

**О. В. Сюттюренко**

### **Направления перспективного развития информационной деятельности ВИНТИ**

*Рассмотрены направления разработки и реализации новой научно-информационной политики развития ВИНТИ РАН. Основная цель – показать перспективные подходы к более глубокой переработке информации, многомерному анализу накопленных данных, возможности создания актуальных видов информационной продукции и услуг, ориентированных на поддержку решений как в научно-технической и промышленной сферах, так и, опосредованно, в экономике в целом.*

**Ключевые слова:** переработка информации, многомерный анализ накопленных данных, актуальные виды информационной продукции, Всероссийский институт научной и технической информации.

В 1953 г. базовой концепцией создания ВИНТИ являлась организация национального центра реферирования мирового потока научно-технической литературы по всем направлениям фундаментальных и прикладных исследований (расширенный, содержательный реферат).

В связи с быстрым развитием информационных технологий, электронных информационных ресурсов, сети Интернет, радикальными изменениями глобальной информационной среды базовая концепция 1953 г. практически исчерпала себя. Предпринятые руководством страны шаги по реформированию Российской академии наук ещё более актуализировали проблему структурно-функциональной модернизации ВИНТИ и развития его информационной деятельности в соответствии с новыми вызовами и задачами создания инновационной экономики России.

Основная системная проблема отечественной информатики, и ВИНТИ в частности, заключается в том, что существующий рынок информационных продуктов и услуг не позволяет в полной мере удовлетворять растущий спрос и расширяющийся спектр информационных потребностей пользователей из сферы управления народным хозяйством, из среды «генерации знаний», инновационного и производственного сектора, сферы образования.

Некоторые шаги в направлении решения этой проблемы как на концептуальном, так и на организационно-технологическом уровне планируются и реализуются за счёт расширения функциональных возможностей Единой технологической базы данных (ЕТБД), развития информационно-аналитической деятельности и информационного моделирования, новых сервисов интернет-доступа к БД ВИНТИ. Определённые результаты уже получены в ходе выполнения Институтом ряда научных проектов в составе федеральных и академических программ.

Цель этого доклада – показать перспективные подходы к более глубокой переработке информации, анализу зафиксированных данных, возможности создания новых видов информационных продуктов и услуг, ориентированных на поддержку исследований и управленческих решений как в научно-технической и промышленной сфере, так и, опосредованно, в экономике в целом.

Основные направления работ по созданию новых информационных продуктов и услуг ВИНТИ показаны на рис. 1.



**Рис. 1. Реферативный банк данных ВИНТИ: постобработка информации с использованием методов анализа данных (С)**

Банк данных ВИНТИ содержит свыше 35 млн записей (с глубиной ретроспективы по некоторым предметным областям до 15 лет). Использование статистических методов при постобработке реферативной и библиографической информации такого объёма представляется весьма перспективным для решения целого ряда задач [1], в числе которых:

- анализ структуры отечественной и мировой науки;
- определение тенденций и процессов, происходящих в мировой и региональной науке;
- выявление наиболее актуальных или, напротив, теряющих свою актуальность научных направлений;
- отслеживание генезиса конкретных научных идей и истории их развития;
- определение продуктивности работы исследователей в конкретной научной области и эффективности материальных затрат в этой области;
- анализ структуры научного сообщества и изучение науки как социального организма.

Социальная, экономическая, демографическая статистика в развитых странах рассматривается как эмпирический базис для управления народным хозяйством и проведения различного рода реформ. На основе анализа распределений в статистике делаются основные выводы и заключения. Постобработка больших массивов научно-технической и технико-экономической информации с использованием статистических методов, методов анализа данных априори позволяет выявлять статистические закономерности, выражающие зависимости между распределениями различных параметров исследуемых систем и процессов и характер изменения распределений во времени [2].

При создании подсистемы постобработки (условно назовем её «Аналитика-статистика») целесообразно применение «коробочных» программных продуктов, предназначенных для статистической обработки и многомерного анализа данных (с использованием алгоритмов факторного, ковариационного, корреляционного анализа данных). Областью применения такой подсистемы должны быть исследования (продукты и услуги), связанные с задачами:

- а) выявления эмпирических закономерностей, объективно существующих для объектов и процессов в науке, образовании, экономике;
- б) прогнозирования динамики изменения показателей многомерных, например технико-экономических, объектов и процессов во времени;
- в) сопоставительного анализа уровня научных исследований, технических и экономических объектов (на основе аппарата теории выбора, в том числе по критерию Парето);
- г) анализа и восполнения пропущенных и сомнительных данных;
- д) визуализации и графического представления результирующих данных постобработки.

Исходной ресурсной базой подсистемы «Аналитика-статистика», помимо реферативного БНД ВИНТИ, могут быть и ресурсы БНД Российского фонда фундаментальных исследований ([www.rfbr.ru](http://www.rfbr.ru)), Роснауки ([www.fcntp.ru](http://www.fcntp.ru)), Интернета, прежде всего Росстата ([www.gks.ru](http://www.gks.ru)), Национального научного фонда США (NSF, [www.nsf.com](http://www.nsf.com)), Института научной информации США (ISI, [www.isinet.com](http://www.isinet.com)), *The Scientific World* ([www.thescientificworld.com](http://www.thescientificworld.com)), Американского химического общества (CAS). База данных CAS (сервисная служба *Chemical Abstracts*, [www.cas.org](http://www.cas.org)) уже содержит свыше 100 млн ссылок.

Следует отметить, что совместная постобработка информации БНД ВИНТИ и, например, данных Росстата (таких как величин: ВВП, произведённой энергии, среднего годового дохода на душу населения, произведённого продукта с использованием высоких технологий и ряда других) – это перспективное множество представляющих практический интерес статистических показателей и распределений. Перечислим некоторые из них:

анализ сравнительного роста:

- а) валового внутреннего продукта;
- б) расходов на образование;
- в) расходов на исследования и разработки;
- г) объёма публикаций российских авторов;

анализ изменений структуры ВВП и структуры публикаций российских авторов;

анализ зависимости роста объёмов инвестиций в народное хозяйство и роста объёмов публикаций (по отраслям народного хозяйства);

анализ зависимости роста выпуска специалистов государственных и муниципальных вузов и роста объёмов публикаций (по отраслям народного хозяйства).

Некоторые результаты постобработки (с использованием данных Росстата) показаны в графическом виде на рис. 2.

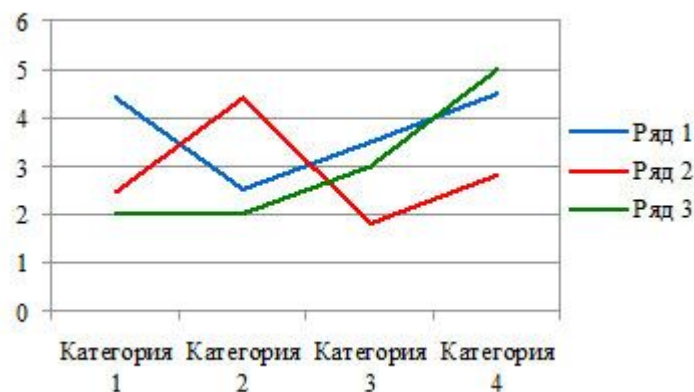


Рис. 2. Результаты постобработки – графики распределений:

- 1 – оценка корреляции изменения бюджета науки и роста числа публикаций по годам (с использованием данных Росстата);
- 2 – оценка зависимости числа публикаций от роста ВВП (по годам);
- 3 – анализ сравнительного роста ВВП и расходов на образование.

В заключение следует подчеркнуть, что целенаправленное использование методов, средств (продуктов и услуг) статистической постобработки информационных ресурсов ВИНТИ стало бы реальным вкладом как в развитие информатики, так и в становление инновационной экономики в нашей стране, а также в перспективе могло бы трансформироваться в новое научное направление «сетевой» наукометрии.

### **Информационные продукты прогнозно-аналитического и обзорного характера (D)**

Необходимо воссоздание на базе новых информационных технологий традиционного для ВИНТИ направления переработки информации с выходными продуктами прогнозно-аналитического и обзорного характера. Например, подготовка ежемесячных выпусков предметно-тематических и /или проблемно-ориентированных экспресс-информационных материалов следующей структуры: краткий обзор (~ 5 с.); библиографическая подборка (40–60 с.).

Ключевые задачи: определение актуальных тематик и создание условий для привлечения к сотрудничеству квалифицированных специалистов, номинация информационных продуктов и услуг, оценка издержек и расчёт ценообразования.

Реализация рассмотренных направлений, помимо статусных и экономических выгод, влияет на расширение возможностей использования результатов прогнозно-аналитической наукометрической деятельности в научно-промышленной сфере и управлении народным хозяйством, а также создаёт реальную основу для: а) анализа структуры отечественной и мировой науки; б) определения тенденций и процессов в научно-технической сфере; в) выявления точек роста, наиболее актуальных и/или стагнирующих научных направлений; г) мониторинга структуры (программ) отечественного научно-промышленного комплекса.

### **Создание распределённой функционально-ориентированной системы информационной поддержки инновационной деятельности (B)**

В результате выполнения комплекса работ должна быть реализована система, ядром которой

является мегапортал, обеспечивающий навигацию и доступ пользователей к серверам информационных систем, хранящих полнотекстовую информацию о результатах исследований и разработок, в том числе тех, которые могут иметь дальнейшую промышленную коммерческую реализацию [3–5]. В рамках первой очереди – комплекс работ с ИС РФФИ, ИС Федеральной целевой научно-технической программы (ФЦНТП), возможно и с ИС Министерства экономического развития РФ.

Наряду с этим в ВИНТИ должна быть разработана интерактивная подсистема, в которую включены следующие элементы: индикативная БД инноваций, БД потенциальных инвесторов, БД предприятий и организаций, заинтересованных в поиске и внедрении тех или иных научно-технических разработок. Концептуальным прототипом этого направления является система *CORDIS* – интерактивная информационная платформа в области европейских инноваций, исследований и разработок. Информационная служба Евросоюза *CORDIS* посредством ряда подсистем, средств и 10 поисковых БД предоставляет пользователям результаты исследований и разработок по всему инновационному циклу. На начало 2012 г. зарегистрировано свыше 300 тыс. пользователей *CORDIS*.

### **Организация информационного обслуживания в среде Интернета на основе полнотекстового банка данных (А)**

Перспективный прецедент такого решения – реализация проекта по созданию Научной электронной библиотеки (1999 г.) по инициативе и при финансовой поддержке РФФИ [2]. В НЭБ ([www.elibrary.ru](http://www.elibrary.ru)) общее число наименований научных журналов – около 30 тыс.; общее число выпусков журналов – 1,2 млн; общее число российских журналов в открытом доступе – 3 тыс.; общее число полнотекстовых статей – 19 млн. Общий объём ресурсов превышает 5 терабайт. Заключены договоры с 14 400 издательствами.

Пользователями НЭБ до последнего времени являлись более 11 666 научных и образовательных учреждений страны (зарегистрированных пользователей свыше 980 тыс.). Среднее количество обращений в сутки – 4 тыс. (пик – 10 тыс.).

В настоящее время ситуация с НЭБ благоприятна для ВИНТИ в том плане, что финансовые средства РФФИ выделяются не на развитие и комплектование НЭБ, а на оплату сетевого доступа научных организаций к зарубежным БД. Следует отметить, что, несмотря на значительный объём информационного обслуживания, ни у НЭБ, ни у РФФИ не было серьёзных осложнений или конфликтных ситуаций с зарубежными издательствами.

Полнотекстовый банк данных ВИНТИ, помимо зарубежных изданий, должен содержать все отечественные научно-технические журналы. Для реализации этого направления необходимо выполнить подготовительные работы значительного объёма (разработка программного обеспечения, договора, правовые вопросы, стоимостная оптимизация и др.). Критической позицией является разработка и/или адаптация надёжно функционирующей биллинговой системы и организация системы взаиморасчётов (с издательствами и другими поставщиками данных).

### **Создание доступной через Интернет БД непубликуемых материалов (Б)**

Также как и для предыдущего направления, основная задача этого – внедрение технологии оцифровки данных с бумажных носителей.

### **Организация информационного обслуживания на базе электронного РЖ (с индикативным рефератом) в сети Интернет (В)**

Основные критические пункты реструктуризации: 1) переориентация на индикативный реферат; 2) широкое использование аннотаций статей; 3) радикальный вариант – по каждой статье даётся реферат на языке оригинала и русский текст названия и аннотация после автоматического (автоматизированного) перевода (для английского, немецкого, французского языков); 4) автоиндексирование статей; 5) минимизация временного лага < 1–1,5 месяцев; 6) реализация режимов: электронного ИРИ, представление данных по произвольным выборкам и срезам, информационного мониторинга (по работам, проектам и/или программам); 7) детальная подготовка и проведение, параллельно с традиционной технологией, пилотного цикла с добавлением в текст реферата рисунков, формул, графики.

### **Создание доступной через Интернет базы данных по производимой и потребляемой промышленной продукции, прежде всего российской и стран СНГ (G)**

Источники комплектования: промышленные каталоги и буклеты, материалы выставок, ресурсы Интернета и др. Эта БД может существенно дополнить информационную поддержку инновационной деятельности.

### **Реализация целевого интернет-СМИ (СМнтИ) нового поколения (M)**

Предпосылки – лавинообразный рост информационных интернет-ресурсов. В 2011 г. только текстовой информации появилось больше, чем за всё существование человечества. Из всех стран мира наибольший рост числа интернет-пользователей за период 2000–2011 гг. наблюдался в России – 1 883%. Из-за огромных объёмов информации в открытом доступе она практически необозрима. Можно с уверенностью сказать, что потенциально любой человек мог бы найти достаточно много интересной, новой и актуальной информации, однако на данный момент отсутствуют механизмы агрегации, обобщения и персонализации информации сети Интернет.

**Задача ВИНТИ** (в русле дифференциации информационных услуг): целевое предоставление информации научно-технического и экономического характера населению из открытых источников Интернета.

**Реализация.** Внедрение системы кластеризации потоковой информации открытых источников с использованием методов построения адаптивных гипермедиа на основе технологии кластеризации неструктурированных данных и обеспечения способа донесения актуальной, лингвистически обработанной информации до различных целевых групп и отдельных пользователей в соответствии с их персональными потребностями и ожиданиями (конвергированное ИРИ нового поколения).

### **Создание доступных через Интернет баз данных (H):**

а) по кабинету фирм (отечественных и зарубежных);

б) по федеральным, ведомственным, межведомственным и региональным научно-техническим и промышленным программам.

В заключение необходимо отметить, что в реализации рассмотренных направлений развития информационной деятельности ВИНТИ помогут значительные заделы и сохранившийся научно-технический потенциал. По каждому из направлений в сжатые сроки можно подготовить развёрнутый план работ и рассчитать объёмы дополнительного финансирования. Безусловно, необходимы определённые шаги по укреплению и обновлению кадровых

ресурсов.

#### **Список источников**

- 1. Борисова Л. Ф., Сюнтюрченко О. В.** Реферативный банк данных ВИНТИ РАН: перспективы постобработки информации с использованием методов анализа данных // НТИ. Сер. 1 – 2007. – № 11. – С. 6–11.
- 2. Симонов О. В., Сюнтюрченко О. В., Черепанов Е. В.** Некоторые автоматизированные процедуры многомерного анализа технико-экономических данных // Техника средств связи. Сер. ТРПА. – 1985. – Вып. 2. – С. 55, 56.
- 3. Борисова Л. Ф., Сюнтюрченко О. В.** Проблемы информационного обеспечения научно-инновационной и промышленной сферы: новые концептуальные подходы // НТИ. Сер. 1. – 2009. – № 4. – С. 9–12.
- 4. Борисова Л. Ф., Сюнтюрченко О. В.** Методологические проблемы создания национальной системы информационной поддержки инновационной деятельности и социальная оценка техники / Труды XIV Международ. конф. «LIBCOM–2010» – «Информационные технологии, компьютерные системы и издательская продукция для библиотек»; п. Ершово. 15–19 ноября 2010. – Москва : ГПНТБ, 2010. – Режим доступа: <http://gpntb.ru/libcom2010>.
- 5. Сюнтюрченко О. В.** Концептуальные положения структурно-функционального развития ВИНТИ РАН / Материалы 8-й Международ. конф. «Актуальные проблемы информационного обеспечения науки, аналитической и инновационной деятельности», Москва 28–30 ноября 2012, ВИНТИ РАН. – Москва, 2012. – С. 154–157.