в статье Н.М. Ермолаевой и Ю.А. Шихановича, опубликованной в «Сообщениях ЛЭМ». Работа была продолжена А.В. Кузнецовым, Е.В. Падучевой и Н.М. Ермолаевой [8-10]. Это была совместная работа логиков и лингвистов. Разрабатываемый информационный язык претендовал на решение задач[18] «следующих типов:

1. Определить, истинно ли данное предложение.

2. Из данных предложений вывести все (не слишком длинные) следствия, говорящие о таких-то понятиях.

3. Сформулировать все возможные теоремы о таких-то понятиях».

Говорилось, что примером такой задачи «является ответ на вопрос: «Что можно сказать о пяти попарно пересекающихся окружностях?» Ясно, что неразумно (да и невозможно) хранить в памяти машины ответы на все такие вопросы в явном виде. Естественно было бы записать более общие теоремы и аксиомы, на которых теоремы относительно пяти попарно персекающихся окружностей и т.п. можно было бы получить в результате логического вывода из теорем и аксиом. Информационный язык, который допускает такие достаточно сложные логические преобразования информации, мы будем называть информационно-логическим».

[15] Иногда этот язык называли «машинным», подчеркивая, что имеется в виду не машинный код, а именно язык для записи информации в памяти машины.

[16] Сообщения Лаборатории электромоделирования. Вып. 1. - М.: Институт научной информации, 1960.

[17] В списке литературы эта статья значится как Успенский [7] (1959), так как еще до публикации в «Сообщениях ЛЭМ» она была опубликована в сборнике «Проблемы кибернетики».

[18] При пересказе содержания этой работы кавычками выделяется прямое цитирование.

Были намечены контуры такого информационно-логического языка (ИЛЯ) для элементарной геометрии. Выбор геометрии в качестве предметной области был не случаен – геометрия является одной из самых формализованных областей математики, давно служившей полигоном для развития логических методов и процедур. Поэтому возникающие логические проблемы казались преодолимыми, по крайней мере, в принципе. Основой для ИЛЯ был язык многосортной логики предикатов, при этом области, относящиеся к разным сортам, могли пересекаться (точнее, устанавливалась иерархия областей по вложению).

Существенную часть работы составляло описание соответствия между формулами ИЛЯ и предложениями русского языка, которые можно было бы назвать записями этих формул. Поскольку исходной задачей было создание языка для информационного поиска, то ставилась задача разработки алгоритма перевода с естественного языка на ИЛЯ и обратно: «построение алгоритма перевода дало бы возможность автоматизировать процесс ввода в машину необходимой информации, задавать “вопросы” машине на естественном языке и получать “ответ” также на естественном языке».

При этом справедливо подчеркивалось, что «проблема перевода с естественного языка на информационный может во многом рассматриваться как самостоятельная и не связанная непосредственно с информационными задачами. Эта проблема в значительной мере аналогична проблеме перевода с одного естественного языка на другой, причем эта разновидность перевода представляет особый интерес, поскольку перевод на логический язык является одновременно выявлением средств выражения естественным языком основных логических отношений».

Авторы [8] отмечали: «Алгоритм перевода с естественного русского языка на ИЛЯ, если он вообще возможен, был бы немыслимо сложным. Поэтому мы начали с построения алгоритма перевода не для русского языка в его естественной форме, а для некоторого стандартизованного языка (СЯ). В СЯ все понятия и отношения... должны быть выражены достаточно ясно и дискретно... В то же время предложения СЯ должны восприниматься как предложения обычного русского языка...».

Эти синонимические трансформации естественных (для русского языка) выражений к выражениям, более или менее близким к формулам ИЛЯ, назывались *толкованиями* исходных синтаксических выражений (по аналогии к толкованиям слов в словарях).

Авторы [8-10] занимались правилами синонимических преобразований (трансформациями) выражений на СЯ, эквивалентных одной и той же формуле ИЛЯ. Целью было получение выражений на СЯ, близких по структуре соответствующим формулам ИЛЯ.

Эти исследования были продолжены в последующих работах Е.В. Падучевой [11, 12] и, по-видимому, послужили «закваской» для ее книги «О семантике синтаксиса» [13].

Было бы интересно сопоставить эти работы с работами Р. Монтегю [14, 15], из которых выросла формальная семантика [16]. При всей разнице в подходах, в этих работах есть много общего. В обоих подходах рассматриваетсся формальный язык, выражениям которого сопоставляются конструкции естественного языка. У Монтегю это язык интенсиональной логики, а в рассматриваемых здесь работах [8-10] – это информационно-логический язык, построенный на основе многосортной логики предикатов. В обоих подходах изучается именно семантика синтаксиса естественного языка (этот термин, введенный Е.В. Падучевой, очень точно соответствует сути обоих подходов). При этом Монтегю выделяет *семантические типы* выражений и *композиционально* строит сложные выражения (и их семантику) из простых. В ИЛЯ не выделяются семантические типы, позволяющие использовать принцип композиционльности. Но там рассматриваются объекты разных *сортов*, что позволяет говорить об иерархии понятий и, по крайней мере, в принципе, описывать сочетаемость при построении сложных выражений.

Но эти сопоставления, конечно, далеко выходят за рамки моего «исторического» текста.

В целом, рассмотренные публикации – это очень интересное направление исследований, не утратившее своей актуальности и теперь. Замечу, что современные работы по онтологии конкретных предметных областей в какой-то мере перекликаются с планами по созданию информационных языков пятидесятилетней давности.

Но, отмечая несомненную научную значимость этих работ, нужно сказать, что они не решали прикладной задачи информационного поиска, которая послужила для них исходным импульсом [17].

Если современные достижения в области вычислительной техники далеко превзошли самые смелые планы Гутенмахера, то современные исследования по синтаксису и семантике естественного языка, несмотря на громадный прогресс в этой области за последние 50 лет, даже сейчас могут не так уж много предложить современным информационным технологиям.

Работы по информационному поиску пошли «другим путем». Вскоре там сформировалось два основных направления – документальный информационный поиск и фактографические информационные системы. Направления эти существенным образом различались по целям и методам. Документальный поиск был нацелен, как и следует из его названия, на поиск документов, в частности, научных статей. И тут быстро стало понятно, что эффективны только сравнительно грубые методы, типа поиска по ключевым словам, а всякого рода попытки использовать сравнительно сложный синтаксис и логико-семантические методы не добавляют эффективности (по крайней мере, при существующем уровне изученности семантики естественного языка). И, конечно, подлинную революцию в документальном поиске совершили системы типа Google и Яндекс, специальным образом организующие пространство документов в сети и также не использующие пока никаких изысканных логико-семантических методов.

В деле организации фактографических информационных систем успех пришел раньше, в конце 1960-х – начале 1970-х гг., с созданием баз данных для разного рода специальных задач и формального описания соответствующих этим задачам предметных областей. Тут

часто использовались логические методы. Например, реляционные базы данных были, по существу, основаны на логике предикатов и реляционной алгебре. Но и здесь логика использовалась не совсем так, как виделось в конце 50-х гг.

Кроме теоретических подразделений, упоминаемых Успенским, в ЛЭМ был создан отдел, занимающийся информационными проблемами в области химии. Руководителем этого отдела был Г.Э. Влэдуц[19], талантливый человек, сыгравший большую роль в осмыслении информационных задач вообще и, в первую очередь, информационных задач в области химии.

Химия была естественной областью для информационных приложений вычислительных машин. Прежде всего, химики имеют дело с громадными массивами информации – миллионами химических соединений, их формулами, названиями, свойствами, реакциями этих соединений, соотношениями фрагментов этих соединений с их свойствами и т.п. Там уже давно была осознана необходимость механизации обработки этих громадных массивов информации для самых разных целей, начиная от издания разного рода указателей, поиска разного рода объектов и кончая разного рода исследовательскими задачами.

Предлагавшаяся Гутенмахером информационная машина часто рекламировалась как информационная машина для химии. Приведу еще одну цитату из работы Успенского: «Выбор из всех наук именно химии для наполнения памяти информационной машины Гутенмахера был вызван тремя причинами. Во-первых, химическая информация первенствует по объему: реферативные журналы по химии – самые толстые. Во-вторых, химическая информация, хотя бы частично, уже формализована в виде химических формул. В-третьих, покровительствующий информационной машине Президент АН СССР А.Н. Несмеянов был химиком (полагаю, кстати, что в силу первых двух причин он ей и покровительствовал)». Добавлю, что академик А.Н. Несмеянов был и основателем ВИНИТИ. Видимо не случайно ЛЭМ в 1957 г. перешла под крыло ВИНИТИ.

Из приведенного оглавления «Сообщений ЛЭМ» видно, что половина статей в этом сборнике посвящена информационным проблемам химии. При этом химики сотрудничали с логиками и лингвистами.

Один из примеров такого сотрудничества – работы М.М. Ланглебен о названиях химических соединений. Химики давно разрабатывали специальные названия химических соединений (типа *циклопропан, изобутан* и т.п.), отражающие в какой-то мере их структуру – так называемую химическую номенклатуру. Названия эти складываются из некоторых элементов (*метра, иза, цикло, мета, ол, ан* и т.п.) и устроены по некоторым правилам, которые можно назвать грамматикой языка химических названий. М.М. Ланглебен занималась этой грамматикой и соотношением этих названий со структурными формулами соответствующих соединений [18-21].

Моя дипломная работа тоже имела отношение к химическим информационным задачам. Но об этом несколько позже, а пока я хочу вернуться к статье о Гутенмахере в Википедии.

Еще о Гутенмахере – мои впечатления того времени

Приведенная статья в Википедии о Гутенмахере как-то не вполне стыкуется с моими собственными воспоминаниями. Причем, не приводимыми там фактами, а скорее стилем, «энциклопедическим» и почти житийным, и, главное, оценками.

В 1958 г., когда я попал в Лабораторию, Гутенмахеру только что исполнилось 50 лет. Мне он казался очень старым[20]. Атмосфера в Лаборатории, по крайней мере в ее технических отделах, призводила двойственное впечатление. Я уже писал, что в Лаборатории работало много ярких и независимых людей. В то же время официально считалось, что всё в Лаборатории делается «под руководством и по идеям Л.И.Гутенмахера».

Может быть раньше, когда Лаборатория была существенно меньше, а сам Гутенмахер был моложе, дело действительно обстояло так или почти так. Небольшие технические лаборатории тогда нередко напоминали феодальные структуры и все научные результаты приписывались их руководителям. Во всяком случае Гутенмахер считался изобретателем и магнитных элементов, и уже упоминавшейся модификации МОЗУ, и ДЕЗУ.

Но в 1958 г. и время было уже другое, да и както незаметно было потока исходящих от него идей. Он был начальником, научным администратором, все контакты с академическим и прочим начальством замыкались на него, он принимал на работу людей и т.п.

В статье в Википедии о Гутенмахере сказано, что «уже в 1950-е годы он занимался компьютерным моделированием когнитивных процессов (таких как логическое мышление, чтение) и математической лингвистикой». Это, мягко говоря, не совсем точно. Не было в 50-е годы такого понятия, как *компьютерное моделирование когнитивных процессов*. И ни математической лингвистикой, ни просто лингвистикой Гутенмахер никогда не занимался. Вообще он не был, на мой взгляд, теоретиком. Хотя делал иногда «философские» высказывания. Я помню одно такое: «Текст – это мысль в одном измерении, а рисунок – мысль в двух измерениях».

Я бы несколько переставил акценты. Гутенмахер действительно понимал, что существует масса информационно-логических задач, которые в принципе можно решать (или пытаться решить) на вычислительных машинах. И тогда это было нетривиально. У него был нюх на такие задачи и людей, которые эти задачи могли бы решать. В конце 1950-х гг. он создал в своей Лаборатории новые подразделения – логики, математической лингвистики, отделы химии и читающих устройств. И можно согласиться с Успен-

[19] О Г.Э. Влэдуце подробно пишет В.А. Успенский в уже цитированной статье (2002) [6, с. 982].

[20] Сейчас, когда я пишу этот текст, мне 75.

ским, что «Льву Израилевичу Гутенмахеру во многом обязана советская семиотика – она начала развиваться под его крылом».[21]

МОЯ ДИПЛОМНАЯ РАБОТА

Мириам Львовна Аврух

Она была руководителем моей дипломной работы. Феликсу (Рохлину) повезло меньше, ему в руководительницы досталась не очень яркая личность. Но в каком-то смысле мы оба были «под крылом» у М.Л. К тому же, как мне кажется, наше пребывание в ЛЭМ делилось на две стадии, преддипломную практику и собственно написание дипломной работы. И в первой стадии мы оба были у М.Л.

Роль, которую она сыграла в нашей судьбе, моей и Феликса, трудно переоценить. Но и сформулировать внятно трудно. Суть, наверное, в том, что наш институт готовил инженеров, причем инженеров для работы на производстве, а вовсе не научных работников. А это совсем разные ментальности, что ли.

А нам хотелось заниматься наукой. Особенно, когда мы попали в ЛЭМ. И именно, к Мириам Львовне. Но мы как-то еще не вполне понимали, что это такое. И вот у М.Л. мы прошли, так сказать, «начальную подготовку».

Я плохо помню, что мы на этой преддипломной практике реально делали. Наверное учились составлять схемы разного рода устройств на «гутенмахеровских» элементах. Осталось только ощущение об очень благожелательной атмосфере, в которой, правда, «спуску» не давалось.

Смутно помню один эпизод. Нам пришла в голову какая-то «гениальная» идея не помню уже чего, мы нарисовали схему, что-то прорисовали подробно, а какую-то часть, которая казалась нам тривиальной, просто обозначили сбоку. И пришли с этим к М.Л. Она нас выслушала. И спросила – а что это вы там сбоку так небрежно нарисовали?

Мы объяснили: что-то простое, кажется из диодов. А сколько их там будет? Мы подумали. Оказалось, что многовато. И это напрочь убивает нашу такую гениальную идею...

Уже после диплома мы поступили в аспирантуру ВИНИТИ. И попали мы туда а только благодаря М.Л.. Во-первых, она убедила Гутенмахера, что мы подходящие кандидаты. Кроме того, надо было преодолеть массу формальных препятствий.

Но это уже другая история.

Тема

Я уже писал, что она имела отношение к химии. Начну с описания задачи.

В память информационной машины предполагалось записать миллионы химических соединений, и этот массив должен был постоянно пополняться – химики ежегодно создают или открывают тысячи новых соединений. Проблема была в том, как эти соединения туда записывать. Химики используют для соединений как разного типа названия, так и формулы тоже разного типа. Названия бывают бытовые, типа *вода*, и упоминаемые выше специальные, относящиеся к той или иной *номенклатуре*, типа *циклопропан* или *изопропилметилкетон*. Формулы также бывают простыми, перечисляющими типы атомов и их количество (*бруттоформулы*), и сложными, фиксирующими все связи между атомами (*структурные формулы* или химические графы).

В данном случае предполагалось записывать в машину структурные формулы[22]. Формулы эти, как правило, нелинейны, причем структурные формулы органических соединений – часто довольно сложные графы с многими циклами и т.п. В памяти машины эти графы надо как-то представлять в виде некоторых *записей*. Задача состояла в том, чтобы разработать правило, сопоставляющее каждой структурной формуле соединения её *каноническую запись*, каноническую в том смысле, чтобы разным формулам соответствовали разные записи, а изоморфным структурным формулам – одинаковые записи. И создать алгоритм, строящий такую каноническую запись по каждой структурной формуле.

Конечно, можно рассматривать самые разные типы записей, правил их канонизации и самые разные алгоритмы построения таких канонических записей для структурных формул.

Рассмотрим один из простых стандартных примеров такого рода записей. Перенумеруем произвольным образом атомы структурной формулы, т.е вершины химического графа. Перечислим множество ребер этого графа. Я не уточняю способ представления ребер, важно только, чтобы вся необходимая информация была так или иначе представлена – номера вершин, символы атомов, типы связи. Для упрощения последующего перебора можно каким-нибудь способом сразу упорядочить этот список ребер. Тогда *записью* формулы можно считать слово в некотором алфавите, представляющее этот список.

Естественно, что существует много нумераций вершин графа. Записи, т.е. слова, представляющие эти нумерации, можно упорядочить (например, лексикографически). Тогда канонической записью структурной формулы можно считать слово, наименьшее (или наибольшее) по этому порядку.

Конечно, для нахождения такой канонической записи потребуется перебор этих возможных записей. Хотя этот перебор можно существенно сократить, используя те или иные алгоритмы, задача построения такого рода канонической записи в принципе является переборной.

Тогда, в 1958 г., Г.Э. Влэдуц и В.К.Финн предложили другой тип записи, основанный на понятии *обхода графа,* и идею алгоритма построения такой за-

[21] Из моих замечаний видно, что я относился к Л.И. Гутенмахеру тогда, мягко говоря, весьма критически. Но вряд ли я был в состоянии тогда составить о нем объективное представление. Он был человеком другого поколения, другого и очень непростого времени и другой ментальности, совсем непохожим на моих тогдашних друзей и знакомых.

[22] Говоря *записывать в машину* я пользуюсь выражением, которое было в ходу в то время. По сути речь шла о *базе данных* химических соединений. Но такого терминна тогда еще не существовало.

писи.[23] Отмечу только, что как и в рассмотренном выше простом примере, рассматривались разные нумерации вершин графа, каждому обходу графа сопоставлялось некоторое слово, все рассмотренные слова лексикографически упорядочивались и самое старшее слово объявлялось канонической записью структурной формулы.

Был разработан алгоритм, строящий такую запись. Результаты были опубликованы в уже упоминавшемся сборнике «Сообщения ЛЭМ» [22].[24]

Я был в этой кампании «младшим автором». Финн предложил выписывать алгоритм как программу для абстрактной машины, которую он для этого случая «сконструировал» – что-то вроде усложненной машины Поста. В машине для данных было две ленты, пять таблиц и несколько отдельных ячеек. В клеточках лент и таблиц записывались символы и числа, по таблицам и лентам бегали головки, связанные с устройством управления. Команды машины могли сравнивать символы в тех или иных клетках, менять их содержимое и двигаться по лентам и таблицам. А потом мы с Финном «программировали» – выписывали алгоритм как программу для этой машины. Это заняло, как мне помнится, не меньше двух месяцев.

Программа содержала 726 команд (!). Она состояла из блоков и подблоков, для каждого блока и подблока были выписаны его исходные данные, результат и краткое описание – что он делает. Приводилось еще несколько страниц содержательных пояснений.

Сейчас я смотрю на текст этой статьи и на саму программу с некоторым ужасом – конечно, никакой читатель (и никакой программист!) не стал бы разбираться в составленной программе. В содержательных пояснениях можно было (не без труда) разобраться и вычленить из них идею правила и алгоритма (сама идея, как я уже говорил, была довольно естественной).

Нас (авторов) несколько извиняло то, что большого выбора языков для записи алгоритмов тогда не было. Языки программирования только начали появляться. Фортран, вроде бы, уже существовал, но в СССР им никто не пользовался. Программировали, в основном, в машинных кодах. Да и доступных нам вычислительных машин не было, если не считать разрабатываемой в Лаборатории и еще неотлаженной ЛЭМ-2. Наша абстрактная машина, по крайней мере, не зависела от конкретной реализации.

Но на самом деле можно было «пойти другим путем». Предложенное понятие обхода графа (как я вижу его сейчас) допускало сравнительно простое формальное описание, опирающееся на несколько четко определенных конструкций. Поэтому достаточно было привести это описание, после чего идея самого алгоритма становилась прозрачной. Можно было ограничиться этим формальным описанием, и оно могло стать естественной основой для программной реализации. Но тогда нам не хватило именно математической культуры (Влэдуц был химиком, Финн – философом, а я – недоучившимся инженером).

Приведу неформальное описание этой конструкции.

Обход графа – динамический процесс и алгебраическая конструкция. Как уже говорилось, каждая запись, получаемая рассматриваемым алгоритмом, представляет *обход* химического графа (структурной формулы). По сути, такая запись строится как линеаризация *дерева путей* в этом графе. Дерево это обладает некоторыми свойствами, которые я рассмотрю далее. Но вначале рассмотрим динамический процесс получения такого дерева, т.е. как *обходится* граф в процессе построения такого дерева. Этот процесс построения (и перестройки) такого дерева является ядром рассматриваемого алгоритма. Ниже я опишу этот процесс, не вдаваясь в технические детали.

Выберем в графе произвольную вершину и будем двигаться от нее по некоторому пути (их может быть много, как и строящихся деревьев), отмечая некоторым образом все возможные развилки и следя за тем, чтобы ребра на этом пути не повторялись. Двигаясь таким образом, мы придем к какой-то вершине, после которой этот путь нельзя будет продолжить, не попадая на уже пройденное ребро. Фиксируем пройденный путь, затем вернемся к последней пройденной развилке и продолжим обход, идя по ответвлению, содержащему непройденные ребра. И будем продолжать этот процесс, пока не останется ответвлений с непройденными ребрами.[25]

Заметим, что строящееся таким образом дерево путей обладает рядом свойств:

t1) оно покрывает, весь граф, перечисляя все его ребра; при этом каждое ребро «проходится» ровно один раз;

t2) его ветви (и ответвления) упорядочены самим порядком построения, так сказать, слева направо, поэтому его можно линеаризовать;

t3) при этом каждая ветвь дерева в некотором смысле максимальна – ее нельзя продолжить, не попав на уже «пройденное» ребро.[26]

Предлагаемая формальная конструкция представляет этот динамический процесс обхода графа, так сказать, в статике. Поскольку путь в графе формально не является графом, то дерево путей описывается как отдельная конструкция – так называемое плоское дерево, т.е. дерево с двумя «перпендикулярными» отношениям порядка: ⇓ («сверху вниз», т.е. от корня

[23] Такой тип записи считался удобным для задач типа поиска нужных соединений по заданному фрагменту.

[24] Два слова о моем участии в этой работе. М.Л. Аврух, руководитель моей дипломной работы, видимо, советовалась с Влэдуцем и Финном, какую работу мне предложить, и было решено, что я приму участие в детальной разработке этого алгоритма, а потом рассмотрю схему специализированного вычислительного устройства для его реализации. Тогда казалось естественным проектировать и строить такие специализированные усройства. Но об этой второй части моего диплома я уже ничего не помню. В лучшем случае это было упражнением по проектированию вычислительных устройств на логических элементах.

[25] В принципе возможен другой вариант такого процесса (и строящегося дерева). Можно идти по каждому ответвлению не до конца, как в рассмотренном варианте, а только до тех пор, пока не встретится уже пройденная вершина (т.е замкнется какой-нибудь цикл в графе) или же до тупиковой вершины.

[26] Свойство t3 выполняется для рассматриваемого варианта. Для варианта, упомянутого в предыдущей сноске, оно формулируется иначе.

к листьям) и ⇒ (слева-направо).[27] Плоские деревья можно естественным образом линеаризовать, в частности, представить в виде слова в подходящем алфавите. Обходом графа назовем вложение (гомоморфизм) плоского дерева в этот граф, удовлетворяющее свойствам, аналогичным сформулированным выше свойствам t1 – t3. Заметим, что в химическом графе и в деревьях обхода вершины помечены символами атомов, формально их можно понимать как унарные отношения.

Точное описание этой конструкции потребовало бы нескольких страниц формальных определений, неуместных в «мемуарной» статье.

Как уже говорилось, деревьев обхода графа и соответствующих им записей может быть много. На них задается порядок, определяемый как свойствами графа, так и порядком, заданным тем или иным образом на множестве атомарных символов. Алгоритм построения канонической записи пользуется этим порядком для организации эффективного перебора, но этой проблемы я здесь касаться не буду.

В феврале 1959 г. я защитил диплом, как и все студенты нашей группы, проходившие практику в ЛЭМ. Кажется в том же году (а может быть и раньше) в ВИНИТИ начала работать аспирантура. Мы с Феликсом Рохлиным получили то ли совет, то ли неформальное предложение поступать туда. И осенью того же года действительно поступили.

ЛЭМ – КОНЕЦ ИСТОРИИ

В 1960 г. Лаборатория электромоделирования была включена в состав Отдела механизации и автоматизации информационных работ ВИНИТИ (ОМАИР). Заведующим этого отдела и самой Лаборатории был назначен А.М. Васильев и фактически Лаборатория перестала существовать как самостоятельное учреждение. Фактически, но не формально. Юридически она еще долго оставалась автономным учреждением АН.[28] Так, меня перевели из ЛЭМ в ВИНИТИ только в 1970 г., видимо, тогда ЛЭМ была окончательно ликвидирована и числившиеся в ней сотрудники были переведены в ВИНИТИ. Гутенмахер перешел в Институт природных газов в 1962 г., причем, видимо, один, практически все сотрудники ЛЭМ остались в ОМАИРе.

* * *

Я благодарен Н.Я. Бирману, Е.В. Падучевой и Ф.З. Рохлину за обсуждения и воспоминания о событиях более чем полувековой давности. Все неточности на моей совести.

СПИСОК ЛИТЕРАТУРЫ

1. Бирман Н.Я. Краткая история Лаборатории электромоделирования (ЛЭМ) АН СССР // Петербургская библиотечная школа. - 2012. - № 2(39). - С. 7-10.
2. Малиновский Б.Н. История вычислительной техники в лицах. - Киев: фирма "КИТ", ПТОО "А.С.К.", 1995. - 384 с. – URL: http://www.lib.ru/MEMUARY/MALINOWSKIJ/0.htm.
3. Головистиков П.П. Первые годы ИТМ и ВТ. – URL: http://www.ipmce.ru/about/history/remembrance/golovistikov/.
4. Бардиж В.В. Магнитные элементы ЭВМ. – URL: http://www.ipmce.ru/about/history/remembrance/bardig_mag/.
5. Гутенмахер Л.И. Электрическое моделирование некоторых процессов умственного труда // Вестник АН СССР. – 1957. - № 10. – С. 88-95.
6. Успенский В. А. Серебрянный век структурной, прикладной и математической лингвистики в СССР и В.Ю. Розенцвейг. Как это начиналось (заметки очевидца); Wiener Slawistisher Almanach. - 1992 - Sonderband 33: Festschrift für Victor Jul'evič Rosencvejg zum 80. - Geburtstag. - P. 119-162. Перепечатано в сб.: Очерки истории информатики в России / ред.-сост. Д.А. Поспелов, Я.И. Фет. - Новосибирск: Научно-изд. центр ОИГММ СО РАН, 1988. - С 273-309; «Серебрянный век структурной, прикладной и математической лингвистики в СССР: Как это начиналось (заметки очевидца)», а также в двухтомнике автора «Труды по НЕ математике». - М.: ОГИ, 2002.- Т.2.- С. 925-1066, под названием «Серебрянный век структурной, прикладной и математической лингвистики в СССР: Как это начиналось (заметки очевидца)».
7. Успенский В. А. К проблеме построения машинного языка для информационной машины // Проблемы кибернетики / под ред. А.А. Ляпунова, вып. 2. - М. : Физматгиз, 1959. - С. 39-50; Перепечатано в сб. Сообщения Лаборатории электромоделирования, вып. 1. - М.: Институт научной информации, 1960. - С. 5-27, под названием «Логико-математические проблемы создания машинного языка для информационной машины», а так же под исходным названием в двухтомнике автора «Труды по НЕ математике». - М.: ОГИ, 2002. - Т.1. - С. 218-233.
8. Кузнецов А.В., Падучева Е.В., Ермолаева Н.М. Об информационном языке для геометрии и алгоритме перевода с русского языка на информационный // Машинный перевод и прикладная лингвистика. Вып. 5-6. - М.: Объединение по машинному переводу,1961.; Так же в сб. Лингвистические исследования по машинному переводу. - М.: ВИНИТИ, 1961.- С. 40-74. (Англ . перевод в : Information storage and retrieval. – 1963. - Vol.1. - P. 147-165).
9. Кузнецов А.В. Логические контуры алгоритма перевода со стандартизованного русского языка на информационно-логический // Тезисы докладов на конференции по обработке информации, машинному переводу и автоматическому чтению текста. - М., 1961.

[27] Такие деревья аналогичны деревьям непосредственных составляющих, давно рассматривавшихся в математической лингвистике.

[28] Подробнее об этом в Успенский (2002) [6 , с. 946, 947, 975-978]

10. Кузнецов А.В. О перспективах разработки алгоритмов для машинного поиска теорем и вывода следствий // Тезисы докладов на конференции по обработке информации, машинному переводу и автоматическому чтению текста. - М., 1961. С. 56-59.
11. Падучева Е.В. Некоторые вопросы перевода с информационно-логического языка на естественный // НТИ. – 1964. - № 6. – С. 43-49.
12. Падучева Е.В. Семантический анализ естественного языка при переводе на языки математической логики // Труды III Всесоюзной конференции по информационно-поисковым системам и автоматизированной обработке научно-технической информации. II том. Семиотические проблемы автоматизированной обработки информации. - М.: ВИНИТИ, 1967. - С. 156-169.
13. Падучева Е.В. О семантике синтаксиса (материалы к трансформационной грамматике русского языка). - М. : Наука ,1974.
14. Montague R. English as a Formal Language / eds. In Bruno Visentini et al. // Linguaggi nella società e nella tecnica. - Milan: Edizioni di Comunità, 1970.- P. 189-224; Reprinted in Montague, 1974. – P. 188-221.
15. Montague R. Formal philosophy : selected papers of Richard Montague / ed. and with an introd. by Richmond H. Thomason. - New Haven : Yale Univ. Pr., 1974.
16. Partee B. H. Formal semantics. In The Cambridge Encyclopedia of the Language Sciences / ed. Patrick Colm Hogan. - Cambridge Cambridge University Press, 2010. – P. 314-317.
17. Влэдуц Г.Э., Налимов В.В., Стяжкин Н.И. Научная и техническая информация как одна из задач кибернетики // Успехи физических наук. - 1959. - Т. 69, №1. – С. 13-56.
18. Ланглебен М.М. О синтезе названий химических соединений // НТИ. – 1965. - № 10. – С. 18-24.
19. Ланглебен М.М. К лингвистическому описанию номенклатуры органической химии // НТИ. – 1967. - № 1. – С. 13-22.
20. Ланглебен М.М. Опыт приспособления лингвистических понятий и лингвистической терминологии к описанию искусственного языка // Информационные поисковые системы и автоматическая обработка научно-технической информации. - М., 1967. - С. 170-224.
21. Ланглебен М.М. Структура номинативных сочетаний в специальном фрагменте русского химического языка: дис. … канд. хим. наук. - М., 1970. - 257 с.
22. Борщев В.Б., Влэдуц Г.Э., Финн В.К. Об алгоритме перевода структурных формул органической химии в каноническую запись // Сообщения лаборатории электромоделирования, вып. 1 – М.: Институт научной информации, 1960. – С. 99-171.

Материал поступил в редакцию 03.10.11.

Сведения об авторе

БОРЩЕВ Владимир Борисович – доктор физико-математических наук
E-mail: borschev@linguist.umass.edu

ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ

УДК 164.03

В.К.Финн

Двенадцать тезисов об аргументационных системах*

Рассматривается разбиение множества высказываний $P=P^a \cup P^b \cup P'$, где P^a, P^b и P', соответственно множества только аргументируемых, аргументируемых и аргументирующих и только аргументирующих (базисных) высказываний. На P определяются две функции выбора аргументов и контраргументов, образующие семантику логик аргументации. Предлагается четырехзначная логика аргументации. Формируется средствами теории графов аргументационные деревья, лес, состоящий из них, а также некоторые преобразования леса такие, что их результат может быть как планарным, так и непланарным графом. Определяются аргументационные системы и их характеризации, использующие аналитические таблицы.

Посредством аргументационных деревьев формализуется уточнение идеи герменевтического («порочного») круга.

Ключевые слова: *логика аргументации, аргументационные деревья, метод аналитических таблиц, аргументационные системы*

Раздел современной логики, в котором исследуются различные средства аргументации, чрезвычайно важен для понимания человеческого интеллекта и его феноменологии, включающей процесс принятия решений, отбора посылок при рассуждении и упорядочении знаний на некотором достаточном основании. В силу этого выявление и формализация средств аргументации является одним из реальных научных средств исследования рациональности.

Исследование и формализация аргументации расширяет возможности логики как науки о рассуждении и инструменте познания, как это формулирует Й. ван Бентем во многих своих работах и, в частности, в статье «Логика и рассуждение: много ли значат факты?» [1][1].

Процедуры индукции, аналогии, абдукции и аргументации расширяют и обогащают арсенал человеческих рассуждений, не сводимый только к дедуктивным доказательствам. А это означает, что изучение человеческого понимания (важного аспекта интеллектуальной деятельности) точными методами становится реальным делом, выразимым в логических языках [2].

Различные аспекты аргументации представлены в книге “Logic and Argumentation” [2]. Отметим также, что обстоятельный обзор исследований аргументации содержится в книге [3].

В настоящей статье рассматриваются средства формализации аргументации некоторого множества высказываний. Предложенная формализация аргументации имеет следующие четыре особенности:

(1) Формируется специальная семантика логик аргументации соответствующего класса посредством двух функций выбора аргументов ($g^+(p)$) и контраргументов ($g^-(p)$), через эти функции определяются и функции оценок высказываний четырехзначных и трехзначных логик (вырожденным случаем является использование ($g^+(p)$) и ($g^-(p)$) и для двузначной аргументации).

(2) Предлагаемые логики аргументации имеют неассоциативные логические связки – конъюнкцию и дизъюнкцию. Введение n-местных $\&$ и $\vee$ ($n \geq 2$) создает возможность выражать целостности (гештальты), что важно для отображения контекста при представлении знаний.

(3) В статье рассматривается различная **глубина** аргументации, а именно аргументы некоторых высказываний в свою очередь могут иметь аргументы, что, возможно, создает условия для отображения рефлексии.

(4) Важным средством представления аргументации являются аргументационные **деревья** и образованный им **лес**. Они порождают представление аргументации средствами теории графов.

Благодаря особенностям ((1) – (4)) предлагаемой формализации аргументации может осуществляться логическая организация знаний для заданного множества высказываний, что является средством его понимания. Эти установки реализуются в предлагаемых ниже двенадцати тезисах.

Тезис 1. Известны три вида организации знаний: (1) массивы знаний без специфической организации

* Работа выполнена при финансовой поддержке РФФИ (проект № 11-07-00618а)

[1] Эта публикация – перевод статьи Johan van Benthem “Logic and Reasoning do the facts matter?” Studia Logica, vol. 88, 2011.

[2] В силу этого рассмотрение и изучение познавательных процессов не является только сокровенным делом философии, как это может показаться при восприятии работ С.Т. Тулмина: Toulmin S. Human Understanding. - Princeton, New Jersey, 1972 и Toulmin S., Ricke R., Janik A. An introduction to reasoning // Macmillan Publishing Co., 1979.

знаний и их упорядочения («хаотическое представление знаний»), (2) открытые теории естествознания (принципы, гипотезы, экспериментальные данные, эвристики), (3) аксиоматические теории.

Каждый из этих видов организации знаний в системах искусственного интеллекта (ИИ) может иметь формализованное представление посредством логики предикатов 1-го порядка, семантических сетей или фреймов. Однако можно предложить ещё один вид организации знаний и его формализованное представление. Таковыми являются (4) аргументационные системы Argsys($\mathcal{P}$), где $\mathcal{P}$ – множество высказываний, которое упорядочивается посредством функций выбора аргументов $g^+(p)$ и контраргументов $g^-(p)$, $p\in\mathcal{P}$, и четырёхзначной логики аргументации $A_{4,1}^{(5)}$, являющейся модификацией логики $A_{4,1}^{(4)}$ [4].

Тезис 2. С этой целью рассмотрим пропозициональную логику аргументации $A_{4,1}^{(5)}$.

Пропозициональные переменные: p, q, r, ... (быть может, с нижними индексами);

логические связки: $\sim$, $\supset$, $\vee$, $\&_n$, $n = 2, 3, \ldots$.

Определение формулы $A_{4,1}^{(5)}$ стандартно.

Истинностные значения: 1, –1, 0, τ – фактические истина, ложь, противоречие, неопределённость.

p	~p
1	–1
–1	1
0	0
τ	τ

$\supset$	1	–1	0	τ
1	1	–1	0	τ
–1	1	1	1	1
0	1	–1	1	τ
τ	1	–1	0	1

$\vee^{(5)}$	1	–1	0	τ
1	1	1	1	1
–1	1	–1	–1	–1
0	1	–1	0	0
τ	1	–1	0	τ

$\&_2^{(5)}$	1	–1	0	τ
1	1	0	0	τ
–1	0	–1	–1	–1
0	0	–1	0	0
τ	τ	–1	0	τ

$\&_2^{(5)}$ – неассоциативная логическая связка: $0 = 1\&_2^{(5)}(\tau\&_2^{(5)}{-}1) \neq (1\&_2^{(5)}\tau)\&_2^{(5)}{-}1 = -1$. Имеется счётное множество конъюнкций $\&_n(p_1, \ldots, p_n)$.

Тезис 3. Семантика $A_{4,1}^{(5)}$:

$\mathcal{P} = \{p_1, \ldots, p_n, \ldots\}$, где p_n – аргументируемые и аргументирующие высказывания;

$g^+: \mathcal{P} \to 2^{\mathcal{P}}$, $g^-: \mathcal{P} \to 2^{\mathcal{P}}$, $\forall p(g^+(p) \cap g^-(p) = \varnothing)$, $g^\sigma(p) \subset \mathcal{P}$, $p\in\mathcal{P}$, $\sigma = +, -$.

Определение атомарной оценки v[p]:

v[p] = 1, если $g^+(p) \neq \varnothing$, $g^-(p) = \varnothing$;

v[p] = –1, если $g^+(p) = \varnothing$, $g^-(p) \neq \varnothing$;

v[p] = 0, если $g^+(p) \neq \varnothing$, $g^-(p) \neq \varnothing$;

v[p] = τ, если $g^+(p) = \varnothing$, $g^-(p) = \varnothing$.

Логическая матрица $A_{4,1}^{(5)}$: $M = \langle\{1, -1, 0, \tau\}, \{1\}, \sim, \supset, \vee^{(5)}, \{\&_n^{(5)}\}_{n\in N}\rangle$, где N – множество целых положительных чисел и $n\geq 2$, «1» – выделенное истинностное значение.

Для $A_{4,1}^{(5)}$ определяется функция оценки $v^{(i)}[\varphi]$ для любой формулы φ.

В том числе, $v^{(i)}[\&_n^{(5)}(\varphi_1, \ldots, \varphi_n)] = 1$, если $v^{(i)}[\varphi_j] = 1$ для всех $j = 1, \ldots, n$;

$v^{(i)}[\&_n^{(5)}(\varphi_1, \ldots, \varphi_n)] = -1$, если $\exists h(v^{(i)}[\varphi_h] = -1\&(\forall j\, v^{(i)}[\varphi_j] \subseteq \{-1, 0, \tau\}))$;

$v^{(i)}[\&_n^{(5)}(\varphi_1, \ldots, \varphi_n)] = 0$, если $\exists h\exists j((v^{(i)}[\varphi_h] = 1\&v^{(i)}[\varphi_j] = -1)\&\forall k(v^{(i)}[\varphi_k] \subseteq \{1, -1, \tau\}))\vee \exists h(v^{(i)}[\varphi_h] = 0\&\forall j\, v^{(i)}[\varphi_j] \subseteq \{0, 1, \tau\}))\vee\exists h\exists j((v^{(i)}[\varphi_h] = 1\&v^{(i)}[\varphi_j] = -1)\&\forall k(v^{(i)}[\varphi_k] \subseteq \{0, -1, 1\}))$;

$v^{(i)}[\&_n^{(5)}(\varphi_1, \ldots, \varphi_n)] = \tau$, если $\exists h(v^{(i)}[\varphi_h] = \tau\&\forall j(v^{(i)}[\varphi_j] \subseteq \subseteq \{1, \tau\}))$.

Для $A_{4,1}^{(5)}$ аналогично [4] формулируется метод аналитических таблиц, учитывающий неассоциативность $\&_2^{(5)}$ (в силу чего имеются $\&_n^{(5)}$ с $n = 2, 3, 4, \ldots$).

Пусть φ – формула $A_{4,1}^{(5)}$, тогда формулы вида $J_\nu\varphi$, где $\nu\in\{1, -1, 0, \tau\}$, а $\mathbf{J}_\nu\phi = \begin{cases} t, \text{ если } v[\phi] = \nu \\ f, \text{ если } v[\phi] \neq \nu \end{cases}$, будем называть помеченными формулами (t и f – истинностные значения двузначной логики). Правила вывода для аналитических таблиц $A_{4,1}^{(5)}$ формулируются аналогично [4].

Обозначим посредством $b_\sim^\nu$, $b_{\vee^{(5)}}^\nu$ и $b_\supset^\nu$, где $\nu\in\{1, -1, 0, \tau\}$, степень ветвления (декомпозиции) правил вывода для помеченных формул $J_\nu\varphi$ с главными логическими связками $\sim$, $\vee^{(5)}$, $\supset$, соответственно. Аналогично обозначим посредством $b^\nu(n)$ степень ветвления правил вывода для формул с главными связками $\&_n^{(5)}$.

$b_\sim^\nu = 1$, $b_{\vee^{(5)}}^1 = 2$, $b_{\vee^{(5)}}^{-1} = 5$, $b_{\vee^{(5)}}^0 = 3$, $b_{\vee^{(5)}}^\tau = 1$; $b_\supset^1 = 4$, $b_\supset^{-1} = 3$, $b_\supset^0 = 2$, $b_\supset^\tau = 2$; $b^1(n) = 1^n - 0^n$, $b^\tau(n) = 2^n - 1^n$, $b^{-1}(n) = 3^n - 2^n$, $b^0(n) = 4^n - 3^n$[3].

Примеры правил вывода

$$\frac{J_{-1}(\phi\supset\psi)}{J_{-1}\phi \mid J_1\psi \mid J_0\phi, J_0\psi \mid J_\tau\phi, J_\tau\psi},$$

$$\frac{J_\tau(\&_2^{(5)}(\phi_1, \phi_2))}{J_1\phi_1, J_\tau\phi_2 \mid J_\tau\phi_1, J_1\phi_2 \mid J_\tau\phi_1, J_\tau\phi_2 \mid},$$

$$\frac{J_\tau(\&_n^{(5)}(\phi_1, \ldots, \phi_n))}{\beta_{11}, \ldots, \beta_{1n} \mid \ldots \mid \beta_{b^\tau(n)1}, \ldots, \beta_{b^\tau(n)n}},$$

где β_{ij} – соответствующие помеченные формулы.

$J_\nu\varphi$, $J_\mu\varphi$, где $\nu\neq\mu$, являются контрарными парами, а $T_{J_\nu\varphi}$, определяемые аналогично [4], есть аналитические таблицы с корнями $J_\nu\varphi$. Замкнутые $T_{J_\nu\phi}$ определяются стандартно.

Тезис 4. В [5] была сформулирована внешняя семантика для четырёхзначных логик аргументации. А именно, $g^+(p)$ и $g^-(p)$ определялись как отображение

[3] Закономерность ветвления для $b^\nu(n)$ обнаружена Д.В. Виноградовым.

$\mathcal{P}$ в 2^A, где A – множество аргументов и контраргументов, заданных **внешним** образом, так как элементы A не являются высказываниями из $\mathcal{P}$.

В настоящей работе g^σ определяются **внутренним** образом, посредством отображения $\mathcal{P}$ в $2^{\mathcal{P}^b \cup \mathcal{P}'}$. Рассмотрим разбиение $\mathcal{P} = \mathcal{P}^a \cup \mathcal{P}^b \cup \mathcal{P}'$, где $\mathcal{P}^a$ – аргументируемые высказывания, $\mathcal{P}^b$ – аргументируемые и аргументирующие высказывания, $\mathcal{P}'$ – только аргументирующие (базисные, «очевидности») высказывания. Если $p \in \mathcal{P}'$, то $g^\sigma(p) = \{p\}$ или $g^\sigma(p) = \varnothing$, где $\sigma \in \{+, -\}$.

Будем рассматривать аргументационные матрицы $\bar{M}$, где $\bar{M} = \langle M, \mathcal{P}^a \cup \mathcal{P}^b \cup \mathcal{P}', g^+, g^- \rangle$, а $g^\sigma(p) = X_1$, $X_1 \subset \mathcal{P}$, если $p \in \mathcal{P}^a \cup \mathcal{P}^b$; $g^\sigma(p) = X_2$, $X_2 \in \{\varnothing, \{p\}\}$, если $p \in \mathcal{P}'$, где $\sigma \in \{+, -\}$. Если $p \in \mathcal{P}'$, то

$v[p] = 1$, если $g^+(p) = \{p\}$, $g^-(p) = \varnothing$;

$v[p] = -1$, если $g^+(p) = \varnothing$, $g^-(p) = \{p\}$;

$v[p] = 0$, если $g^+(p) = \{p\}$, $g^-(p) = \{p\}$;

$v[p] = \tau$, если $g^+(p) = \varnothing$, $g^-(p) = \varnothing$.

Очевидно, что $g^+(p) \cap g^-(p) = \varnothing$, только если $p \in \mathcal{P}^a \cup \mathcal{P}^b$.

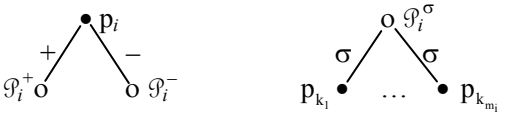
Имеют место следующие типы аргументации:

Таблица 1

	P^a	P^b	P'	Тип аргументации
(1)	+	+	+	$P^a \cup P^b \cup P'$
(2)	+	+	–	$P^a \cup P^b$
(3)	+	–	+	$P^a \cup P'$
(4)	+	–	–	P^a
(5)	–	+	+	$P^b \cup P'$
(6)	–	+	–	P^b
(7)	–	–	+	P'
(8)	–	–	–	$\varnothing$

Невырожденными типами аргументации являются (1) – (3), (5), (6). Далее рассматриваются невырожденные типы.

Тезис 5. Определим аргументационные деревья (аргдеревья) $T(p_j)$. Пусть $U_1 \subseteq \mathcal{P}$, $U_2 \subseteq 2^{\mathcal{P}}$, из U_1 образуются вершины вида • с пометкой $p_i \in U_1$, а из U_2 образуются вершины вида o с пометкой $\mathcal{P}_i^\sigma = g^\sigma(p_i)$, $\mathcal{P}_i^\sigma \in U_2$, где $\sigma \in \{+, -\}$. Таким образом, имеются два вида вершин • p_i и o $\mathcal{P}_i^\sigma$ в аргдеревьях $T(p_j)$ с корнем p_j. $T(p_j)$ имеют два типа рёбер •——o и o——•, которые имеют пометки σ, где $\sigma \in \{+, -\}$. Рёбра •——o и o——• с пометками σ представляют, соответственно, пары $\langle p_i, \mathcal{P}_i^\sigma \rangle$ и $\langle \mathcal{P}_i^\sigma, p_i \rangle$. Аргдерево $T(p_j)$ образовано поддеревьями



$d(\mathcal{P}_i^\sigma) = m_i$, $m_i \geq 1$. $d(\mathcal{P}_i^\sigma) = 0$, если $\mathcal{P}_i^\sigma = \varnothing$, где $\mathcal{P}_i^\sigma = \{p_{k_1}, \ldots, p_{k_{m_i}}\}$ или $\mathcal{P}_i^\sigma = \varnothing$, а $d(\mathcal{P}_i^\sigma)$ – степень вершины $\mathcal{P}_i^\sigma$.

Концевыми вершинами $T(p_j)$ являются синглетоны $\{p_i\}$, где $p_i \in \mathcal{P}'$, или $\varnothing$.

Обозначим посредством $\mathfrak{E}_1$, $\mathfrak{E}_2$ и $\mathfrak{F}_1$, $\mathfrak{F}_2$ множества рёбер вида •——o и o——• и отображения $\mathfrak{E}_1$ в $U_1 \times U_2$, $\mathfrak{E}_2$ в $U_2 \times U_1$, соответственно. Определим аргдерево $T(p) = \langle \{p\}, U_1, U_2, \mathfrak{E}_1, \mathfrak{E}_2, \mathfrak{F}_1, \mathfrak{F}_2 \rangle$ как структуру, удовлетворяющую условиям (1) – (9):

(1) $p \in \mathcal{P}^a \cup \mathcal{P}^b$,

(2) $U_1 \subseteq \mathcal{P}$,

(3) $U_2 \subseteq 2^{\mathcal{P}}$,

(4) $\mathfrak{F}_1$: $\mathfrak{E}_1 \to U_1 \times U_2$, где $\mathfrak{F}_1(e_1^\sigma) = \langle p_i, g^\sigma(p_i) \rangle$, $e_1^\sigma \in \mathfrak{E}_1$, $p_i \in U_1$, $g^\sigma(p_i) \in U_2$, $\sigma \in \{+, -\}$.

(5) $\mathfrak{F}_2$: $\mathfrak{E}_2 \to U_2 \times U_1$, где $\mathfrak{F}_2(e_2^\sigma) = \langle g^\sigma(p_i), p_i \rangle$, $e_2^\sigma \in \mathfrak{E}_2$, $p_i \in U_1$, $g^\sigma(p_i) \in U_2$, $\sigma \in \{+, -\}$.

(6) $\forall x((x \in U_1) \supset \exists \varepsilon^+ \exists \varepsilon^- ((\varepsilon^+, \varepsilon^- \in \mathfrak{E}_1) \& (\mathfrak{F}_1(\varepsilon^+) = \langle x, g^+(x) \rangle) \& (\mathfrak{F}_1(\varepsilon^-) = \langle x, g^-(x) \rangle)))$, где x, ε^+, ε^- – соответствующие переменные,

(7) $\forall x((x \in U_1) \supset \exists \varepsilon^+ \exists \varepsilon^- ((\varepsilon^+, \varepsilon^- \in \mathfrak{E}_2) \& (\mathfrak{F}_2(\varepsilon^+) = \langle g^+(x), x \rangle) \& (\mathfrak{F}_2(\varepsilon^-) = \langle g^-(x), x \rangle)))$, где x не является корнем,

(8) концевыми вершинами (если они существуют) являются $\varnothing = g^\sigma(p_i)$, $\sigma \in \{+, -\}$, $p_i \in U_1$, или $\{p_i\} = g^\sigma(p_i)$, где $p_i \in \mathcal{P}'$,

(9) $|U_1| = l_1$, $|U_2| = l_2$, $|\mathfrak{E}_1| = k_1$, $|\mathfrak{E}_2| = k_2$, где $|\ |$ – число элементов, соответствующих множеств, а $l_1 + l_2 = k_1 + k_2 + 1$.

Переменные x, y, z и X, Y, Z (быть может, с нижними индексами) имеют области определения U_1 и U_2, соответственно[4].

Для последовательностей рёбер x_i•——oX_j (с пометкой σ_j), X_jo——•x_k (с пометкой σ_{j+1}), представимых парами $\langle x_i, X_j \rangle$, $\langle X_j, x_k \rangle$, соответственно, определим предикаты пути из вершины x в вершину X и т.д.: $\Pi(x, X)$, $\Pi(X, x)$, $\Pi(x, y)$, $\Pi(X, Y)$.

Если x_0 – корень аргдерева $T(x_0)$, а X – концевая вершина, то предикат $\Pi(x_0, X)$ представляет **максимальный путь** из x_0 в X, называемый **ветвью**. Ветви $\theta(X)$ соответствует множество её вершин $Set(\theta(X))$. Заметим, что ветвь θ в $T(x_0)$ может не иметь концевой вершины, если $\mathcal{P}' = \varnothing$. Если же θ имеет концевую вершину, то $|Set(\theta)| < \infty$.

Рассмотрим пример аргдерева $T(p_1)$ с типом аргументации (6):

$\mathcal{P}^a = \varnothing$, $\mathcal{P}^b \neq \varnothing$, $\mathcal{P}' = \varnothing$, где $\mathcal{P}^b = \{p_1, p_2, p_3, p_4\}$, а $g^+(p_1) = \{p_2\}$, $g^-(p_1) = \varnothing$, $g^+(p_2) = \{p_3, p_4\}$, $g^-(p_2) = \varnothing$, $g^+(p_3) = \{p_1\}$, $g^-(p_3) = \varnothing$, $g^+(p_4) = \{p_3\}$, $g^-(p_4) = \varnothing$. (Рис. 1)

Из определения аргдеревьев следует, что они конечно-порождённые, так как степени вершин конечны: $d(x) = 2$, а $d(X)$ есть некоторое число m, $m \geq 1$. В силу неограниченного повторения вершин с пометкой $\{p_2\}$ $|T(p_1)| = \infty$, а потому, в силу леммы Кёнига [7], в $T(p_1)$ существуют бесконечные ветви, а именно $|Set(\theta_1)| = \infty$ и $|Set(\theta_2)| = \infty$. Пунктиром в аргдереве обозначены бесконечные ветви θ_1 и θ_2.

[4] Используется язык двусортной логики предикатов 1-го порядка [6]

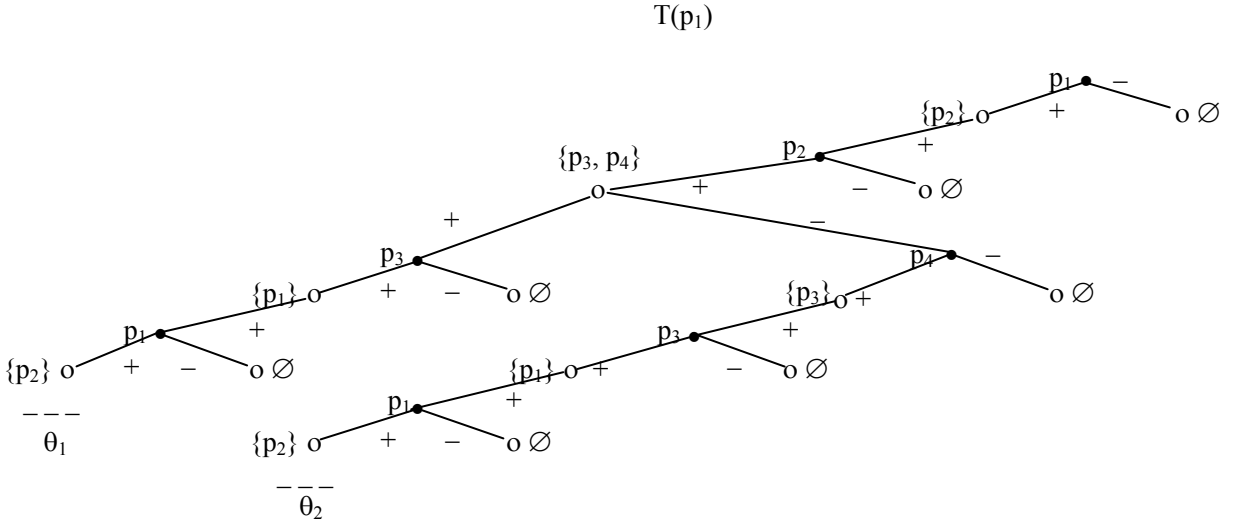


Рис. 1.

В силу конечно-порождённости аргдеревьев в любом аргдереве $T(p_i)$ существует лишь конечное число ветвей $\theta_1, \ldots, \theta_k$. Рассмотрим множества вершин $Set(\theta_j)$, $j = 1, \ldots, k$, тогда $Set(T(p_i)) = \bigcup_{j=1}^{k} Set(\theta_j)$ будет конечным множеством, если выполняется следующее условие (C_1): $\exists x_1((x_1 \in Set(\theta_j)) \;\&\; (g^+(x_1)=\{x_1\} \vee g^+(x_1)=\varnothing) \;\&\; (g^-(x_1)=\{x_1\} \vee g^-(x_1)=\varnothing)\&\ldots\&\; \exists x_k((x_k \in Set(\theta_k))\& \;(g^+(x_k)=\{x_k\} \vee g^+(x_k)=\varnothing)\& \;(g^-(x_k)=\{x_k\} \vee g^-(x_k)=\varnothing))$ для $j = 1, \ldots k$.

Введём в рассмотрение метапредикат $!_1$ конечности аргдерева $T(x_0)$: $!_1T(x_0)$. $!_1T(x_0)$ тогда и только тогда, когда имеет место (C_1). $T(x_0)$ бесконечно тогда и только тогда, когда имеет место отрицание (C_1), то есть $\neg!_1T(x_0)$. Очевидно, что $\neg!_1T(x_0)$, если в $T(x_0)$ существует бесконечная ветвь θ. Следовательно, в $Set(\theta)$ не существует вершин x таких, что $x \in \mathcal{P}'$.

Тезис 6. Из определения предиката пути Π в аргдереве следуют свойства его транзитивности:

$\forall X \forall x \forall Y((\Pi(X, x)\&\Pi(x, Y)) \supset \Pi(X, Y))$,

$\forall x \forall X \forall y((\Pi(x, X)\& \Pi(X, y)) \supset \Pi(x, y))$,

$\forall x \forall y \forall X((\Pi(x, y)\&\Pi(y, X)) \supset \Pi(x, X))$,

$\forall X \forall Y \forall x((\Pi(X, Y)\&\Pi(Y, x)) \supset \Pi(X, x))$.

Очевидно также следующее

Утверждение 1. $\forall x(\Pi(x, x) \supset (x \in \mathcal{P}^b))$.

Из определения $\Pi(x, x)$ следует, что существуют последовательности рёбер $x\bullet$——oX, …, Xo——$\bullet x$, $x\bullet$——oX, где $g^\sigma(x)=X$, следовательно, $x \in \mathcal{P}^b$, так как x аргументируемо и аргументирует.

Утверждение 2. $\exists x \Pi(x, x) \supset \exists X \Pi(X, X)$.

Утверждение 3. $\exists X \Pi(X, X) \supset \exists x \Pi(x, x)$.

Докажем **Утверждение 2**.

Так как при $\Pi(x, x)$ имеется последовательность рёбер $x\bullet$——oX_1, …, X_ko——$\bullet x$, но $g^\sigma(x)=X$, следовательно, добавляется ребро $x\bullet$——oX, а потому в силу транзитивности Π получаем $\Pi(X, X)$. Аналогично доказывается **Утверждение 3**.

Следующее утверждение формулирует необходимое и достаточное условие бесконечности аргдеревьев.

Утверждение 4. $\neg!_1T(p_i)$ тогда и только тогда, когда $\exists x \Pi(x, x)$.

(1) Необходимость: если $\neg!_1T(p_i)$, то $\exists x \Pi(x, x)$.

Так как $|\, T(p_i)\, | = \infty$, то по лемме Кёнига [7] существует в $T(p_i)$ ветвь θ такая, что $|Set(\theta)| = \infty$. Но $\mathcal{P}$, из которого имеются пометки типа • вершин $T(p_i)$, является **конечным** множеством, следовательно, в θ существует вершина x такая, что она **повторяется**, то есть имеет другое вхождение в θ. Следовательно, в силу определения предиката Π существует путь из x в другое вхождение x в θ, т.е. $\exists x \Pi(x, x)$.

(2) Достаточность: если $\exists x \Pi(x, x)$, то $\neg!_1T(P_i)$.

Так как $\exists x \Pi(x, x)$, то $x \in \mathcal{P}^b$ в силу **Утверждения 1**. Следовательно, вершине x соответствует ребро, представленное парой $\langle x, X\rangle$, где $X = g^\sigma(x)$, а потому из вершины X имеется путь в x. Продолжая этот процесс, получим бесконечную ветвь θ, а следовательно, $\neg!_1T(p_i)$.

Из **Утверждения 4** и **Утверждения 2** выводится

Утверждение 5. $\neg!_1T(p_i)$ тогда и только тогда, когда $\exists X \Pi(X, X)$[5].

Легко вывести также следующее

Утверждение 6. (*a*) Если $\neg!_1T(p_i)$, то $\mathcal{P}^b \neq \varnothing$; (*b*) если $\mathcal{P}^b \neq \varnothing$ и $\mathcal{P}' = \varnothing$, то $\neg!_1T(p_i)$.

Тезис 7. Сформулируем уточнение определения аргдерева $T(p_i)$. Дело в том, что вершины типа $x \bullet$ (с пометкой x) могут иметь несколько вхождений в $T(p_i)$, а потому им следует придать пометки j и m, которые являются номерами вхождений ($j \neq m$). Таким образом, в $T(p_i)$ вершины типа $x \bullet$ представимы посредством $x \bullet j$, $x \bullet m$. Аналогично представимы и вершины X o k.

Следовательно, делаются замены $U_1 \subseteq \mathcal{P}$, $U_2 \subseteq 2^{\mathcal{P}}$ на $\tilde{U}_1$ и $\tilde{U}_2$ посредством использования кратных вхождений $x \bullet j$ и X o k. Соответственно, определим $\tilde{\mathfrak{F}}_1 : \tilde{\mathfrak{S}}_1 \to \tilde{U}_1 \times \tilde{U}_2$, $\tilde{\mathfrak{F}}_2 : \tilde{\mathfrak{S}}_2 \to \tilde{U}_2 \times \tilde{U}_1$. Очевидно, что для ар-

[5] Заметим, что **Утверждения 4** и **5** являются уточнением и формализацией герменевтического («порочного») круга [8].

гументационных матриц $\bar{M}$ имеет место условие: $\forall x(g^{\sigma}(x \bullet j) = g^{\sigma}(x \bullet m))$, где $x \in \mathcal{P}$, а $\sigma \in \{+, -\}$.

Рассмотрим $\bar{M}$ с $\mathcal{P} = \mathcal{P}^a \cup \mathcal{P}^b \cup \mathcal{P}'$ и $U = \{p_1, \ldots, p_n\}$ такие, что $p_i \in U_1^{(i)}$, $p_i \in U$, $U_1^{(i)} \subseteq \mathcal{P}$, $U_2^{(i)} \subseteq 2^{U_1^{(i)}}$, $i = 1, \ldots, n$; $\mathcal{P} = \bigcup_{i=1}^{n} U_1^{(i)}$, а $U \subseteq \mathcal{P}^a \cup \mathcal{P}^b$.

Определим аргдеревья $T(p_i)$ с корневыми вершинами p_i: $T(p_i) = \langle \{p_i\}, \tilde{U}_1^{(i)}, \tilde{U}_2^{(i)}, \tilde{\mathcal{E}}_1^{(i)}, \tilde{\mathcal{E}}_2^{(i)}, \tilde{\mathfrak{F}}_1^{(i)}, \tilde{\mathfrak{F}}_2^{(i)} \rangle$, где $\tilde{\mathcal{E}}_1^{(i)}$, $\tilde{\mathcal{E}}_2^{(i)}$ – множества соответствующих рёбер.

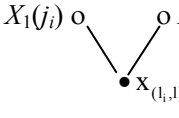
Множество аргдеревьев $T(p_i)$ с $p_i \in U$ образует **лес** Wd, определяемый следующим образом: $Wd = \langle U, \mathcal{P}, \{T(p_i)\}_{p_i \in U} \rangle$[6].

Так как каждой вершине $\bullet\, p_j$ в $T(p_i)$ соответствуют два ребра вида $\langle p_j, g^{+}(p_j) \rangle$, $\langle p_j, g^{-}(p_j) \rangle$, то $\bullet\, p_j$ взаимнооднозначно может быть сопоставлено $J_\nu p_j$, где $\nu \in \{1, -1, 0, \tau\}$. Поэтому $T(p_i)$ может быть взаимнооднозначно сопоставлен его J-образ – дерево $JT(p_i)$ с соответствующими помеченными вершинами $J_\nu p_j$ и приписанными им номерами кратных вхождений.

Будем говорить, что вершины $J_\nu p_j(l_i)$ в $JT(p_i)$ и $J_\nu p_j(l_k)$ в $JT(p_k)$ **аргументационно равносильны** в Wd. Будем также говорить, что вершина $\bullet\, p_j$ **аргументационно зависима** от вершины $\bullet\, p_h$ в $T(p_i)$, если пара $\langle p_j(m_1), p_h(m_2) \rangle$ выполняет предикат $\Pi(x, y)$. Аргументационно равносильные и аргументационно зависимые вершины в аргдеревьях $\{T(p_i)\}_{p_i \in U}$, образующих лес Wd, будем называть **аргументационно релевантными** вершинами.

Тезис 8. Определим преобразование леса Wd в **заросли** Br, используя релевантные вершины аргдеревьев $T(p_i)$. Рассмотрим два вида преобразований ∇ и # множества $\{T(p_i)\}_{p_i \in U}$ в Br. Каждую пару $\langle p_j, p_h \rangle$ аргументационно зависимых вершин в $T(p_i)$ соединим ребром $e_{j,h}$. Применяя эту процедуру # ко всем $T(p_i)$ из Wd, преобразуем $\{T(p_i)\}_{p_i \in U}$ в несвязный граф $\Gamma = \{\Gamma_i\}_{p_i \in U}$ со связными компонентами $\Gamma_i = \#(T(p_i))$.

Рассмотрим $T(p_i)$ и $T(p_k)$ из Wd. Пусть e_{j_i,l_i} и e_{j_k,l_k} – рёбра из $T(p_i)$ и $T(p_k)$, соответственно, где $\tilde{\mathfrak{F}}_2^{(i)}(e_{j_i,l_i}) = \langle X_1(j_i), x(l_i) \rangle$, $\tilde{\mathfrak{F}}_2^{(k)}(e_{j_k,l_k}) = \langle x_2(j_k), x(l_k) \rangle$. Определим преобразование ∇ множества $\{T(p_i), T(p_k)\}$ в связный граф $\Gamma^* = \nabla\{T(p_i), T(p_k)\}$ следующим образом: ребро e_{j_i,l_i} (в $T(p_i)$) и ребро e_{j_k,l_k} (в $T(p_k)$) и соответствующие им вершины преобразуем в подграф



, где вершина $x_{(l_i,l)_k}$ – результат совмещения аргументационно равносильных вершин $x(l_i)$ и $x(l_k)$, преобразуемых в $x_{(l_i,l)_k}$.

Распространим это преобразование на любое подмножество из $\{T(p_i)\}_{p_i \in U}$ и получим определение преобразования ∇: $\nabla(\{T(p_{h_1}), \ldots, T(p_{h_m})\}) = \Gamma^*$, где $\{T(p_{h_1}), \ldots, T(p_{h_m})\} \subseteq \{T(p_i)\}_{p_i \in U}$, а $2 \le m \le |U|$.

Br(Wd) будем называть **зарослями**, если имеют место следующие равенства:

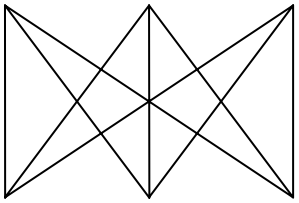
(*a*) $Br(Wd) = \#(\{T(p_i)\}_{p_i \in U}) = \{\#(T(p_i))\}_{p_i \in U}$,

(*b*) $Br(Wd) = \nabla(\{T(p_i)\}_{p_i \in U})$,

(*c*) $Br(Wd) = \#(\nabla(\{T(p_i)\}_{p_i \in U}))$,

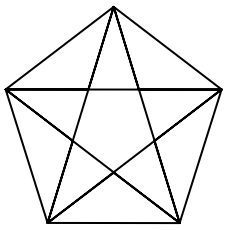
(*d*) $Br(Wd) = \nabla(\#(\{T(p_i)\}_{p_i \in U}))$.

Тезис 9. Лес Wd является несвязным планарным графом, а заросли Br(Wd) могут быть **непланарным** графом Γ^*, непланарность которого устанавливается посредством теоремы Куратовского-Понтрягина [9]: граф Γ^* является планарным тогда и только тогда, когда он не содержит подграфов, гомеоморфных одному из следующих подграфов:



«колодец»

Рис. 2



«звезда»

Рис. 3

Приведём пример леса Wd такого, что ему соответствуют **непланарные** заросли Br(Wd). Рассмотрим аргументационную матрицу $\bar{M}$ с $\mathcal{P} = \mathcal{P}^a \cup \mathcal{P}^b \cup \mathcal{P}'$, где $\mathcal{P}^a = \{p_4, p_5, p_6\}$, $\mathcal{P}^b = \varnothing$, $\mathcal{P}' = \{p_1, p_2, p_3, p_7, p_8, p_9\}$, такую, что она представима лесом Wd, содержащим деревья $T(p_4)$, $T(p_5)$ и $T(p_6)$, которые определяются посредством $g^{\sigma}(p_i)$, где $1 \le i \le 9$, а $\sigma \in \{+, -\}$.

$T(p_4)$:
$g^{+}(p_4) = \{p_1, p_2, p_3, p_7\}$, $g^{-}(p_4) = \varnothing$;
$g^{+}(p_1) = \{p_1\}$, $g^{-}(p_1) = \varnothing$;
$g^{+}(p_2) = \{p_2\}$, $g^{-}(p_2) = \varnothing$;
$g^{+}(p_3) = \{p_3\}$, $g^{-}(p_3) = \varnothing$;
$g^{+}(p_7) = \{p_7\}$, $g^{-}(p_7) = \varnothing$.

$T(p_5)$:
$g^{+}(p_5) = \{p_1, p_2, p_3, p_8\}$, $g^{-}(p_5) = \varnothing$;
$g^{+}(p_1) = \{p_1\}$, $g^{-}(p_1) = \varnothing$;
$g^{+}(p_2) = \{p_2\}$, $g^{-}(p_2) = \varnothing$;
$g^{+}(p_3) = \{p_3\}$, $g^{-}(p_3) = \varnothing$;
$g^{+}(p_8) = \{p_8\}$, $g^{-}(p_8) = \varnothing$.

[6] Заметим, что множество, состоящее из двух аргдеревьев, считается лесом.

$T(p_6)$:
$g^+(p_6) = \{p_1, p_2, p_3, p_9\}$, $g^-(p_6) = \varnothing$;
$g^+(p_1) = \{p_1\}$, $g^-(p_1) = \varnothing$;
$g^+(p_2) = \{p_2\}$, $g^-(p_2) = \varnothing$;
$g^+(p_3) = \{p_3\}$, $g^-(p_3) = \varnothing$;
$g^+(p_9) = \{p_9\}$, $g^-(p_9) = \varnothing$[7].

Применяем ∇ и получаем $Br(Wd) = \nabla(\{T(p_4), T(p_5), T(p_6)\})$ такие, что соответствующим им граф Γ^* содержит подграф Γ', являющийся «колодцем» (см. рис. 2)

$\{p_1, p_2, p_3, p_7\}$ $\{p_1, p_2, p_3, p_8\}$ $\{p_1, p_2, p_3, p_9\}$

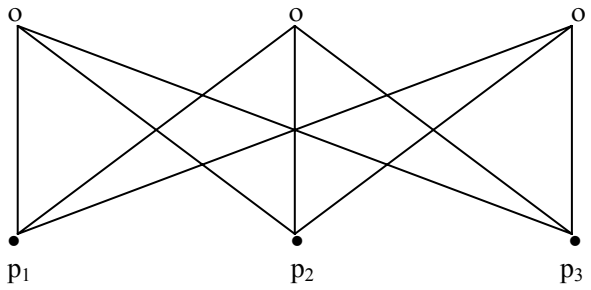
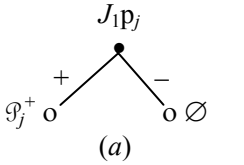


Рис.4

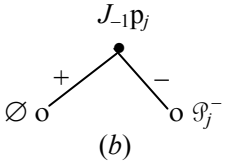
Для сравнительно простого типа аргументации (3) с $\mathcal{P}^a \neq \varnothing$, $\mathcal{P}^b = \varnothing$ и $\mathcal{P}' \neq \varnothing$ построены непланарные заросли, являющиеся его **графическим представлением**.

Введём метапредикат ! для распознавания планарности (непланарности) Br(Wd). !Br(Wd) означает, что зарослям Br(Wd) соответствует планарный граф Γ^*. Неэкономный алгоритм распознавания непланарности, т. е. установление истинности $\neg$!Br(Wd), сводится к перечислению подграфов Γ^* и поиску наличия его подграфов Γ' таких, что они гомеоморфны «колодцу» (рис. 2) или «звезде» (рис. 3). Отсутствие таковых означает истинность !Br(Wd).

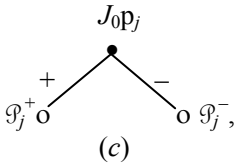
Тезис 10. Рассмотрим J-образы $J\mathrm{T}(p_i)$ аргдеревьев $T(p_i)$. В $J\mathrm{T}(p_i)$ имеются подграфы



(*a*)



(*b*)



(*c*)

где $\mathcal{P}_j^\sigma = g^\sigma(p_j) \neq \varnothing$ или $\mathcal{P}_j^\sigma = \varnothing$.

Рёбра e^σ такие, что $J_\nu p_j$ •——o $\mathcal{P}_j^\sigma$ (с пометкой σ) и $\mathcal{P}_j^\sigma \neq \varnothing$, $\sigma \in \{+, -\}$, будем называть **информативными**.

Пусть $\mathcal{P}_j^\sigma = \{p_{l_1}, \ldots, p_{l_h}\}$, $|\mathcal{P}_j^\sigma| = h$, $\&_h^{(5)}(\mathcal{P}_j^\sigma) = \&_h^{(5)}(p_{l_1}, \ldots, p_{l_h})$. Вершинам вида J_1p_j и рёбрам e^+ с $\mathcal{P}_j^+ \neq \varnothing$ сопоставим формулу $J_1(p_j \supset \&_h^{(5)}(\mathcal{P}_j^+))$, вершинам вида $J_{-1}p_j$ и рёбрам e^- с $\mathcal{P}_j^- \neq \varnothing$ сопоставим формулу $J_1(\&_h^{(5)}(\mathcal{P}_j^-) \supset p_j)$, вершинам вида J_0p_j с рёбрами e^+ и e^- с $\mathcal{P}_j^+ \neq \varnothing$ и $\mathcal{P}_j^- \neq \varnothing$ сопоставим формулы $J_1(p_j \supset \&_h^{(5)}(\mathcal{P}_j^+))$ и $J_1(\&_h^{(5)}(\mathcal{P}_j^-) \supset p_j)$, соответственно. Используемые импликации $A_{4,1}^{(5)}$ выражают **аргументационные зависимости** вершин вида • и o.

Рассмотрим в $J\mathrm{T}(p_i)$ конечную ветвь θ. На θ определим множество формул, такое, что оно состоит из всех помеченных вершин типа $\bullet\, J_{\nu_j} p_j$ и всех формул, представляющих аргументационные зависимости помеченных вершин типа • и типа o, представимых определёнными выше импликациями логики $A_{4,1}^{(5)}$. Множество таких формул, соответствующих ветви θ_m, обозначим посредством $\mathrm{Set}(\theta_m)$. Пусть $\theta_1, \ldots, \theta_s$ – все ветви $J\mathrm{T}(p_i)$ такие, что их концевые вершины не являются $\varnothing$ (эти ветви образованы только информативными рёбрами). **Описанием** аргдерева $T(p_i)$ будем называть $\mathrm{Des}(T(p_i)) = \mathrm{Set}(\theta_1) \cup \ldots \cup \mathrm{Set}(\theta_s)$, а $\mathrm{Des}(\theta_m) = \mathrm{Set}(\theta_m)$, $m = 1, \ldots, s$, будем называть **описанием** ветви θ_m.

Заметим, что если $\neg!_1 T(p_i)$, то его описание $\mathrm{Des}(T(p_i))$ – конечное множество формул (из-за повторения вершин типов • и o).

Пусть $\Sigma_m = \mathrm{Des}(\theta_m)$, $m = 1, \ldots, s$; $\Sigma = \mathrm{Des}(T(p_i)) = \bigcup_{m=1}^{s} \Sigma_m$. построим средствами $A_{4,1}^{(5)}$ аналитическую таблицу $\mathfrak{T}_\Sigma$ для множества Σ, началом которой является Σ. Σ – **непротиворечиво**, если $\mathfrak{T}_\Sigma$ содержит открытую ветвь θ. Σ является **противоречивым**, если все ветви $\mathfrak{T}_\Sigma$ замкнуты [4, 5]. Соответственно, введём обозначения $\mathrm{Consis}(\Sigma)$ и $\neg\mathrm{Consis}(\Sigma)$. Определим метапредикат $!_2$ непротиворечивости для $\mathrm{Des}(T(p_i))$: $!_2(T(p_i)) \rightleftharpoons \mathrm{Consis}(\Sigma_i)$, где $\Sigma_i = \mathrm{Des}(T(p_i))$. Соответственно, $\neg!_2(T(p_i)) \rightleftharpoons \neg\mathrm{Consis}(\Sigma_i)$. Приведём пример описания аргдерева $T(p_1)$ с $\mathcal{P}^a = \{p_1\}$, $\mathcal{P}^b = \{p_2\}$ и $\mathcal{P}' = \{p_3\}$.

Для $T(p_1)$ легко восстановить аргументационную матрицу $\bar{M}$ и построить $J\mathrm{T}(p_1)$. Информативными ветвями $T(p_1)$ являются θ_1 и θ_4: $\mathrm{Des}(\theta_1) = \{J_1p_1, J_1(p_1 \supset \&_2^{(5)}(p_2, p_3)), J_1p_2, J_1(p_1 \supset p_3), J_1p_3, J_1(p_3 \supset p_3)\}$, $\mathrm{Des}(\theta_4) = \{J_1p_1, J_1(p_1 \supset \&_2^{(5)}(p_2, p_3)), J_1(p_3 \supset p_3)\}$.

Для $\Sigma_1 = \mathrm{Des}(\theta_1) \cup \mathrm{Des}(\theta_4)$ построим аналитическую таблицу $\mathfrak{T}_{\Sigma 1}$ в $A_{4,1}^{(5)}$. В $\mathfrak{T}_{\Sigma 1}$ существует открытая ветвь, а, следовательно, $!_2(T(p_1))$.

[7] Представляем аргдеревья $T(p_i)$, $i = 4, 5, 6$, упрощённо, без пометок рёбер и номеров вершин для вхождений $p_j \in \mathcal{P}$. Это означает, что используются $U_1^{(i)}$ и $U_2^{(i)}$, а не $\tilde{U}_1^{(i)}$, $\tilde{U}_2^{(i)}$.

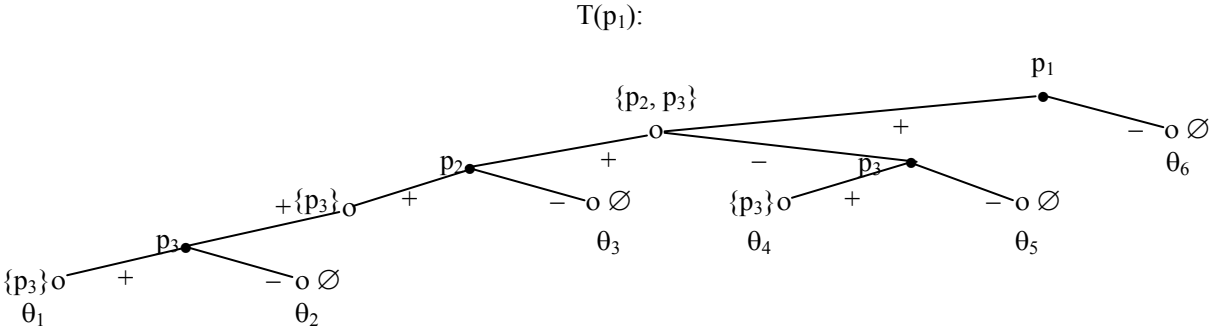


Рис. 5

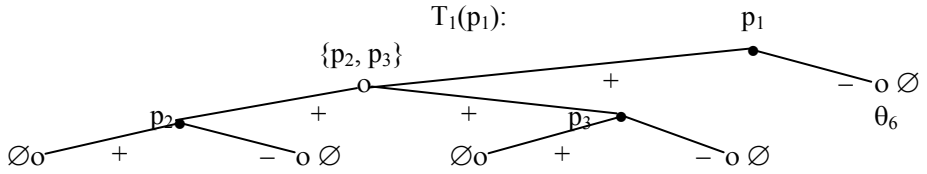


Рис. 6

$g^+(p_1) = \{p_2, p_3\}$, $g^-(p_1) = \varnothing$;
$g^+(p_1) = \varnothing$, $g^-(p_1) = \varnothing$;
$g^+(p_2) = \varnothing$, $g^-(p_2) = \varnothing$.
$\Sigma_1 = \mathrm{Des}(T(p_1)) = \{J_1p_1, J_1(p_1 \supset \&_2^{(5)}(p_2, p_3)), J_\tau p_2, J_\tau p_3\}$.

Построим аналитическую таблицу $\mathfrak{T}_{\Sigma 1}$ для Σ_1, соответствующую $T_1(p_1)$:

$$
\begin{array}{ccccc}
 & & J_1p_1 & & \\
 & & J_\tau p_2 & & \\
 & & J_\tau p_3 & & \\
 & & J_1(p_1 \supset \&_2^{(5)}(p_2, p_3)) & & \\
J_{-1}p_1 & | & J_1(\&_2^{(5)}(p_2, p_3)) & | & J_0p_1 \quad | \quad J_\tau p_1 \\
* & | & J_1p_2 & | & J_0(\&_2^{(5)}(p_2, p_3)) \; | J_\tau(\&_2^{(5)}(p_2, p_3)) \\
 & & J_1p_3 & | & * \quad | \quad * \\
 & & * & &
\end{array}
$$

* – указатель замкнутости ветви, а J_1p_1, $J_{-1}p_1$; $J_\tau p_2$, J_1p_2; J_1p_1, J_0p_1; J_1p_1, $J_\tau p_1$ – контрарные пары на соответствующих ветвях $\mathfrak{T}_{\Sigma 1}$. Следовательно, $\mathfrak{T}_{\Sigma 1}$ замкнута, а потому $\neg!_2(T_1(p_1))$: $\mathrm{Des}(T(p_1))$ противоречиво.

Вершины аргдеревьев типа • такие, что они непосредственно предшествуют концевым вершинам, будем называть **предконцевыми**. Имеет место

Утверждение 7. Если $J_\mu p_i$ – корень аргдерева $T(p_i)$, $\mu \in \{1, -1, 0\}$, а все предконцевые вершины имеют вид $J_\tau p_j$, то $\neg!_2(T(p_i))$.

Тезис 10. Пусть дано аргдерево $T(p_i)$, θ_1, …, θ_s – все его ветви, а $\mathrm{Set}^*(\theta_j)$, где $j = 1, \ldots, s$, – соответствующие им множества вершин типа $\bullet x$ (значениями x являются p_j, где $p_j \in \mathcal{P}$). Обозначим посредством $\mathrm{Set}^*(J\theta_1)$, …, $\mathrm{Set}^*(J\theta_s)$ множества вершин $JT(p_i)$ типа $\bullet x$, где $J_\nu p_j$ являются значениями x, а $\nu \in \{1, -1, 0, \tau\}$. Пусть $\mathrm{Set}^*(J\theta_k) = \{ J_{\nu_{i_1}} p_{i_1 k}, \ldots, J_{\nu_{i_l}} p_{i_l k} \}$, где i_j – номер значения x вершины типа $\bullet x$, $j = 1, \ldots, l$, а k – номер ветви, $k = 1, \ldots, s$. Пусть далее $J_\mu p_h$ – помеченная корневая вершина $JT(p_h)$. Будем говорить, что информативная ветвь θ_k аргдерева $T(p_h)$ семантически корректна, если и только если $\nu[\&_l^{(5)}(p_{i_1 k}, \ldots, p_{i_l k})] = \mu$, где $\mu \in \{1, -1, 0, \tau\}$, а $\nu[p_{i_j k}] = \nu_{i_j}$, $j = 1, \ldots, l$. Аргдерево $T(p_h)$ будем называть **семантически корректным**, если и только если все его информативные ветви семантически корректны.

Для представления семантической корректности аргдеревьев введём метапредикат $!_3$. Утверждения о семантической корректности и некорректности $T(p_h)$ будем обозначать посредством $!_3T(p_h)$ и $\neg!_3T(p_h)$, соответственно. Рассмотрим $JT(p_1)$ для аргдерева $T(p_1)$ (Рис. 4). В нём имеются две информативные ветви – θ_1 и θ_4: $\mathrm{Set}^*(\theta_1) = \{J_1p_1, J_1p_2, J_1p_3\}$, $\mathrm{Set}^*(\theta_4) = \{J_1p_1, J_1p_3\}$. Так как для $J\theta_1$ и $J\theta_4$ в $JT(p_1)$ имеют место $\&_3^{(5)}(1, 1, 1) = 1$ и $\&_2^{(5)}(1, 1) = 1$, а J_1p_1 – корень, то $!_3T(p_1)$.

Для $T(p_1)$ истинно утверждение $!_1T(p_1)\&!_2T(p_1)\&!_3T(p_1)$ о том, что $T(p_1)$ – конечно, его описание $\mathrm{Des}(T(p_1))$ – непротиворечиво, и $T(p_1)$ – семантически корректно.

Конъюнкцию $(!_1T(p_h))^{\sigma_1} \&(!_2T(p_h))^{\sigma_2} \&(!_3T(p_h))^{\sigma_3}$, где σ_i есть либо 0, либо 1 (i =1, 2, 3), будем называть **характеризацией** аргдерева $T(p_h)$. Здесь $(!_iT(p_h))^{\sigma_i} = !_iT(p_h)$, если σ_i =1; $(!_iT(p_h))^{\sigma_i} = \neg!_iT(p_h)$, если σ_i =0. Приведём все формально возможные характеризации аргдеревьев $T(p_h)$:

Таблица 2

$!_1T(p_h)$	$!_2T(p_h)$	$!_3T(p_h)$
+	+	+
+	+	–
+	–	+
+	–	–
–	+	+
–	+	–
–	–	+
–	–	–

+ и – обозначают истинность и ложность $!_iT(p_h)$, соответственно. Лес Wd, состоящий из n аргдеревьев $T(p_h)$, где $h = 1, \ldots, n$, имеет 2^{3n} характеризаций.

Очевидно, что возможными характеризациями зарослей Br(Wd) являются конъюнкции $(!Br(Wd))^{\sigma}\&(!_1T(p_h))^{\sigma_1^{(i)}}\&(!_2T(p_h))^{\sigma_2^{(i)}}\&(!_3T(p_h))^{\sigma_3^{(i)}}$, $\sigma = 0, 1$; $\sigma_j^{(i)} = 0, 1$; $j = 1, 2, 3$, $p_h \in U$, $h = 1, \ldots, n = |U|$

Введём теперь основное определение аргументационной системы Argsys($\mathcal{P}$). **Аргументационной системой** будем называть $\mathrm{Argsys}(\mathcal{P}) = \langle \bar{M}, A_{4,1}^{(5)}, U, \{T(p_i)\}_{p_i \in U}, \Sigma, !, !_1, !_2, !_3\rangle$, где $U \subseteq \mathcal{P}^a \cup \mathcal{P}^b$, $\Sigma = \bigcup_{p_i \in U} \mathrm{Des}(T(p_i))$, а $!, !_1, !_2, !_3$ – ранее определённые метапредикаты, используемые в утверждениях $!_kT(p_i)$, $!Br(Wd)$, $k = 1,2, 3$. Лес Wd и заросли Br(Wd) являются графическими представлениями Argsys($\mathcal{P}$).

Тезис 11. Важным обстоятельством является **сохранение структур** $T(p_i)$ и Br(Wd) при **обеднении** множества истинностных значений $\{1, -1, 0, \tau\}$ логики $A_{4,1}^{(5)}$. Оно может быть получено исключением «τ» для трёхзначной логики $A_{3,1}^{(6)}$ и последующим исключением «0» для двузначной логики $A_{2,1}^{(7)}$. Отметим, что при этих обеднениях $A_{4,1}^{(5)}$ сохраняется принцип аргументационной семантики с функциями выбора g^+ и g^-, заданных внутренним образом на организованном множестве высказываний $\mathcal{P}$.

Для $A_{3,1}^{(6)}$ с $M = \langle\{1, -1, 0\}, \{1\}, \sim, \supset, \vee^{(6)}, \{\&_n^{(6)}\}_{n\in N}\rangle$ сохраняется неассоциативность $\&_2^{(6)}$, так как $(1\&_2^{(6)}-1)\&_2^{(6)}-1 = 0\&_2^{(6)}-1 = -1$, а $1\&_2^{(6)}(-1\&_2^{(6)}-1) = 1\&_2^{(6)}-1 = 0$, где

p	~p
1	–1
–1	1
0	0

$\supset$	1	–1	0
1	1	–1	0
–1	1	1	1
0	1	–1	1

$\vee^{(6)}$	1	–1	0
1	1	1	1
–1	1	–1	–1
0	1	–1	0

$\&_2^{(6)}$	1	–1	0
1	1	0	0
–1	0	–1	–1
0	0	–1	0

v[p] = 1, если $g^+(p) \neq \varnothing$, $g^-(p) = \varnothing$;

v[p] = –1, если $g^+(p) = \varnothing$, $g^-(p) \neq \varnothing$;

v[p] = 0, если $(g^+(p) \neq \varnothing \& g^-(p) \neq \varnothing)$ или $(g^+(p) = \varnothing \& g^-(p) = \varnothing)$.

Для $A_{3,1}^{(6)}$ формулируется метод аналитических таблиц, $\bar{M}$, $T(p_i)$, Wd, Br(Wd) и Argsys($\mathcal{P}$).

Аналогично получаем двузначную логику с аргументационной семантикой такую, что v[p] = 1, если $g^+(p) \neq \varnothing$, $g^-(p) = \varnothing$; v[p] = –1, если $(g^+(p) = \varnothing \& g^-(p) \neq \varnothing)$ или $(g^+(p) \neq \varnothing \& g^-(p) \neq \varnothing)$ или $(g^+(p) = \varnothing \& g^-(p) = \varnothing)$. Однако для двузначной логики с аргументационной семантикой & получается не обеднением $\&_2^{(6)}$, а заданием

&	1	–1
1	1	–1
–1	–1	–1

Соответственно, определяется Argsys($\mathcal{P}$) для двузначной логики $A_{2,1}^{(7)}$.

Тезис 12. Сформулируем в заключение возможные сферы применения Argsys($\mathcal{P}$) и некоторые проблемы их развития.

1°. В **Тезисе 1** было отмечено, что возможно построить способ организации знаний посредством логик аргументации и их семантики. Этот подход предполагает задание множества высказываний $\mathcal{P} = \mathcal{P}^a \cup \mathcal{P}^b \cup \mathcal{P}'$ с соответствующими типами аргументации.

Возможными областями для подобной формализации организации знаний являются так называемые «науки о культуре» (точнее, о человеке и обществе; прежде всего – социология, история, юриспруденция, антропология, филология, а также медицина, в том числе, психиатрия).

Для систем ИИ этот подход может быть полезен при разработке систем представления знаний в интеллектуальных системах.

2°. Для филологии и истории Argsys($\mathcal{P}$) являются формальным аппаратом **герменевтики** [8], значение которой для ИИ неоднократно отмечается в работах Д.А. Поспелова. Важной идеей в Argsys($\mathcal{P}$), получившей точное выражение, является характеризация герменевтического («порочного») круга посредством $\neg!_1(T(p_i))$. Это означает, что надо стремиться при задании $\mathcal{P}$ к непустоте базисного множества $\mathcal{P}'$.

3°. Для социологии использование Argsys($\mathcal{P}$) означает применение **глубоких** (многоуровневых) средств аргументации при выборе респондентами решений, что означает усиление имитации рациональности мнений (ранее была рассмотрена одноуровневая система аргументации: у аргументов не было аргументов [10]). Для Argsys($\mathcal{P}$) с большим объёмом $\mathcal{P}$ может при реализации в компьютерных системах потребоваться применение суперкомпьютеров и использование параллелизма в вычислениях.

4°. Argsys($\mathcal{P}$) можно истолковать как **семиотическую систему**, реализующую процесс семиозиса в смысле Ч.С. Пирса (C.S. Peirce). Это соображение обусловлено тем, что **синтаксическая** характеризация $\mathcal{P}$ представима посредством утверждений $!_2(T(p_i))$ или $\neg!_2(T(p_i))$, **семантическая** же характеризация представима $!_3(T(p_i))$ или $\neg!_3(T(p_i))$. И, наконец, **прагматическая** характеризация представима посредством утверждений $!_1(T(p_i))$ или $\neg!_1(T(p_i))$, а также $!Br(Wd)$ или $\neg!Br(Wd)$. Последняя характеризация

отображает сложность восприятия массива знаний посредством аргументации интерпретатора.

5°. Первой серьёзной проблемой развития Argsys($\mathcal{P}$) является использование для представления $\mathcal{P}$ языка логики предикатов 1-го порядка без индивидных переменных и кванторов. Это усиление выразительных средств Argsys($\mathcal{P}$) создаёт условия применения к $\mathcal{P}$ правдоподобных рассуждений, включающих индукцию, что создаёт новые когнитивные возможности формирования аргументации, включающей порождённые **гипотезы**.

6°. Ещё одним значительным усилением выразительной силы языка Argsys($\mathcal{P}$) является использование языка логики предикатов 1-го порядка. Однако это требует специального рассмотрения.

СПИСОК ЛИТЕРАТУРЫ

1. Бентем Й. ван. Логика и рассуждение: много ли значат факты? // Вопросы философии. – 2011. - №12. - С. 63-76.
2. Logic and Argumentation / ed. by J. van Benthem, F.H. van Eemeren, R. Grootendorst, F. Veltman. – Amsterdam: North-Holland, 1996.
3. Вагин В.Н., Головина Е.Ю., Загорянская А.А., Фомина М.В. Достоверный и правдоподобный вывод в интеллектуальных системах. – М. : Физматлит, 2008.
4. Финн В.К. Стандартные и нестандартные логики аргументации // в кн.: Финн В.К. Искусственный интеллект: методология применения, философия. - М.: КРАСАНД, 2011. - С. 312 – 338.
5. Финн В.К. Об одном варианте логики аргументации // НТИ. Сер. 2. – 1996. - № 5- 6. - С. 3 – 19.
6. Hao Wang. Logic, computers, and sets // Chapter XII. Many-sorted Predicate Calculi. - NY: Chelsea Publishing Company, 1970. - P. 322 – 333.
7. König D. Sur les correspondences multivoques des ensembles // Fundamenta math. -1926. - No 8. - P. 114 – 134.
8. Гадамер Х.-Г. Истина и метод. – М.: Прогресс, 1988.
9. Харари Ф. Теория графов. – М.: Мир, 1973.
10. Finn V.K., Mikheyenkova M.A. Plausible Reasoning for the Problems of Cognitive Sociology // Logic and Logical Philosophy. – 2011. - Vol. 20. - P. 113 – 139.

ПРИЛОЖЕНИЕ

Логика аргументации $A_{4,1}^{(7)}$

Рассмотрим четырёхзначную логику аргументации с обобщёнными законами Де Моргана для счётного множества конъюнкций $\{\&_n^{(5)}\}_{n\in N}$ и счётного множества дизъюнкций $\{\vee_n^{(7)}\}_{n\in N}$ таких, что для любого n ($n\geq 2$) имеют место тавтологии $\sim\&_n^{(5)}(p_1, \ldots, p_n) \equiv \vee_n^{(7)}(\sim p_1, \ldots, \sim p_n)$, $\sim\vee_n^{(7)}(p_1, \ldots, p_n) \equiv \&_n^{(5)}(\sim p_1, \ldots, \sim p_n)$, где $(p\equiv q) \rightleftharpoons \&_4^{(5)}((p\supset q), (\sim p\supset\sim q), (q\supset p), (\sim q\supset\sim p))$, $(p\equiv q) = \begin{cases}1, \text{если } v[p]=v[q]\\ 0, \text{если } v[p]\neq v[q]\end{cases}$, $\vee_2^{(7)}$ – неассоциативная логическая связка, так как $\vee_2^{(7)}(1, \vee_2^{(7)}(-1, \tau)) = \vee_2^{(7)}(1, \tau) = 1$, $\vee_2^{(7)}(\vee_2^{(7)}(1, -1), \tau)) = \vee_2^{(7)}(0, \tau) = 0$.

$\vee_2^{(7)}$	1	–1	0	τ
1	1	0	1	1
–1	0	–1	0	τ
0	1	0	0	0
τ	1	τ	0	τ

Логическая матрица $A_{4,1}^{(7)}$ есть $M = \langle\{1, -1, 0, \tau\}, \{1\}, \sim, \supset, \{\vee_n^{(7)}\}_{n\in N}, \{\&_n^{(5)}\}_{n\in N}\rangle$, где $\sim$ и $\supset$ – логические связки $A_{4,1}^{(5)}$ и $A_{4,1}^{(4)}$ ($i = 1, 2$)

Определим функцию оценки $v[\varphi]$ для формул $\vee_n^{(7)}(\varphi_1, \ldots, \varphi_n)$ логики $A_{4,1}^{(7)}$. Ранее она была определена для $\sim\varphi$, $\varphi\supset\psi$ и $\&_n^{(5)}(\varphi_1, \ldots, \varphi_n)$, где $n\in N$.

1°. $v[\vee_n^{(7)}(\varphi_1, \ldots, \varphi_n)] = -1$, если $\forall i(v[\varphi_i] = -1)$;

2°. $v[\vee_n^{(7)}(\varphi_1, \ldots, \varphi_n)] = \tau$, если $\exists i(v[\varphi_i] = \tau)\&\forall j(v[\varphi_j]\subseteq\{-1, \tau\})$;

3°. $v[\vee_n^{(7)}(\varphi_1, \ldots, \varphi_n)] = 1$, если $\exists i(v[\varphi_i] = 1)\&\forall j(v[\varphi_j]\subseteq\{1, 0, \tau\})$;

4°. $v[\vee_n^{(7)}(\varphi_1, \ldots, \varphi_n)] = 0$, если $\exists i\exists j(v[\varphi_i] = 1\&v[\varphi_j]=-1)\vee(\exists k(v[\varphi_k] = 0)\&\forall m(v[\varphi_m]\subseteq\{-1, 0, \tau\}))$.

Для 4° эквивалентным условием является $\exists i\exists j((v[\varphi_i] = 1\&v[\varphi_j]=-1)\vee\exists k(v[\varphi_k] = 0)\&\forall m(v[\varphi_m]\neq 1))$.

Из определения $\vee_n^{(7)}$ следует, что $\vee_n^{(7)}$ двойственна $\&_n^{(5)}$ для $n\in N$, а $\sim 1 = -1$, $\sim(-1) = 1$, $\sim 0 = 0$, $\sim\tau = \tau$. Поэтому число разветвлений в правилах вывода аналитических таблиц (то есть вхождений $\nu = 1, -1, 0, \tau$ в истинностных таблицах) $b^{-1}_{\vee_n^{(7)}}(n) = b^{1}_{\&_n^{(5)}}(n) = 1^n - 0^n$, $b^{\tau}_{\vee_n^{(7)}}(n) = b^{\tau}_{\&_n^{(5)}}(n) = 2^n - 1^n$, $b^{1}_{\vee_n^{(7)}}(n) = b^{-1}_{\&_n^{(5)}}(n) = 3^n - 2^n$, $b^{0}_{\vee_n^{(7)}}(n) = b^{0}_{\&_n^{(5)}}(n) = 4^n - 3^n$.

Используя $b^{\nu}_{\sim}$, $b^{\nu}_{\supset}$, $b^{\nu}_{\&_n^{(5)}}(n)$, $b^{\nu}_{\vee_n^{(7)}}(n)$, где $\nu = 1, -1, 0, \tau$, определим правила вывода для аналитических таблиц логики $A_{4,1}^{(7)}$ с неассоциативными $\&_2^{(5)}$ и $\vee_2^{(7)}$.

Посредством φ, ψ (быть может, с нижними индексами) обозначим произвольные формулы $A_{4,1}^{(7)}$.

$$\frac{J_1(\sim\phi)}{J_1\phi}, \frac{J_{-1}(\sim\phi)}{J_1\phi}, \frac{J_0(\sim\phi)}{J_0\phi}, \frac{J_\tau(\sim\phi)}{J_\tau\phi}.$$

$$\frac{J_1(\phi\supset\psi)}{J_{-1}\phi \mid J_1\psi \mid J_0\phi, J_0\psi \mid J_\tau\phi, J_\tau\psi},$$

$$\frac{J_{-1}(\phi\supset\psi)}{J_1\phi, J_{-1}\psi \mid J_0\phi, J_{-1}\psi \mid J_\tau\phi, J_{-1}\psi},$$

$$\frac{J_0(\phi\supset\psi)}{J_1\phi, J_0\psi \mid J_\tau\phi, J_0\psi}, \frac{J_\tau(\phi\supset\psi)}{J_1\phi, J_\tau\psi \mid J_0\phi, J_\tau\psi}.$$

$$\frac{J_1(\&_n^{(5)}(\phi_1,\ldots,\phi_n))}{\alpha_{11},\ldots,\alpha_{b^1_{\&_n^{(5)}}(n)n}},$$

где $\alpha_{ij} = J_1\varphi_j$, $b^1_{\&_n^{(5)}}(n) = 1$, $i = 1, j = 1, \ldots, n$.

Таким образом, $\dfrac{J_1(\&_n^{(5)}(\phi_1,...,\phi_n))}{J_1\phi_1,...,J_1\phi_n}$.

$\dfrac{J_\tau(\&_n^{(5)}(\phi_1,...,\phi_n))}{\beta_{11},...,\beta_{1n}\,|...|\,\beta_{b^\tau_{\&_n^{(5)}}(n)1},...,\beta_{b^\tau_{\&_n^{(5)}}(n)n}}$, где $b^\tau_{\&_n^{(5)}}(n) = 2^n - 1^n$, а β_{ij} соответствуют истинностным значениям φ_j, где $j = 1, \ldots, n$ для $v[\&_n^{(5)}(\varphi_1, \ldots, \varphi_j, \ldots, \varphi_n)] = \tau$ с $v[\varphi_j]$, $j = 1, \ldots, n$; $i = 1, \ldots, b^\tau_{\&_n^{(5)}}(n)$.

$\dfrac{J_{-1}(\&_n^{(5)}(\phi_1,...,\phi_n))}{\gamma_{11},...,\gamma_{1n}\,|...|\,\gamma_{b^{-1}_{\&_n^{(5)}}(n)1},...,\gamma_{b^{-1}_{\&_n^{(5)}}(n)n}}$, где $b^{-1}_{\&_n^{(5)}}(n) = 3^n - 2^n$, а γ_{ij} соответствуют истинностным значениям φ_j, где $j = 1, \ldots, n$ для $v[\&_n^{(5)}(\varphi_1, \ldots, \varphi_j, \ldots, \varphi_n)] = -1$ с $v[\varphi_j]$, $j = 1, \ldots, n$; $i = 1, \ldots, b^{-1}_{\&_n^{(5)}}(n)$.

$\dfrac{J_0(\&_n^{(5)}(\phi_1,...,\phi_n))}{\delta_{11},...,\delta_{1n}\,|...|\,\delta_{b^0_{\&_n^{(5)}}(n)1},...,\delta_{b^0_{\&_n^{(5)}}(n)n}}$, где $b^0_{\&_n^{(5)}}(n) = 4^n - 3^n$, а δ_{ij} определяется аналогично β_{ij} и γ_{ij}.

Таким образом, каждому блоку (элементу разветвления в правилах вывода) взаимно однозначно соответствуют векторы $\vec{\nu}^{(i)} = \langle \nu_{i1}, \ldots, \nu_{in} \rangle$, где $i = 1, \ldots, 4^n$, а $\nu_{ij} \in \{1, -1, 0, \tau\}$, $j = 1, \ldots, n$.

Рассмотрим пример правил вывода для $n = 2$, тогда $b^{-1}_{\vee_2^{(7)}}(2) = b^1_{\&_2^{(5)}}(2) = 1^2 - 0^2 = 1$, $b^\tau_{\vee_2^{(7)}}(2) = b^\tau_{\&_2^{(5)}}(2) = 2^2 - 1^2 = 3$, $b^1_{\vee_2^{(7)}}(2) = b^{-1}_{\&_2^{(5)}}(2) = 3^2 - 2^2 = 5$,

$b^0_{\vee_2^{(7)}}(2) = b^0_{\&_2^{(5)}}(2) = 4^2 - 3^2 = 7$.

$$\frac{J_1(\&_2^{(5)}(\phi_1,\phi_2))}{J_1\phi_1, J_1\phi_2};$$

$$\frac{J_{-1}(\&_2^{(5)}(\phi_1,\phi_2))}{J_{-1}\phi_1, J_{-1}\phi_2 \,|\, J_{-1}\phi_1, J_0\phi_2 \,|\, J_0\phi_1, J_{-1}\phi_2 \,|\, J_{-1}\phi_1, J_\tau\phi_2 \,|\, J_\tau\phi_1, J_{-1}\phi_2};$$

$$\frac{J_0(\&_2^{(5)}(\phi_1,\phi_2))}{J_1\phi_1, J_{-1}\phi_2 \,|\, J_{-1}\phi_1, J_1\phi_2 \,|\, J_1\phi_1, J_0\phi_2 \,|\, J_0\phi_1, J_1\phi_2 \,|\, J_0\phi_1, J_0\phi_2 \,|\, J_0\phi_1, J_\tau\phi_2 \,|\, J_\tau\phi_1, J_0\phi_2};$$

$$\frac{J_\tau(\&_2^{(5)}(\phi_1,\phi_2))}{J_1\phi_1, J_\tau\phi_2 \,|\, J_\tau\phi_1, J_1\phi_2 \,|\, J_\tau\phi_1, J_\tau\phi_2}.$$

$$\frac{J_{-1}(\vee_2^{(7)}(\phi_1,\phi_2))}{J_{-1}\phi_1, J_{-1}\phi_2};$$

$$\frac{J_1(\vee_2^{(7)}(\phi_1,\phi_2))}{J_1\phi_1, J_1\phi_2 \,|\, J_1\phi_1, J_0\phi_2 \,|\, J_0\phi_1, J_1\phi_2 \,|\, J_1\phi_1, J_\tau\phi_2 \,|\, J_\tau\phi_1, J_1\phi_2};$$

$$\frac{J_0(\vee_2^{(7)}(\phi_1,\phi_2))}{J_1\phi_1, J_{-1}\phi_2 \,|\, J_{-1}\phi_1, J_1\phi_2 \,|\, J_1\phi_1, J_0\phi_2 \,|\, J_0\phi_1, J_1\phi_2 \,|\, J_0\phi_1, J_0\phi_2 \,|\, J_0\phi_1, J_\tau\phi_2 \,|\, J_\tau\phi_1, J_0\phi_2};$$

$$\frac{J_\tau(\vee_2^{(7)}(\phi_1,\phi_2))}{J_1\phi_1, J_\tau\phi_2 \,|\, J_\tau\phi_1, J_1\phi_2 \,|\, J_\tau\phi_1, J_\tau\phi_2}.$$

Аналитические таблицы для $A^{(7)}_{4,1}$ определяются стандартно посредством применения правил вывода к корню дерева вывода и к последующим вершинам. Контрарными парами являются помеченные формулы $J_\nu\varphi$, $J_\mu\psi$, где $\nu \neq \mu$. Ветвь аналитической таблицы θ называется замкнутой, если $\mathrm{Set}(\theta)$ содержит контрарную пару. Аналитическая таблица называется замкнутой, если все её ветви замкнуты.

Формула φ доказуема в $A^{(7)}_{4,1}$, если аналитические таблицы $\Im_{J_{-1}\phi}$, $\Im_{J_0\phi}$ и $\Im_{J_\tau\phi}$ замкнуты, где $J_{-1}\varphi$, $J_0\varphi$, $J_\tau\varphi$ – корни соответствующих деревьев. Если φ доказуема, то будем писать $\vdash\varphi$.

φ называется тавтологией $A^{(7)}_{4,1}$, если $\forall v(v[\varphi] = 1)$. $\models\varphi$ означает, что φ – тавтология.

Для $A^{(7)}_{4,1}$ имеют место теоремы о корректности и полноте, соответственно: если $\vdash\varphi$, то $\models\varphi$; если $\models\varphi$, то $\vdash\varphi$. Доказательство этих утверждений аналогичны соответствующим теоремам в [2] для $A^{(4)}_{4,1}$.

Рассмотрим классификацию помеченных формул $A^{(7)}_{4,1}$.

α-формулы: $J_\nu(\sim\varphi)$, где $\nu = 1, -1, 0, \tau$;

α-формулы типа $\alpha(n)$: $J_1(\&_n^{(5)}(\varphi_1, \ldots, \varphi_n))$, $J_{-1}(\vee_n^{(7)}(\varphi_1, \ldots, \varphi_n))$.

$\beta(n)$-формулы: $J_\tau(\&_n^{(5)}(\varphi_1, \ldots, \varphi_n))$, $J_\tau(\vee_n^{(7)}(\varphi_1, \ldots, \varphi_n))$.

$\gamma(n)$-формулы: $J_{-1}(\&_n^{(5)}(\varphi_1, \ldots, \varphi_n))$, $J_1(\vee_n^{(7)}(\varphi_1, \ldots, \varphi_n))$.

$\delta(n)$-формулы: $J_0(\&_n^{(5)}(\varphi_1, \ldots, \varphi_n))$, $J_0(\vee_n^{(7)}(\varphi_1, \ldots, \varphi_n))$.

ε-формулы: $J_\tau(\varphi\supset\psi)$, $J_0(\varphi\supset\psi)$.

μ-формулы: $J_1(\varphi\supset\psi)$.

η-формулы: $J_{-1}(\varphi\supset\psi)$.

Обозначим числа разветвлений для этих типов формул посредством b^α, $b^\beta(n)$, $b^\gamma(n)$, $b^\delta(n)$, b^ε, b^μ, b^η, где $b^\alpha = 1^n - 0^n$, $b^\beta(n) = 2^n - 1^n$, $b^\gamma(n) = 3^n - 2^n$, $b^\delta(n) = 4^n - 3^n$, $b^\varepsilon = 2$, $b^\mu = 4$, $b^\eta = 3$.

Для элементарных формул $J_\nu p$ и формул указанных выше типов определяются множества Хинтикки аналогично [4, 5] и доказывается лемма Хинтикки, посредством которой доказывается теорема о полноте.

Обеднением истинностных таблиц логики $A^{(7)}_{4,1}$ посредством устранения τ получим истинностные таблицы трёхзначной логики аргументации $A^{(8)}_{3,1}$:

p	~
1	–1
–1	1
0	0

$\supset$	1	–1	0
1	1	–1	0
–1	1	1	1
0	1	–1	1

$\vee_2^{(8)}$	1	–1	0
1	1	0	1
–1	0	–1	0
0	1	0	0

$\&_2^{(6)}$	1	–1	0
1	1	0	0
–1	0	–1	–1
0	0	–1	0

$\vee_2^{(8)}$ и $\&_2^{(6)}$ – неассоциативные логические связки, так как $\vee_2^{(8)}(1, \vee_2^{(8)}(-1, -1)) = \vee_2^{(8)}(1, -1) = 0$, но $\vee_2^{(8)}(\vee_2^{(8)}(1, -1), -1)) = \vee_2^{(8)}(0, -1) = -1$; аналогичное утверждение имеет место и для $\&_2^{(6)}$.

Легко можно получить правила вывода для аналитических таблиц $A^{(8)}_{3,1}$, вычёркивая из истинностных таблиц $A^{(7)}_{4,1}$ блоки ветвлений, содержащие вхождения τ. Определимы также числа ветвлений этих пра-

вил (т. е. число блоков) $b^{\nu}_{\sim} = 1$, где $\nu = 1, -1, 0$; $b^{1}_{\supset} = 3$, $b^{-1}_{\supset} = 2$, $b^{0}_{\supset} = 1$, $b^{-1}_{\vee} = 1^n - 0^n$, $b^{1}_{\vee} = 2^n - 1^n$, $b^{0}_{\vee} = 3^n - 2^n$, $b^{1}_{\&} = 1^n - 0^n$, $b^{-1}_{\&} = 2^n - 1^n$, $b^{0}_{\&} = 3^n - 2^n$.

Правила вывода для $n = 2$.

$$\alpha: \frac{J_1(\sim\phi)}{J_1\phi}, \frac{J_{-1}(\sim\phi)}{J_1\phi}, \frac{J_0(\sim\phi)}{J_0\phi},$$

$$\frac{J_0(\phi\supset\psi)}{J_1\phi, J_0\psi}, \frac{J_{-1}(\vee_2^{(8)}(\phi_1,\phi_2))}{J_{-1}\phi_1, J_{-1}\phi_2}, \frac{J_1(\&_2^{(6)}(\phi_1,\phi_2))}{J_1\phi_1, J_1\phi_2}.$$

$$\gamma(n): \frac{J_{-1}(\&_2^{(6)}(\phi_1,\phi_2))}{J_{-1}\phi_1, J_{-1}\phi_2 \mid J_{-1}\phi_1, J_0\phi_2 \mid J_0\phi_1, J_{-1}\phi_2},$$

$$\frac{J_1(\vee_2^{(8)}(\phi_1,\phi_2))}{J_1\phi_1, J_1\phi_2 \mid J_1\phi_1, J_0\phi_2 \mid J_0\phi_1, J_1\phi_2}.$$

$$\delta(n): \frac{J_0(\&_2^{(6)}(\phi_1,\phi_2))}{J_1\phi_1, J_{-1}\phi_2 \mid J_{-1}\phi_1, J_1\phi_2 \mid J_1\phi_1, J_0\phi_2 \mid J_0\phi_1, J_1\phi_2 \mid J_0\phi_1, J_0\phi_2}.$$

$$\mu: \frac{J_1(\phi\supset\psi)}{J_{-1}\phi \mid J_1\psi \mid J_0\phi, J_0\psi}.$$

$$\eta: \frac{J_{-1}(\phi\supset\psi)}{J_1\phi, J_{-1}\psi \mid J_0\phi, J_{-1}\psi}.$$

$\vdash\varphi$, если аналитические таблицы $\Im_{J_{-1}\phi}$ и $\Im_{J_0\phi}$ являются замкнутыми.

Формулы χ_1 и χ_2, где $\chi_1 = (\sim\vee_2^{(8)}(\varphi, \psi)\supset\&_2^{(6)}(\sim\varphi, \sim\psi))$ и $\chi_2 = (\sim\&_2^{(6)}(\varphi, \psi)\supset\vee_2^{(8)}(\sim\varphi, \sim\psi))$, доказуемы в $A^{(7)}_{3,1}$, так как аналитические таблицы $\Im_{J_{-1}\chi_1}$, $\Im_{J_0\chi_1}$ и $\Im_{J_{-1}\chi_2}$, $\Im_{J_0\chi_2}$ являются замкнутыми.

Материал поступил в редакцию 21.09.12.

Сведения об авторе

ФИНН Виктор Константинович – доктор технических наук, профессор, заслуженный деятель науки Российской Федерации, зав. сектором интеллектуальных информационных систем ВИНИТИ РАН, зав. отделением интеллектуальных систем в гуманитарной сфере РГГУ, Москва.
E-mail: finn@viniti.ru

АВТОМАТИЗАЦИЯ ОБРАБОТКИ ТЕКСТА

УДК 004.912

Г.Г. Белоногов, Р.С. Гиляревский, А.А. Хорошилов

Проблемы автоматической смысловой обработки текстовой информации

Рассматриваются основные проблемы автоматической смысловой обработки текстовой информации. Утверждается, что главным условием успеха при решении этих проблем являются правильные представления о семантико-синтаксической структуре текстов и опора на исторически сложившееся фразеологическое богатство естественных языков.

Ключевые слова: *поиск информации, смысловая обработка текстов, автоматическое индексирование, автоматическое реферирование, машинный перевод, автоматизация составления словарей*

ЕДИНИЦЫ СМЫСЛА В ЯЗЫКЕ И РЕЧИ

Современный этап человеческой цивилизации характеризуется широким применением автоматизированных информационных технологий. Однако развитие этих технологий происходит весьма неравномерно: если современный уровень электронной вычислительной техники и средств связи поражает воображение, то в области смысловой обработки информации успехи значительно скромнее. Эти успехи зависят, прежде всего, от достижений в изучении человеческого мышления, процессов речевого общения между людьми и от умения моделировать эти процессы на ЭВМ. А это – задача чрезвычайной сложности.

Успешному развитию методов смысловой обработки текстовой информации мешают также имеющие хождение неправильные представления о природе информации и о смысловой структуре текстов. Так, в настоящее время широко распространена точка зрения, что информацию следует рассматривать как некоторую характеристику *степени неопределенности* или *степени разнообразия объектов*. А некоторые философы и их последователи пытаются утверждать, что информация является атрибутом материи и изначально присуща всем объектам живой и неживой природы. При этом **главное свойство информации – быть носителем смысла – игнорируется** [1-3].

До сих пор бытует господствовавшая веками неправильная точка зрения на средства обозначения наименований понятий в текстах: считается, что основным средством их обозначения являются отдельные слова, а не устойчивые фразеологические словосочетания. Такой взгляд на смысловую структуру текстов в прошлом веке задержал развитие систем машинного перевода текстов с одних языков на другие как минимум на несколько десятилетий. И даже в настоящее время некоторые разработчики систем машинного перевода продолжают еще считать, что при переводе текстов следует ориентироваться на “значения слов”.

Та истина, что устойчивые фразеологические словосочетания являются основным средством выражения наименований понятий в текстах и что они используются в такой роли в сотни раз чаще, чем отдельные слова, была впервые установлена в 27 ЦНИИ МО СССР и в ВИНИТИ РАН в процессе масштабных статистических исследований современных русских и английских текстов, проводившихся в течение ряда десятилетий. Она была многократно подтверждена в процессе разработки и эксплуатации ряда систем автоматической обработки текстовой информации – систем автоматического индексирования документов, их автоматической классификации и поиска, систем автоматического перевода текстов с одних естественных языков на другие.

Как известно, естественный язык является универсальным средством общения между людьми – средством восприятия, накопления, хранения и передачи информации. Более того, он является инструментом мышления человека [2, 4-6]. Психологи считают, что естественный язык представляет собой вторую сигнальную систему человека, функционирующую на основе первой сигнальной системы, которая, в свою очередь, работает как система врожденных безусловных рефлексов, возникающих под воздействием сигналов, получаемых от зрительных, слуховых, тактильных и других рецепторов [5, 7]. Языковые сигналы лишь инициируют мыслительные процессы, происходящие во внутреннем духовном мире человека, в его “душе”, но не определяют их полностью. По мне-

нию психологов, интерпретация речевых сигналов человеком (их понимание) происходит с учетом опыта, накопленного им в течение всей его жизни.

В *лингвистике* язык рассматривается как некоторая *знаковая система* [8]. По мнению Ф. де Соссюра – одного из создателей современной науки лингвистики и науки семиотики - *языковые знаки состоят из двух компонент: из означающего и означаемого.* ***Означающее*** *- это звуковой или графический образ знака, а* ***означаемое*** *– соответствующее ему* ***понятие****.*

Понятие - это социально значимый мыслительный образ, за которым в языке закреплено его наименование в виде отдельного слова или, значительно чаще, в виде устойчивого фразеологического словосочетания. Под устойчивыми фразеологическими словосочетаниями мы будем понимать не только идиоматические выражения и терминологические словосочетания, но и любые повторяющиеся отрезки связных текстов длиной от двух до десятипятнадцати слов (более длинные устойчивые словосочетания встречаются редко).

В развитых языках мира (русском, английском, немецком, французском и др.) *количество различных наименований понятий достигает нескольких сотен миллионов. Большинство из них обозначаются словосочетаниями, смысл которых не сводим к смыслу составляющих их слов. Слова, входящие в состав словосочетаний, обозначают лишь некоторые признаки понятий, позволяющие отличать их друг от друга, но не исчерпывающие их содержания.* Содержание понятий в полном объеме интерпретируется только в "душе" человека - в его внутреннем мире, где "все связано со всем" [1, 5].

При создании систем автоматической обработки текстовой информации очень важно исходить из правильных представлений о смысловой структуре языка и речи. По современным представлениям *наиболее устойчивыми единицами смысла являются* ***понятия*** [1, 8]. Они занимают центральное место в языке и речи и являются теми базовыми строительными блоками, на основе которых формируются смысловые единицы более высоких уровней. *Второй по значимости единицей смысла является* ***предложение****. Из предложений формируются различного рода* ***сверхфразовые единства***, которые представляются в виде последовательностей предложений (связного текста).

Основной чертой предложений является их ***предикативность*** – т. е. то их свойство, что в них утверждается наличие у объектов определенных признаков и их отношений [1, 9]. Свойством предикативности обладают и высказывания, формулируемые на формализованных языках. Это позволяет сделать вывод, что *в основе и предложений на естественном языке, и формализованных логических высказываний лежит* ***предикатно-актантная структура****, компонентами которой являются понятия-предикаты (признаки, отношения) и понятия-актанты, выступающие в роли описываемых объектов.*

В естественных и в формализованных языках предикатно-актантные структуры являются теми смысловыми инвариантами, которые позволяют осуществлять автоматический перевод текстов с естественных языков на формализованные и с формализованных на естественные. Они также позволяют осуществлять автоматический перевод текстов с одних естественных языков на другие.

СМЫСЛОВАЯ ОБРАБОТКА ТЕКСТОВОЙ ИНФОРМАЦИИ

Смысл текстовых документов выражается с помощью единиц смысла, входящих в их состав. Как уже было указано, в естественных языках базовыми единицами смысла являются понятия. Поэтому центральной процедурой любых систем автоматической смысловой обработки текстов должна быть процедура их семантико-синтаксического концептуального анализа. По нашему мнению, она должна быть реализована как процедура фразеологического концептуального анализа на основе мощных словарей наименований понятий. При этом следует опираться на адекватные семантико-синтаксические модели текстов, в которых понятия представляются преимущественно фразеологическими словосочетаниями (например, синтаксическая модель "дерево зависимостей" для этой цели непригодна, так как в этой модели текст расчленяется на отдельные слова, в результате чего его понятийная структура разрушается).

В современном обществе потребности в автоматической **с**мысловой обработке текстовой информации весьма многообразны, и с течением времени они растут. В рамках настоящей статьи мы не имеем возможности учесть все их в полном объеме. Поэтому ограничимся рассмотрением только наиболее популярных процедур обработки текстовой информации. На наш взгляд, такими процедурами являются:

1. Автоматическая классификация, индексирование и реферирование документов.
2. Автоматическое распознавание смысловой близости документов.
3. Машинный перевод текстов с одних естественных языков на другие.
4. Поиск информации в одноязычных и многоязычных базах данных по запросам, сформулированным на естественном языке.

Мы начнем наше рассмотрение с комплексной процедуры автоматического распознавания смысловой близости документов, в состав которой входят процедуры их автоматического индексирования и реферирования. При этом мы будем исходить из того, что процедура распознавания смысловой близости документов должна выполняться путем сравнения смыслового содержания некоторого заданного документа (документа-запроса) с содержанием множества других документов, представленных в виде поискового массива. При этом для всех документов поискового массива и для документа-запроса предварительно составляются их формализованные описания – концептуальные (понятийные) образы документов (КОДы).

Концептуальные образы документов создаются на основе их автоматического концептуального анализа – выделения в них наиболее значимых наименований понятий. Для этого необходимо располагать достаточно мощным словарем наименований

понятий, включающим в свой состав наиболее информативные понятия поискового массива. Это достигается путем составления по текстам поискового массива частотного словаря наименований понятий и его последующего редактирования.

При составлении частотного словаря в него целесообразно включать все слова и словосочетания длиной от 2-х до 10-15 слов (как мы уже говорили, устойчивые фразеологические словосочетания длиной более 15-ти слов встречаются редко). Затем составленный словарь должен быть подвергнут редактированию - автоматическому и ручному.

При автоматическом редактировании словаря из него исключаются отдельные слова, которые являются местоимениями, предлогами, союзами и частицами, а также фразеологические словосочетания, имеющие в своем составе местоимения или начинающиеся и оканчивающиеся союзами или предлогами. Из словаря исключаются также все редкие слова и словосочетания (например, все слова и словосочетания с частотой встречаемости f<=3), так как эти лексические единицы практически не оказывают влияния на процесс распознавания смысловой близости документов, но занимают значительную часть объема словаря (более 90%). В процессе ручного редактирования частотного словаря из него исключаются наиболее частые малоинформативные слова и словосочетания, которые не были исключены на этапе автоматического редактирования.

Важнейшей характеристикой наименований понятий является их длина – количество входящих в их состав слов. Более многословные словосочетания бывают обычно более информативны - они в большей степени отражают смысловое содержание документов, чем более короткие словосочетания. Вторым по важности показателем является частота встречаемости наименований понятий в текстах документов. Повторение одних и тех же наименований понятий в тексте (так называемые лексические повторы) служит средством выражения смысловых связей между входящими в его состав предложениями и характеризует его общую смысловую направленность.

Что касается других возможных характеристик значимости наименований понятий (например, их принадлежности к категориям географических названий, фамильно-именных групп и др.), то в рамках настоящей статьи мы их рассматривать не будем, так как они по-разному оцениваются при решении различных прикладных задач.

После составления частотного словаря наименований понятий и его редактирования следующим шагом при построении системы распознавания смысловой близости документов должно быть их автоматическое индексирование. Оно должно заключаться в концептуальном анализе текстов документов и отборе наиболее значимых наименований понятий.

Мы будем оценивать значимость наименований понятий их “весами”. В общем случае “вес” P наименования понятия можно определять по формуле:

$$P = f\, n^3, \qquad (1)$$

где f – частота встречаемости наименования понятия в тексте документа, n – количество слов в наименовании.

Однако при этом следует учитывать, что однословные наименования понятий повторяются в текстах значительно чаще, чем словосочетания, хотя они бывают, как правило, менее информативными. Поэтому нельзя допускать, чтобы числовые оценки “весов” однословных наименований понятий были равны или, тем более, превосходили третью степень количества слов в двухсловных словосочетаниях (число 8). По этой причине мы считаем, что всем однословным наименованиям понятий с частотой $f > 7$ нужно присваивать частоту 7. После присвоения “весов” наименованиям понятий они сортируются по их убыванию, и в составе концептуального образа документа оставляются только наиболее значимые наименования понятий.

По завершении процесса индексирования документов поискового массива и документа-запроса составляются их рефераты. Это необходимо для быстрой оценки результатов поиска документов-аналогов по документу-запросу. В рефератах должны быть представлены заголовки документов и несколько их наиболее значимых предложений. Значимость предложений определяется как сумма “весов” наименований понятий, входящих в их состав.

Идея автоматического распознавания смысловой близости документов заключается в том, что концептуальный образ документа-запроса (его КОД) сравнивается с концептуальными образами документов (КОДами) поискового массива. При этом для каждой пары сравниваемых КОДов определяется коэффициент их смысловой близости. В упрощенной постановке задачи этот коэффициент можно было бы положить равным отношению суммы весов наименований понятий КОДа документа-запроса, совпавших с наименованиями понятий КОДа из поискового массива, к сумме весов всех наименований понятий КОДа документа-запроса. Однако такой подход был бы слишком упрощенным а, следовательно, и некорректным.

Некорректным потому, что обычно в разных текстах одни и те же объекты и процессы могут описываться с различной степенью обобщения и с помощью различных языковых средств. Поэтому при решении задач автоматического распознавания смысловой близости документов необходимо в той или иной мере учитывать такие явления как словоизменение, словообразование, синонимия, гипонимия (родо-видовые отношения), разнообразие средств выражения межфразовых связей.

Явление словоизменения может быть учтено путем применения процедуры автоматического морфологического анализа слов и отождествления различных форм слов по их основам. Для учета явления словообразования можно использовать словарь словообразовательных вариантов слов. Для учета явлений синонимии и гипонимии необходимо использовать словарь синонимов, гипонимов (более узких по объему понятий) и гиперонимов (более широких по объему понятий) как на уровне отдельных слов, так и на уровне фразеологических словосочетаний.

Сложнее дело обстоит с учетом таких явлений, как вариации обозначений одних и тех же понятий в связных текстах. Дело в том, что в связном тексте

наименование понятия, выраженное некоторым фразеологическим словосочетанием, может быть сначала представлено в своей исходной форме, но затем, в последующих предложениях, - в сокращенных вариантах. Оно может быть также заменено на наименование родового понятия или на местоимение.

Отрицательное влияние этих явлений на процесс распознавания смысловой близости документов частично сглаживается тем, что различные варианты наименований одних и тех же понятий будут включены в частотный словарь наименований понятий при его составлении и, как следствие, – в поисковые образы документов.

При сопоставлении наименований понятий концептуальных образов документа-запроса и документов из поискового массива необходимо в первую очередь учитывать явления словоизменения, синонимии и гипонимии, так как эти явления оказывают наибольшее влияние на полноту распознавания смысловой близости документов. При этом информацию о различных вариантах синонимов, гипонимов и гиперонимов слов и словосочетаний следует вносить в концептуальный образ документа-запроса, а не в концептуальные образы документов поискового массива, так как во втором случае это привело бы к существенному увеличению поискового массива.

При сопоставлении однословных наименований понятий они считаются связанными по смыслу, если выполняется следующее условие: либо они полностью совпадают, либо являются синонимами, либо находятся в родо-видовых отношениях. При сопоставлении фразеологических словосочетаний к ним предъявляются те же требования, что и к однословным наименованиям понятий, но если указанное выше условие не выполняется, тогда предпринимается попытка пословного сопоставления словосочетаний. В этом случае словосочетания считаются связанными по смыслу, если в обогащенном КОДе документа-запроса для каждого слова словосочетания из КОДа документа поискового массива находится хотя бы один синоним, гипоним или гипероним.

По окончании процесса сопоставления всех наименований понятий КОДа документа-запроса и КОДа документа из поискового массива определяется коэффициент смысловой близости этих документов. При этом производится суммирование “весов” исходных наименовании понятий КОДа документа-запроса, связанных по смыслу с наименованиями понятий КОДа документа из поискового массива, и полученная сумма делится на сумму весов всех исходных наименований понятий КОДа документа-запроса. Документы считаются связанными по смыслу, если коэффициент их смысловой близости превышает заданный порог значимости.

После обработки всех документов поискового массива документы, связанные по смыслу с документом-запросом, упорядочиваются по убыванию значений их коэффициентов смысловой близости документу-запросу. Пользователю выдаются рефераты наиболее значимых документов-аналогов. Этот результат может быть представлен в виде нескольких эшелонов выдачи, содержащих информацию о документах с различной степенью их смысловой близости заданному документу-запросу.

Описанная процедура автоматического распознавания смысловой близости документов была проверена авторами на представительном массиве общественно-политических текстов и оказалась весьма эффективной [10].

Под автоматической классификацией документов мы будем понимать процесс распознавания их принадлежности к тематическим классам некоторой системы классификации. Этот процесс аналогичен процессу распознавания смысловой близости документов с тем лишь отличием, что в качестве документов-запросов здесь будут выступать классифицируемые документы, а в качестве поискового массива – массив предметных рубрик классификатора. При этом для классифицируемых документов и рубрик классификатора необходимо будет составить их формализованные концептуальные образы (КОДы).

В процессе классификации документа его КОД сравнивается с КОДами предметных рубрик и определяется степень его смысловой близости с этими рубриками. Сравнение проводится с учетом иерархической структуры классификатора: сначала просматриваются только понятия самого верхнего (первого) уровня иерархии, затем - только понятия второго уровня, подчиненные релевантному понятию первого уровня, затем – понятия третьего уровня, подчиненные релевантному понятию второго уровня, и т.д. Документ считается принадлежащим к той рубрике, коэффициент смысловой близости с которой превышает заданное пороговое значение. Таких рубрик может оказаться несколько. Формализованные описания тематических рубрик классификаторов могут создаваться путем автоматического концептуального анализа ранее расклассифицированных массивов документов.

Автоматический (машинный) перевод текстов с одних языков на другие – давнишняя мечта человечества. Об этом мечтал, в частности, известный английский математик Чарльз Бэббидж (1791-1871) – автор проекта механической аналитической машины (по существу - первого универсального программируемого компьютера*)*. Однако реальная возможность создания систем машинного перевода появилась только в середине прошлого века, когда были построены первые ЭВМ.

В 1954 г. в США был продемонстрирован перевод с помощью ЭВМ небольшого текста с русского языка на английский (Джорджтаунский университет, г. Вашингтон). После этого в ряде ведущих стран мира были начаты работы по машинному переводу.

Но задача оказалась не простой, и быстро решить ее не удалось. Более того, через 12 лет после Джорджтаунского эксперимента был опубликован доклад Консультативного комитета по автоматической обработке текстовой информации, созданного при Национальной академии наук США, в котором был сделан вывод о бесперспективности работ по машинному переводу [11]. А еще через 16 лет к аналогичному выводу пришел руководитель японской государственной программы по машинному переводу профессор Макото Нагао из университета Киото. Он

заявил, что существующие подходы к решению проблемы машинного перевода бесперспективны и дело рано или поздно зайдет в тупик. В качестве выхода из тупика он предложил осуществлять машинный перевод текстов по аналогии с переводами, выполненными людьми [12]. Но он не указал конкретных способов решения этой задачи.

В 1975 г., за девять лет до выступления Макото Нагао, сотрудник 27 ЦНИИ МО СССР профессор Г.Г. Белоногов предложил свою концепцию машинного перевода текстов с одних естественных языков на другие, которая была им опубликована в предисловии к книге Д.А. Жукова "Мы переводчики" [13]. Он назвал ее *концепцией семантико-синтаксического преимущественно фразеологического машинного перевода* (кратко – *концепцией фразеологического машинного перевода)*.

В 1993 г. эта концепция была реализована в ВИНИТИ РАН в виде промышленной версии системы русско-английского и англо-русского перевода, получившей название RETRANS (Russian-English TRANslation System) [1, 14]. Система сразу получила признание в ряде государственных учреждений России, США, Франции и Англии, а некоторые правительственные учреждения США и Франции в течение нескольких лет финансировали исследования и разработки, направленные на совершенствование этой системы.

Главной отличительной чертой системы RETRANS была ее ориентация на представление наименований понятий в словарях преимущественно фразеологическими словосочетаниями. Так в одной из ее последних версий базовые политематические словари (англо-русский и русско-английский) имели общий объем более 5 млн словарных статей, а дополнительные тематические словари – более 1 млн 200 тыс. Во всех этих словарях и русские, и английские наименования понятий были на 80% представлены фразеологическими словосочетаниями длиной от двух до 17 слов и только на 20% – отдельными словами. Были в этой системе и другие нововведения (например, в процедурах морфологического и синтаксического анализа текстов использовался метод аналогии). Более полные сведения о системе RETRANS можно найти в публикациях [1, 14, 15].

В конце 80-х гг. прошлого столетия на свет появились программные системы под названием Translation memory (память переводов) или, другое их название,– Sentence memory (память предложений) [16]. В России они получили название "память переводчика". Эти системы предназначались для учреждений, где систематически выпускалась и переводилась на иностранные языки однородная по содержанию техническая документация.

Такая документация сначала переводилась вручную, после чего результаты перевода представлялись в виде пар предложений на исходном и на иностранном языках и записывались в память ЭВМ ("память переводчика"). В последующих выпусках документации тексты документов членились на предложения, которые искались в массиве ранее переведенных предложений, и для найденных предложений выбирались их переводы на иностранный язык. Если же какое-то предложение на исходном языке не находилось среди ранее переведенных предложений, то оно переводилось вручную, и полученная пара разноязычных предложений заносилась в "память переводчика". При ручном переводе делалась попытка найти для фрагментов "новых" предложений их переводы среди ранее переведенных предложений. Позднее стали создавать гибридные системы, состоящие из "памяти переводчика" и обычной системы машинного перевода (для перевода "новых" предложений).

Системы типа "память переводчика" безусловно полезны. Они позволяют существенно сократить трудозатраты на перевод документов, близких по содержанию. Но эти системы непригодны для перевода политематических текстов. Они также непригодны и для перевода текстов, принадлежащих к одной и той же тематике, если в них мало совпадающих предложений (а это обычно так и бывает).

Поиски более эффективных методов машинного перевода все время продолжаются. Так, в последнее время усиленно пропагандируется так называемый "статистический машинный перевод". С некоторыми оговорками его можно рассматривать как попытку реализовать упомянутое выше предложение японского ученого Макото Нагао осуществлять машинный перевод текстов по аналогии с переводами, выполненными людьми [12]. Его можно также рассматривать и как косвенное признание бесперспективности метода семантико-синтаксического преимущественно пословного перевода, культивировавшегося в мире в течение ряда десятилетий, начиная с конца 50-х гг. прошлого столетия.

Согласно этому методу, перевод текстов с одного языка на другой предлагалось осуществлять преимущественно как перевод "значений" слов с оформлением результатов перевода в соответствии с грамматическими нормами выходного языка. При этом фразеологические словосочетания исходного текста должны были представляться на выходном языке как последовательности переводов "значений" слов, входящих в состав этих словосочетаний. Такие искусственно созданные последовательности слов часто искажали смысл исходного текста и не соответствовали устоявшимся фразеологическим нормам выходного языка.

К сожалению, у большинства современных ученых-лингвистов и у разработчиков систем машинного перевода и в настоящее время нет еще должного понимания той истины, что при переводе текстов с одного языка на другой **нельзя формировать фразеологические единицы выходного текста как последовательности переводов "значений" слов входного текста, а нужно опираться на исторически сложившееся фразеологическое богатство выходного языка.**

Следует заметить, что в современных системах "статистического перевода" огромные массивы двуязычных текстов используются только для поиска переводных эквивалентов отдельных слов и, значительно реже, для двух- и трехсловных словосочетаний. А желательно бы перед обращениям к массивам двуязычных текстов провести концептуальный ана-

лиз исходных текстов, чтобы искать в двуязычных текстах не произвольные последовательности слов исходного текста, а фразеологические словосочетания, являющиеся наименованиями понятий.

Для проведения автоматического концептуального анализа исходных текстов необходимо располагать мощным словарем наименований понятий, для начала хотя бы объемом в полтора-два десятка миллионов словарных единиц. Наш опыт показывает, что такой словарь можно составить за два-три месяца.

У нас вызывает сомнение правильность принятого в настоящее время общего порядка проведения "статистического машинного перевода": ведь статистический анализ огромного массива двуязычных текстов здесь повторяется для каждого переводимого текста. А почему бы вместо этого не составить по двуязычным текстам один двуязычный политематический словарь наименований понятий и ряд дополнительных тематических словарей? И для этого совсем не нужны суперЭВМ и потребуется всего несколько месяцев (были бы только исходные двуязычные тексты!). Не нужны суперЭВМ и для перевода текстов, если будут использоваться системы фразеологического перевода [1, 13-15]. Достаточно будет персональных компьютеров.

В начале нашей статьи мы среди наиболее популярных проблем автоматической обработки текстовой информации указали проблему ее поиска в одноязычных и многоязычных базах данных по запросам, сформулированным на естественном языке. Работы по этой проблеме ведутся уже более полувека.

Первые автоматизированные информационно-поисковые системы появились в конце 40-х – начале 50-х гг. прошлого столетия. Так, в 1951 г. американская фирма IBM (International Business Machine Corporation) сообщила о том, что ею построены два варианта автоматизированных документальных поисковых систем. В одном из них формализованные описания документов были представлены в "прямой" форме, в другом – в "инверсной". В *прямой форме представления* номерам документов ставились в соответствие перечни номеров понятий-дескрипторов, описывающих содержание этих документов. В *инверсной форме,* наоборот, номерам дескрипторов ставились в соответствие перечни номеров документов, в описания которых эти дескрипторы входили. Термины *прямая и инверсная форма представления информации* используются и в настоящее время.

За прошедшие более чем полвека системы автоматизированного поиска документов претерпели значительные изменения. По мере развития электронной вычислительной техники росли объемы информации, хранимой в памяти ЭВМ, и совершенствовались языковые и программные средства этих систем. Сначала поиск документов велся на основе их поисковых образов (ПОДов), представленных числовыми кодами понятий. Затем числовые коды понятий стали заменяться на их наименования. Позднее в поисковые массивы стали вводиться заголовки и рефераты документов, и, еще позднее, появилась возможность вводить в ЭВМ полные тексты документов. В настоящее время в порядок дня поставлен вопрос о поиске документов в полнотекстовых базах данных.

Вопрос этот весьма не простой. Ведь в текстах одни и те же объекты и явления могут описываться в терминах различной степени общности с привлечением различных выразительных средств естественных языков. Возникает проблема адаптации традиционных форм представления речевой информации к возможностям электронной вычислительной техники (электронных "алгоритмических" машин). При этом приходится учитывать противоречивые требования к форме хранения информации в памяти ЭВМ: с одной стороны, требования удобства для человека при поиске информации, с другой стороны - удобства выполнения процедур поиска информации.

Для удовлетворения этих противоречивых требований пришлось, наряду с исходными текстами, хранить в ЭВМ и некоторые формализованные структуры. Вначале роль таких структур выполняли поисковые образы документов, позднее, наряду с ними, стали использоваться инверсная форма представления текстов и гипертекст.

Понятие "гипертекст" (hypertext) обычно определяется как технология работы с текстовыми данными, позволяющая устанавливать ассоциативные связи – "гиперсвязи" между отдельными терминами, фрагментами документов и статьями в текстовых массивах и благодаря этому допускающая не только последовательную, линейную работу с текстом, как при обычном чтении, но и произвольный доступ к информации и ее ассоциативный просмотр в соответствии с установленной структурой связей.

Гипертекстовые связи представляют собой по существу перекрестные ссылки, которые дают возможность быстрого обращения к нужным фрагментам информации. Эти связи наиболее эффективны тогда, когда они используются при поиске в больших массивах информации, расчлененных на множество мелких ассоциированных по смыслу фрагментов, и когда пользователю в каждый данный момент требуются только небольшие объемы информации. Гипертекст наиболее эффективно используется в мультимедийных коммерческих вычислительных системах.

Обобщая приведенные характеристики гипертекста, можно утверждать, что он представляет собой семантическую сеть, узлы которой соответствуют некоторым блокам информации, а дуги – ассоциативным связям между ними. Узлом гипертекста может быть фрагмент текста, рисунок, фотография, движущееся или мультипликационное изображение, звуковая речь или музыкальное произведение и даже выполняемая программа. Если часть данных не является текстовой, то о конечном продукте говорят как о мультимедийной системе (multimedia, hypermedia).

Инверсные файлы и гипертекстовое представление информации часто используются совместно, в одной и той же поисковой системе. При этом инверсные файлы обеспечивают начальное обращение к фрагментам текстов по запросам, а гипертекст дает возможность продолжать поиск, используя ассоциативные связи между этими фрагментами.

На наш взгляд, гипертекстовая структура в ее нынешнем состоянии имеет два существенных недостатка: 1) здесь поиск информации можно вести только по тем связям, которые были установлены

при создании поисковых массивов; 2) установление ассоциативных связей между текстами и их фрагментами осуществляется вручную, а их полнота и точность зависит от квалификации индексаторов. Преодоление указанных недостатков должно идти по пути совершенствования поискового аппарата инверсного представления текстов и гипертекста и по пути большей интеграции этих структур, чем это имело место до сих пор.

Как мы уже говорили, инверсные файлы и гипертекст являются формализованными надстройками над текстом, отражающими его семантико-синтаксическую структуру. При этом в инверсных файлах акцент делается на облегчение доступа к отдельным словам при сохранении информации о порядке их следования в тексте, а в гипертексте – на ассоциативные связи между текстами и их фрагментами. Таким образом, в первом случае четко выделяется только одна единица смысла - слово, а во втором случае - только сверхфразовые единства (тексты и их фрагменты). А нужно, чтобы при поиске информации "работали" единицы смысла всех уровней. Это легче всего можно осуществить в случае инверсных файлов, поскольку там практически полностью представлен лексический состав текстов и есть возможность использовать при поиске парадигматические связи между словами и словосочетаниями.

На протяжении более чем полувековой истории развития информационно-поисковых систем неоднократно предпринимались попытки автоматического поиска информации в текстах по запросам на неформализованных "естественных" языках и диалогового общения с ЭВМ на таких языках. Но пока что эта задача еще далека от своего полного решения.

И не удивительно. Ведь речь здесь идет ни много ни мало как об автоматическом распознавании "смысла" запросов и о последующем сопоставлении этого "смысла" со "смыслом" текстов, в которых ведется поиск. А средства выражения этого "смысла" весьма многообразны: здесь и многообразие словоизменительных и словообразовательных форм слов, и явление лексической полисемии, синонимии и гипонимии, и многообразие синтаксических возможностей представления одного и того же смысла (синтаксическая синонимия), и явление эллипсиса, и наличие ассоциативных парадигматических связей между фразеологическими словосочетаниями как целостными единицами смысла, не сводимых к парадигматическим связям между составляющими их словами, и еще многое другое.

На наш взгляд, поиск информации в полнотекстовых базах данных можно было бы реализовать на основе тех принципов, которые мы изложили выше, рассматривая процедуру автоматического установления смысловой близости документов. Здесь, как и в любых системах автоматической смысловой обработки текстовой информации, в основу должна быть положена процедура ее автоматического концептуального анализа.

При решении задачи автоматизированного поиска текстовой информации возникает еще одна важная проблема – преодоление языковых барьеров между странами и народами. В мировом масштабе это очень сложная проблема. Действительно, в мире существует около 2500 различных языков, а число их парных сочетаний превосходит 3 млн. Но если бы даже различных языков было не более сотни, то и тогда для перевода текстов с любого языка на любой другой потребовалось бы около пяти тысяч систем перевода. А это трудно выполнимая задача.

Выходом из этого затруднения мог бы быть отказ от построения систем машинного перевода с любого языка на любой другой, и вместо этого выполнение перевода с помощью *языка-посредника*. Тогда можно было бы существенно сократить число разрабатываемых систем перевода. Например, в случае 100 различных языков вместо 4950 систем перевода пришлось бы создавать только 99 (т. е. в 50 раз меньше!).

Идея языка-посредника была высказана еще на рубеже конца 50-х и начала 60-х гг. прошлого столетия. Но она тогда не была реализована, так как для этого не было необходимых условий. Однако в настоящее время, в связи с улучшением качества машинного перевода, к этой идее можно было бы вернуться.

Среди различных предложений по языку-посреднику, выдвинутых пионерами машинного перевода, было предложение использовать в качестве такого языка искусственный язык Esperanto. На наш взгляд это неразумно, так как любой искусственный язык имеет более бедную систему понятий, чем естественные языки, и не годится в качестве языка-посредника. В таком качестве может выступать только один из естественных языков с достаточно богатой системой понятий (например, русский, английский, немецкий или французский).

Скорее всего, развитие машинного перевода пойдет по пути разработки двуязычных систем перевода в интересах наиболее развитых стран мира. А по мере их создания постепенно будет появляться возможность перевода текстов и между новыми парами языков, не обеспеченными изначально системами перевода, через посредство имеющихся в наличии систем. И, возможно, только на более позднем этапе развития будет достигнуто соглашение о едином языке-посреднике или о нескольких таких языках.

ЗАКЛЮЧЕНИЕ

Подводя итоги нашим рассуждениям о перспективах развития систем автоматической смысловой обработки текстовой информации, мы хотели бы подчеркнуть важность первоочередного решения следующих проблем.

Проблема первая – выявление понятийного и фразеологического богатства естественных языков. Это важно потому, что понятия являются базовыми и наиболее устойчивыми единицами смысла и в мышлении, и в языке, и в речи. Кое-что в этом направлении уже сделано, но предстоит сделать еще больше, особенно в части выявления фразеологического богатства естественных языков.

Проблема вторая – выявление наиболее устойчивых внеконтекстных ассоциативных смысловых связей между понятиями. Эта проблема по существу является частью первой, так как речь здесь идет об

описании смыслового содержания понятий, а оно наиболее полно раскрывается в системе их ассоциативных связей друг с другом.

Проблема третья – разработка базовых процедур семантико-синтаксического анализа и синтеза текстов на основе их фразеологического и концептуального анализа и синтеза.

Перечисленные проблемы являются наиболее приоритетными, так как они возникают при решении любых достаточно сложных задач автоматической обработки текстовой информации.

В самой общей постановке системы автоматической смысловой обработки текстовой информации – это системы класса "искусственный интеллект". Над их созданием висит "проклятие размерности". Ведь, по мнению психологов, человек даже при простой интерпретации (понимании) смыслового содержания текстовых сообщений опирается на весь свой жизненный опыт. А если речь пойдет о более сложных мыслительных процессах?

Наверное справедлив известный философский тезис, что логика идей определяется логикой вещей. Но мир вещей, окружающий человека, бесконечно сложен. И хотя мир идей – это, прежде всего, мир обобщающих понятий и представлений, он тоже очень сложен. И быстрого решения даже самых важных проблем смысловой обработки текстовой информации ожидать не приходится. Они будут решаться постепенно.

СПИСОК ЛИТЕРАТУРЫ

1. Белоногов Г.Г. Теоретические проблемы информатики. Т. 2. Семантические проблемы информатики / под общ. ред. К.И. Курбакова.– М.: РЭА им. Г.В. Плеханова, 2008. – 215 с.
2. Белоногов Г.Г., Гиляревский Р.С. Еще раз о гносеологическом статусе понятия "информация" // НТИ. Сер. 2. – 2010. - № 2. – С. 1-6.
3. Белоногов Г.Г., Хорошилов Ал-др А., Хорошилов Ал-сей А. Единицы языка и речи в системах автоматической обработки текстовой информации // НТИ. Сер. 2. – 2005. - № 11. – С. 21-29.
4. Новая философская энциклопедия. Т. 1-4. – М. : Мысль, 2000.
5. Максименко С.Д. Общая психология. – Киев : Рефл-бук : Ваклер, 2000. – 533 с.
6. Спиркин А.Г. Философия (издание второе). - М.: Гардарики, 2006. – 735 с.
7. Гиляревский Р.С. Информационный менеджмент: управление информацией, знаниями, технологией. – М.: Профессия, 2009.
8. Соссюр Фердинанд де. Курс общей лингвистики. – М. : Прогресс, 1977.
9. Звегинцев В.А. Предложение и его отношение к языку и речи. - М.: Изд-во Моск. ун-та, 1976.
10. Белоногов Г.Г., Гиляревский Р.С., Хорошилов Ал-сей А., Хорошилов-мл. Ал-сей А. Автоматическое распознавание смысловой близости документов // НТИ. Сер. 2. – 2011. - № 7. – С. 15-22.
11. Черный А.И. Всероссийский Институт Научной и Технической Информации: 50 лет служения науке. – М.: ВИНИТИ, 2005. – 298 с.
12. Nagao M. A framework of a mechanical translation between Japanese and English by analogy principle, in Artificial and Human Intelligence / ed. A. Elithorn, R. Banerji. - North Holland, 1984. - P. 173-180.
13. Жуков Д.А. Мы переводчики. – М. : Знание, 1975.
14. Белоногов Г.Г., Хорошилов Ал-др А., Хорошилов Ал-сей А., Козачук М.В., Рыжова Е.Ю., Гуськова Л.Ю. Каким быть машинному переводу в XXI веке // Перевод: традиции и современные технологии. – М. : ВЦП, 2002.
15. Белоногов Г.Г., Хорошилов Ал-др А., Хорошилов Ал-сей А. Автоматизация составления англо-русских двуязычных фразеологических словарей по массивам двуязычных текстов (билингв) // НТИ. Сер. 2. – 2010. - № 5. – С. 1-8.
16. Webb L. E. Advantages and Disadvantages of Translation Memory: a Cost/Benefit Analysis. - San Francisco State University, 1992.

Материал поступил в редакцию 11.07.12.

Сведения об авторах

БЕЛОНОГОВ Герольд Георгиевич – доктор технических наук, профессор, главный научный сотрудник компании RETRANS Technologies, Москва

E-mail: belonogov@mail.ru

ГИЛЯРЕВСКИЙ Руджеро Сергеевич – доктор филологических наук, профессор, зав. отделением ВИНИТИ РАН, Москва

E-mail: giliarevski@viniti.ru

ХОРОШИЛОВ Александр Алексеевич - доктор технических наук, генеральный директор компании RETRANS Technologies, Москва

E-mail: khorochilov@mail.ru

Москва, ВИНИТИ РАН
28–30 ноября 2012 года

***8-я Международная конференция* «НТИ-2012»,**
посвященная 60-летию ВИНИТИ РАН

«АКТУАЛЬНЫЕ ПРОБЛЕМЫ ИНФОРМАЦИОННОГО ОБЕСПЕЧЕНИЯ НАУКИ, АНАЛИТИЧЕСКОЙ И ИННОВАЦИОННОЙ ДЕЯТЕЛЬНОСТИ»

Для участия **в «НТИ-2012»** приглашаются специалисты в области информационных технологий, телекоммуникаций, создатели и потребители информационных продуктов и услуг, ученые и специалисты РАН, вузовской и отраслевой науки, работники информационных центров и библиотек, служб распространения информационных продуктов и услуг.

Доклады или тезисы докладов, которые будут опубликованы в специальном сборнике, направлять в Оргкомитет конференции «НТИ-2012».

Планируется проведение пленарных заседаний, круглых столов и работа по секциям.

Адрес: Россия, 125190, Москва, ул. Усиевича, 20, Всероссийский институт научной и технической информации (ВИНИТИ РАН), ОНИИР, Оргкомитет **«НТИ-2012».**
Тел: (495) 155-44-22, 155-44-29, 152-64-41,
Факс: (495) 152-54-92, 943-00-60
E-mail: conf@viniti.ru , market@viniti.ru
http://www.viniti.ru

Российская академия наук

Федеральное государственное бюджетное учреждение науки
ВСЕРОССИЙСКИЙ ИНСТИТУТ НАУЧНОЙ И ТЕХНИЧЕСКОЙ ИНФОРМАЦИИ
РОССИЙСКОЙ АКАДЕМИИ НАУК

предлагает научным работникам, аспирантам и другим специалистам в области естественных, точных и технических наук, желающим быстро и эффективно опубликовать результаты своей научной и научно-производственной деятельности, использовать способ публикации своих работ через *систему депонирования.*

«Депонирование (передача на хранение) – особый метод публикации научных работ (отдельных статей, обзоров, монографий, сборников научных трудов, материалов научных мероприятий – конференций, симпозиумов, съездов, семинаров) узкоспециального профиля, разрешенных в установленном порядке к открытому опубликованию, которые нецелесообразно издавать полиграфическим способом печати, а также работ широкого профиля, срочная информация о которых необходима для утверждения их приоритета. Депонирование предусматривает прием, учет, регистрацию, хранение научных работ и обязательное размещение информации о них в специальных информационных изданиях».

Подготовка и передача на депонирование научных работ происходит в соответствии с «Инструкцией о порядке депонирования научных работ по естественным, техническим, социальным и гуманитарным наукам» (М., 2003).

Результатом депонирования является публикация информации о депонированных научных работах в информационных изданиях ВИНИТИ РАН – реферативном журнале и аннотированном библиографическом указателе «Депонированные научные работы».

В соответствии с “Положением о порядке присуждения ученых степеней”, утвержденным Постановлением Правительства Российской Федерации от 30.01.2002 № 74 (в ред. Постановлений Правительства РФ от 20.04.2006 № 227, от 02.06.2008 № 424, от 20.06.2011 № 475), научные работы, депонированные в организациях государственной системы научно-технической информации, признаны публикациями, учитывающимися при защите кандидатских и докторских диссертаций.

Подать научную работу на депонирование можно обратившись в Отдел депонирования ВИНИТИ РАН по адресу:

125190, Москва, ул. Усиевича, 20.
ВИНИТИ РАН, Отдел депонирования научных работ.
Тел.: 8 (499) 155-43-28, Факс: 8 (499) 943-00-60.
e-mail: dep@viniti.ru

С инструкцией о порядке депонирования можно ознакомиться на сайте ВИНИТИ РАН: **http://www.viniti.ru**