

АЛГОРИТМ АВТОМАТИЗИРОВАННОГО АНАЛИЗА ПРИЧИННО-СЛЕДСТВЕННЫХ СВЯЗЕЙ С ИСПОЛЬЗОВАНИЕМ ГЕОИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ И ПРОГНОЗИРОВАНИЕ ДИНАМИКИ ЗДОРОВЬЯ НАСЕЛЕНИЯ

Комбарова М.Ю.¹, Рембовский В.Р.¹, Красильников И.А.², Радиллов А.С.¹, Дулов С.А.¹

¹*Федеральное государственное унитарное предприятие «Научно-исследовательский институт гигиены, профпатологии и экологии человека» Федерального медико-биологического агентства, 188663, Ленинградская обл., Всеволожский район, г.п. Кузьмолловский, ст. Капитолово, Т/факс (812) 449-61-77; (812) 449-61-71; e-mail: gpech@fmbamail.ru; [kombarova@rihophe.ru](mailto:kambarova@rihophe.ru).*

²*ООО «Стратег», 197341, Санкт-Петербург, пос. Парголово, Выборгское ш., д. 503, корп.3, лит.А, оф. 428, Тел.: +7 (921) 846-32-31, e-mail: igor.kras@hotmail.com*

Резюме: В статье представлена краткая характеристика прототипа информационно-аналитической автоматизированной системы учета, хранения и обработки медико-гигиенической информации о состоянии производственной среды и здоровья персонала химически опасных объектов, здоровья населения и среды его обитания в районах расположения данных объектов. Апробация работы программы проведена на примере г.Тихвина Ленинградской области.

Ключевые слова: химическая безопасность, химически опасный объект, информационно-аналитическая автоматизированная система, статистическое наблюдение, интерфейс, база данных, хранилище данных, гигиено-эпидемиологический анализ

ALGORITHM OF AUTOMATED EXPERT ANALYSIS OF CAUSE-EFFECT RELATIONSHIPS USING GEOINFORMATION TECHNOLOGIES AND PROGNOSIS OF POPULATION'S HEALTH DYNAMICS

Kombarova M.Yu.,¹ Rembovskiy V.R.,¹ Krasil'nikov I.A.,² Radilov A.S.¹, Dulov S.A.¹

¹*Research Institute of Hygiene, Occupational Pathology and Human Ecology Federal State Unitary Enterprise, Federal Medical Biological Agency, 188663, bulid. 93, Kapitolovo St., G/P Kuz'molovsky, 188663 Leningrad Region.*

Phone/fax (812) 449-61-77; (812) 449-61-7; e-mail: gpech@fmbamail.ru.

²*Strateg OOO, 197341, St. Petersburg, Pargolovo, Vyborgskoe av. 503, build. 3, lit. A, office 428*

Tel +7 (921) 846-32-31, E-mail: igor.kras@hotmail.com

Abstract: The paper provides a brief characteristic of a prototype informational analytical automated system for registration, storage, and processing of medical hygienic information on the state of occupational environment and health of the personnel of chemically hazardous objects, health of the population and its life environment in the vicinity of such objects. The system was tested on an example of the Tikhvin Town, Leningrad Region.

Key words: chemical safety, chemically hazardous object, informational analytical automated system, statistical monitoring, interface, database, hygienic epidemiological analysis.

Введение

В настоящее время исследования по оценке состояния здоровья в зависимости от загрязнения окружающей среды имеют весьма широкое распространение. Комплекс факторов окружающей среды оказывает воздействие на формирование популяционного здоровья населения, особенно в связи с изменением социально-экономических условий, сопровождающихся ослаблением контроля за качеством среды обитания, ухудшением демографической ситуации, изменением структуры питания населения. Анализ количественных зависимостей в системе «среда-здоровье» получил развитие в разработке критериев и методов количественной оценки воздействия факторов окружающей среды на здоровье населения

Оценка влияния экологических факторов на здоровье населения является очень сложной и наукоемкой проблемой, требующей проведения объемных исследований.

Среди множества факторов, формирующих здоровье населения, существенную роль играет качество среды обитания. Антропогенное загрязнение окружающей среды оказывает мощное воздействие на формирование популяционного здоровья населения. Деятельность человека по масштабам своего влияния на распространенность химических элементов в окружающей среде возрастает настолько, что становится в ряд с природными геохимическими факторами.

Для оценки влияния антропогенного загрязнения среды обитания на здоровье человеческой популяции большинство авторов выделяет в качестве основных объектов наблюдения атмосферный воздух и питьевую воду, а также депонирующие среды, в том числе почву.

Весьма актуальным стало сочетание экологического мониторинга с мониторингом здоровья населения, под которым понимается динамическое слежение за состоянием здоровья населения с пространственной интерпретацией данных для обоснования региональных мероприятий, направленных на устранение риска в конкретных районах [1].

Большое значение при проведении медико-экологического мониторинга и изучения влияния факторов риска на здоровье населения имеет полнота учета этих факторов. Недооценка корреляции между факторами среды обитания может привести к ошибочным результатам. Между переменными (случайными величинами) может существовать функциональная связь, проявляющаяся в том, что одна из них определяется как функция от другой. Кроме того, могут быть общие случайные факторы, влияющие на обе переменные. К сожалению, во многих работах, посвященных анализу факторов риска, этому обстоятельству уделяется недостаточное внимание. Решение проблемы использования в анализе коррелированных переменных может быть частично решено путем использования факторного анализа с сокращением числа переменных и замены групп переменных на так называемые главные факторы (главные компоненты) [2].

Серьезной проблемой остается действенный контроль за состоянием здоровья работающего населения. Влияние факторов, воздействующих на здоровье человека на рабочих местах, требует постоянного мониторинга с оценкой качественных и количественных изменений в годовой и многолетней динамике.

Для обоснования потенциально эффективных мероприятий по охране здоровья населения и работников промышленных предприятий, особенно с вредными условиями производства, необходимо проведение комплексных аналитических исследований на основе сведений, аккумулируемых в информационных системах «Производственная среда - здоровье персонала», «Среда обитания – здоровье населения».

Цель исследования

Целью данной работы явилась апробация прототипа информационно-аналитической автоматизированной системы (ИАС) учета, хранения и обработки медико-гигиенической информации о состоянии производственной среды и здоровья персонала особо опасных предприятий Ленинградской области, здоровья населения и среды его обитания в районах расположения указанных предприятий в целях обеспечения химической безопасности в регионе на примере г. Тихвина.

Материалы и методы исследования

Для апробации алгоритма автоматизированного экспертного анализа причинно-следственных связей в системах «производственная среда - здоровье персонала», «среда обитания – здоровье населения» в разработанном прототипе ИАС были использованы следующие материалы:

1. Результаты аттестации 235 рабочих мест ТФЗ, проведенной в декабре 2008 г. (150 рабочих мест) и в декабре 2010 г. (85 рабочих мест) по показателям оценки условий труда по степени вредности и опасности с градациями 1, 2, 3.1, 3.2, 3.3, 4.

2. Сведения о 2322 случаях заболеваний у 202 работников ТФЗ и примерно соответствующих им по возрастно-половой характеристике и месту проживания 194 жителей г.Тихвин.

3. Сведения о среде обитания в г. Тихвине преимущественно за период 2007-2012 гг., ..

Таблица 1 - Пробы, характеризующие среду обитания в г. Тихвине в 2007-2012 гг.

Фактор	Число проб
Вода питьевая централизованного водоснабжения	13 799
Вода поверхностная для питьевого и хозяйственно-бытового водоснабжения, а также для водоснабжения пищевых предприятий	2 622
Вода поверхностная для рекреационного водопользования, а также в черте населенных мест	1 170
Воздух атмосферный	29 304
Почва - близкая к нейтральной (суглинистая и глинистая), рН КСl>5,5	846
ВСЕГО	47 741

4. Результаты лабораторных исследований на микроэлементы волос детей г. Тихвина, посещающих детские сады в условно «грязной» (42 ребенка) и условно «чистой» (74 ребенка) зонах города.

5. Общая оценка здоровья населения осуществлялась по отчетам о числе заболеваний населения, обслуживаемого МУЗ «Тихвинская центральная районная больница» за период 2008-2012 гг. и сведений о смертности населения Тихвинского района за этот же период (всего в ИАС введено 7774 показателя).

6. Влияние среды обитания на здоровье населения оценивалось на основе данных о 5062 случаях заболеваний детей дошкольного возраста и школьников, посещающих детские сады и школы в условно-грязной (263 ребенка) и условно-чистой (274 ребенка) зонах города.

Основная аналитическая работа была ориентирована на использование технологии OLAP (On Line Analytical Processing), реализованной на платформе MS Office Professional Plus 2013 с применением PowerPivot, последней версией сводных таблиц.

Углубленный математико-статистический анализ выполнялся в пакете Statistica for Windows. В исследовании использованы методы изучения предпосылок болезней и различные виды статистического анализа. Проведено математико-картографическое моделирование с использованием ГИС-технологий [3-7] .

Результаты исследования

При выполнении исследования создан прототип автоматизированной информационной системы, предназначенной для сбора, хранения и обработки данных о здоровье людей и влияющих на него факторов. В основу ИАС положена разработанная база данных (БД) оригинальной структуры. Подготовлена необходимая для функционирования ИАС техническая документация и методические рекомендации по ее использованию.

Созданная ИАС позволила ввести для последующей обработки в упорядоченном виде сведения о производственной среде, здоровье населения и среде обитания всего более 240 тыс. записей в 44 таблицах БД.

В процессе работ сделан выбор в пользу свободно распространяемой реляционной системы управления базами данных (далее – СУБД) MySQL v.5.6. В качестве архитектурного решения использовалась облачная технология с доступом к базе данных (далее – БД) через Интернет.

При создании БД, в которой должны храниться все сведения, необходимые для автоматизированного анализа причинно-следственных связей, необходимо было предусмотреть возможность однозначной идентификации всех характеристик с целью обеспечения возможности их безошибочного извлечения для последующего анализа, а также ввод первичных сведений о людях и влияющих на них факторов (транзакционная БД).

БД носит реляционный характер, где все сведения распределяются по плоским двумерным таблицам, в которых каждая запись (строка) однозначно идентифицируется так называемым первичным ключом, а в таблицах, где содержатся соответствующие связанные сведения, имеются внешние ключи. Таким образом, создается весьма сложный комплекс таблиц, между которыми существуют связи (отношения, реляции) типа «один-ко-многим», «многие-к-одному», «многие-ко-многим».

В ИАС можно выделить три информационных блока:

- люди и их индивидуальные характеристики;
- состояние среды обитания;
- популяционные характеристики населения.

Для ввода данных разработан интерфейс, обеспечивающий возможность подключения к БД ИАС тонкого клиента. Интерфейс позволяет вводить и редактировать как собственно данные, так и классификаторы и справочники.

Обработка данных включает несколько этапов. Первый этап традиционно включает статистическое наблюдение, обеспечивающее сбор сведений, необходимых для дальнейшего анализа. Затем эти сведения вводятся в транзакционную БД (этап 2).

После заполнения БД данными статистического наблюдения третий традиционный шаг – создание Хранилища данных (Data Warehouse). В хранилище с определенной периодичностью импортируются те данные из транзакционной БД, которые с высокой вероятностью не будут в дальнейшем меняться. В зависимости от задач, решаемых с помощью информационной системы, обновление данных в Хранилище данных может происходить с разной периодичностью – с интервалом времени от нескольких минут до нескольких дней и даже месяцев.

Одна из основных проблем, решаемых с помощью Хранилища – это упрощение структуры хранения данных, что существенно облегчает работу специалиста-аналитика. Такое упрощение возможно в связи с тем, что в Хранилище данные хранятся в неизменяемом виде, что снимает необходимость придерживаться принципов нормализации БД. Таблицы в Хранилище обычно структурированы в виде звезды (центральная таблица фактов в окружении нескольких таблиц-справочников) или снежинки (когда таблицы справочники имеют, в свою очередь, подчиненные им периферические таблицы).

Однако, упрощение хранения данных не решает задачу ускорения их обработки. Для решения этой задачи в 90-е годы прошлого столетия была разработана технология OLAP (On Line Analytical Processing). На 4-м этапе данные хранилища подвергаются дополнительной обработке. По специально разработанным алгоритмам в фоновом режиме, например, в ночное время, осуществляются многочисленные расчеты и на основе данных Хранилища формируются показатели, которые, с высокой вероятностью, могут потребоваться при проведении анализа. Абстрактно это можно представить как формирование некоего виртуального куба данных, все ячейки которого заполнены уже готовыми показателями, а срезы граней представляют собой измерения какого-либо признака. Результатом такой подготовительной работы является высочайшая скорость получения результатов запросов, которые на миллионах строк таблиц Хранилища выполняются практически мгновенно.

Для визуализации полученных результатов к OLAP-кубу на 5-м этапе подключают какие-либо внешние приложения. Это могут быть как многофункциональные бизнес-приложения класса Business Intelligence (например, Oracle BI или Crystal Reports, ArcGIS), так и более простые и распространенные программы, например, MS Excel, рисунок 1.

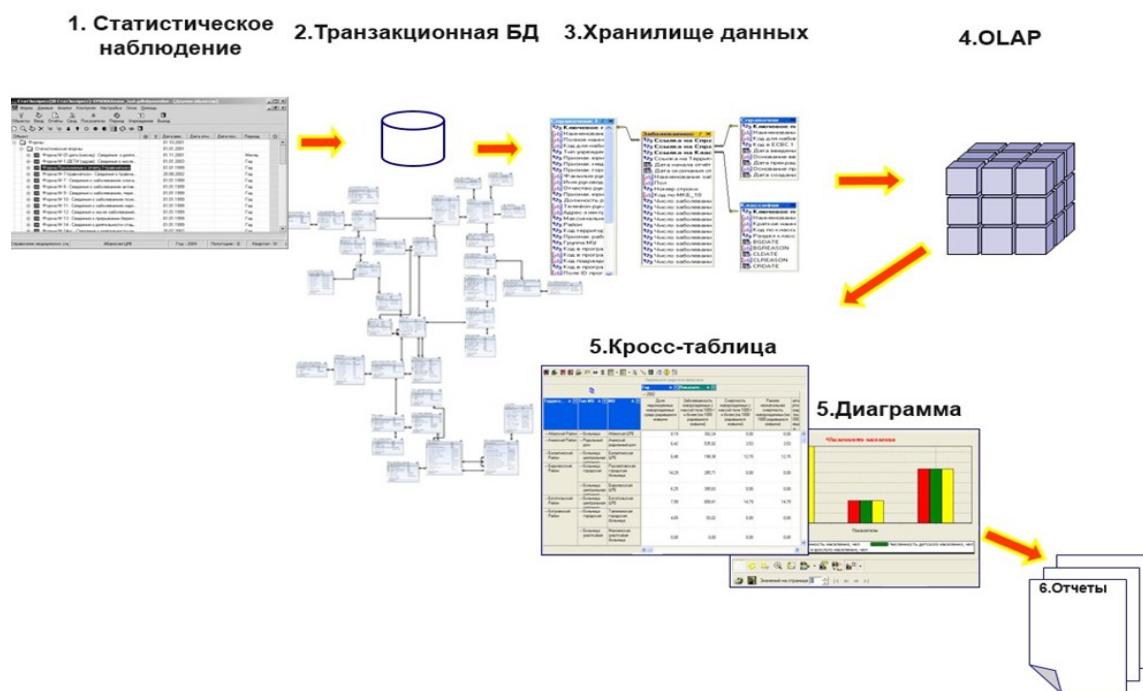


Рисунок 1. - Алгоритм анализа данных с использованием Хранилищ данных и OLAP (On Line Analytical Processing)

В последние годы появились решения, которые для относительно небольших БД объемом в несколько гигабайт (несколько миллионов записей) позволяют существенно упростить алгоритм анализа. В первую очередь речь идет о так называемой надстройке COM для MS Excel – PowerPivot (MS Office Professional Plus 2013).

Функциональность, заложенная в PowerPivot 2013, в том числе развитый, но весьма простой язык запросов DAX, позволили использовать ее в ИАС, выполняя в одном приложении сразу три этапа обработки данных: создание Хранилища данных, формирование OLAP-куба и визуализация полученных результатов в виде динамических перекрестных таблиц и диаграмм, рисунок 2.

Сбор и обработка данных в ИАС нацелены на проведения комплексного гигиено-эпидемиологического анализа состояния производственной среды и здоровья персонала, а также состояния здоровья населения и среды его обитания. Обобщенный алгоритм такого анализа предполагает знание характеристик среды обитания (часть которых может быть факторами риска развития патологии у человека) и популяционных характеристик здоровья населения, живущего и/или работающего в этой среде.

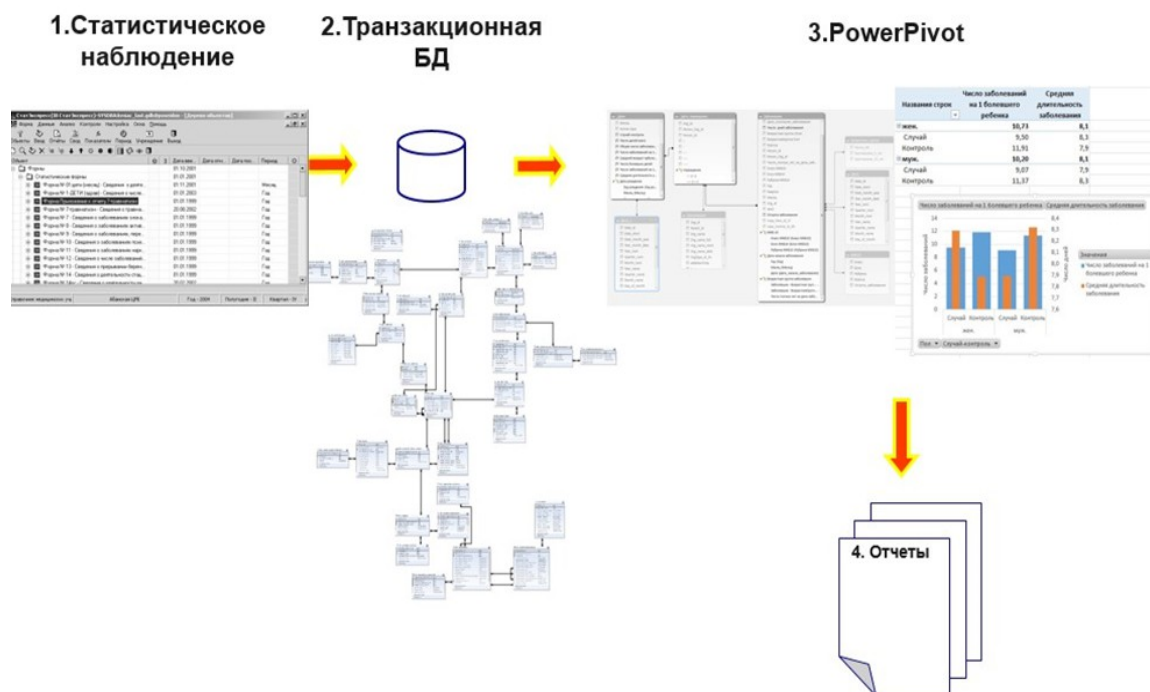


Рисунок 2. - Алгоритм анализа данных с использованием PowerPivot

При наличии таких сведений, обычно представленных в виде матриц данных с разбивкой по территориям (местам работы), группам населения (персонала предприятий) и времени, необходимо поэтапно:

- установить наличие или отсутствие статистических корреляционных связей между показателями здоровья и предполагаемыми факторами риска;
- при наличии корреляционных связей обосновать наличие или отсутствие причинно-следственных связей;
- при наличии причинно-следственных связей вскрыть механизм их действия и постараться определить конкретные элементы этого механизма, на которые можно воздействовать потенциально-эффективными профилактическими мероприятиями.

С помощью ИАС можно подготовить данные, которые потребуются для проведения первого этапа – установления статистических корреляционных связей, в том числе статистических моделей с использованием корреляционного, регрессионного, факторного, кластерного и других методов многомерного анализа данных. Для автоматизации последующих этапов гигиено-эпидемиологического анализа требуется

использование систем искусственного интеллекта (например, экспертных систем), разработка которых вполне возможна, но выходит за рамки настоящей работы.

Алгоритмы обработки данных в системах «Производственная среда – здоровье персонала» и «Среда обитания – здоровье населения» реализуются в ИАС в виде сводных таблиц с использованием PowerPivot. При этом пользователи имеют возможность после дополнительного обучения модифицировать существующие или разрабатывать новые сводные таблицы для решения стоящих перед ними задач.

В PowerPivot можно одновременно загрузить несколько таблиц, а затем, используя визуальный интерфейс, установить между ними связи, необходимые для однозначной идентификации анализируемых данных.

Полученные на основе PivotPivot сводные таблицы позволяют проанализировать как распределение результатов измерений факторов на рабочих местах (среде обитания) практически с любыми наложенными условиями, так и оценить влияние совокупности факторов на здоровье человека.

В результате проведенной работы выполнен комплексный гигиено-эпидемиологический анализ состояния производственной среды и здоровья персонала ЗАО «ТФЗ», а также состояния здоровья населения и среды его обитания в районе расположения указанного объекта в динамике за период 2007-2012 годы в ИАС, включая ГИС анализ [8-11].

Установлен перечень основных специальностей, составляющих «группу риска» по степени вредности и опасности трудового процесса (бригадиры, рабочие основных цехов, вспомогательные технические работники); перечень основных цехов по вредности и опасности (плавильный корпус, склад шихты с шихтоподачей); «лидирующая» патология у группы риска (болезни органов дыхания, болезни системы кровообращения, болезни костно-мышечной системы, болезни кожи и подкожной клетчатки, болезни глаза и его придаточного аппарата).

В результате исследования «копи-пара» сделан вывод о повышенном риске развития у работников ЗАО «ТФЗ», как у женщин, так и у мужчин, болезней глаза и его придаточного аппарата и более высокая вероятность развития психических расстройств и расстройств поведения. Необходимо также обратить внимание на весьма высокий риск заболеваний мочеполовой системы у женщин и костно-мышечной и соединительной ткани у мужчин.

Установлено, что ведущими вредными факторами, способными оказать неблагоприятное влияние на состояние здоровья населения являются:

- загрязнение атмосферного воздуха озоном, свинцом, диоксидами азота и серы. За весь анализируемый период максимальные значения среднесуточных концентраций для диоксида азота составили 2,4 ПДКсс, взвешенных веществ – 9,5 ПДКсс, озона – 2,6 ПДКсс, диоксида серы – 4,1 ПДКсс, оксида азота – 1,2 ПДКсс.

- неудовлетворительное качество питьевой воды, содержащей повышенное количество хлорорганических соединений (хлорметана) и алюминия, образовавшихся на БОС при обработке воды из р. Тихвинки, а также повышенное количество железа и высокая мутность воды из-за неудовлетворительного состояния водопроводных сетей города;

- загрязнение почвы бенз(а)пиреном. Выявлена наметившаяся в последние годы тенденция к росту загрязнения почвы цинком, свинцом, кадмием и медью.

Отмечено ежегодное снижение показателя уровня общей смертности населения Тихвинского района.

Основными причинами смертей населения Тихвинского района на протяжении последних пяти лет (как и в предшествующие годы) являлись болезни системы кровообращения, новообразования, травмы, отравления и некоторые другие последствия воздействия внешних причин.

Итоговый уровень общей заболеваемости населения Тихвинского района достоверно превышал общеобластной показатель за весь период наблюдения, однако имел тенденцию снижения.

К «лидирующим» формам патологии взрослого, детского населения и подростков г. Тихвина отнесены болезни органов дыхания, болезни уха и сосцевидного отростка, болезни глаза и его придаточного аппарата, болезни кожи и подкожной клетчатки, а также болезни костно-мышечной системы.

В группу «лидирующей» патологии детей района «условного риска» входят болезни глаза и его придаточного аппарата, болезни уха и сосцевидного отростка, болезни кожи и подкожной клетчатки, болезни костно-мышечной системы и соединительной ткани, а также психические расстройства и расстройства поведения.

Отмечено динамическое снижение показателей, т.е. первичная регистрация заболеваний среди детей имела положительный тренд.

Выявлена тесная корреляционная связь между заболеваниями органов пищеварения и болезнями костно-мышечной системы у детей из г. Тихвина с повышенным содержанием в питьевой воде железа и алюминия.

Установлен дефицит содержания в организме детей кальция, магния, селена, меди и кобальта (у 34,8 – 88,9% детей), участвующих в процессах формирования и функционирования костно-мышечной системы у человека.

Анализ результатов биотестирования детей детских дошкольных учреждений г.Тихвина (детские сады «Зоренька», «Россияночка», «Улыбка» и «Чайка») показал, что частота обнаружения металлов из группы «жизненно-важные вещества» в градации -3 - -4 (недостаток) или +3 - +4 (избыток) по большинству показателей выше у детей, посещающих детский сад «Улыбка» («условно чистый» микрорайон) .

В целом по заболеваемости и смертности населения г. Тихвина дан краткосрочный благоприятный прогноз.

С учетом ранее выделенных «условно грязных» и «условно чистых» микрорайонов г. Тихвина и полученных результатов ГИС анализа влияние выбросов ЗАО «ТФЗ» на заболеваемость населения, проживающего в районе расположения предприятия не выявлено.

Необходимо отметить преимущества и отличия созданного прототипа ИАС:

1. Методологическое единство предметной области;
2. Открытая архитектура системы, позволяющая обновлять информацию и запросы по требованию пользователя системы;
3. Комплексное проектирование системы, включающее полный жизненный цикл системы: проектирование, реализацию, внедрение в эксплуатацию;
4. Унификация используемой нормативной базы (классификаторов, кодификаторов, словарей, картографических основ и т.д.);
5. Оперативная аналитическая обработка данных (On-Line Analytical Processing - OLAP);
6. Максимальное использование имеющегося научного, информационного, технического, программно-технологического и кадрового потенциала;
7. Быстрая скорость ввода и простота обработки данных.

В рамках развития ИАС рекомендовано шире использовать возможности современных информационных систем, созданных в структурах здравоохранения и обязательного медицинского страхования. В них накоплены детальные сведения об оказанных медицинских услугах, что позволяет перейти от статистических показателей в разрезе муниципальных образований к статистике, привязанной к любым произвольно взятым участкам жилой застройки, в том числе примыкающим к точкам контроля окружающей среды.

Для совершенствования ИАС можно выделить несколько основных направлений: более глубокая интеграция ИАС с геоинформационными системами; организация «корпоративного» доступа с использованием Active Directory и LDAP; разработка модулей пакетной загрузки и обработки данных, а также модулей пакетной выгрузки данных в различных форматах – XML, XLS, PDF и т.п.

В рамках развития геоинформационного направления логично добавить функционал создания и хранения геоинформационных объектов в ИАС с их последующим использованием при обработке и анализе данных.

Отдельным интересным направлением развития системы является создание приложения для мобильной платформы Android, что позволит осуществлять сбор, обработку и анализ данных «на местах» с использованием всех преимуществ мобильных технологий.

Наличие API позволяет в перспективе осуществлять обмен данными между ИАС и различными федеральными информационными системами.

ИАС предназначена для специалистов Федерального медико-биологического агентства, органов и учреждений Роспотребнадзора Российской Федерации, научно-исследовательских институтов, территориальных комитетов экологии и природных ресурсов, осуществляющих надзор и контроль за объектами окружающей среды и санитарно-эпидемиологической безопасностью населения, а также для управлений территориями, администрации химически опасных объектов.

Литература:

1. Прохоров Б.Б. Экология человека: учеб. для студ. высш. учеб. заведений / Б.Б.Прохоров. – 5-е изд., стер. - М.:Издат. центр «Академия», 2010. - 320с.
2. Электронный учебник по статистике. StatSoft, Inc. (2012). – М.: StatSoft. - WEB:<http://www.statsoft.ru/home/textbook/default.htm>
3. Малхазова СМ. Медико-географический анализ территорий: картографирование, оценка, прогноз. - М.: Научный мир, 2001. - 240с.
4. Барышников И.И., Вишневский Е.П., Нагорный С.В. Программно-математические решения задач картографирования и управления данными в интересах медико-географических исследований // Медико-географические аспекты оценки уровня здоровья населения и состояния окружающей среды. Сб. науч. трудов.– СПб. 1992.– С. 46-52.

5. Макаров В.З., Новаковский Б.А., Чумаченко А.Н. Эколого-географическое картографирование городов.– М.: Научный мир, 2002.– 196с.
6. Жуков В.Т., Новаковский Б.А., Чумаченко А.Н. Компьютерное геоэкологическое картографирование.– М.: Научный мир, 1999.– 128с.
7. Заруцкая И.П., Красильникова Н.В. Картографирование природных условий и ресурсов.– М.: Недра, 1988.– 300с.
8. Ревич Б.А., Авалиани С.Л., Тихонова Г.И. Основы оценки воздействия загрязненной окружающей среды на здоровье человека. Пособие по региональной экологической политике. - М.: Акрополь, ЦЭПР, 2004. - 268с.
9. Зейлер М. Моделирование нашего мира.– М.: МГУ им. Ломоносова, 2001 – 254с.
10. Комплексная гигиеническая оценка степени напряженности медико-экологической ситуации различных территорий, обусловленной загрязнением токсикантами среды обитания населения: Методические рекомендации. – Утв. Гл. госсанврачом РФ 30.07.97 г., № 2510/5716-97-32. – М., 1997. – 28с.
11. Маймулов В.Г., Нагорный С.В., Шабров А.В. Основы системного анализа в эколого-гигиенических исследованиях. – СПб.: СПбГМА им. И.И. Мечникова, 2000. – 342с.