

Ю. Р. Валькман, д.т.н., зав. отд., Р. Ю. Валькман, н. с., Л. Р. Исмаилова, н. с.,
Международный научно-учебного центр информационных технологий и
систем НАН и МОН Украины

СВОЙСТВА ГИПОТЕЗЫ КАК МОДЕЛИ ЗНАНИЙ

Введение. Данная работа продолжает серию публикаций «*На пути к автоматизации синтеза и анализа гипотез*». Результаты соответствующих исследований, опубликованы в [1-6].

Объектом исследования являются процессы порождения и обоснования гипотез. Предмет исследования – свойства гипотез как моделей. Цель исследования – разработка моделей и методов генерации гипотез в исследованиях сложных систем. Ожидаемые результаты – разработка интеллектуальной компьютерной технологии, поддерживающей синтез и анализ гипотез исследуемых процессов и объектов.

Само слово «*гипотеза*» всем известно: в обиходе оно обозначает любое *предположение*. Обычно предположение так или иначе связано с *проблемной ситуацией*, т. е. с вопросом или с группой взаимообусловленных вопросов, возникающих в ходе какой-то деятельности. Не на всякие вопросы могут быть сразу даны однозначные ответы в виде категорических суждений; иной раз ответы принимают форму *предположений, догадок, допущений*. В психологическом плане предположению соответствуют субъективные состояния, включающие в себя известное сомнение, неуверенность в том, что дело обстоит так-то и так-то. *Именно поэтому предположения нуждаются в проверке*. Иногда они легко и быстро проверяются какими-то несложными действиями. Скажем, чтобы проверить возникшую догадку (если угодно, можно назвать ее гипотезой), что знакомый голос в соседней комнате принадлежит тете Тане, достаточно открыть дверь; примерно так же просто проверяются предположения (и их при желании можно назвать гипотезами) об источнике шума за окном, о том, что причиной болевых ощущений в ноге является гвоздь в подошве, что экран телевизора погас из-за сгоревшего предохранителя и т. д. Другие предположения (гипотезы) проверить сложнее. Именно об этих гипотезах (чаще, научных) мы будем говорить.

Проверка подобных предположений требует определенных, достаточно длительных, и, по существу, исследовательских действий (включающих наблюдение, отбор материала и др.), направленных на то, чтобы исключить случайный результат.

Но, в любом случае, в гипотезе отражаются наши предположения, догадки о решении соответствующей задачи, проблемы (что это, как это сделать, как оно устроено, почему происходит и т. д.). В гипотезе, поскольку она является формой приобретения *нового знания*, изначально заложена идея развития. Действительно, по самой сути дела гипотеза не может быть

самоценной. Она ведь и задумана всего лишь как предположение, рассчитана на то, чтобы, сыграв преходящую роль в становлении некоего фрагмента знания, затем сойти со сцены. Формулирование гипотез является одной из важных интеллектуальных операций. Сложность этой операции осознают все учёные. Недаром *постановка проблемы считается в науке существенной частью любого исследования.* Гипотеза - это научно обоснованное предположение о причинах или закономерных связях каких-либо явлений или событий природы, общества, мышления. Специфика гипотезы - быть **формой развития знаний** - предопределяется основным свойством мышления, его постоянным движением - углублением и развитием, стремлением человека к раскрытию новых закономерностей и причинных связей, что диктуется потребностями практической жизни.

Для построения компьютерных технологий генерации и анализа гипотез, нам необходимо выяснить, *в какой форме представлены соответствующие знания и как их использовать.*

1. Концепция трех миров

Видимому, целесообразно начать с вопроса – «*где находится знание?*». То есть с проблемы его локализации. Эта проблема не настолько тривиальна, как может показаться. Для этого лучше всего начать с классики. Великий славянский философ восемнадцатого века Григорий Сковорода в своем сочинении «*Потоп змиин*» еще в 1780-х годах (см. в [7]) описал концепцию трех миров.

Согласно его описанию, в первый мир входит окружающая нас реальность (по английски *Reality – R*).

Во второй мир входит внутренний мир человека (английский термин *Mind* или *M*).

Третий мир, согласно Г. Сковороде, есть мир знаков (*signs – или S*).

Действительно, этот мир не пересекается ни с одним из первых двух миров. Трудно сомневаться в автономности существования мира знаков в эпоху Интернета. Особенно, если это было предсказано более двухсот лет назад. Схема трех миров Г. Сковороды представлена на рис. 1.

Независимо от Сковороды к открытию *третьего мира* пришли некоторые западные философы XX века – например Карл Поппер. Этот мир он даже называл «*третьим миром*» [8]. Эта парадигма давно уже не является периферийной, полузабытой или невостребованной. Философия Карла Поппера входит в аспирантский минимум [8]. Концепция К. Поппера представлена в верхней части рис. 2.

Можно продолжить рассуждение о том, как эта теоретическая конструкция может послужить информационным технологиям, в частности, более ясному пониманию того, *что есть знание.* И становится ясно, более актуальной и плодотворной является проблема не первичности, а принадлежности.

Действительно, данные – это явления реального мира (**R**). Это намагниченные участки дисков, бороздки грампластинок и CD, модулированные радиоволны и т.д. Их можно измерить, а часто и увидеть или пощупать. Однако эти данные несут в себе закодированную

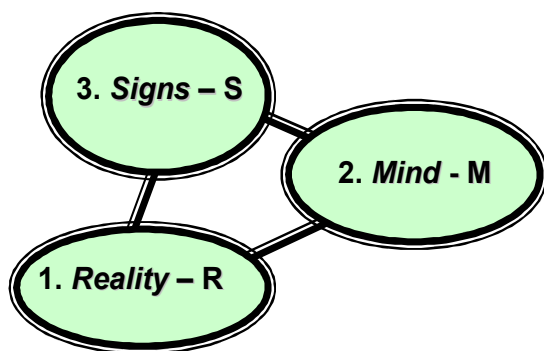


Рис. 1. Схема миров Г. Сковороды

информацию, которая может быть воспринята только если она преобразуется в **знаки** (**S**) – например, визуальные, звуковые. На экране компьютера мы видим их знаковое содержание. Через динамик мы слышим закодированные звуки с грампластины или переданные по радио. Это не является открытием, а констатацией факта. *То*

есть информация имеет знаковую природу и принадлежит миру знаков. Аналогичное рассуждение приводит к выводу, что знание принадлежит внутреннему миру человека (M). В полном объеме только в сознании человека происходят сложные и простые когнитивные процессы – от сложных доказательств теорем до Аристотелевых силлогизмов.

Итак, *данные* – это объекты материального мира.

Информация – это знаки, полученные при преобразовании данных в сознании человека или в процессоре компьютера.

Знания же пока есть только у человека или в самом элементарном виде в кибернетических системах.

Здесь же уместно вспомнить о триаде «данные-информация-знания» и о семиотическом треугольнике Г. Фреге (см., например, в [10]).

Вполне очевидно, что миры **M** и **S** – это системы моделей; т. к. соответствующие объекты являются отражением сущностей реального мира **R**. При этом любая знаковая модель **S** – это представление (на некотором носителе и языке) наших мыслей **M** об объектах **R**. И, хотя гипотеза рождается в **M**, она является результатом анализа **R** (или **S(R)**, например, измерительные матрицы [2,4]). А для ее анализа, обсуждения, доказательства («превращения» в теорию), нам необходимо построить знаковую модель **S(R)** или **S(M)**, или **S(M(R))**.

Возможно, целесообразно ввести в рассмотрение еще один классификатор моделей: *гипотезы и теории*.

Заметим, таким образом, фактически, модели играют роль своеобразного посредника между субъектом и реальностью.

2. Миры в ДСМ-методе В. К. Финна

Интересную спецификацию мира 1 у Поппера (или мира **R** у Сковороды) предложил [9] В. К. Финн в своем ДСМ-методе автоматического порождения гипотез.

С точки зрения методологии науки утверждение о том, что мир 1 изучается в различных науках посредством соответствующих понятийных систем и исследовательских методов, оказывается вполне тривиальным. Более содержательным подходом была бы классификация типов "подмиров", являющихся частными случаями мира 1. Следующие спецификации мира 1 могут быть выделены в соответствии с характеристикой универсумов, содержащих изучаемые события.

- *Мир 1-1* состоит из случайных событий, изучаемых вероятностными методами.
- *Мир 1-2* состоит из событий, между которыми имеются причинно-следственные отношения.
- *Мир 1-3* объединяет характеристики мира 1-1 и мира 1-2, а именно, в мире 1-3 имеются как причинно-следственные отношения между событиями, так и случайные изменения событий, влияющие на заключительные состояния этого мира.

Обобщенная схема миров К. Поппера и В. К. Финна представлена на рис.2. Эту структуру подмиров Финн построил, что бы, в частности, при генерации гипотез методы индуктивных выводов отделить от методов теории вероятностей и математической статистики. Мы не согласимся с В. К. Финном в этой части. С нашей точки зрения, методы статистических обобщений в отношении порождения гипотез «равноправны» с методами Д. С. Милля, абдуктивными методами, аналогией и др. средствами классических и неклассических логик.

Все определяется целями исследований и форматами представления исходных данных. Статистические обобщения действительно являются особым видом умозаключений неполной индукции связанные с анализом массовых событий. К ним относятся, например, массовые транспортные перевозки пассажиров и грузов, рождаемость и смертность людей, распространение заболеваний, транспортные происшествия, динамика преступлений и многие другие. Учитывая трудности выявления причинных зависимостей, анализ таких массовых событий позволяет установить устойчивое распределение интересующих исследователя случайных признаков. *Количественная информация*, выражающая устойчивые тенденции развития, имеет важное практическое значение для правильной организации обслуживания населения, профилактических мероприятий, борьбы с преступностью и т.п. Анализ массовых событий ведется чаще всего путем не сплошного, а выборочного исследования отдельных групп или образцов и логического переноса полученных результатов на все их множество. Вывод в этом случае протекает в форме статистического обобщения.

Таким образом, *статистическое обобщение* – это умозаключение неполной индукции, в котором установленная в посылках количественная информация о частоте определенного признака в исследуемой группе (образце) переносится в заключении на все множество явлений этого рода.

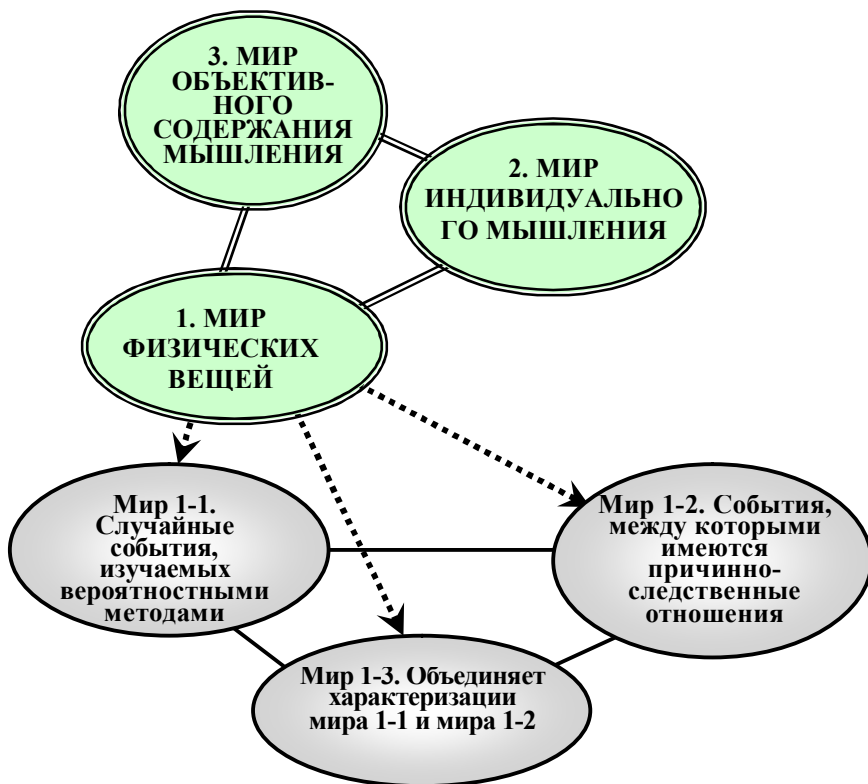


Рис.2. Обобщенная схема миров К. Поппера и В. К. Финна

В отличие от *индукции через перечисление* при отсутствии противоречащего случая в посылках статистического умозаключения фиксируется следующая информация:

- (1) *общее число составляющих исследуемую группу, или образец случаев;*
- (2) *число случаев, в которых присутствует интересующий исследователя признак;*
- (3) *частота появления интересующего признака.*

Здесь же целесообразно кратко остановиться на природе случайности, которая характерна не только для подмира 1-1, но для подмира 1-2, т. к. мы не можем априори быть уверены, что для данного процесса характерен

детерминизм. Поэтому, выделение подмира 1-2 из Попперовского мира 1 всегда затруднено или тривиально.

Существует несколько точек зрения на природе случайности, причем каждая из них имеет достаточные основания.

1. Согласно первой точке зрения, *случайным нам представляется нечто такое, в чем мы пока не уловили закономерности*. По мере познания явления в последнем остается все меньше и меньше случайного. Ярким выразителем такой позиции был Лаплас, считавший, что случайность не присуща самим объектам, а связана только с незнанием, в принципе устранимым.

2. Прогивоположная точка зрения состоит в том, что случайность является объективным свойством всех явлений. Более ста лет назад О. Курно писал [12], что "*случайность вмешивается во все, что творится на свете*", что "миром управляет случай, или, говоря точнее, случай имеет свою часть, и притом весьма значительную, в управлении вселенной".

3. Промежуточная позиция признает *как существование вполне детерминированных явлений, так и в принципе случайных*, описываемых статистическими закономерностями (законы Менделя; статистические законы физики, химии, термодинамики; законы квантовой механики и т. д.). Таким образом, случайность признается объективным свойством лишь некоторых явлений.

4. В последние годы представители школы И. Пригожина развивают подход, согласно которому *случайные и детерминированные периоды сменяют друг друга в истории любой системы*. Детерминированные процессы постепенно сменяются процессами, все более удаленными от равновесия, все более хаотическими, пока в период сильной неравновесности случайность не становится определяющей причиной того, в какое из возможных новых равновесных состояний придет система.

Далее мы детальнее остановимся на проблемах случайности, вероятности, нечеткости в приложении к генерации гипотез.

На рис. 3 представлена условная схема порождения и корректировки гипотез (см. описание технологии в [2] и схему эволюционной эпистемологии К. Поппера в [3]). Действительно, мы сначала определяем (в большей мере неформальными методами – мир M) какие свойства связаны между собой (y, x_1, x_2, x_3, \dots - *исходные системы*), затем собираем значения соответствующих параметров (строим измерительные матрицы - *системы данных* - мир S), затем строим зависимости (как эти свойства связаны – определяем $y = f(x_1, x_2, x_3, \dots)$) – *порождающие системы* - мир S), потом проверяем построенную гипотезу и, при необходимости синтезируем новую гипотезу. Многогочия на рисунке соответствуют вербальной трактовке вводимых обозначений. Заметим, именно на этом на этом уровне математики, как правило, *«растаются с семантикой проблемной области и начинают играть в синтаксические игры»*, считая, что теперь в формализованных структурах уже представлен ***весь необходимый и (главное!) достаточный для выводов***

смысл. Приведенную схему можно интерпретировать как (на высоком уровне абстрактивности) технологию порождения и корректировки гипотез.

Фактически, на рис. 3 представлены три модели одного объекта (мир R) различных форматов и различной информативности. Процесс генерации гипотезы можно рассматривать как технологию трансформации моделей.

Следует отметить, что гипотеза всегда обладает большим содержанием и большей логической силой, чем те данные, на которых она основана (см. рис. 3 системы данных и порождающие системы). Поэтому систему 2 и назвал [11] *порождающей*.

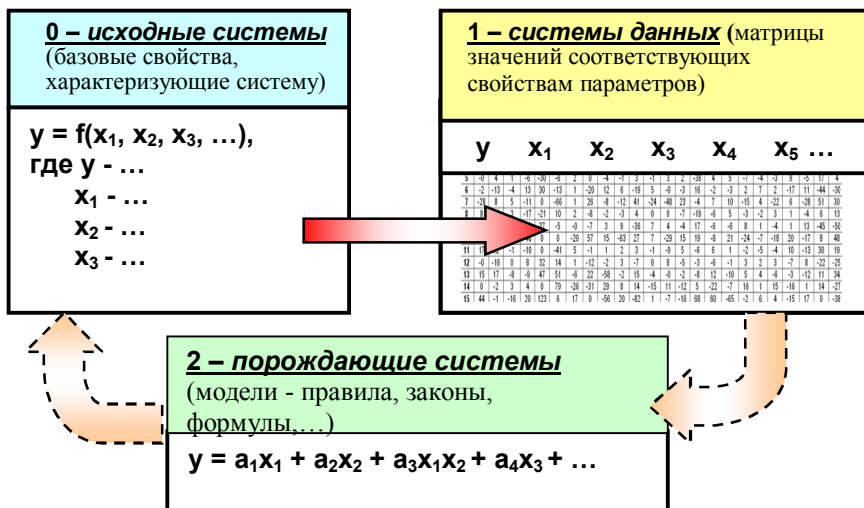


Рис. 3. Условная схема порождения и корректировки гипотез

Поскольку гипотеза не относится к единичным суждениям опыта, а всегда превосходит их по содержанию, ее нельзя обосновать, исходя только из данных. Эмпирические данные могут лишь опровергнуть гипотезу, но не подтвердить ее. Гипотеза ставится под сомнение уже в том случае, когда вступает в противоречие хотя бы с одним фактом или при ее логической проверке убеждаются, что гипотетические способы действия не приводят к цели. Но каждая новая гипотеза, как правило, не отбрасывает целиком содержание прежних гипотез, а использует все рациональное. Новая гипотеза в основе своей выступает как усовершенствованная предыдущая (поэтому К. Поппер и назвал «свою» эпистемологию *эволюционной*).

3. Гипотеза как модель

Любая гипотеза является моделью. Обратное верно для многих разновидностей моделей. Даже для макетов зданий, фотомodelей и т.д. Но не

верно для многих знаковых и/или символьных моделей: *денежных знаков, условных обозначений* и т. п.

Здесь мы рассматриваем абстрактные модели. Основные свойства таких моделей определяются в следующем дескриптивном определении:

Модель есть целевое отображение оригинала:

- *познавательное или прагматическое;*
- *статическое или динамическое;*
- *конечное, упрощенное, приближенное;*
- *имеющее наряду с безусловно истинным, условно истинное и безусловно ложное содержание,*
- *адекватное;*
- *появляющееся и развивающееся в процессе практического использования.*

Заметим, модели предназначены для обеспечения всех видов деятельности человека; *без моделирования вообще невозможно что-либо сделать*. Эта особенность очевидна на примере целенаправленной деятельности. Ведь цель субъекта - это образ желаемого будущего, т.е. модель того, чего нет, но что хотелось бы осуществить. Далее, для достижения поставленной цели составляется план действий, но ведь это не сами действия, а их описание, т.е. модель. Итак, моделирование не является таким действием, которое можно делать, а можно и не делать; *моделирование есть неизбежная, неотъемлемая, обязательная часть любой деятельности человека* (и не только человека, но об этом разговор отдельный). Таким образом, в процессе своей жизнедеятельности мы непрерывно строим гипотезы.

1. На рис. 4. представлена классическая схема метасистемы моделирования. Обратим внимание на ее соответствие схеме миров Сковороды или Поппера (см. рис. 1 и 2). На основе анализа процессов построения моделей можно сделать вывод, что отображение моделирования («объект → модель») является бинарным. Однако, оно, как минимум, тернарно: «объект → создатель → модель» [10].

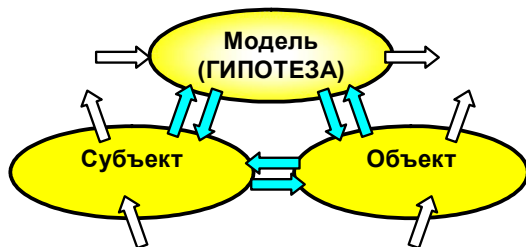


Рис. 4. Метасистема моделирования

У любой модели (гипотезы) всегда есть автор (создатель). С одной стороны, этот «конструктор гипотезы» преследует свои цели, посредством выбора наиболее существенных свойств моделируемого объекта, аппарата синтеза модели. При этом учитываются и

планируемые операции анализа модели и интеграции ее с другими моделями

(других агрегатов, подсистем, процессов функционирования сложного изделия или моделируемого сложного явления). *С другой стороны*, часто модели создаются одним исследователем, а используются другим, который преследует свои цели при интерпретации модели. Все это предъявляет особые требования к *контекстам* моделируемых в гипотезах процессами явлениям.

2. Необходимо различать *познавательное* или *прагматическое* моделирование. В первом случае мы строим гипотезы для исследуемых процессов, явлений, объектов (они уже существуют). Во втором случае строятся гипотезы, «формальные мысли», о том, что проектируется, создается, планируется; соответствующих процессов, явлений (объектов еще нет).

3. Важно различать модели *статики* и *динамики* процессов, объектов, явлений. Модели статических структур трудно представлять в форме дифференциальных уравнений, полиномов и т. п., больше подходят аппарат теории графов, решетки, семантические сети. Для гипотез описания динамики процессов – наоборот.

4. Как и модель, любая гипотеза представляет собой *конечное, упрощенное, приближенное* описание моделируемого объекта. В любой гипотезе мы можем отражать только некоторые свойства и характеристики изучаемых явлений. Поэтому, любая гипотеза – это некоторое приближение и упрощение. Заметим, только упрощенные представления помогут выявить главные закономерности исследуемых процессов.

5. В любой гипотезе всегда присутствует наряду с *безусловно истинным, условно истинное и безусловно ложное содержание*. Это – прямое следствие того, что любая гипотеза является моделью.

6. Модель, с помощью которой успешно достигается поставленная цель, называется адекватной. При этом, для познавательных моделей *истинность*, фактически, является синонимом *адекватности*, а для прагматических моделей адекватной может быть ложная модель. Например, гипотеза устройства солнечной системы Птолемея ложна, но мореплаватели тех времен достигали своих целей. Поэтому, в практической деятельности мы можем успешно применять и ложные гипотезы! Целесообразность важнее.

7. В классической науке общие свойства объектов исследований отображаются четко сформулированными законами. Закон в науке имеет характер абсолютной категории на данном уровне знаний. Он может быть либо безусловно верен, либо безусловно неверен и тогда просто отвергается. Нельзя говорить о хороших и плохих законах – такое утверждение просто лишено смысла. Точно так же нельзя говорить, что одно и то же явление можно объяснить двумя или несколькими слегка различными законами. Именно поэтому, В. В. Налимов подчеркивал [13], что в настоящее время понятие закона в науке заменяется более расплывчатым понятием модели. Таким образом, один объект может представляться с помощью различных гипотез, в зависимости от поставленных целей.

8. И, конечно, гипотезы являются «живыми организмами». Они *появляются и развиваются в процессе практического использования.*

Заключение. Представляется целесообразным заметить, часто логику определяют как науку о мышлении или правильном/правилах мышления (т. е. об операциях в мире M). С нашей точки зрения – это не совсем верно. Мы, пока, мало знаем о том, как мы размышляем. А когда говорим об этом, то мы рассуждаем. Поэтому, *логика – это наука о рассуждениях*, т. е. правилах, принципах построения корректных выводов (при переходе от мира M к миру S и операциях в мире S). А мир S – это мир знаковых моделей. Таким образом, индукция, дедукция, абдукция, аналогия и другие логические выводы – это модели операций нашего мышления. Поэтому, так важно исследовать свойства гипотезы как модели объектов мира R , перед тем, как автоматизировать процесс ее генерации и проверки.

1. Валькман Ю. Р., Быков В. С. Дедуктивные и недедуктивные аспекты в моделировании образного мышления. // Моделювання та інформаційні технології, Збірник наукових праць ІПМЕ, Київ, 2006, Випуск 35, - с. 87 - 96.
2. Валькман Ю. Р., Дембовский О. Ю. Процессы порождения и обоснования гипотез: индукция, аналогия, абдукция и дедукция. // Моделювання та інформаційні технології, Збірник наукових праць ІПМЕ, Київ, 2008, Випуск 47, - с. 97 - 107.
3. Валькман Ю. Р. Гипотезы: определения, структура, классификация, жизненные циклы // Моделювання та інформаційні технології, Збірник наукових праць ІПМЕ, Київ, 2008, Випуск 48, - с. 91- 98.
4. Валькман Ю. Р., Рыхальский А. Ю. Компьютерные технологии индуктивного формирования знаний // Моделювання та інформаційні технології, Збірник наукових праць ІПМЕ, Київ, 2009, Випуск 49, - с. 86- 92.
5. Валькман Ю. Р. Индукция – индуктивные модели - индуктивное формирование знаний // Сборник тезисов САИТ, XI международная научно-техническая конференция УНК ИПСА НТУУ «КПИ», 2009, - с. 29.
6. Валькман Ю. Р., Валькман Р. Ю. Гипотезы: определения, свойства, структура, классификация, жизненные циклы// Сборник тезисов САИТ, XI международная научно-техническая конференция УНК ИПСА НТУУ «КПИ», 2009, - с. 60.
7. Софронова Л.А. Три мира Григория Сковороды. – Москва: Индрик, 2002. – 462 с.
8. Поппер К.Р. Логика и рост научного знания. - Москва: Прогресс, 1983. – 134 с.
9. Финн В.К. Правдоподобные рассуждения в интеллектуальных системах типа ДСМ. // Итоги науки и техники. Информатика, Т.15, Москва: 1991. – с. 25-49.
10. Валькман Ю.Р. Интеллектуальные технологии исследовательского проектирования: формальные системы и семиотические модели. - Киев: Port-Royal, 1998. - 250 с.
11. Клир Дж. Системология. Автоматизация решения системных задач. - М.: Радио и связь, 1990. - 544 с.
12. Курно О. Основы теории шансов и вероятностей. – Москва: Наука, 1970 г. - 384 с.
13. Налимов В.В. Реальность нереального. Вероятностная модель бессознательного. Москва: "Мир идей", АО АКРОН, 1995.- 432 с.