

УДК 02:001.8
ББК 78.30в6+72в

СМЫСЛОВОЙ АНАЛИЗ ТЕКСТОВ

© Г. Ф. Гордукалова, 2010

*Санкт-Петербургский государственный университет культуры и искусств,
191186, г. Санкт-Петербург, Дворцовая набережная, 2*

Дается обзор подходов к семантическому анализу текстов. Предлагается авторский термин и подход – «смысловой анализ текстов», который включает разноуровневые «концепты», в том числе сугубо ситуационные, связанные не только с научными, но и деловыми, справочными текстами, публикациями о досуге. Предлагается методическая схема использования данного подхода в деятельности библиографа-аналитика.

Ключевые слова: анализ текстов, смысл, свертывание информации, библиограф-аналитик.

The review of approaches to the semantic analysis of texts is presented. The author's term and approach «texts sense analysis» is offered. It includes «concepts» of different levels, including situational, connected not only with scientific, but also business, reference texts, publications on leisure. The methodical scheme of this approach use in the bibliograph-analyst activity is given.

Key words: text analysis, sense, compression of information, bibliograph-analyst.

При изучении данной проблемы важно ответить на многие вопросы: *Зачем нужен анализ смыслов в библиотеке? Есть ли такой метод? Нужно ли его формировать? Почему выбрано наименование «смысловой анализ»? Как его осуществить? Где будет использоваться?*

Выявлением, анализом и освоением смыслов были заняты «все-во-все» времена. Цели анализа различны – от профессиональных, исследовательских до учебных и бытовых. Авторы и редакторы сформировали определенный инструментарий для выделения важнейших идей, выводов в печатных текстах. Электронная среда обогатила возможности использования шрифтов, цветов, ссылок и даже звуков для фиксации важнейших положений в тексте, но одновременно привнесла и ряд отрицательных свойств в огромное число слабо редактируемых текстов:

- избыточность, высокая степень дублирования;
- недостоверность фрагментов;
- отсутствие данных об авторах, времени подготовки и размещения документов в сети;
- компилирование, «заимствования» из чужих текстов без ссылок и др.

Это значительно осложняет отбор документов в процессах комплектования, библиографирования, обслуживания читателей.

Именно у библиографа, аналитика информационных ресурсов (ИР) есть специальная ПРОФЕССИОНАЛЬНАЯ задача – выделить, учесть, библиографировать и диагностировать тексты, значи-

мые для конкретной темы, запроса. Она решалась всегда, но являлась без специализированных методик трудновыполнимой «сверхзадачей». Качество ее зависело от опыта и таланта библиографа.

Есть ли такой метод? Можно найти много советов и даже методик для смыслового анализа, но нет специального метода. Безусловно, в отечественной информатике достаточно широко проводились и теоретически обсуждались подобные исследования текстов. Наиболее яркие, широко цитируемые работы в мире – Ю. А. Шрейдера, Г. Э. Влэдуца и В. А. Успенского, С. Г. Карамурзы, А. В. Соколова. Несколько веков назад шел спор – может ли библиограф, составляя указатель книг, «проникать в существо оных». При этом отбор и систематизация материала в указателе, конечно, уже активно требуют компоненты смыслового анализа.

Наши «соседи» – работники архивов, лингвисты, разработчики программного обеспечения анонсируют многие разновидности семантического анализа – от терминологического до дискурсивного – в той или иной мере решают когнитивные задачи. Их перечень достаточно широк: интен-анализ, контент-анализ, казуальный, категориальный и др. Специалисты различно называют этот метод-подход-уровень анализа текста:

- концептографический,
- концептуальный,
- креативный,

- смысловой,
- онтологический,
- когнитивный.

Почему предлагается выбрать наименование «смысловой анализ текстов»? Приведенные понятия не являются полными синонимами, не равны по своему объему. Рассмотрим их отличия.

Концептографический анализ. Термин меньшего объема в прямой его трактовке, который активно продвигался А. В. Соколовым и Д. И. Блюменау по двум важным причинам – он «скромен», нацеливает только на описание концепций, заложенных автором в текст. Конечно, он предполагал и анализ гипотез, прогнозов, новшеств, т. е. искусственно расширялся его объем – описание концепций. Главная причина для позитивного отношения к данному термину заключается в том, что авторы предложили целостную теорию с логичным, классически выдержанным и реалистичным выделением трех уровней информационного обслуживания:

- документографического,
- фактографического,
- концептографического.

Для реализации концептографического анализа Д. И. Блюменау предложил:

- концепции-предсказания (гипотезы, прогнозы, предположения),
- концепции-теории (объясняющие или обосновывающие какое-либо явление с определенной достоверностью),
- регулятивные концепции (рекомендации, принципы и методики, предписания, нормы, запреты),
- концепции-оценки, проблемные концепции.

Концептуальный анализ. В 1970-е гг. понятие «концепция» было наиболее обобщающим в числе обозначающих развитие научного знания. Очевидно, что данный производный термин, иногда упоминаемый в классической информатике, не равноценен «концептографическому» и «смысловому» анализу. Имеет иной оттенок, аспект, означая уровень научного анализа объекта в большей мере, чем текста. Сейчас чаще употребляют сочетание анализ «концептов» – общих понятий, а не концепций – системы серьезных и самостоятельных взглядов, ведущей мысли.

Креативный анализ – «создающий», более «настроен» на творческий подход к делу, нежели на целевой процесс выделения смыслов. Он может быть употреблен для библиографа-аналитика, его индивидуального мастерства.

Онтологический анализ – «современное» и модное понятие, которое более затеняет, чем облегчает работу. Кроме того, представленные «онтологии» как результат анализа напоминают дескрипторные словари 1960-х гг., либо результаты категориального анализа предметного поля объекта.

Термин «онтологический» был предложен в 1613 г. Р. Гоклениусом в его «Философском словаре». Как раздел философии онтология определяется «наукой о бытии», наукой о сущем. Наиболее близкая современная трактовка – формализованное описание предметной области. Идеальная онтологическая модель – это конечное множество понятий предметной области, отношений между понятиями и конечное множество функций интерпретации. Такую задачу библиографы издавна решали при создании библиотечно-библиографических классификаций. В рамках аналитики она уместна при информационном моделировании и глубинном анализе состояния узкой тематической области.

Когнитивный анализ. Этим термином можно обозначать лишь те методики, которые обеспечат получение существенно нового выводного знания из текстов.

Смысловой анализ текстов. Первые упоминания сочетания «смысловой анализ» в информационной практике обнаружены в статьях В. Г. Овчинникова [1] и Г. Б. Преображенской [2]. Эти работы редко упоминаются, но это не снижает их провидческого потенциала для современной системы коммуникаций. Теория смыслового анализа естественных текстов В. Г. Овчинникова предназначена для определения смысловой близости двух текстов и более. Цель – автоматизация процессов построения рубрикаторов, индексирования и реферирования текстов. Были заложены основные требования к смысловому анализу:

- доступность средств для исполнителей невысокой предметной квалификации,
- более или менее однозначные результаты,
- удовлетворительное качество анализа текстов с позиций специалистов.

Главной проблемой В. Г. Овчинников справедливо называл «разработку средств снижения неоднозначной интерпретации текста». Для этого вводился «информационный язык» – формализованные обозначения отношений между именами объектов в виде предложенных двух типов ориентированных графов описания смыслов текста.

Смысловой анализ текстов в 1980–1990-е гг. в качестве самостоятельного метода не анонсировался, но важен цикл работ профессора А. В. Соколова о «смыслах коммуникаций», заложивших понятие «смысл» для всех типов письменной и устной передачи информации. Выход в свет книг Б. А. Кулика «Логические основы здравого смысла» (1997) и «Логика естественных рассуждений» (2001), издание журнала «Здравый смысл» философским факультетом МГУ и другие публикации с разных позиций, но привели к движению, которое можно обозначить как «назад к здравому смыслу».

В настоящее время термин «смысл» пока не принят библиотечным и информационным сообществом в качестве профессионального. При всех его «недостатках» (высокий уровень обобщения; невозможность строгого определения как и для родовых понятий «информация», «знание»; широкое его использование в бытовой речи, усложненные семиотические трактовки), его употребление как специального в библиотечно-библиографической работе может принести немало полезного. В частности, мы будем вынуждены разрабатывать собственные методики, реализовывая оценочную функцию библиографии, тем самым облегчим собственную жизнь в глобальной сети, где отбор документов по ключевым словам и формальным признакам уже явно недостаточен для качества нашей продукции, фондов, выполнения запросов. Главное преимущество термина в том, что он включает разноуровневые «концепты», в том числе сугубо ситуационные, связанные не только с научными, но и деловыми, справочными текстами, публикациями о досуге.

Не так важно приживется ли предложенный термин, или будет придуман новый. Важно показать разные по назначению произведения, достойные глубокого содержательного анализа. Локализованная задача диагностики смыслов уберезет библиографа и аналитика от обширных обзорных текстов, «пересказывающих» авторскую позицию. Это позволит сохранить профессиональную позицию, нацеленную на осторожное свертывание текста.

Определим примерное содержание метода с позиций информационного анализа текста: *смысловой анализ текста – это метод информационного картирования авторизованных высказываний в тексте, значимых для целей анализа.*

Он включает разметку концептуальных положений, их формально-логический и содержательный сравнительный анализ. Самое важное – научиться выделять «единицы анализа».

Понятие «смысл» было введено уже античными стоиками. Под смыслом слова понималась выраженная в нем мысль. Изначально ставилась задача понимания текста, а позднее появилась первая теоретическая парадигма – герменевтика. Она возникла как учение о толковании скрытых текстов Священного Писания. Постепенно было осознано различие между значением и смыслами слов, выражений, в зависимости от контекста их использования и восприятия (Матиас Флавиус Иллирийский, XVI в.), а интерпретация сложных текстов получила более широкое распространение. Близкая трактовка была воспринята и в семантике, но понятия смысл и значение ТЕРМИНА рассматриваются в ней как синонимы. Семиотики лишь сейчас «робко» перешли от слов и предложений к малым текстам.

Комплекатор, библиограф, аналитик ИР же работают со множеством однотемных текстов. Им нужны особые методики. Кроме того, основное наше профессиональное орудие – ключевые слова. При анализе и диагностике смыслов речь идет, прежде всего, о ключевых точках содержания. Поэтому условимся считать «смыслами» глубинные замыслы автора, ради которых и был написан текст. Это не расходится с общепринятым восприятием слова: смысл – целостная мысль, представление о явлениях, свойствах и отношениях между объектами, фактами их существования, условиями развития, средствами преобразования или вариантами использования.

Как его осуществить в библиотеке? Среди многообразия «смыслов» – замыслов, концептов, позиций, предположений и т. д. можно выделить группы разного уровня общности:

- замысел автора, идея, система идей, гипотеза, предположение (ИСТОКИ),
- аспекты, проблемы, суть объяснений, оценки, прогнозы (РЕЗУЛЬТАТЫ),
- обстоятельства, ситуации, средства реализации, алгоритмы, методики, рекомендации (РЕШЕНИЯ).

Все группы-уровни могут быть представлены в одном тексте, но с разной степенью детализации. Если акцент делается на одной, то иные лишь обозначаются. Поэтому аналитик ИР должен уже на этапе анализа запроса выделить главную смысловую группу, значимую для пользователя. Смысловые уровни по-разному организуют текст.

Если анализируется массив текстов одного или нескольких авторов, то возникает задача различения авторских концептуальных суждений. Это требует работы уже не столько с ключевыми словами, сколько с размытыми понятиями – авторскими интонациями в тексте – служебными словами, глаголами, синонимами. Методические средства разнообразны, требуют освоения и заимствования.

Аристотель показал, что правильные рассуждения подчиняются небольшому числу законов, независимых от частной природы объектов. Приемы поиска «подобного» и отличительного в знании, получения «соотнесенного» знания, построения выводного знания – основной инструментарий смыслового анализа. Поиск отрицаний, выявление скрытых аналогий и противоречий – это перспективный путь развития методик смыслового анализа.

Конечно, проблем больше, чем решений. Когнитивная (концептуальная, смысловая) структура текста в явном виде не фиксируется в тексте, и даже не выделяется лингвистами в качестве самостоятельной. План выражения может не совпадать с планом содержания. Мысль как таковая или мысль полезная может отсутствовать, либо не диагностироваться.

Структурная дифференциация смысловых позиций: вначале новое, неопределенное, важное, срочное. Далее развертывание текста может быть последовательным по уровням смыслов (ИСТОКИ–РЕЗУЛЬТАТЫ–РЕШЕНИЯ), либо автором предлагается детализация смыслов одного из уровней.

Кроме того, следует принять и базовые достижения смежных областей деятельности. В частности, важны выводы семиотики, компьютерной лингвистики.

Смыслы могут размещаться в тексте по трехчленной классификации У. Чейфа «данное – доступное – новое». Различна степень активированности смыслов – активная информация, полуактивная и неактивная. Смысловые единицы в статусе «данное» обозначаются слабоакцентированными местоимениями или нулем, а имеющие статус «доступное» или «новое» – ударными полными именными группами и т. д. У. Чейф сформулировал также весьма жесткое ограничение: один элемент новой информации в одной «интеллектуальной единице». Когнитивная причина – невозможность активации более одного элемента информации в рамках одного фокуса сознания.

Фиксируется и структурное распределение смысловой информации в тексте. Например, по результатам У. Чейфа – «легкая» информация объединяет «данное» (в 81% случаев в выборке), «доступное» (в 16% случаев) и несущественное новое (лишь в 3%). Близкие показатели были получены и другими специалистами. Нами изучалась степень дублирования информации в работах ведущих отечественных литературоведов (1985 г.) Не дублирующаяся в ранних работах, значимая информация в среднем не превышала 4% на публикацию. Примерно такую долю информации можно извлекать в процессе смыслового анализа уже отобранных по теме текстов.

Интеллектуальная работа с текстами на уровне выявления их смыслов ведется сейчас в рамках скорочтения и мнемотехники [3], в библиотечном обслуживании [4 и др.]. Выявляются приемы построения «алгоритмов абзаца», «карты текста» и многие другие образно-логические приемы распознавания смыслов в тексте.

Аналитик ИР самостоятельно накапливает личный опыт смыслового анализа, обогащает его схему, используя внешние достижения в анализе разных внутренних структур текстов. Начинать обучение рекомендуется на массиве машиночитаемых текстов одной тематики.

Примерная методическая схема

- **Постановка задачи.** Аналитик выделяет смысловую уровень, который следует диагностировать в соответствии с запросом.

- **Детализация задачи.** Составляется перечень семантических единиц, раскрывающих этот уровень. Возможна фиксация наиболее очевидных лексических единиц, сопровождающих подобные смысловые фрагменты.

- **Картирование смыслов.** Осуществляется семантическая разметка текста с помощью цвета, условных символов или увеличением размера шрифта.

- *Размышление.* *Общая оценка текста, зрительная фиксация выделенных смысловых фрагментов, сокращение их числа.*

- **Визуализация связей.** Установление внутритекстовой связи между фрагментами (мысленное, при сложной структуре – графическое в виде карты памяти, графа, кластера связей).

- **Смысловое «аннотирование» фрагментов – семантическое свертывание:**

- снимается избыточная информация,
- диагностируются коммуникативные цели высказываний (явные и скрытые),
- свертываются фрагменты текста до высказывания.

- **Интерпретация смыслов, слабых связей (фрагментов, терминов, оценок) в тексте.**

- **«Чтение между строк».** Восполнить смысловую неполноту фрагмента текста, отдельного высказывания введением родового понятия, уточнением аспекта, включением иноязычного синонима или русификацией понятия, его определением. Подчеркнуть неявную (неуверенную, предположительную) авторскую оценку ситуации выделением использованного им глагола, наречия и др.

- **«Библиографирование» смысловых высказываний – перенесение их в синтезированный текст, включение в таблицу, построение графа связей.**

- **Ситуативное знание о тексте.** «Осмысление смыслов» с явно выраженной оценкой аналитика (полезно–бесполезно–неопределенно). При значительном числе компилированных высказываний возможна оценка каждой или основных единиц.

- **Выводное знание.** Выделение только «полезных» и «возможно полезных» смысловых единиц текста.

Подобную методику аналитик должен апробировать многократно, на текстах разного назначения – научных, деловых, публицистических. В ней главное – это воспроизводимость результатов анализа, выполненного разными исполнителями. В текущем году был проведен малый «эксперимент»: под руководством дипломницы Марины Тиминой студенты анализировали одну подборку текстов – социально-политические о кризисе, литературоведческие, юридические. Фиксировался и временной показатель. Сходные результаты были получены при выделении смысловых единиц, но разные – при интент-анализе текстов – эмоциональной оценке событий в текстах.

Методическая схема будет уточняться и упрощаться в зависимости от цели анализа текстов. Использовать смысловой анализ можно при *создании электронных коллекций, обзоров литературы по теме, выполнении сложных запросов, отборе перспективных проектов в библиотеке, при поиске наилучших практик, при подготовке путеводителей по электронным ресурсам, выборе и т. д.*

Смысл же данной публикации состоит в том, что современная библиотека – центр интеллектуальной жизни региона. Только в случае активной позиции библиотеки смысловой анализ текстов станет полезным и развивающимся направлением.

Сведения об авторе: *Гордукалова Галина Феофановна – доктор педагогических наук, профессор, заведующий кафедрой информационной аналитики*

Список литературы

1. *Овчинников В. Г.* Об одном подходе к смысловому анализу естественных текстов // Науч.-техн. информ. Сер. 2. – 1968. – № 7. – С.17–23.
2. *Преображенская Г. Б.* Смысловой анализ ссылок в статьях по металлургии // Науч.-техн. информ. Сер. 2. – 1969. – № 10.
3. *Васильева Е.* Суперпамять, или Как запомнить, чтобы вспомнить? – Краснодар : Сов. Кубань, 2003. – С. 89–316.
4. *Бородин В. А.* Теория и технология читательского развития в отечественном библиотековедении. – М. : Шк. б-ка, 2006. – С. 198–199.

Материал поступил в редакцию 12.11.2009 г.