

В.Б. БАРАХНИН¹:
**«ОБРАБОТКА ИНФОРМАЦИИ – ЭТО ПРОЦЕСС, КОТОРЫЙ
ПРОИСХОДИТ ПРЕЖДЕ ВСЕГО ПРИ УЧАСТИИ ЧЕЛОВЕКА»**

Ю.Ю. Черный: Владимир Борисович, поздравляю Вас с успешной защитой докторской диссертации.

В.Б. Барахнин: Спасибо.

Ю.Ю. Черный: Ваша диссертация называется «Программные системы информационного обеспечения научной деятельности: модели, структуры и алгоритмы». Специальность 05.13.17 – Теоретические основы информатики.



В.Б. Барахнин: Защита была в Московском государственном университете печати имени Ивана Федорова.

Ю.Ю. Черный: Почему именно в этом месте?

В.Б. Барахнин: Там создан Совет по теоретическим основам информатики. В Сибирском отделении РАН Совета по этой специальности нет.

¹ Барахнин Владимир Борисович – доктор технических наук, доцент, старший научный сотрудник Института вычислительных технологий СО РАН, заведующий кафедрой информационных технологий Высшего колледжа информатики НГУ. Интервью состоялось 1 марта 2011 г. Беседовал Ю.Ю. Черный.

А так как по паспорту специальность весьма широка, то в тех новосибирских вузах, где она есть, информатика воспринимается более технически.

Ю.Ю. Черный: А в университете печати?

В.Б. Баракнин: Достаточно широко. В том числе там есть специалисты по семантической информации, по обработке текстов на естественном языке. Одним из моих оппонентов был Владимир Федорович Хорошевский, зав. сектором ВЦ РАН – специалист по искусственному интеллекту. Другой оппонент – Светлана Валентиновна Мальцева, она сейчас декан «Вышки», работает на факультете бизнес-информатики.

Ю.Ю. Черный: Значит, есть еще и такая отраслевая дисциплина, как бизнес-информатика?

В.Б. Баракнин: Если Вас это интересует на уровне не столько концептов, сколько денотатов, есть даже журнал «Бизнес-информатика».

Кстати, Google – это неисчерпаемый источник сведений статистики. Когда набираешь «Бизнес-и», то четвертая строка – бизнес-информатика: бизнес-идея, бизнес-инкубатор, бизнес-инфо и бизнес-информатика. Как распространен этот термин в ином контексте – не так как в «Вышке»? 2 миллиона 960 тысяч ссылок на русском языке.

На мой взгляд, та информатика, которая была в ВИНТИ, имеет большее право пользоваться этим термином. А в школьном учебнике по информатике учат тому, как пользоваться персональным компьютером. Конечно, это немного не то. Понятно, что каждый пытался «натянуть» термин под себя...

Хочу Вам показать одну концепцию информации. Это многоуровневая модель информации Вернера Гитта.

Ю.Ю. Черный: Как Вы о ней узнали?

В.Б. Баракнин: Во время одной из конференций в Переславле-Залесском в местном отеле продавались разные книги. В том числе книги некоего российского лютеранского издательства по очень умеренному креационизму. Не такому, как у американских радикалов. В одной из них была описана эта модель.

Ю.Ю. Черный: В русском переводе?

В.Б. Баракнин: Да, это был перевод со ссылкой на Гитта. Потом я, естественно, нашел пятиуровневую модель информации. У Гитта показано, что нижний уровень модели соответствует шенноновскому значению термина «информация», три последующих – семиотической триаде (синтактика – семантика – прагматика), а верхний носит философский характер.

Ю.Ю. Черный: Недавно профессор А.В. Соколов из Санкт-Петербурга выпустил книгу «Философия информации». К.К. Колин тоже издал новую книгу. Но Гитт в их работах не упоминается. Я не уверен в том, что о Гитте знает Р.С. Гиляревский.

В.Б. Баракнин: Гитт достаточно известен – можно посмотреть «Википедию». Но его достаточно серьезно критикуют за пятый уровень.

Ю.Ю. Черный: Очень интересно, что есть такая концепция! Мне кажется, Владимир Борисович, что Вы в своей работе соединяете подходы «научно-информационной» и «компьютерной» информатик.

В.Б. Баракнин: Я начинал с создания информационных систем (специальность – 05.25.05). Потом по мере выполнения работ все это осмысливалось в гуманитарном плане – как оно будет восприниматься потребителями.

Ю.Ю. Черный: Это родилось изнутри или сложилось под влиянием обстоятельств?

В.Б. Баракнин: Изнутри. У меня есть склонность к таким делам, хоть я и заканчивал матфак.

Ю.Ю. Черный: И когда началась такая рефлексия?

В.Б. Баракнин: Лет восемь назад. Я взял в библиотеке подшивку второй серии сборника ВИНТИ «Научно-техническая информация», начиная с 1965 года, и нашел для себя много интересного. Там много публиковался Ю.А. Шрейдер. Его работы оказали на меня очень сильное влияние. Я пытался установить преемственность, за что мне иногда и доставалось. Во всех отзывах оппонентов было отмечено, что автор мало знаком с современными работами по Semantic Web. Может быть, и мало. Я не ставил перед собой такой цели. Потому что наши алгоритмы направлены, прежде всего, на слабо структурированную информацию.

Ю.Ю. Черный: Можно ли изложить идею Вашей докторской диссертации простым языком?

В.Б. Баракнин: Сверхзадача заключалась в создании алгоритма, позволяющего включить в информационный процесс слабоструктурированные документы.

Глядя на обложку какой-нибудь книги или журнала, мы видим, где авторы, где название, где аннотация, а где ключевые слова. Для распознавания этих элементов мы подсознательно используем имеющиеся знания. Чтобы научить делать то же самое и машину, в идеальном случае эти элементы должны находиться в постоянных местах. Но в большинстве публикаций не пишутся ни слово «авторы», ни слово «название». Документы, у которых

метаданные присутствуют, но не являются элементами заданных словарей, называются слабоструктурированными. Мы хотим создать такие алгоритмы, чтобы большие массивы слабоструктурированных документов могли обрабатываться автоматически. А дальше эти документы автоматически будут вовлекаться в научный процесс.

Зачем это надо делать, понятно – для поиска. Потому что, если мы вводим слово «Пушкин», нужно определиться. Пушкин – это автор? Пушкин в заголовке документа? Пушкин в аннотации документа? Или, может быть, это место издания документа – город Пушкин? То есть, в научно-информационный процесс информация должна вовлекаться уже в структурированном виде. Но если она исходно была слабоструктурированной, наши алгоритмы позволяют сделать так, чтобы ей было удобнее пользоваться.

Ю.Ю. Черный: У Вас получилось?

В.Б. Барахнин: С какими-то ограничениями – да. Удалось доказать, что так сделать можно, и кое-что уже работает.

Ю.Ю. Черный: Очень интересно! Итак, Вы стали читать журнал «НТИ» с 1965 года...

В.Б. Барахнин: Да, но и не только его. Сильное впечатление произвели монографии ВИНТИ, начиная от «Основ научной информации» вплоть до «Инфосферы» и коллективной монографии 2006 года «Информатика – наука об информации» под редакцией Р.С. Гиляревского.

Ю.Ю. Черный: И что Вам дали эти работы?

В.Б. Барахнин: Я понял, что обработка информации – это процесс, который происходит прежде всего при участии человека. Об этом говорили большие ученые – А.Н. Колмогоров, А.А. Ляпунов, Ю.А. Шрейдер. Никогда не отрицалось, что у информации есть не только источник, но и потребитель, и ее ценность во многом определяется свойствами потребителя. Но обычно это как-то забывается.

У нас компьютерной информатикой занимаются люди, почти все вышедшие из математики. Особой склонности к гуманитарным предметам они не имеют. Есть, конечно, исключения – как, например, В.С. Диев – декан философского факультета НГУ. Или О.В. Шарыпов – доктор физико-математических наук и доктор философских наук. Но обычно те, кто занимается компьютерной информатикой, к гуманитарным задачам испытывают некоторое предубеждение.

С другой стороны, научно-информационная информатика не всегда в состоянии разрабатывать алгоритмы и опирается во многом на то, что ей предлагает компьютерная информатика. Кроме того, до последнего времени

мейнстримом считалась Semantic Web. Но внедрять Semantic Web в России – все равно, что на какую-то нашу дорогу вывести спортивный автомобиль и сказать: «Езжай!». Конечно, на автобане это может быть и хорошо, но по нашим дорогам лучше ездить на джипах.

Ю.Ю. Черный: Владимир Борисович, не кажется ли Вам, что настоящей науки об информации еще нет, она только создается?

В.Б. Баракнин: Можно начать с того, какой смысл мы вкладываем в понятие науки. С чем работает наука? – С объективной реальностью. Все-таки мы в данном разговоре не стоим на позициях конвенционализма Куна и Поппера. Информатика изучает и некую объективную реальность, и интересубъективную реальность, созданную человеком. Кстати, информационную среду можно сравнить с архитектурой.

Ю.Ю. Черный: Согласен. Будем считать, что это не только объективная, но и интересубъективная реальность.

В.Б. Баракнин: С другой стороны, после закрепления на материальных носителях она становится уже объективированной. В этом смысле интересно исследование И.М. Зацмана.

Можно долго развивать тему, почему электронные носители аудио уже практически вытеснили аналоговые, а бумажные книги пока активно издаются. Книга – это непосредственный источник информации, ее можно взять всегда с полки. С электронной книгой могут случиться определенные проблемы в силу ряда причин, связанных с двойным кодированием, Та энергия, которая сосредоточена в твердой копии, принципиально отличает ее от электронного носителя.

Ю.Ю. Черный: Понимаю. Повторю свой вопрос в более конкретном виде: «Не наблюдаем ли мы сегодня рождение нового качества информатики в силу переплетения ее документалистской и компьютерной ветвей?»

В.Б. Баракнин: Носителей первого подхода в чистом виде я представляю мало, потому что сейчас все библиотечные работники так или иначе участвуют в консорциумах, связанных с электронным документооборотом. Недавно на одной из конференций выступал директор БЕН РАН Н.Е. Каленов. Из новосибирских специалистов этим занимаются Н.С. Редькина, С.Р. Баженов.

С другой стороны, чтобы работать с документом, как это, например, предлагает Semantic Web, его надо предварительно обработать. Вспоминается ответ Толстого на вопрос, о чем его роман Анна Каренина: «Если бы можно было ответить короче, я бы и написал короче». То же самое и с документом. Чтобы его обработать, надо проделать не менее квалифицированную работу, чем это сделали авторы.

Ю.Ю. Черный: Приведу пример из нашей институтской практики. Иногда рефераты книг получаются интереснее первоисточников, потому что референт обладает бóльшим кругозором, понимает проблему шире, нежели сам автор.

В.Б. Барахнин: Совершенно верно. Но к каждому интернет-документу такого референта не приставишь, и потому возникают большие проблемы – как будут создаваться документы в плане Semantic Web. В рамках консорциума w3 было предложено создать специальный метатег, чтобы указывать в нем название, автора и так далее. Конечно, уже тогда, в середине 90-х годов, рассчитывали, что поисковые системы, «заглядывая» при просмотре интернет-документов в этот метатег, будут воспринимать его как структурированный текст. Но будем говорить честно – реально никто этого не сделал. Тем не менее консорциум w3 в 2001 году предложил: «Давайте-ка еще углубим этот подход!» В итоге создается интересный проект с, может быть, какими-то важными результатами, но все это повисает в воздухе...

Ю.Ю. Черный: Почему?

В.Б. Барахнин: Потому что это реально оторвано от того, как сделан документ. Как говорится, если теория противоречит фактам, тем хуже для фактов. Мой подход исходит из реальности. Наши алгоритмы, конечно, менее сложны с математической точки зрения. Мы опираемся на «научно-информационную» информатику. Нам это кажется более эффективным.

Ю.Ю. Черный: Р.С. Гиляревский проводит различие между данными и информацией, говоря о том, что данные – это нечто материальное, а информация – это смысл, возникающий при интерпретации данных в сознании человека. Как Вы относитесь к этой идее?

В.Б. Барахнин: Считаю, что различие между информацией и данными проведено в этом случае абсолютно корректно. Данные соответствуют синтаксическому уровню сообщения, информация в узком смысле – семантическому, а знания – прагматическому. В принципе, данные тоже зависят от восприятия всего этого системой. Но, конечно, если рассматривать информацию на уровне понимания, то это семантика. В схеме Гитта нет понятий «данные», «информация» и «знание», но мы обозначили в ней эти три уровня.

Ю.Ю. Черный: Еще в некоторых работах, в частности, в книге А.В. Соколова «Философия информации» встречается мысль о том, что научная информация есть не что иное, что научное знание, рассматриваемое в коммуникационном аспекте.

В.Б. Барахнин: В принципе, да. Можно задаться вопросом, что такое факт и отличается ли документальный поиск от библиографического поиска.

Возьмем простейший пример – кто убил Кеннеди. Вроде бы, считается, что Освальд. Тем не менее группа видных ученых доказывала, что это не так. Можно ли считать утверждение, что Кеннеди убил Освальд, фактом? Вывод здесь таков: библиографический поиск от документального ничем не отличается.

Более того, в связи с этим возникла ситуация в контексте моей защиты. Меня спросили: «Позволяет ли Ваша система, Ваш метод отделять правильный факт от неправильного? Сможет ли она выявить статью, сгенерированную машиной? (Наверное, Вы помните скандал с такой статьей, которая была опубликована в ВАКовском журнале?)» Я ответил, что нет. Все-таки мы работаем с некоторой информацией как с данностью. Она уже стала достоянием научной среды. Возможность сравнивать эти данные с другими данными относится к другому уровню работы. Это методы Data Mining, искусственный интеллект. Конечно, искусственный интеллект тоже применяется в этих исследованиях. Но все же это – не проблема информатики как науки об информации...

Ю.Ю. Черный: А Вы информатику понимаете именно так?

В.Б. Барахнин: У каждой науки должны быть свои объект, предмет и рабочее определение. По большому счету, это близко к тому (может быть, с некими коррективами) к определению Михайлова, Черного и Гиляревского.

Ю.Ю. Черный: То есть, информатика – наука о сборе, переработке, хранении и распространении информации, закрепленной в научных документах?

В.Б. Барахнин: Именно так. То, как она делается научной, какую она проходит экспертизу – немного другой вопрос. Но если, пройдя соответствующие процедуры, она стала достоянием научного сообщества, то она стала научной информацией. С ней мы и работаем.

Ю.Ю. Черный: И последний вопрос, Владимир Борисович. Работая над этой темой, связанной с информатикой, я постоянно думаю о том, что должно произойти в будущем: например, одна из «информатик» должна отказаться от этого наименования – скажем, стать «компьютерной наукой», а информатика в смысле ВИНИТИ остаться? Или наоборот – информатика ВИНИТИ станет «информационной наукой», а вторая, «компьютерная» информатика останется с прежним названием. А, может быть, должна быть создана какая-то единая наука, включающая в себя обе ветви?

В.Б. Барахнин: Сильно сомневаюсь, что это может быть сделано. Ставить задачу «связать» науку – это было бы слишком смело. Мы движемся и размышляем. Будет ли создана такая наука? Я всегда с осторожностью

относился к подобным интегральным проектам, но считаю, что двигаться в эту сторону надо. Надеюсь, моя и Ваша работа будут этому способствовать.