

Н.А. Герте, Д.С. Курушин, Н.М. Нестерова

Пермский национальный исследовательский политехнический университет

СВЕРТЫВАНИЕ ИНФОРМАЦИИ В ПРОЦЕССЕ РЕФЕРИРОВАНИЯ: МЕТОДЫ И ВОЗМОЖНЫЕ ПУТИ ФОРМАЛИЗАЦИИ

Рассмотрен процесс реферирования, характеризующий его как один из самых сложных видов речемыслительной деятельности. Особое внимание уделено исследованию основных методов реферирования, а также механизмов свертывания содержания с целью ее формализации.

Ключевые слова: реферирование, свертывание информации, фасетное и поаспектное реферирование, денотатная структура текста, денотатный граф, автореферирование.

В связи с интенсивным увеличением научно-технических публикаций, в том числе и на иностранных языках, и развитием информатики как особой научно-практической деятельности, одной из наиболее существенных проблем стала задача их смыслового свертывания и компактного представления содержащейся в них информации. Среди различных видов аналитико-синтетической обработки первичных документов реферирование является наиболее востребованным и в то же время более сложным по сравнению с аннотированием и индексированием.

В информатике в соответствии с понятиями об информации и метаинформации выделяется два основных вида свертывания: информационное свертывание, «которое предполагает создание ряда документов, основная цель которых – служить непосредственным источником информации при решении определенных задач», и метаинформативное свертывание, которое предполагает «создание ряда документов, основная цель которых – в той или иной степени раскрыть тему и содержание других документов». В этих документах, по терминологии Ю.А. Шрейдера, содержится «информация об информации» [1]. Применительно к рефератам можно сказать, что к первому типу документов относятся информативные рефераты, ко второму – индикативные.

Вид и степень свернутости первичного документа, соответственно, и объем реферата зависят от реферируемого первоисточника, а именно – от его содержания и в значительной мере от его доступности для массового читателя. В частности, в случае реферирования иностранных первоисточников наиболее оптимальным представляется информативный реферат, достаточно полно отражающий содержание первоисточника, семантически адекватный

ему, так как именно реферат иноязычного документа заменяет оригинал для значительной части читателей. Такой взгляд на реферирование иностранных публикаций подтверждается, в частности, требованиями, содержащимися в инструкции для референтов Chemical Abstracts, одного из самых популярных американских реферативных журналов.

Говоря о реферировании, необходимо отметить, что несмотря на его востребованность и долгую историю (еще во времена Шумерской цивилизации писцы создавали краткие записи на глиняных табличках), реферирование как один из видов свертывания научной информации до последних десятилетий развивалось чисто эмпирически. Однако ввиду роста значения вторичных документов в информационном обслуживании возникла необходимость создания теоретических основ и методики процесса реферирования с целью его совершенствования и улучшения качества рефератов. В связи с этим и появились первые теоретические и методические работы, в которых делалась попытка предложить оптимальную и научно обоснованную методику составления реферата. Так, было предложено несколько методик, которые можно разделить на две большие группы. Первую составляют методы, основанные на анкетировании. К ним относятся фасетное и поаспектное реферирование. Вторая группа представляет собой методики, в основе которых лежит анализ смысловой структуры первоисточника.

Из методик первой группы наиболее известен метод *поаспектного реферирования*, который был разработан В.И. Соловьевым [2]. Данный метод предполагает семантический анализ текста с помощью его разбиения на отдельные содержательные аспекты (сетки). К таким аспектам можно отнести проблему и тему исследования, область знания, цель работы и полученные результаты, методы и условия проведенного исследования и другие. По мнению автора данной методики, наличие таковой обеспечивает как семантическую адекватность первичного и вторичного текстов, так и эффективное восприятие информации. Данная схема, включающая перечень основных семантических аспектов, может быть принята как универсальная, поскольку названные аспектные признаки присущи почти всем научным документам. Соответственно, такой подход позволяет сделать процесс реферирования более формализованным, а также стандартизировать сами рефераты.

Другой метод реферирования был предложен А.И. Жолковой, которая обозначила его как *фасетное реферирование* [3]. Он представляет собой анализ содержания конкретной сферы научной деятельности, заключающийся в накладывании на текст сетки фасетов, соответствующей числу категорий, которые раскрывают специфику данной области в различных ее аспектах. Основным понятием предложенной методики реферирования необходимо считать элементарные логические сообщения, создающие табличный (структурированный) реферат, в котором каждому фасету отведена верти-

кальная колонка. Элементарные логические сообщения располагаются рядом по горизонтали, что создает стратегию поиска информации из упорядоченного набора избирательно сочетающихся между собой и частично повторяющихся в разных сообщениях ключевых слов [4].

Автор фасетного реферирования справедливо считала, что табличная форма реферата – это один из способов индексирования документов и структуризации фактографической информации. Табличный реферат не был предназначен для замены реферата традиционного типа. Можно сказать, что фасетное реферирование в некоторой степени близко поаспектному, так как обе методики основываются на «накладывании» сетки определенных аспектов или фасетов на оригинальный текст. Именно такой подход позволяет выявить в тексте те или иные темы, играющие существенную роль в определенной области знания и, таким образом, необходимые для отражения во вторичном тексте [5].

Вышеназванные методы можно рассматривать как своего рода анкетирование: референту предлагается «анкета» с «вопросами», которые помогут ему найти в тексте первоисточника соответствующую информацию. Данные виды реферирования – это попытки создания моделей вторичных документов, которые используются в различных областях знаний [5]. Главным достоинством анкетных видов реферирования является особая стандартизованность в рамках априорно заданной схемы, которая отражает структуру предметной области, потребности специалиста и другие парадигматические задачи. С другой стороны, у референта могут возникнуть некоторые трудности. Как утверждает Блюменау, главным недостатком этих методов является обращение к интуиции референта, который должен выделить «существенную информацию», а также стремление к детальности самой методики [1]. Наличие такой детальной методики может значительно усложнить процесс реферирования. Представляется очевидным, что методы, основанные на анкетировании, в большей степени ориентированы на создание профильных, специализированных рефератов, поскольку выделяется та информация, которая удовлетворяет потребности специалистов конкретной области знания.

Другую большую группу методов реферирования представляют собой *методы выделения основного содержания первоисточника на основе анализа реферируемого текста*. К данной группе можно отнести, например, логико-психологический анализ, предложенный А.А. Вейзе [6]. К этой же группе относится и так называемое информационно-классификационное реферирование, в основу которого лег критерий точности информирования. Толчком для его развития послужил метод поаспектного реферирования, а именно – его подход к разделению информации по степени значимости и полезности для потребителей и их использованию в реферате. Предложенный метод реферирования опирается на содержательный и логический анализ

первоисточника, что позволяет оценивать важность и приоритетность одних элементов информации по отношению к другим и делает методику достаточно универсальной. В связи с этим данный подход стал широко применяться в центрах информации, и он достаточно популярен в настоящее время. Суть этого метода свертывания информации заключается в том, что на основе реферативно-аналитического чтения происходит структуризация первоисточника, а затем с использованием критериев значимости выделяется непосредственно та информация, которая создаст основу вторичного текста.

Однако, независимо от того, какой метод реферирования выбран референтом, основной задачей является установление семантического соответствия, то есть семантической адекватности оригинала и реферата. Именно семантическая адекватность текстов говорит о точности реферирования, о качестве полученных вторичных текстов: семантически адекватным является такой реферат, в котором в сокращенной форме, но точно, без искажений и интерпретации, воспроизводится основное содержание первичного документа.

Вопрос семантической адекватности непосредственно связан с проблемой «основного содержания» текста. В свою очередь, эта проблема может быть решена на основе анализа внутренних закономерностей процесса понимания, который является ведущим в любом виде смыслового преобразования, в том числе при реферировании. Понимание предполагает оценку текста, связанную с приданием каждому смысловому блоку определенной значимости через установление отношений между частями текста и соотнесение частей с целым [7]. Результаты экспериментальных данных показывают, что смысловое восприятие текста и его понимание есть некий переход от воспринимаемых языковых средств к единицам внутреннего языка [5]. Этот переход сопровождается выделением из текста ключевых элементов, которые А.Н. Соколов называет «смысловыми вехами». Такими «опорными пунктами» могут быть любые слова, содержащиеся в тексте, однако именно они представляют во внутренней речи большие смысловые отрезки, которые объединяются в «целостное мыслительное образование», несущее смысл первичного документа [5]. Таким образом, понимание – это сложный аналитико-синтетический процесс, включающий в себя анализ (расчленение) и синтез (объединение). Так, процесс понимания становится процессом смыслового свертывания за счет способности ключевых элементов нести в себе семантику текста оригинала. Следовательно, ввиду того, что при понимании происходит активное смысловое преобразование информации, этот процесс становится ведущим для различного рода преобразований текста, в первую очередь для реферирования.

Согласно теории текста, разработанной школой Н.И. Жинкина и А.И. Новикова, критерием и условием понимания текста является переход от его внешней формы к внутренней, то есть к денотатной структуре текста, представляющей совокупность иерархически связанных между собой денота-

тов. Такой подход обеспечивает возможность эксплицитно выразить содержание в виде денотатного графа, методика построения которого была предложена А.И. Новиковым. Денотатный граф, который можно рассматривать как «перекодирование линейного текста в целостную, иконическую схему», создается на основе денотативного анализа всего текста [8]. Именно данная методика позволила говорить о формализованном выделении содержания текста. Подобное графовое представление материала вызывает полное преобразование текста, когда происходит определенный отбор денотатов – устраняются те, которые не имеют отношения к моделируемой ситуации. Иными словами, устраняется языковая избыточность, свойственная первичному документу, определяется иерархия денотатов и выявляется их соответствие различным уровням текста. Таким образом, денотатный граф – это свернутое эксплицитное отображение структуры содержания текста, которой могут соответствовать различные языковые формы [5].

Денотативный анализ позволяет представить реферирование как переход $T1 - ДС - T2$: сначала происходит переход от внешней языковой структуры текста к структуре его содержания, то есть его денотатной структуре, и затем переход от денотатной структуры к новой внешней форме вторичного текста. Несомненно, предложенная схема смыслового преобразования является условной и отражает процесс в общем виде. Стоит отметить также, что выделенные этапы взаимозависимы и взаимообусловлены: уровень понимания (как результат первого этапа) определяет вторичный текст с точки зрения его адекватности оригиналу. Установка на определенный вид смыслового преобразования влияет на процесс осмысления, и каждый такой вид имеет свои внутренние закономерности и специфику. Реферирование, в частности, характеризуется особыми отношениями между $T1$ и $T2$, которые можно обозначить как отношения «коммуникативной неравноценности» [5]. Другими словами, в процессе реферирования происходит содержательное уподобление текстов при полном отсутствии формального. Такая неоднозначность вновь возвращает нас к основной проблеме смыслового свертывания, а именно – к проблеме семантической адекватности разных по объему первичных и вторичных текстов. Связано это различие с теми ограничениями, которые накладываются на текст реферата, где денотатная структура исходного текста должна быть представлена в сжатом виде. Другими словами, происходит речевое сжатие, суть которого состоит в членении денотатной структуры на крупные фрагменты и в их лексическом и синтаксическом «означивании». Это обеспечивает компрессированное внешнее выражение денотатной структуры. В.И. Проворотов, занимающийся изучением деятельностной стороны реферирования, указывал на единство его трех составляющих: сокращения, замены, введения нового. Они в свою очередь связаны с усложнением и уп-

рошением: сокращение предполагает упрощение, замена – грамматическое и лексическое перефразирование путем использования слов с более емкой семантикой, и введение новых элементов включает в себя как перефразирование, так и переструктурирование содержания [9].

Таким образом, реферирование можно рассмотреть как вербальную актуализацию понимания, в результате которого порождаются новые смыслы в форме «контртекста», или «встречного текста» [10]. При этом стоит упомянуть, что главные составляющие процесса реферирования (как понимание, осмысление и преобразование текста) разделены лишь условно. В данном контексте проблема понимания рассматривается во взаимосвязи с понятием смысл. А.И. Новиков подчеркивал, что эти два понятия «оказываются комплиментарными, то есть одно предполагает другое» [10]. По словам ученого, именно проблема смысла текста «фокусирует в себе все остальные проблемы, связанные с изучением основных закономерностей устройства текста и его функционирования» [11]. Другими словами, проблема смысла неотделима от проблемы текста, и именно категория смысла определяет признаки текста, в том числе психолингвистические, как связность, внутренняя целостность, информативность, компрессивность. Именно эти свойства делают возможным как смысловое свертывание текста, так и речевую компрессию.

Итак, можно сказать, что общим для всех существующих подходов к реферированию является, во-первых, стремление создать методику, позволяющую наиболее полно и адекватно отражать основное содержание реферируемого текста, во-вторых, попытки каким-то образом формализовать процесс.

В настоящее время формализация реферирования стала одной из важнейших задач компьютерной обработки текстов. Сегодняшние технологии позволяют уже много больше, чем в то время, когда создавались методики реферирования, о которых речь шла выше. Исследователи, занимающиеся вопросом автоматизированного реферирования, выделяют несколько способов обработки информации с помощью компьютера, каждый из которых характеризуется своим методом выделения материала: это так называемое квазиреферирование и краткое изложение содержания первичного текста [12].

Квазиреферирование заключается в выделении наиболее информативных фраз и формировании квазирефератов на их основе. Сегодня различают следующие методы отбора информации:

1. *Статистические методы*, основанные на оценке уровня информативности элементов первичного текста по частоте их появления. Именно частота появления служит ключевым критерием информативности слов или целых фрагментов.

2. *Позиционные методы*, основной идеей которых можно считать предположение о том, что информативность слова (или предложения) в тексте зависит от его позиции.

3. *Индикаторные методы*, позволяющие выявить наличие специальных слов в документе, то есть те, которые наиболее значимы на содержательном уровне.

Другой способ автоматизированного реферирования – это краткое изложение содержания, базисом которого является выделение наиболее важной информации с помощью автоматизированного качественного контент-анализа. Такой анализ проходит в три стадии. На первой стадии происходит сведение исходного текста к заданному числу фрагментов, то есть единиц значения. На второй стадии подключается процесс поиска связей между выбранными единицами значения, после чего начинается третья, последняя, стадия – формирование выводов и обобщений, которые ложатся в основу текста реферата в виде связанных смысловых единиц.

Для выполнения всех этих преобразований используются манипуляции логическими предположениями и выделяются шаблоны в имеющейся базе данных. В результате рождается концептуальная структура текста, то есть концептуальные «выжимки» из текста [12].

Безусловно, такой вид смыслового свертывания, как автоматизированное реферирование, дает возможность ускорить процесс и исключить субъективность при изложении содержания оригинала. Однако необходимо признать, что методы автоматизированного реферирования, которые применяются сегодня, не удовлетворяют всем требованиям реферирования, в связи с чем качество вторичных текстов значительно ухудшается. При выделении текстовых блоков не учитываются связи и отношения между ними, что приводит к написанию бессвязных рефератов: некоторые предложения могут быть опущены либо в них могут присутствовать слова, которые невозможно понять без контекста. Очень часто референты прибегают к исключению таких предложений из рефератов, реже делают ссылки с помощью методов лингвистического анализа.

Однако мы считаем, что задача получения семантически адекватного вторичного текста может быть решена при условии создания удовлетворительной вычислительной модели денотативного анализа текста. В этом случае процесс порождения вторичного текста будет основан не на статистических или логических правилах, а на отображении текста-оригинала (Т1) в денотативную структуру (ДС) и, затем, отображении ДС в Т2 (вторичный текст). Этот процесс в большей степени соответствует естественному текстопорождению и, при условии создания такой модели, позволит получать семантически адекватные рефераты. Ключом к решению этой задачи является создание семантически адекватной модели предметной области. Теоретически, современные методы вычислительной лингвистики позволяют строить такие модели, но этот процесс является достаточно трудоемким. Выход видится в создании самообучающейся системы и подготовке корпуса текстов («учебника»)

для ее первоначального обучения. В этом случае в распоряжении реферирующего алгоритма окажется достаточно полная семантически адекватная модель предметной области текста. Предполагается, что реферирующий алгоритм будет анализировать T1, устанавливать соответствия между его структурными элементами и моделью предметной области, получая таким образом «подграф», соответствующий данному тексту. Затем алгоритм переходит в фазу текстопорождения, выполняя обход подграфа таким образом, чтобы удовлетворить критериям, заданным пользователем — заказчиком реферата: минимальная и максимальная длина текста, обязательное освещение или неосвещение тех или иных фактов. Важным моментом является то, что денотатная модель текста инвариантна относительно языков T1 и T2. Это позволяет говорить о возможности реферативного перевода с высокой степенью семантической адекватности.

Список литературы

1. Блюменау Д.И. Проблемы свертывания научной информации. – Л.: Наука, 1982.
2. Соловьев В.И. Поаспектный метод реферирования // НТИ. Сер. 2. – 1971. – № 2. – С. 14–17.
3. Жолкова А.И. К вопросу о реферировании научно-технической периодики в реферативном журнале ВИНТИ // НТИ. Сер. 1. – 1970. – № 5. – С. 24–25.
4. Жолкова А.И. Фасетный метод реферирования как проблема библиографирования технической литературы: автореф. дис. ... канд. пед. наук. – Л.: ЛГИК им. Н.К. Крупской, 1985.
5. Новиков А.И., Нестерова Н.М. Реферативный перевод научно-технических текстов. – М., 1991.
6. Вейзе А.А. Реферирование текста. – Минск: Изд-во Белорус. гос. ун-та им. В.И. Ленина, 1978.
7. Нестерова Н.М., Наугольных А.Ю., Наугольных Е.А. «Сказать мало, но хорошо»: деятельностная характеристика реферативного перевода // Филологические науки. Вопросы теории и практики: в 2 ч. – Тамбов: Грамота, 2012. – № 7 (18), ч. I. – С. 146–149.
8. Жинкин Н.И. Сенсорная абстракция // Проблемы общей, возрастной и педагогической психологии – М., 1978.
9. Проворотов В.И. Аннотирование и реферирование как методические приемы в обучении при работе над специальными текстами // Ученые записки: электронный научный журнал Курского государственного университета. – 2006. – № 2. – С. 129–135.

10. Новиков А.И. Текст и «контртекст»: две стороны процесса понимания // Вопросы психолингвистики. – 2003. – № 1. – С. 64–76.

11. Новиков А.И. Текст и его смысловые доминанты. – М.: Азбуковник, 2007.

12. Автоматическое реферирование [Электронный ресурс]. – URL: http://www.webground.su/services.php?param=book&part=chapter%203_2_3.htm (дата обращения: 20.05.2013).

Получено 25.05.2013

N.A. Gerte, D.S. Kurushin, N.M. Nesterova

**COMPRESSION OF INFORMATION IN THE SUMMARY
TRANSLATION: METHODS AND POSSIBLE WAYS
OF ITS FORMALISATION**

The article describes the process of summarizing making it one of the most difficult types of verbal and cogitative activity. Particular attention is paid to the research of the main summarizing methods and the mechanism of the content compression and its following formalization.

Keywords: *summarizing, content compression, faceted and aspect-oriented summarizing, denotation structure, denotation graph, computerized summarizing.*