

УДК 004.622

АВТОМАТИЗАЦИЯ ПРОВЕРКИ ДАННЫХ О ЛЕЧЕНИИ ОНКОЛОГИЧЕСКИХ ЗАБОЛЕВАНИЙ.

Каратецкая М.Ю. (Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург),

Суренков Н.С. (Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург),

Рыжков Н.М. (Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург),

Сулейманов Р.И. (Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург)

Научный руководитель – Береснев А.Д.

(Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург)

Аннотация

В докладе рассматривается архитектура информационной системы для автоматической проверки данных о лечении онкологических заболеваний и иной информации связанных с данным процессом.

Введение.

По статистике в России за последние 10 лет количество случаев онкологических заболеваний выросло на 23%. С увеличением зарегистрированных случаев увеличивается поток поступающих данных, который нужно обработать как для определения текущего состояния в онкологических центрах (количество больных, назначение лечение и другое), так и для прогнозирования, например, для оценки количества закупаемого оборудования и лекарств, определения необходимого размера финансирования, расположения ЛПУ и др. В настоящее время информация о ходе лечения, диагнозах, использованных лекарственных препаратах и оказанной высокотехнологичной помощи содержится во множестве информационных систем ОСМ, МИС и ЗАГС. Записи в них часто вносятся вручную и могут содержать ошибки. Соответственно использовать эти данные при условии их успешной агрегации для составления достоверных прогнозов не представляется возможным.

В национальном медицинском исследовательском центре (НМИЦ) онкологии им. Н.Н. Петрова стекается вся информация из онкологических центров в северо-западном регионе России. Объем этих данных значителен, поэтому для оптимизации работы и валидации данных институту требуется информационная система, которое позволит осуществлять проверку данных для формирования достоверного набора данных для прогнозирования.

Основная часть.

Исходными данными для работы являлись европейский стандарт на качество данных оказания услуг лечения онкологических заболеваний, результаты интервьюирования экспертов НМИЦ им.Петрова, результаты анализа специфики данных и нормативно-справочной информации.

Было установлено, что агрегируемые данные могут содержать как формальные (незаполненные поля, неверный формат записей и т.п.), так и смысловые ошибки (например, несоответствие значений одних данных допустимым значениям других полей). Из вышеописанного можно сделать вывод, что требуется осуществлять проверки по сложному набору правил, позволяющие проверять зависимости значений полей данных в их взаимосвязи.

Была построена типология правил проверки. Среди прочих установлено что необходимо производить проверку: корректность заполнения по типам данных (возраст, даты, числовые показатели), корректность постановления диагноза и лечения на основе имеющегося анамнеза больного, обязательность заполнения полей (в одиночных полях и в группе полей), возможны зависимости некоторых полей между собой, возможны и исключаящие комбинации

заболеваний (например, есть типы рака, которые могут возникнуть только в определенном возрасте). Ввиду сложности предметной области и отсутствия возможности сформировать окончательный оптимальный набор правил, то требуется обеспечить возможность для врача эксперта удобно модифицировать и изменять набор правил.

Таким образом проектируемая информационная система:

- 1) позволит составлять, хранить и обрабатывать наборы правил, всех необходимых типов,
- 2) обеспечит поддержку версионирование наборов правил,
- 3) предоставит эксперту возможность изменять наборы правил,
- 4) осуществит проверку данных по набору правил с целью обнаружения формальных и смысловых ошибок и неточностей,
- 5) осуществит их классификацию по уровням достоверности,
- 6) и автоматизирует подготовку отчетов о качестве данных.

Разрабатываемая информационная система будет включать в себя несколько модулей, каждый из которых служит для решения определенных задач, но в данной работе рассматривается модуль, предназначенный для автоматизации проверки данных.

Выводы.

В результате выполненной работы будут получены данные, которые в дальнейшем будут использованы для выполнения главной задачи проекта - проведение анализа и составление прогнозов. Для качественной аналитики необходимы чистые данные, не содержащие в себе неточностей и ошибок, поэтому большое внимание уделено именно их проверке.

Каратецкая М.Ю. (автор)

Суренков Н.С. (соавтор)

Рыжков Н.М. (соавтор)

Сулейманов Р.И. (соавтор)

Береснев А.Д. (научный руководитель)