

М. А. Жданович

ГЕНЕРАЦИЯ СЛОВАРНОЙ СТАТЬИ ДЛЯ ТЕРМИНОЛОГИЧЕСКОГО СЛОВАРЯ: ВОЗМОЖНОСТИ ИИ

В представленной работе рассматривается процесс разработки словарной статьи для терминологического словаря с использованием современных инструментов искусственного интеллекта (ИИ). Нами были изучены возможности и ограничения чат-бота DeepSeek при создании словарной статьи для англоязычной терминологической единицы 'code', принадлежащей предметной области «Информационные технологии».

Для проведения эксперимента в чат-бот был отправлен следующий запрос: «*Make up a dictionary entry for the term 'code' for a terminological dictionary of the IT domain. Pay attention to all the meanings of this word.*». Полученная в ответ на данный запрос информация, была сгруппирована в зависимости от зоны словарной статьи, которую она заполняет. Полученный результат представлен в таблице 1.

Т а б л и ц а 1

Сгенерированная словарная статья

Code	
Зона грамматических характеристик	<i>Noun</i>
Зона фонетических характеристик	/koud/

Зона семантизации	дефиниция	пример
	<p>1. Software Code: a structured set of instructions written in a formal programming language to define the behavior, logic, and functionality of a computer program or system.</p> <p>a. Source code: Human-readable text written in high-level languages (e.g., Python, Java, C#).</p> <p>b. Machine code: Low-level binary instructions executed directly by a CPU.</p> <p>c. Bytecode: intermediate code executed by virtual machines (e.g., Java Virtual Machine).</p>	<p>The team refactored the Python code to improve its readability and performance.</p>
	<p>2. Data Encoding A system of rules or algorithms for converting information into a specific format for storage, transmission, or processing.</p>	<ul style="list-style-type: none"> - Character encoding: ASCII, Unicode (UTF-8, UTF-16) for text representation. - Binary code: Data represented as sequences of 0s and 1s. - Graphical codes: QR codes, barcodes for embedding information visually.
	<p>3. Specialized IT Contexts</p> <ul style="list-style-type: none"> - Error codes: numeric or alphanumeric identifiers for diagnosing issues - Cryptographic codes: algorithms for encryption/decryption - Version control: unique identifiers to track changes in codebases. - Hardware codes: firmware or microcode embedded in devices 	<ul style="list-style-type: none"> - HTTP status codes: 404 «Not Found», 500 «Internal Server Error» - AES, RSA, SHA-256. - Git commit hashes - BIOS/UEFI.
Зона парадигматических отношений	<i>algorithm, syntax, compiler, interpreter, debugger, cipher, repository</i>	
Зона этимологической справки	From Latin <i>codex</i> («book of laws»), later evolving to mean systems of rules or symbols. Adopted into computing in the mid-20th century.	
Зона идиоматического значения	<p>Codebase: The entire collection of source code for a project.</p> <p>Code of conduct: Rules governing behavior in tech communities or organizations.</p> <p>Spaghetti code: Poorly structured, tangled software code.</p>	

Следует отметить, что чат-бот DeepSeek также предоставил информацию энциклопедического характера. Так, для первого значения термина *'code'* были определены стадии разработки, а именно: *'writing'*, *'compiling/interpreting'*, *'debugging'*, *'deploying'*. Кроме того, в конце словарной статьи содержалась заметка о том, что данный термин в широком значении обозначает любой системный метод представления или преобразования информации, например, генетический код, азбука Морзе, в то время как в области ИТ он связан с программированием и репрезентацией данных.

В целом можно утверждать, что посредством обращения к ИИ пользователь имеет возможность быстро и легко получить общее представление о месте термина в терминосистеме. Но если речь идет о создании терминологического словаря, то данная словарная статья требует существенных доработок. Например, проведенный анализ словарных статей термина *'code'* в 6 словарях предметной области ИТ (Butterfield, A. Oxford Dictionary of Computer Science, Collin, S. M. H. Dictionary of Computing, Collin, S. M. H. Dictionary of ICT, Downing, D. A. Dictionary of Computer and Internet Terms, Microsoft Computer Dictionary, Rigdon, J. C. Dictionary of Computer and Internet Terms) показал, что все значения анализируемого термина можно объединить в три группы:

- 1) преобразование информации;
- 2) представление информации;
- 3) шифрование.

В сгенерированной словарной статье значения также делятся на три группы, но третья группа включает ряд значений, объединенных неким специализированным контекстом.

Кроме того, необходимо обратить внимание на отсутствие примеров употребления изучаемого термина, четко отражающих все его значения, а также на то, что термин *'code'* может выполнять роль глагола, и в данном случае тоже является многозначным. Следовательно, нужно просить чат-бот доработать словарную статью или завершить работу вручную.

Проведенный анализ работы чат-бота DeepSeek по генерации словарной статьи демонстрирует, что ИИ становится эффективным инструментом, но для повышения качества сгенерированной словарной статьи необходимо тщательно подходить к написанию запроса и отражать в нем как можно более детальное описание зон и требований к содержащейся в них информации.

Таким образом, интеграция ИИ в процесс создания терминологических словарей поможет получить большой пласт информации, характеризующей искомую терминологическую единицу с разных сторон. Сбор информации является самым трудоемким процессом, поэтому преимущества ИИ на данном этапе неоспоримы. Но нельзя полностью доверять истинности данных, поэтому всю информацию необходимо подвергнуть дополнительной проверке. Следовательно, ИИ ускоряет процесс создания словарей, в частности, словарных статей, но в их редактировании должны принимать участие лингвисты и эксперты в исследуемых областях.