

Информатические описания.

Опус 13. Информационный поиск как разновидность аналитико-синтетической переработки информации

Исходя из понимания аналитико-синтетической переработки информации как профессионального процесса библиотечно-библиографического творчества, информационный поиск рассматривается в качестве одной из её разновидностей.

Ключевые слова: информационный поиск, аналитико-синтетическая переработка информации, библиотечно-информационная деятельность, машиночитаемая каталогизация, оцифровка, Национальный библиотечно-информационный фонд.

Задача этого описания – выяснить логическое соотношение между понятиями *информационный поиск* и *аналитико-синтетическая переработка информации*. Такая потребность возникла в связи с определением содержания курса «Аналитико-синтетическая переработка информации», который в соответствии с Федеральным государственным образовательным стандартом третьего поколения включён в учебный план подготовки бакалавров по направлению «Библиотечно-информационная деятельность».

Возможны два взаимоисключающих решения: 1) информационный поиск и аналитико-синтетическая переработка информации (АСПИ) – самостоятельные рабочие процессы, поэтому обозначающие их понятия логически независимы; 2) информационный поиск – разновидность АСПИ, т.е. соответствующие понятия связаны родовидовыми отношениями, и, следовательно, всякий информационный поиск представляет собой АСПИ.

Это формально-логическое упражнение имеет методологическое значение для педагогической практики библиотечно-информационных факультетов. Согласно первому решению, рассматривать информационно-поисковые проблемы в рамках курса АСПИ нелогично, а второе решение предписывает учитывать все виды аналитико-синтетической переработки, в том числе – информационный поиск. Конечно, учебные планы строятся не по логическому эталону, и содержания учебных дисциплин часто дублируют друг друга, а иногда вступают в противоречие. Но всё-таки хотелось бы минимизировать нелепости учебного процесса.

Обратимся, как советовали древние философы, к истокам.

В классической библиотечно-библиографической терминологии не было терминов *аналитико-синтетическая переработка информации* и *информационный поиск*, в ходу были выражения «обработка литературы (новых поступлений)» и «библиографическое разыскание». Существовали *научная* обработка – составление каталожного описания, систематизация, предметизация и *техническая* обработка – инвентаризация, штемпелевание, простановка авторских знаков и расстановочных шифров, подготовка формуляров и пр.

Библиографическое разыскание мыслилось как «логически обоснованные и рационально проводимые целенаправленные поиски сведений о произведениях печати, опирающиеся на высокую идейно-политическую, общекультурную, источниковедческую и книговедческую подготовку библиографа» [1. С. 44].

Очевидно, что научная обработка и библиографическое разыскание представляют собой

операции умственного труда, требующие понимания различных текстов, анализа их содержания и синтезирования новых текстов.

В 1950-е – 1960-е гг. библиотековеды и библиографоведы не использовали информационный подход и не называли анализ и синтез текстов «переработкой информации», а библиографическое разыскание «информационным поиском». Истоки интересующих нас понятий обнаруживаются в информатике.

1. Исходное понятие об аналитико-синтетической переработке информации

Термин *аналитико-синтетическая переработка информации* появился в научном лексиконе с лёгкой руки основоположников научной информатики А. И. Михайлова, А. И. Чёрного, Р. С. Гиляревского. Правда, сначала в своих монографиях в 1960-х гг. они говорили об аналитико-синтетической переработке не информации, а *документов*, называя в качестве основных видов такой переработки «библиографическое описание документов, их классификацию (индексирование), аннотирование, реферирование, перевод с одного языка на другой, а также составление обзоров» [2. С. 164; 9. С. 158].

В 1976 г., рассматривая проблемы «анализа и синтеза научной информации», тот же перечень рабочих процессов они предпочли именовать «аналитико-синтетической переработкой научных документов и информации» [3. С. 289].

В дальнейшем информационные работники, практикующие информационную, а никак не документационную деятельность, решили избавиться от упоминания о документах, и в научно-информационной среде получил признание термин *аналитико-синтетическая переработка информации*.

Практически мыслящие библиотековеды первоначально неодобрительно отнеслись к термину *переработка*. С 1993 г. в МГУКИ читался курс «Аналитико-синтетическая обработка документов» (АСОД). АСОД трактовалась так же, как переработка документов в монографиях классиков информатики, а именно как «совокупность взаимосвязанных, взаимозависимых процессов формирования библиографической записи посредством анализа и синтеза. При анализе документа выявляются его основные библиографические сведения, индексационные, аннотационные, реферативные признаки, которые синтезируются в библиографическом описании, заголовке, классификационном индексе, предметной рубрике, ключевом слове, аннотации или реферате и т.д.» [4].

В результате в системе документальной коммуникации на основе фондов первичных документов порождаются (генерируются) потоки вторичных документов в виде библиографических (каталожных) описаний, реферативной и обзорной информации. Возникла терминологическая неувязка: библиотекари и библиографы считали создание вторичных документов «обработкой» первоисточников, а информационные работники усматривали в этом «переработку» первичной информации во вторичную. Строго говоря, оба термина небезупречны: каталогизация литературы – это явно «обработка», а не «переработка» книг; в то же время аннотирование книги логично считать аналитико-синтетической «переработкой», а не «обработкой» её содержания.

В 2000-е гг. формирование глобальных информационных сетей и многообразных интернет-приложений способствовало распространению представления об особом (виртуальном) мире информации. Приобрела актуальность задача не только учётно-регистрационной обработки циркулирующих в этом мире информационных потоков, но и содержательной их переработки. Этой задаче довольно точно соответствует понятие *аналитико-синтетическая*

переработка информации. Оно получило признание педагогического библиотечно-информационного сообщества, посчитавшего в 2003 г. необходимым образовать в библиотечно-информационной школе курс «Аналитико-синтетическая переработка информации».

В 2008 г. вышел в свет первый учебник «Аналитико-синтетическая переработка информации», предназначенный для студентов вузов, обучающихся по специальности «Библиотечно-информационная деятельность» [5]. Авторы поставили задачу «представить всю совокупность взаимосвязанных процессов обработки документа в едином технологическом цикле с учётом главных изменений, произошедших в этой области за последние десятилетия» [5. С. 9]. Основное внимание уделялось формированию библиографических записей в человекочитаемой и машиночитаемой форме; охарактеризованы также такие важные процессы смысловой переработки документов, как предметизация, систематизация, координатное индексирование, аннотирование и реферирование документов. В учебнике впервые представлены в общем контексте тематически сходные и логически взаимосвязанные разделы библиотековедения, библиографоведения, информатики и приняты следующие логические соотношения между понятиями:

1. Понятие *аналитико-синтетическая переработка информации* есть *родовое* (обобщающее) по отношению к таким *видам* библиотечно-информационной деятельности, как Описание документов, Предметизация, Систематизация, Аннотирование, Реферирование, Подготовка обзоров.

2. Понятие *информационный поиск* не является разновидностью АСПИ. Следовательно, ни библиографическое разыскание, ни библиографический поиск, ни другие библиотечно-библиографические поисковые операции с текстами, например комплектование документных фондов, АСПИ считать нельзя.

Невозможно согласиться с подобной логической структурой, потому что она не учитывает главную особенность АСПИ – *творческий характер* этого процесса. Дело в том, что переработка информации – это не превращение одного материального предмета в другой, вроде размола зерна в муку или выпечки хлеба из муки. Речь идёт об умственных (интеллектуальных) процессах, нацеленных на получение новых, вторичных, текстов путём анализа и синтеза смыслового содержания исходных, первичных, текстов. Исключение из логической структуры АСПИ интеллектуальной операции информационного поиска сильно обедняет профессиональное библиотечно-информационное творчество. Естественно, возникает вопрос: что понимается под «творчеством» в этом случае?

2. Понятие о библиотечно-информационном творчестве

В русском языке слово «творчество» понимается двояко: во-первых, как деятельность, порождающая инновацию (новшество) – качественно новую и общественно полезную ценность; во-вторых, как совокупность ценностей, сотворённых данным субъектом (индивидом или социальной группой). Способность к творчеству – неотъемлемый атрибут человека, который проявляется в труде. В библиотечно-информационной деятельности результатами индивидуального творчества являются библиографические и фактические справки, рефераты и обзоры, подготовленные креативными сотрудниками.

Коллективное творчество библиотечного сообщества воплощено в системе библиотек и библиографических фондах, являющихся важнейшей составной частью национального культурного наследия.

Всякая общественно организованная трудовая деятельность нуждается как в инновациях, так

и в воспроизведении положительного опыта, поэтому на практике сосуществуют и тесно переплетаются компоненты творчества и нетворческого исполнительства. Типичный компонент творчества – внедрение новых средств и методов, заменяющих привычные. Нетворческие компоненты заключаются в репродуцировании (повторении, воспроизведении) положительно зарекомендовавших себя приёмов и операций. Следовательно, процессы библиотечно-информационной деятельности делятся на *творческие* (продуктивные) и *нетворческие* (технические, рутинные). Так, ксерокопирование журнальной статьи – операция нетворческая, а реферирование той же статьи – творческая переработка информации; расстановка карточек в библиотечном каталоге – нетворческая, а поиск информации по запросу в том же каталоге – творческая.

В процессе библиотечно-библиографической деятельности часто приходится сортировать массу документов (у библиотекарей порой вырабатывается навык профессионального скорочтения, вплоть до «чтения пальцами»), но это ещё не АСПИ. Даже обращение со справочным запросом к ресурсам Интернета и автоматическую распечатку полученных данных нельзя признать АСПИ. Принципиальное различие между творческой и нетворческой деятельностью заключается в том, что результат последней предопределён заранее, нужно только строго следовать предписанным алгоритмам, а результаты творчества непредсказуемы, потому что они зависят от субъективных качеств исполнителя.

Большой плюс формализованных (автоматизированных) нетворческих операций – *однозначность* (воспроизводимость) их результатов, обусловленная независимостью от человеческого фактора. Отсюда – стремление к формализации (автоматизации) библиотечно-информационных технологий. Если при реферировании статьи специалистами мы получаем столько же разных рефератов, сколько было референтов, то автоматическое реферирование неизменно выдает один и тот же результат. Однако важно подчеркнуть, что нельзя считать критерием творчества только новизну или однозначность, забывая об этическом и эстетическом аспектах, о ценностных и нравственных оценках. Творчество только тогда является подлинно человеческим, когда согласуется с идеалами истины, добра и красоты.

В завершение экскурса в философию творчества процитирую определение Л. В. Яценко: «Творчество – это социально обусловленная целенаправленная созидательная деятельность, в которой полнее всего представлены продуктивные компоненты, присущие человеческому труду; эта деятельность направлена на порождение принципиально, качественно новых, нравственно оправданных и эстетически совершенных культурных ценностей. Творчество – процесс, благодаря которому происходит восхождение человека к высшим ступеням духовности, нравственности и свободы, развитие личности и культуры» [6. С. 21].

Теперь, получив представление о категории творчества, можно сформулировать понятие библиотечно-информационного творчества. *Библиотечно-информационное творчество* – это целенаправленное создание в процессе индивидуальной или коллективной библиотечно-информационной деятельности качественно новых общественно полезных продуктов и услуг. Анализ и синтез играют важнейшую роль во всех творческих процессах, поэтому есть основание рассмотреть АСПИ как компонент библиотечно-информационного творчества.

3. Аналитико-синтетическая переработка информации как область творчества

Чтобы понять творческий потенциал АСПИ, нужно уяснить: существуют анализ, синтез, информация. Начнём с методологии. Анализ и синтез – важнейшие *методы познания*, т.е. получения истинного знания, которые используются во всех отраслях науки и общественной практики.

Анализ (analysis) – начальный этап познания, заключающийся в мысленном расчленении

(дезинтеграции) познаваемого объекта на составные части (элементы, аспекты) и изучении строения, свойств, отношений каждой из частей.

Синтез (synthesis) – завершающий этап познания, состоящий в мысленном воссоединении (интеграции) проанализированных частей в единое целое и получении нового знания об изучаемом объекте. Результаты аналитико-синтетического познания – всестороннее описание и комплексная оценка изучаемых объектов, раскрытие их структуры и функций, построение классификаций и теоретических моделей, нахождение новых оригинальных решений в практической и теоретической деятельности.

Объектами анализа и синтеза могут быть всевозможные материальные и идеальные предметы, начиная с астрономических светил и заканчивая древнеславянским фольклором. В случае АСПИ объектом является *информация*, которая представляет собой не вещь или воспринимаемый органами чувств атрибут вещи, а умопостигаемый идеальный предмет. Понятие информации используется во всех отраслях знания, но в различных значениях, поэтому необходимо уточнить его содержание. В Опусе 5, посвящённом природе и сущности информации, предложены общенаучное определение информации и трактовки различных типов информации [7]. Напомню некоторые исходные положения.

Один из основоположников информатики академик В. М. Глушков (1923–1982) ровно полвека тому назад писал: «Информация в самом общем её понимании представляет собой меру неоднородности распределения материи и энергии в пространстве и во времени, меру изменений, которыми сопровождаются все протекающие в мире процессы. Совершенно не обязательно непременно связывать с понятием информации требование её осмысленности, как это имеет место при обычном, житейском понимании этого термина. Информацию несут в себе не только испещрённые буквами листы бумаги или человеческая речь, но и солнечный свет, складки горного хребта, шум водопада, шелест листвы и т.д.» [8]. Ясно, что столь широкая, философская трактовка информации не годится в нашем случае, ибо библиотечно-информационная деятельность имеет дело с осмысленными сообщениями, а не с неоднородностями распределения материи и энергии.

Библиотечно-информационная деятельность представляет собой деятельность по сбору, обработке, хранению, поиску и распространению стабильных сообщений (документов), представленных коммуникабельными, т.е. пригодными для использования в социальной коммуникации, знаками (устная или письменная речь, изображение, музыка, материальный экспонат). Здесь информация выступает как средство выражения социальных смыслов (знаний, умений, эмоций, волевых побуждений, фантазий) коммуникабельными знаками и именуется *семантической (смысловой) информацией*.

Технология АСПИ в этом случае заключается в переработке семантической информации в виде документов, включённых в библиотечно-информационную деятельность, в документы иной формы и содержания. Исходные документы называются *первичными документами (первичной информацией)*, а документы, полученные в результате их аналитико-синтетической переработки, – *вторичными документами (вторичной информацией)*. Творческий характер переработки информации в данном случае обусловлен креативностью и профессионализмом авторов, создающих в результате умственного труда *новый и общественно полезный* информационный продукт.

Следовательно, можно принять следующее определение: *аналитико-синтетическая переработка информации – профессиональный процесс библиотечно-библиографического творчества, заключающийся в создании на основе потока первичных документов нового потока вторичных документов, востребованных социально-культурной практикой, а также оригинальных средств и методов переработки информации.*

Оригинальными средствами и методами АСПИ, созданными в результате индивидуального (авторского) творчества, являются: авторские библиографические пособия, универсальные и специальные библиотечно-библиографические классификации и рубрикаторы информационных изданий; методики предметизации и словари предметных рубрик, включая различные авторитетные файлы; стандарты и машиночитаемые форматы библиографических (каталожных) описаний документов; общие и частные методики аннотирования, реферирования, составления обзоров. Коллективное библиотечно-библиографическое творчество в области АСПИ воплощено в потоках библиографической информации, системе библиотечных каталогов и справочных аппаратов, в виде структурированных библиотечных фондов.

В соответствии с определением АСПИ, её разновидностями, несомненно, являются все виды описания, систематизации и свертывания документов. Однако вовсе не очевидно, что информационный поиск можно поставить в один ряд с аннотированием или подготовкой обзоров в качестве разновидности АСПИ. Где в поисковых процессах поток первичных документов и поток вторичных документов? Очевидно, что в результате поиска в книгохранилище библиотекарь выдаёт первичный, а не вторичный документ. Если же библиограф ведёт тематический поиск в массиве библиографических описаний, то первичные документы отсутствуют вообще. А разве является полученная в результате библиографического разыскания фактическая справка «вторичным документом»? Разберёмся с этими «недоуменными» вопросами.

4. Анализ – Релевантность – Синтез

Прежде всего, уточним понятие *информационный поиск* и технологическую схему поиска информации. В классической монографии по информатике содержится следующая формулировка: «*Информационный поиск* – это некоторая последовательность операций, выполняемых с целью отыскания документов (статей, научно-технических отчётов, описаний к авторским свидетельствам, книг и т.д.), содержащих определённую информацию (с последующей выдачей самих документов или их копий), или с целью выдачи фактических данных, представляющих собой ответы на заданные вопросы» [9. С. 248]. Это определение правомерно назвать «библиотечно-библиографическим», ибо оно соответствует классической технологии библиотечного и библиографического обслуживания.

Более «информатической» выглядит поздняя дефиниция, предложенная сотрудниками ВИНТИ: «*Информационный поиск (information retrieval)* – процесс нахождения и выбора (выдачи) требуемой (т.е. определённой заранее заданными признаками) информации из отдельного текста, документа, совокупности документов или вообще из запоминающего устройства любой физической природы» [10. С. 150].

Наконец, вовсе технократически звучит утверждение, что информационный поиск – это «действия, методы и процедуры, позволяющие осуществлять отбор определённой информации из массива данных» (ГОСТ 7.73–96 Поиск и распространение информации. Термины и определения).

Если первые две формулировки распространяются на библиографическую эвристику, творческий характер которой детально рассмотрен П. Н. Берковым [11], то последняя ориентирована на автоматизированные системы обработки данных, действующие по алгоритмическим программам. Следовательно, возможны *два принципиально разных понимания* информационного поиска: как творческого процесса, требующего новаторских решений, или как механической процедуры с предрешённым результатом.

Творческая природа поиска библиографической информации никогда не вызывала сомнений у наших библиографов. Приведу примеры библиографических разысканий, потребовавших

творческой переработки информации, заимствованные из книги М. А. Садовой [12]. В библиотеку поступило требование читателя: «Маяковский В. Светочи прошлого – творцы будущего. 1919». Такого сочинения Маяковского найти не удалось. Тогда библиограф методом «проб и ошибок» попытался заменить одну букву в фамилии автора другой. И нашлось произведение: Мияковский В. Радищев. Очерк жизни и деятельности. Для самообразования. – Пг. : Огни, 1919 (Светочи прошлого – творцы будущего).

Другой пример. Читатель запросил книгу: Шабанов. Научно-методические указания по лечению и диагностике гипертонической болезни. В алфавитном каталоге были просмотрены все карточки на Шабановых, а также Шебановых и Шибановых. В систематическом каталоге под несколько изменённым названием «Научно-методические указания по диагностике и лечению ...» нашлась работа, принадлежащая доктору медицины И. И. Сперанскому. Когда библиограф взял книгу, он обнаружил на титульном листе надпись «Утверждаю. Зам. министра здравоохранения СССР. А. Шабанов».

Приведённые примеры иллюстрируют новаторскую новизну, непредсказуемость и неоднозначность решений, принимаемых в процессе библиотечно-библиографического поиска.

О решающей роли человеческого фактора в технологии поиска информации вполне определённо свидетельствует И. С. Пилко, которая видит суть информационного поиска в «выделении в некотором множестве (информационном массиве) подмножества релевантных документов или данных, отвечающих запросу потребителя» [13. С. 66]. Технологию информационного поиска «на самом общем уровне» она представила в виде последовательности следующих операций: 1) Формулировка запроса, выделение в его структуре основных поисковых признаков: ключевых слов и понятий, предметов и аспектов поиска; 2) Идентификация данных: сравнение поисковых признаков с данными в информационном (поисковом) массиве; 3) Отбор: проверка выявленного подмассива документов или данных на соответствие заданным критериям поиска; 4) Структурирование (упорядочение) документов или данных в соответствии с логикой запроса.

Можно интерпретировать эту технологическую схему как АСПИ. Тогда операция 1) будет представлять собой *анализ* задания на поиск; операции 2) и 3) – это *анализ* поискового массива (первичного множества документов) и отбор релевантных документов (вторичного подмножества документов); операция 4) – *синтез* результатов поиска в виде ответа на запрос.

Складывается типичная структура творческой АСПИ: преобразование первичного массива документов (исходный объект) во вторичную совокупность документов (ответ на запрос) при активном участии профессионального библиотекаря-библиографа. Однако эта структура не отражает специфических особенностей переработки информации в реальных информационно-поисковых системах, представляющих собой совокупность массива документов, информационно-поискового языка, правил (методов) индексирования текстов, критериев релевантности (смыслового соответствия) документов и запросов, а также материально-технических средств (картотека ручного поиска, персональный компьютер, телекоммуникация и т.д.) и обслуживающего персонала.

Специфическая особенность АСПИ в случае поиска заключается в том, что смысловое соответствие (релевантность) определяется не между первичным и вторичным документом, что характерно для всех остальных видов АСПИ (библиографическое описание, аннотирование, реферирование и др.), а между запросом и выданными в результате поиска документами. Именно на основании оценки релевантности производится отбор подлежащих выдаче документов, поэтому понятие *релевантность* имеет ключевое значение в теории информационного поиска.

Важно подчеркнуть, что объективную оценку релевантности текстов запроса и документа может осуществить только коллектив экспертов (арбитражная комиссия), но ни в коем случае не автор запроса (субъективные суждения автора запроса – это оценки «по pertinентности», а не «по релевантности»). Причём оценки по релевантности плохо поддаются формализации и являются творческими решениями экспертов, порой неожиданными и интуитивными. Учитывая эту особенность информационного поиска, целесообразно представить его в качестве разновидности АСПИ, осуществляемой по схеме «Анализ – Релевантность – Синтез».

Достоинство этой схемы в том, что она детализирует процессы переработки информации в процессе поиска. «Анализ» в данном случае – это операция *индексирования* документов и запросов средствами информационно-поискового языка. Напомню, что ИПЯ представляет собой искусственно созданный язык, предназначенный для описания содержания и формы первичных документов с целью их последующего поиска по запросам. ИПЯ обладает лексикой (словарем лексических единиц, или индексов), синтаксисом (правила построения предложений) и служит для составления поисковых образов документов и запросов. Поисковые образы – не что иное, как тексты, т.е. *вторичные документы на ИПЯ*. Поэтому индексирование иногда трактуют как «перевод на поисковый язык».

В результате индексирования формируется лингвистическое пространство, в котором осуществляются логические операции дизъюнкции, конъюнкции, импликации, отрицания, служащие для варьирования поисковых предписаний, критериев релевантности, эшелонирования выдачи и пр. Эти логические операции, дополняемые иногда вычислениями различных весовых коэффициентов, представляют собой программу переработки информации, которая имитирует экспертную оценку релевантности вторичных документов (поисковых образов первичных документов) имеющемуся запросу. Завершающий поиск «Синтез» заключается в выдаче первичных документов (или их адресов), поисковые образы которых были признаны релевантными. Таким образом, детальное рассмотрение процедуры поиска информации подтверждает вывод о том, что информационный поиск представляет собой АСПИ.

Вероятно, не случайно творческие усилия наших классиков – М. Дьюи, Ч. Кеттера, П. Отле, Ш. Ранганатана, многочисленных отечественных творцов и реформаторов библиотечного дела и библиографии были сосредоточены преимущественно в области средств и методов информационного поиска. Разработанные ими идеи децимализма, словарного индексирования, фасетного анализа стали основой теории информационно-поисковых языков. Поэтому исключение информационного поиска я расценил как недопустимое обеднение арсенала АСПИ.

Наконец, вернёмся к «недоуменным вопросам», которые возникли в конце раздела 3. Обратим внимание на относительность порядковых числительных «первичный» и «вторичный», присваиваемых документам, которые участвуют в АСПИ, и использованных в определениях аналитико-синтетической переработки информации. Отбросим предрассудок, что книга – всегда первичный документ, а её аннотация – вторичный. Ведь рукопись книги «первичнее» её издания, а переиздание – «вторичнее» оригинала. Всякая переработка имеет начальный объект (то, что перерабатывается) и конечный результат (то, что получилось). Строго говоря, между начальным объектом и конечным результатом существует отношение «материал – изделие», а не «первичный – вторичный». Книга может быть изделием (произведением печати), а может быть исходным материалом для написания обзора, поэтому характеристики первичности или вторичности – понятия для документов не имманентные, а ситуативные и неопределённые.

Тем не менее, когда речь идёт о переработке информации, неопределённость снимается контекстом: понятно, например, что книга – первична, а её аннотация – вторична. Что же

получается в случае информационного поиска? Исходным объектом является массив документов (он первичен), а конечным результатом – фрагмент массива или отдельный документ, которые признаны релевантными запросу (они вторичны). Выходит, что при поиске в книгохранилище вторичными документами будут книги; при поиске в библиографической ИПС – библиографические описания с аннотациями или без них; при обращении к фактографической справочной системе – фактическая справка.

5. Автоматизированные системы переработки информации

Аббревиатуру АСПИ можно расшифровать двояко: как «Аналитико-синтетическая переработка информации» и как «Автоматизированная система переработки информации». Это совпадение символично, поскольку выводит нас на проблему автоматизации творческих процессов АСПИ в библиотечно-библиографических условиях. Проблема заключается в том, что за уникальные новаторские результаты приходится расплачиваться трудоёмкостью, субъективизмом, спонтанностью библиотечного творчества. Библиотечное творчество, как и любое другое, *нетехнологично*, т.е. несовместимо с технологиями массового производства. В то же время для переработки многотысячных потоков первичных документов требуются как раз высокопроизводительные технологии, подразумевающие использование современных технических средств. Появились проекты автоматизации творческих операций анализа и синтеза информации, начало которым было положено ещё в 1950-е гг., когда энтузиасты кибернетики принялись создавать системы машинного перевода.

Первое поколение «переводчиков-автоматов» разработано на основе предельно простой технологии *пословного перевода*. Допустим, в память компьютера загружается англо-русский словарь. Английский текст преобразуется в русский путём последовательной замены английских слов русскими эквивалентами. Теоретической основой подобной технологии было понимание естественного языка как кода, а процедуры перевода как перекодирования текста из одного набора кодов в другой. Никаких сложностей не предвиделось, и осуществление автоматического перекодирования казалось элементарной операцией. Однако первые же попытки неадаптированного пословного перевода обескуражили. В компьютер была введена фраза из Библии: *The spirit is strong, but body is weak*, что означает «Дух крепок, а плоть слаба». Компьютер, пользуясь находившимся в его памяти англо-русским словарем, перевёл: «Виски крепкое, а мясо протухло». Это объясняется полисемией естественного языка: *spirit* имеет несколько значений, в том числе «дух» и «виски»; *strong* можно перевести и как «крепкий» (вино), и как «сильный» (человек); *body* – это «тело» и «мясо», а *weak* – «слабый» и «испорченный».

Попробовали ввести в компьютер комментарии, которые позволили бы алгоритмически выбирать значения слов (кодов), уместные в том или ином контексте. Оказалось, чтобы решить эту задачу, нужно загрузить в машинную память практически необозримый объём информации, моделирующий знания, которые содержатся в личностном и социальном интеллекте. Кроме того, в живой речи полно двусмысленностей, противоречий, иносказаний, неопределённостей, которые невозможно сделать доступными пониманию машинного интеллекта. Короче говоря, для получения практически адекватного перевода нужно учитывать экстралингвистическую информацию, т.е. сведения, которые не содержатся в переводимом тексте, но предположительно известны человеку, воспринимающему этот текст.

В итоге разработчики систем пословного перевода, чаще всего математики и инженеры, отказались от концепции языка как кода и обратились за помощью к лингвистам. В языкознании возникло новое направление – структурная (математическая) лингвистика, изучающая проблематику АСПИ. Лингвисты занялись построением информационно-лингвистических моделей, имитирующих понимание (анализ и синтез) текстов. Однако, несмотря на многолетние усилия энтузиастов, так и не удалось добиться «человекоподобного» перевода.

С 1960-х гг. основное внимание лингвистов-информатиков было направлено на автоматизацию информационного поиска, ключевой задачей которой была разработка лингвистического обеспечения в виде дескрипторных языков, семантических кодов, информационно-поисковых тезаурусов. Некоторые из автоматизированных ИПС достигли стадии промышленного внедрения, демонстрируя преимущества и ограничения формализованного анализа и синтеза семантической информации. Развитием идеи автоматизированного поиска стали фундаментальные и прикладные исследования в области автоматизации индексирования и реферирования. Привлекательность автоматизации этих процессов виделась в оперативной обработке больших документных потоков, значительной глубине индексирования, исключении субъективной непоследовательности индексаторов и референтов. Технологии алгоритмического анализа и синтеза текстов первичных документов были несложными и соответствовали возможностям вычислительной техники 1970-х – 1980-х гг.

Так, *автоматическое индексирование* в дескрипторной ИПС «Электротехника» заключалось в распознавании и пословном переводе ключевых слов, введённых в систему рефератов, на дескрипторный язык с помощью дескрипторного словаря, хранившегося в памяти системы. В результате получался поисковый образ документа в виде перечня дескрипторов. В более совершенных ИПС учитывались синтаксическая структура индексируемого текста и грамматические отношения между дескрипторами. Автоматизация *реферирования* состояла в отборе тех фраз текста, в которых имелось большее число дескрипторов. При отборе предпочтение отдавалось предложениям в начале и конце текста как наиболее значимым, а также учитывались заданные списком слова-маркеры, такие как «задача», «цель», «метод», «результат», «вывод», «итог» и т.п.

Вторичный текст, составленный из формально извлечённых (экстрагированных) из текста предложений, не обладал логической связностью, но давал некоторое представление о содержании первичного документа. Поэтому его называли «квазирефератом». Системы автоматического реферирования до сих пор не вышли из стадии экспериментирования.

Практическая реализация автоматизированных ИПС показала два присущих им «узких места»: во-первых, ввод первичных документов (произведений письменности и печати), во-вторых, доставка найденных документов территориально удалённым потребителям информации. Суть первой проблемы, известной как *машиночитаемая каталогизация*, заключается в аналитико-синтетической переработке текста человекочитаемого первичного документа в машиночитаемый вторичный документ, служащий объектом автоматизированного библиографического поиска.

Технология АСПИ в данном случае сводилась к заполнению каталогизатором формата (рабочего листа), исчерпывающе характеризующего форму и содержание первичного документа, перенесению данных формата на машиночитаемый носитель и включению машиночитаемого каталожного описания в электронный каталог. Становилось необходимым межбиблиотечное сотрудничество при заполнении машиночитаемых форматов. Это было реализовано Библиотекой Конгресса США в системе MARC. Как известно, в нашей стране идею корпоративной (точнее – кооперативной) системы машиночитаемой каталогизации с 1998 г. воплощает Общероссийская информационно-библиотечная компьютерная сеть ЛИБНЕТ.

Машиночитаемая каталогизация неразрывно связана с *дигитализацией* (оцифровыванием) – преобразованием всех видов человекочитаемой информации (текстовой, аудиовизуальной) в цифровую форму, что делает возможной компьютерную обработку. Благодаря дигитализации стремительно расширяются пределы электронной коммуникации за счёт документных каналов научной периодики, официальной документации и т.д., несмотря на юридические ограничения, связанные с авторским правом. Оцифровывание фондов обеспечивает создание

электронных библиотек с дистанционным библиографическим поиском и электронной доставкой документов, преодолевая второе «узкое место» автоматизированных систем АСПИ.

Дигитализация используется в библиотеках при конверсии (конвертировании) традиционных каталогов из карточной формы в машиночитаемую электронную, а также для создания электронных коллекций редких и ценных книг, плакатов, гравюр и т.п. На международной арене компания *Google* инициировала ряд амбициозных образовательных и книжных проектов, в том числе *Мировая электронная библиотека*, начатый в 2006 г. совместно с Библиотекой Конгресса США и объединяющий сегодня ресурсы многих национальных библиотек.

Государственная программа «Информационное общество (2011–2020)», принятая Правительством России в октябре 2010 г., рассматривает оцифровывание культурного наследия в качестве «ключевой задачи». В Программе предусмотрено создание *национального библиотечного ресурса* с унифицированным каталогом на базе оцифрованных фондов Российской государственной библиотеки, Российской национальной библиотеки, Президентской библиотеки имени Б. Н. Ельцина, библиотек государственных академий наук Российской Федерации, а также государственных и муниципальных публичных библиотек.

В русле Государственной программы построения информационного общества находится разработанная РГБ концепция *Национального библиотечно-информационного фонда* [14]. Национальный фонд должен включать, разумеется в оцифрованной форме, все виды документов, произведённых на территории Российской Федерации: печатные издания (текстовые, нотные, картографические, изоиздания), аудиовизуальные документы всех видов, электронные издания, компьютерные программы и базы данных, неопубликованные (непубликуемые) документы, патентные документы и стандарты, специальные издания для слепых. Участники Национального фонда – крупнейшие библиотеки, книжные палаты, архивы, информационные центры, хранилища специальных видов документов (Гостелерадиофонд, Госфильмофонд, Роспатент, Росстандарт и др.). Национальный фонд презентует себя как «российский сегмент всемирных документных ресурсов, результирующих научные и культурные достижения земной цивилизации в её историческом развитии».

Итак, в нашей стране и во всем мире энергично и целенаправленно реализуются программы автоматизации АСПИ, включающие проекты построения электронных библиотек, сводных каталогов, корпоративных библиотечных сетей, служб виртуального библиографического обслуживания и т.п. При этом единогласно признаётся, что магистральный путь человечества в информационное общество – глобальная сеть Интернет, которая ознаменовала наступление в XXI в. эпохи электронной коммуникации. Благодаря Интернету традиционные каналы документальной коммуникации были дополнены каналами электронной коммуникации, влияние которых на образ жизни и мировоззрение людей достаточно очевидно.

Несомненна зависимость будущего библиотечной профессии от развития электронной коммуникации. Машиночитаемая дигитализация библиотек означает, что они изменяют свою коммуникационную сущность: выходят из документальной коммуникации и становятся элементом электронной коммуникации. В компьютерных библиотечных и информационных сетях аналитико-синтетическая переработка массовых потоков первичной информации производится не квалифицированными референтами-библиографами, а персональными терминалами и серверными узлами, осуществляющими автоматическую обработку данных и информационный поиск по адресным, тематическим, фактографическим запросам. Энтузиасты информатизации уверяют, что в информационном обществе культурное наследие воплотится не в архаичные документные фонды, а в электронные (оцифрованные) информационные ресурсы глобального масштаба с возможностями доступа из любого места и в любое время.

6. Нерешённые проблемы автоматизации АСПИ обусловлены тем, что современные компьютеры не способны творчески, т.е. неалгоритмически, перерабатывать семантическую (смысловую) информацию, потому что они оперируют не смыслами (знаниями, умениями, эмоциями, желаниями, фантазиями), а машиночитаемыми данными, имитирующими человеческие смыслы. По этой причине не удаётся преодолеть информационный кризис, почти век удручающий учёных и педагогов. По-прежнему человеку не по силам освоить адресованные ему потоки новейшей информации (кризис текущего информирования). По-прежнему мы «не знаем, что мы знаем», поскольку не располагаем надёжными средствами информационного поиска.

Интернет уже сегодня обеспечивает библиографический поиск и электронную доставку релевантных запросам документов более полно и оперативно, чем традиционные библиотечно-библиографические учреждения. Беда в том, что количественный рост вторичной информации не уменьшает, а увеличивает информационную нагрузку на человека, которая давно превзошла все гигиенические нормы. Складывается парадоксальная ситуация: автоматизация не облегчает информационный поиск, она усугубляет кризис текущего информирования. Как устранить этот парадокс?

Д. И. Блюменау справедливо заметил: «Одна из важнейших задач, которая ждёт своего решения, заключается в том, чтобы передать компьютеру процедуру автоматического свертывания входного потока документов с целью формирования пакетов самостоятельных фрагментов, упорядоченных по тематике и их аспектной принадлежности. Образовавшаяся в результате такого свертывания база знаний будет информационной основой для формирования различных синтезируемых документов типа дайджестов, квазиобзоров, квазиконспектов и квазихрестоматий» [15. С. 186]. Сомнительно, что алгоритмически синтезированные «квазивторичные документы» смогут заменить продукты живого человеческого творчества. Таким образом, выявляется *первая нерешённая проблема автоматизации АСПИ*: перейти от автоматического свертывания текстов к переработке (оценке, обобщению, ранжированию) смыслов.

Не менее сложно преодолеть кризис ретроспективного поиска, заключающийся в том, что «мы не знаем, что мы знаем». В принципе возможно обеспечить контроль за содержанием оцифрованных фондов библиотек и архивов. Интеллектуальный робот-библиограф на основе информационно-поисковых тезаурусов сможет обнаружить все концепции (гипотезы, теории, законы), относящиеся к данной тематической области. Но исключить беспокоящую учёных ситуацию, когда «гениальные открытия сделаны, опубликованы и похоронены в фондах библиотек», таким путём не удастся. Камень преткновения таится в том, что о степени «гениальности» открытия могут судить только люди, а никак не интеллектуальные компьютеры. Отсюда *вторая нерешённая проблема автоматизации АСПИ*: овладеть информационно-логическими методами хотя бы на уровне логического вывода «Петя кушает яблоко → (значит) мальчики едят фрукты».

Перспективы автоматизации библиотечно-информационного анализа и синтеза информации зависят от прогресса в области автоматической обработки текстов, главной проблемой которой с самого начала был и до сих пор остаётся машинный перевод. После «романтической прамбулы» 1950-х гг. сменились три поколения систем машинного перевода, и качественные показатели значительно улучшились, но «серьёзное сдерживающее значение для автоматизации и компьютеризации процессов перевода» имеет то обстоятельство, что «для правильного понимания предложения и его перевода надо учитывать весь окружающий текст». Следовательно, требуется, чтобы автоматический переводчик «каким-то образом сверял перевод каждого предложения со всем переведённым текстом. Но это действие требует достаточно развитого интеллекта, которым ещё не обладает ни одна компьютерная система» [16. С. 18].

Способность определять содержательное соответствие (релевантность) первичных и вторичных текстов необходима всем автоматизированным системам поиска и свертыwania семантической информации. «Автоматическое понимание текстов» – это *третья нерешённая проблема автоматизации АСПИ*.

Учитывая чрезвычайную сложность проблематики автоматизации АСПИ, следует признать, что задача заключается не в имитации с помощью информационных технологий продуктов человеческого интеллекта, а в разумном сочетании возможностей человека и техники. Конечная цель развития аналитико-синтетической переработки информации – человеческое творчество, а помогать нам в этом призвана компьютерная техника. При этом нужно иметь в виду, что творчество нетехнологично, но всякая компьютерная технология – результат творчества человека.

Список источников

1. **Библиография**. Общий курс / под ред. М. А. Брискмана и А. Д. Эйхенгольца. – Москва : Книга, 1969.
2. **Михайлов А. И., Чёрный А. И., Гиляревский Р. С.** Основы научной информации. – Москва : Наука, 1965.
3. **Михайлов А. И., Чёрный А. И., Гиляревский Р. С.** Научные коммуникации и информатика. – Москва : Наука, 1976.
4. **Зупарова Л. Б.** АСОД или АСПИ? Что стоит за переименованием курса? // Библиотеки и информационные ресурсы в современном мире науки, культуры, образования, бизнеса : материалы междунар. конф. «Крым». [Электрон. ресурс]. – Москва : ГПНТБ России, 2005.
5. **Зупарова Л. Б.** Аналитико-синтетическая переработка информации : учеб. / Л. Б. Зупарова, Т. А. Зайцева, под ред. Ю. Н. Столярова. – Москва : ФАИР, 2008. – 400 с. (Специальный издательский проект для библиотек).
6. **Новая философская энциклопедия** : в 4 т. / Ин-т философии РАН. Т. 4. – Москва : Мысль, 2010.
7. **Соколов А. В.** Информатические опусы. Опус 5. Природа и сущность информации // Науч. и техн. б-ки. – 2011. – № 2. – С. 5–27.
8. **Глушков В. М.** Мышление и кибернетика // Вопросы философии – 1963. – № 1. – С. 36–48.
9. **Михайлов А. И., Чёрный А. И., Гиляревский Р. С.** Основы информатики. – Москва : Наука, 1968.
10. **Словарь терминов по информатике на русском и английском языках** / Г. С. Жданова, Е. С. Колобродова, В. А. Полушкин, А. И. Чёрный. – Москва : Наука, 1971.
11. **Берков П. Н.** Библиографическая эвристика. К теории и методике библиографических разысканий. – Москва : Всесоюз. книж. палата, 1960. – 173 с.
12. **Садова М. А.** В поисках книги. Библиографические разыскания при выполнении читательских требований / под ред. М. А. Брискмана. – Москва : Всесоюз. книж. палата, 1963. – 84 с.

13. **Пилко И. С.** Информационные и библиотечные технологии : учеб. пособие. – С.-Петербург : Профессия, 2006.
14. **Национальный** библиотечно-информационный фонд страны: новые рубежи / Ю. А. Гриханов, А. А. Джиго, Е. Ю. Елисина, Н. З. Стародубова, Л. М. Толчинская // Библиотековедение. – 2011. – № 1. – С. 14–21.
15. **Блюменау Д. И.** Информационный анализ / синтез для формирования вторичных документов : учеб.-метод. пособие. – С.-Петербург : Профессия, 2002.
16. **Семенов А. А.** Современные информационные технологии и перевод. – Москва : Академия, 2008.