



УДК 616-07:004

ПРОБЛЕМА КОРЕФЕРЕНТНОСТИ И МОДЕЛЬ КОДИФИКАЦИИ КЛИНИЧЕСКОЙ ИНФОРМАЦИИ

О.М. КУЗЬМИНОВ¹
Н.Н. ШАЛАМОВА²
О.В. МУРОМЦЕВА¹

¹⁾ *Белгородский государственный
национальный исследовательский
университет*

²⁾ *МБУЗ «Городская клиническая
больница № 1», г. Белгород*

e-mail: o-kiselva@mail.ru

В статье рассмотрены прикладные задачи информатизации лечебно-диагностического процесса, исследованы особенности клинических симптомов и синдромов (нозологических форм) как информационных объектов баз данных, обоснован оптимальный алгоритм кодификации клинической информации в аспекте решения проблемы кореферентности.

Ключевые слова: кореферентность, формализация, симптомы, синдромы (нозологические формы).

Широкое внедрение информационных технологий в практическое здравоохранение диктует необходимость ввода, поиска, извлечения, хранения и обмена данных между различными информационными системами и пользователями. Объединение медицинских информационных систем в единое пространство предполагает обеспечение кореферентности клинической информации, связанной с сопоставлением данных, получаемых от различных источников.

Кореферентность или референциональное тождество — отношение между именами — компонентами высказывания, в котором имена ссылаются на один и тот же объект (ситуацию) внеязыковой действительности (референт) [1, 2]. Например, имя «ревматизм» кореферентно имени «ревматическая лихорадка» или «ослабляющая лихорадка» кореферентно «ремитирующей лихорадке» и так далее.

Проблема кореферентности решается в рамках научного направления разработки искусственного интеллекта и компьютерной лингвистики [1, 9, 10, 11, 12]. Сущность проблемы заключается в трудностях определения семантической совместимости различных текстовых структур друг с другом при компьютерной обработке электронных записей. Как отмечает В.Л. Бердник, необходимость определения семантической эквивалентности двух и более символьных имен сущности возникает в таких задачах, как исключение семантически дублирующих записей таблиц баз данных, перенос сведений между ними в виде электронных или печатных документов [1]. В медицине особенную актуальность решение этой проблемы приобретает в связи с переходом к электронному документообороту, созданием глобальных систем хранения медицинских данных и необходимостью их компьютерного анализа [9, 6]. Для решения проблемы кореферентности создаются специальные программные системы, обеспечивающие идентификацию (референцию) семантической сущности явлений, описанных многообразными языковыми средствами.

Большое значение для решения проблемы кореферентности приобретает формализация медицинской информации. Формализация включает два направления: кодификацию информации (представление в наиболее абстрактной форме — в виде числовых или знаковых символов) и стандартизацию медицинских терминов (создание различных номенклатур). В качестве примера реализации данного подхода можно привести Международную классификацию болезней. Однако она предназначена в основном для статистических целей и не решает многообразных клинических задач. До сих пор отсутствуют общепринятые номенклатуры симптомов, синдромов и нозологических форм. Следует заметить, что в настоящее время предложены классификаторы осложнений и исходов заболеваний, а также классификатор фармакотерапевтических лекарственных групп для «установления единых требований к порядку диагностики и лечения больных с различными заболеваниями» [3]. Очевидно, что дальнейшее создание номенклатур медицинских терминов, синдромов и симптомов, разработка алгоритмов кодификации медицинской информации является необходимым условием дальнейшего внедрения информационных технологий в клиническую практику. Обоснование принципов формализации клинической информации является одной из актуальных задач совершенствования лечебно-диагностического процесса. Как отмечает В.М. Тавровский, «на хорошо структурированной ос-



нове можно делать все: проверять внутреннюю логику истории болезни, давать подсказки врачу, делать кое-что за врача» и многое другое [7, 8].

Цель работы. Обоснование и разработка принципов кодификации клинических симптомов, синдромов (нозологических форм) как информационных объектов баз данных для прикладных медицинских информационных систем.

Материалы и методы. С использованием методов системного анализа исследованы прикладные задачи информатизации лечебно-диагностического процесса, особенности клинических симптомов и синдромов (нозологических форм) как информационных объектов баз данных, обоснован оптимальный алгоритм кодификации клинической информации в аспекте решения проблемы кореферентности.

Формализация информации непосредственно зависит от тех задач, которые она обеспечивает. При информатизации лечебно-диагностического процесса проблема семантической совместимости может возникнуть при формировании документов (протоколов лечебно-диагностического процесса, «истории болезни») с использованием текстовых шаблонов. Это могут быть вербально-формализованные шаблоны в протоколах описания клинических симптомов и синдромов (нозологических форм) при формировании электронных медицинских записей. Кроме того, поиск сведений в электронных медицинских архивах (накопление и использование клинического опыта), обеспечение поддержки диагностических решений, использование клинической информации в медицинской статистической отчетности, клиническом управлении, в дидактических целях также требует решения проблемы кореферентности.

В различных базах данных семантическая совместимость информации может быть обеспечена с помощью формализованных атрибутов – кодов. Кодирование клинической информации является важным элементом ее формализации и стандартизации. Главной целью кодификации информации, помимо обеспечения идентификации информационного объекта, является реализация принципа нормализации данных. Принцип нормализации данных – это исключение повторения и дублирования информации в предметной области.

Кодирование клинической информации должно согласовываться с отраслевым стандартом N 91500.16.0001-2001 «Спецификация файла данных государственного статистического наблюдения и формализованных показателей социально-гигиенического мониторинга для обмена информацией с помощью электронных (машинных) носителей и в системах информационных коммуникаций» [5].

«Кодирование представляет собой образование по определенным правилам кодов и присвоение их объекту или группе объектов, позволяющее заменить несколькими знаками (символами) наименования этих объектов. С помощью кодов обеспечивается идентификация объектов максимально коротким способом, т.е. с помощью минимального числа знаков. Минимизация количества знаков, идентифицирующих объекты, способствует повышению эффективности сбора, учета, хранения, обработки (анализа) информации» [4].

В основе алгоритма кодирования информации лежит ее классификация по иерархическому, фасетному или смешанному принципу: когда определяется множество объектов, подлежащее классификации для решения конкретных задач; определяются признаки разделения на подмножества; выбирается оптимальное количество уровней деления подмножеств, обеспечивающее удобство практического использования предметной области [4].

Конкретный алгоритм кодификации клинических информационных объектов зависит от назначения классификатора, специфики информации и характера решаемых прикладных задач предметной области. Так, базы данных симптомов заболеваний в основном предназначены для автоматизации формирования электронных записей клинических протоколов и создания систем поддержки диагностических решений. Синдромы (нозологические формы), как информационные объекты, обеспечивают предметную область формирования развернутого клинического диагноза, решение задач медицинской статистики и клинического управления. В связи этим формализация клинических симптомов и синдромов (нозологических форм) для информационных систем может иметь свои особенности.

Проведенный анализ прикладных задач информатизации лечебно-диагностического процесса позволил обосновать общие принципы и модель кодификации клинических информационных объектов для организации баз данных. Общая модель и алгоритм кодификации могут быть представлены следующим образом. Объектом кодификации являются клинические симптомы и синдромы (нозологические формы). Признаками классификации симптомов оптимально определить способы их выявления, а также отношение к основным функционально-анатомическим системам и органам. Признаками классификации синдромов (нозологическим формам) – отношение к этиологии и патогенезу, позволяющее идентифицировать заболевание

как отдельную форму и соотносить его с международной классификацией, отношение к частным особенностям проявления патологии для организации лечебно-диагностического процесса. Анализ задач, которые решаются с помощью кодификации симптомов и синдромов (нозологических форм), а также принцип минимальной достаточности позволяет обосновать в качестве наиболее оптимальной структуры восьмизначный, десятичный цифровой код. Формула кода: XX.XX.XX.XX (рис. 1)



Рис. 1. Общая модель кодификации симптомов и синдромов (нозологических форм)

Выводы:

- Кодификация клинической информации позволяет решить отдельные проблемы ко-референтности.
- Представленная модель восьмизначного четырехуровневого кода позволяет вместить достаточно обширное количество информационных объектов и резерв для кодирования новых объектов.
- Модель имеет минимальное количество знаков, обеспечивая достаточно подробную и пригодную для пользователя однозначную идентификацию объектов по определенным реквизитам.
- Модель обеспечивает возможность контроля ошибок при вводе в компьютерные системы.
- С использованием предложенных модели и алгоритма кодификации создана база данных основных симптом в терапии для обеспечения клинических и дидактических предметных областей.

Литература

1. Бердник, В. Л. Модели и методы семантического сравнения строк символов в коллекции документов [Текст] : автореф. дис. ... канд. техн. наук / В. Л. Бердник. – Волгоград : Стигма, 2008. – 24 с.
2. Лингвистический энциклопедический словарь [Текст]: / под ред. В.Н. Ярцева. – М. : Советская энциклопедия, 1990. – 688 с.
3. ОСТ ПВБ N 91500.09.0001-1999. Протоколы ведения больных. Общие положения. – Введ. 1999-09-01. – М., 1999. – 25 с.
4. ОСТ ТО N 91500.01.0003-2000. Принципы и порядок построения классификаторов в здравоохранении. Общие положения. – Введ. 2000-09-11. – М., 2000. – 12 с.
5. ОСТ ТО N 91500.16.0001-2001. Спецификация файла данных государственного статистического наблюдения и формализованных показателей социально-гигиенического мониторинга для обмена информацией с помощью электронных (машинных) носителей и в системах информационных коммуникаций. – Введ. 2001-02-01. – М., 2001. – 23 с.
6. Павлова, О. Информатизация российского здравоохранения только в начале пути [Текст] / О. Павлова // PC Week Review: ИТ в медицине. – 2011. – № 3. – С. 13-14 .
7. Тавровский, В.М. Основы автоматизации лечебно-диагностического процесса в стационаре и поликлинике : метод. рекомендации / В.М. Тавровский, А.И. Макаров. – Киров ; Тюмень, 2009. – 54 с. [Электронный ресурс] – Режим доступа: <http://vmtavr2.narod.ru/>
8. Тавровский, В.М. К чему должна привести информатизация здравоохранения: попытка спроектировать будущее / В.М. Тавровский, А.В. Гусев // [Электронный ресурс] – Режим доступа: <http://vmtavr2.narod.ru/>



9. Tian, Ye He. Coreference Resolution on Entities and Events for Hospital Discharge Summaries/ Y.H. Tian // Massachusetts Institute of Technology. – 2007. – 78 p. [Электронный ресурс]. – Режим доступа: <http://groups.csail.mit.edu/mede/ftp/hetian/Tian-He-SM-Final.pdf>
10. Surdeanu, M. Combination strategies for semantic role labeling./ M. Surdeanu, M. Lluís, C. Xavier, Pere R. Comas. // Journal of Artificial Intelligence Research. – 2007. – № 29. – P. 105-151.
11. Recasens, M. BLANC: Implementing the Rand index for coreference evaluation / Marta Recasens and Eduard Hovy // Natural Language Engineering. – 2011. – № 17(4). – P. 485-510.
12. Lee, H. Joint Entity and Event Coreference Resolution across Documents/ Heeyoung Lee, Marta Recasens, Angel Chang, Mihai Surdeanu, Dan Jurafsky // Stanford University, Stanford. – 2012. [Электронный ресурс]. – Режим доступа: <http://www.stanford.edu/~jurafsky/emnlp2012-coref.pdf>

THE PROBLEM OF COREFERENCE REFERENTIAL IDENTITY AND CLINICAL INFORMATION CODIFICATION

O.M. KUZMINOV¹
N.N. SHALAMOVA²
O.V. MUROMTZEVA¹

*¹⁾ Belgorod National
Reserch University*

²⁾ Belgorod City Clinical Hospital №1

e-mail: o-kiselva@mail.ru

The paper under study deals with the applied task of informational support in diagnostic and treatment process, researches of clinical symptoms and syndromes as data base objects (disease entity) and justifying of optimal page in clinical information codification in terms of coreference referential identity problem solving.

Key words: coreference, formalization, symptoms, syndromes (disease entity).