

**АНАЛИЗ СТРУКТУРЫ НАУЧНОГО ТЕКСТА И ПОКАЗАТЕЛИ
ЕГО ПЕРЕВОДИМОСТИ**
**ANALYSIS OF A SCIENTIFIC TEXT STRUCTURE AND INDICATORS
OF ITS TRANSLATABILITY**

Аннотация: В статье приводятся данные о распределении релевантной для реферата информации по разделам научной статьи и лингвистических явлениях, осложняющих перевод текста статьи.

Ключевые слова: автоматическое реферирование; автоматический перевод; научный текст.

Abstract: In this paper we give the data about distribution of the relevant information for the abstract in sections of the scientific article and the linguistic phenomena complicating translation of the article.

Keywords: automatic summarization; machine translation; scientific text.

На настоящий момент объем научных публикаций постоянно увеличивается. По данным торговой ассоциации издателей научной литературы [7, с. 5], ежегодно в мире только в рецензируемых журналах публикуется около 1,8 миллиона научных статей. Для ознакомления с новейшими публикациями используются рефераты статей. Реферат статьи представляет собой краткое изложение основных результатов работы. Автоматическое реферирование – создание рефератов статей с помощью компьютера – позволяет оперативно обрабатывать большие объемы публикаций.

Как правило, для публикации статьи часто требуется предоставить ее реферат на английском языке. Учитывая объем публикаций, очень сложно оперативно переводить вручную рефераты, и для перевода часто используются системы автоматического перевода. Качество автоматического перевода уступает переводу, выполненному человеком, так как перевод с одного естественного языка на другой представляется трудной задачей для автоматизации.

При автоматическом переводе рефератов часто возникают дополнительные сложности, связанные с языком написания. Часто авторы статей используют сложные синтаксические конструкции, большое количество вводных слов, опускают члены предложения. В данной статье мы рассмотрим, из каких структурных частей статьи следует отбирать информацию в реферат, и лингвистические явления, осложняющие перевод научного текста.

К структуре научной статьи предъявляются различные требования. В [6, с. 522] отмечается, что общая структура научной статьи включает в себя сле-

дующие разделы – введение, методы, результаты и их обсуждение. В [1, с. 9] авторы выделяют следующие типовые части научной статьи – введение, основную часть, заключение, примечания.

В качестве материала для исследования мы отобрали 107 научных статей и авторских рефератов на русском языке по тематике «математика и математическое моделирование» общим объемом 203729 словоформ. Статьи были взяты из таких журналов и сборников статей, как «Вестник Южно-Уральского государственного университета. Серия: Математическое моделирование и программирование», «Математическое моделирование», «Вестник Томского государственного университета. Математика и механика», «Математические заметки», «Вестник Ивановского государственного энергетического университета», «Известия Челябинского научного центра УрО РАН».

Статьи либо представляли собой сплошной текст, либо были структурированы и включали такие разделы, как «Введение», «Заключение», «Выводы», а также разделы, озаглавленные в соответствии с рассматриваемой в них проблемой. Для неструктурированных статей мы ввели условные разделы «Начало» (первые два абзаца статьи), «Конец» (последние два абзаца) и «Середина» (текст статьи, который находится между «Введением / Началом» и «Заключением / Концом»). Количество структурированных статей составило 93, неструктурированных – 14.

По требованиям ГОСТа [2, с. 5] в реферате можно выделить следующие информационные части – тему, цель, методы, результаты. Мы провели сопоставительный анализ предложений авторских рефератов и соответствующих статей, чтобы выявить, из каких структурных частей статьи следует отбирать информацию для реферата. В таблице 1 мы привели распределение информации для включения в реферат по разделам статей.

Таблица 1

Распределение релевантной для реферата информации по разделам статьи

Раздел статьи с информацией для реферата	Кол-во предложений, %
Введение / Начало	57,4
Середина	30,8
Заключение / Конец	11,8
Предложений всего	100

Мы выявили, что значительная часть информации для реферата (около 31 %) приводится в основном тексте статьи, поэтому для создания реферата необходимо обрабатывать полный текст статьи.

Научный текст характеризуется определенной синтаксической сложностью [3, с. 8]. Тем не менее, для улучшения результатов автоматического перевода текст должен обладать определенными характеристиками переводимости. В [4, с. 148] переводимость текста определяется как «свойство текста, определяемое возможностью быть переведенным на другой язык (в другую систему симво-

лов)». Авторы работы [5, с. 235] определяют переводимость текста для машинного перевода как меру, отражающую качество перевода данного предложения текста с помощью определенной системы машинного перевода. В работе [8, с. 364] авторы приводят следующие параметры, негативно влияющие на качество машинного перевода: слишком длинные (более 25 слов) либо короткие (менее 3 слов) предложения, придаточные предложения, эллипсис, однородные члены, неоднозначная лексика.

При анализе материала мы обнаружили следующие недостатки научного текста, усложняющие его перевод.

Большое количество причастных оборотов в предложении:

В силу этого, для расчета физических полей в квазифрактальных средах, представимых вкраплениями одной среды в другую (заполненных флюидом пор в скелете, зерен одного вещества в другом и т. п.), могут быть использованы классические алгоритмы, учитывающие большое количество мелких анизотропных включений, генерируемых процедурами, реализующими при построении принцип фрактальности.

Большая длина предложения:

Указанное представление служит основой нового декомпозиционного подхода к перечислению минимальных разрезов графа, состоящего, во-первых, из поиска только неприводимых минимальных разрезов в графе и, во-вторых, из синтеза всего множества минимальных разрезов по частично упорядоченному подмножеству неприводимых разрезов в дистрибутивной решетке минимальных разрезов.

Математические формулы в научном тексте могут выполнять синтаксические роли:

Нашей целью является изучение однозначной разрешимости задачи Коши для уравнения (1), а также устойчивости решений уравнения (1) в окрестности точки нуль в случае, когда оператор L необратим, в частности, его ядро $\ker L \neq \{0\}$.

Эллипсис:

В перспективе результаты работы позволят, с одной стороны, перейти к рассмотрению полулинейных эволюционных уравнений с памятью, а с другой стороны – исследовать уравнения Соболевского типа с памятью.

Библиографический список

1. Беляева, Л. Н. Научная статья: подготовка к публикации : методические рекомендации / Л. Н. Беляева, Н. Л. Шубина – СПб. : Книжный Дом, 2009. – 32 с.
2. ГОСТ 7.9-95. Система стандартов по информации, библиотечному и издательскому делу. Реферат и аннотация. Общие требования. – М. : Изд-во стандартов, 1995. – 8 с.
3. Котюрова, М. П. Культура научной речи. Текст и его редактирование : учебное пособие / М. П. Котюрова, Е. А. Баженова – 2 е изд., перераб. и доп. – М. : Флинта : Наука, 2008. – 280 с.

4. Нелюбин, Л. Л. Толковый переводоведческий словарь / Л. Л. Нелюбин. – 3-е изд., перераб. – М. : Флинта : Наука, 2003. – 320 с.
5. Automatic Rating of Machine Translatability / K. Uchimoto, N. Hayashida, T. Ishida, H. Isahara // Machine Translation Summit (MTSummit X). – P. 235–242.
6. Peh, W. C. G Basic structure and types of scientific papers / W. C. G. Peh, K. H. Ng // Singapore Medical Journal. – 2008. – Vol. 49 Issue 7. – P. 522–525.
7. The STM Report. 2012. An overview of scientific and scholarly journal publishing. – http://www.stm-assoc.org/2012_12_11_STM_Report_2012.pdf.
8. Underwood, N. L. Translatability Checker: A Tool to Help Decide Whether to Use MT / N. L. Underwood, B. Jongejan // Proceedings of MT Summit VIII 18th-22nd September 2001, Santiago de Compostela. – P. 363–368.