УДК 005.7; 681.5; 614.2

РАЗВИТИЕ СИСТЕМЫ РЕПОЗИТАРИЯ ДАННЫХ И ЗНАНИЙ АВТОМАТИЗИРОВАННОГО ГЕНЕРАТОРА СИСТЕМНО-ОБОСНОВАННОГО ТЕХНИЧЕСКОГО ЗАДАНИЯ НА МИС

Грицюк Е.М.[1], Гольдштейн С.Л.[2], Шипигусев А.А. [2]

[1] ГБУЗ СО ДКБВЛ НПЦ «Бонум», г. Екатеринбург

[2] ФАГОУ ВПО УрФУ, г. Екатеринбург

Разработаны структура и алгоритмы функционирования системы репозитария данных и знаний в составе автоматизированного генератора системно-обоснованного технического задания на медицинскую информатику.

Ключевые слова: репозитарий, сохранение информации, хранилище информации, ранжирование информации, модель, структура, алгоритмы.

Development of the data and knowledge repository system of automated generator system-informed technical specifications

[1] Scientific and Practical Center "Bonum", Ekaterinburg

[2] Ural Federal University, Ekaterinburg

Gritsyuk E.M. [1], Goldshtein S.L. [2], Shipigusev A.A. [2]

Developed structure and algorithms of the system repository of data and knowledge as part of an automated generator system- informed technical specifications for medical informatics.

Key words: repository, saving of information, storage of information, information ranging, model, structure, algorithms.

Актуальность и постановка задачи

В ЭВМ первых поколений каждая прикладная программа снабжалась своими собственными данными, которые вводились в ЭВМ вместе с программой. Пока на ЭВМ решались отдельные, не связанные друг с другом задачи, такое положение было естественным. Однако начиная с 60-х годов и особенно в 70-е годы основная масса применений ЭВМ получила другое, так называемое, системное направление. В системных применениях задачи, решаемые на ЭВМ, взаимосвязаны воедино общей целью [1].

Не будем рассматривать традиционный, т.е. наиболее известный технический подход к информации, ее количеству и передаче в виде сигналов по К.Шеннону. Для объяснения сущности информации этот подход не очень продуктивен, так как оперирует с сигналами, как носителями информации, т.е. узок. Нас больше интересует непосредственно информация, как качественный феномен или явление, т.е. определенная сущность, без которой невозможно разрешение ситуаций в научно-практической структуре на основе управления знаниями и управления на основе знаний [2].

При создании системно-обоснованного технического задания (СО ТЗ) на информационную систему, в частности, на медицинскую [3], необходимо рациональное хранение информации с обеспечением ее безопасности (в т.ч. возможности восстановления); ранжирования по очередности ее использования в разных системах автоматизированного генератора системно-обоснованного технического задания (АГ СО ТЗ); накопления промежуточных продуктов, готовых ТЗ, опыта и отчетности; взаимодействия других подсистем генератора с репозитарием; а также - различного уровня доступа пользователей; прежде всего, разработчика и заказчика.

В данной статье поставлена и решена задача развития системы репозитария данных и знаний в составе АГ СО ТЗ.

Прототип системы репозитария

Репозитарий – одна из самых главных систем АГ СО ТЗ (рис.1) [4].

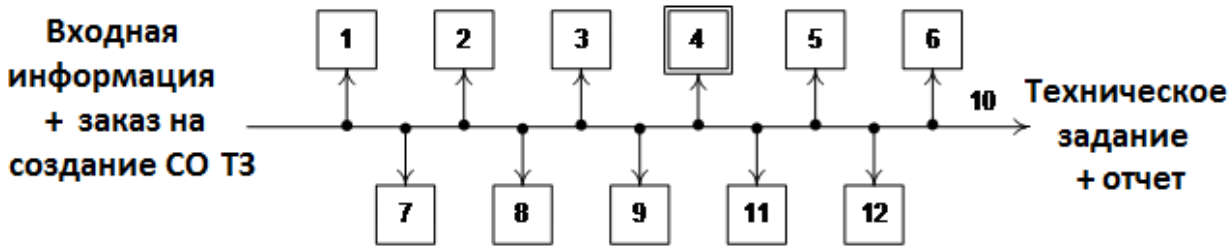


Рис.1. Системно-структурная модель АГ СО ТЗ

(Системы: 1 – электронизации входной информации, 2 - работы с заказчиком и разработчиком, 3 - генератора моделей, 4 - репозитария данных и знаний ТЗ, 5 - составления аудит-отчета о предпроектном обследовании шаблона ТЗ, 6 - интегратора ТЗ: форматирования, верстки и утверждения текста, 7 – управления генератором, 8 - визуализации информации, 9 - печати, 11 - настройки на специфику, 12 - системно-интеграционной поддержки, 10 – интерфейсов).

В качестве прототипа системы 4 нами выбран библиографический менеджер Zotero [5] - программа с открытым исходным кодом, дополнение (плагин) к браузеру Firefox. Zotero позволяет собирать цитаты с указанием источников в виде перечня ссылок на PDF-

документы, веб-страницы, видео, другие файлы и прочие интересные материалы в Сети, таким образом создавать персональные библиографии по темам.

Функции прототипа: интеграция с MS Office и openoffice.org, 40 языков (включая русский, английский, немецкий, французский), синхронизация с сервером, перенос на другой компьютер, сохранение библиотеки на переносных носителях, создание отчётов из выбранного источника или группы источников, а также - иметь хронологическую шкалу источников, сохранять библиографическую информацию on-line с сайтов google scholar, google books, amazon.com, sciencedirect, springerlink и др. источников (веб-страниц, .pdf, видео, звуковых и др.).

Программа разработана и поддерживается *Center for History and New Media* в George Mason University, выпускается под GPL, как и FireFox, работает на платформах Windows, Linux и Mac OS X.

Zotero выбран прототипом из-за схожести решаемых задач, а именно загрузки, хранения и сортировки различной информации и возможности интеграции с MS Office.

Недостатки прототипа: нет автоматизированной сортировки и ранжирования (осуществляется «вручную» пользователем); отсутствует возможность восстановления исходных документов; не заложен способ хранения и извлечения знаний (в т.ч. правила обработки информации); не отображается механизм оценки качества исходных данных, промежуточных и готовых продуктов (не предусмотрено хранение критериев оценки).

Развитие структуры системы репозитария.

На основании прототипа [5] и его критики нами предложена следующая системно-структурная модель системы репозитария (рис.2).

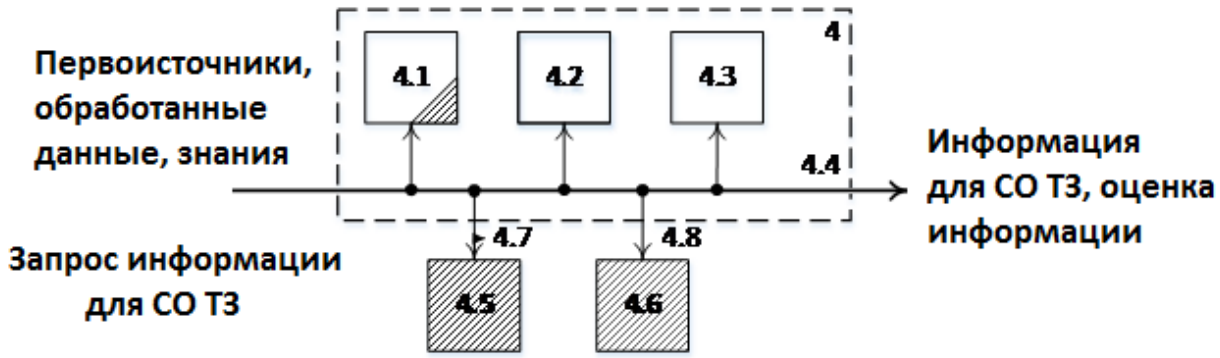


Рис. 2. Системно-структурная модель системы 4 «Репозитарий» по прототипу и предложенному решению – фон, уголки (подсистемы: 4.1 - сортировщика данных, 4.2 - репозитария исходных данных, 4.3 - репозитария обработанных данных, 4.5 - репозитария знаний, 4.6 – оценки качества, 4.4, 4.7, 4.8 – интерфейсов).

Предложенное решение состоит во введении двух новых подсистем (4.5, 4.6), а также – в модернизации подсистемы 4.1.

Сначала информация в репозитарии исходных данных (подсистема 4.2) собирается с помощью системы электронизации входной информации [6] в разделе оригинальной информации, затем ранжируется по очередности ее использования, чем пополняется раздел «Ранжируемая информация». Для обеспечения безопасности хранения документов в разделе «Резервные копии» осуществляется накопление различных вариантов (версий) одного и того же файла исходной информации после внесения каждого изменения, что позволяет восстановить его после каких-либо ошибочных действий. В репозитарии содержатся промежуточные продукты каждой системы АГ СО ТЗ (модели и правила их построения, шаблоны ТЗ, фрагменты готового ТЗ и др.), а также в этот раздел включена и сама база данных АГ СО ТЗ. Предусмотрено хранение готовых продуктов: СО ТЗ, отчета по его написанию и опыту работы (например, переписка заказчика и разработчика с замечаниями по ходу составления ТЗ). В репозитарии заложена подсистема 4.6 оценки (критерии оценки работы каждой подсистемы и правила их применения) пользователей, степени готовности промежуточных и окончательных продуктов. Кроме того, в процессе создания СО ТЗ идет накопление знаний о качестве проектируемой медицинской информационной системы в виде:

$$T = a1*k1+a2*k2+…+an*kn ,$$

где Т – полнота описания проектируемой информационной системы, k – коэффициенты – характеристики ТЗ, которые должны быть учтены в процессе написания, а – веса коэффициентов.

При создании ТЗ возможно добавление новых характеристик или удаление старых, то есть совершенствование формулы СО ТЗ.

Алгоритм функционирования системы репозитария

На рис.3-6 представлены алгоритмы функционирования системы репозитария данных и знаний АГ СО ТЗ и ее подсистем. На рис.3 представлен основной алгоритм функционирования. На вход системы репозитария поступают ранжированные данные, промежуточные и конечные продукты систем АГ СО ТЗ. Имеется 2 цикла: по времени и по документам. Первоначально данные ранжируются подсистемой 4.1 по степени обработки. После окончания ранжирования начинают работать параллельно подсистемы 4.2 и 4.3. Затем пользователю предлагается оценить качество работы системы и, если качество удовлетворяет, система репозитария заканчивает свою работу. Если оценка

«неудовлетворительно», то включается в работу подсистема 4.5 и заново начинаются циклы по времени и документам до тех пор, пока качество не будет на должном уровне.

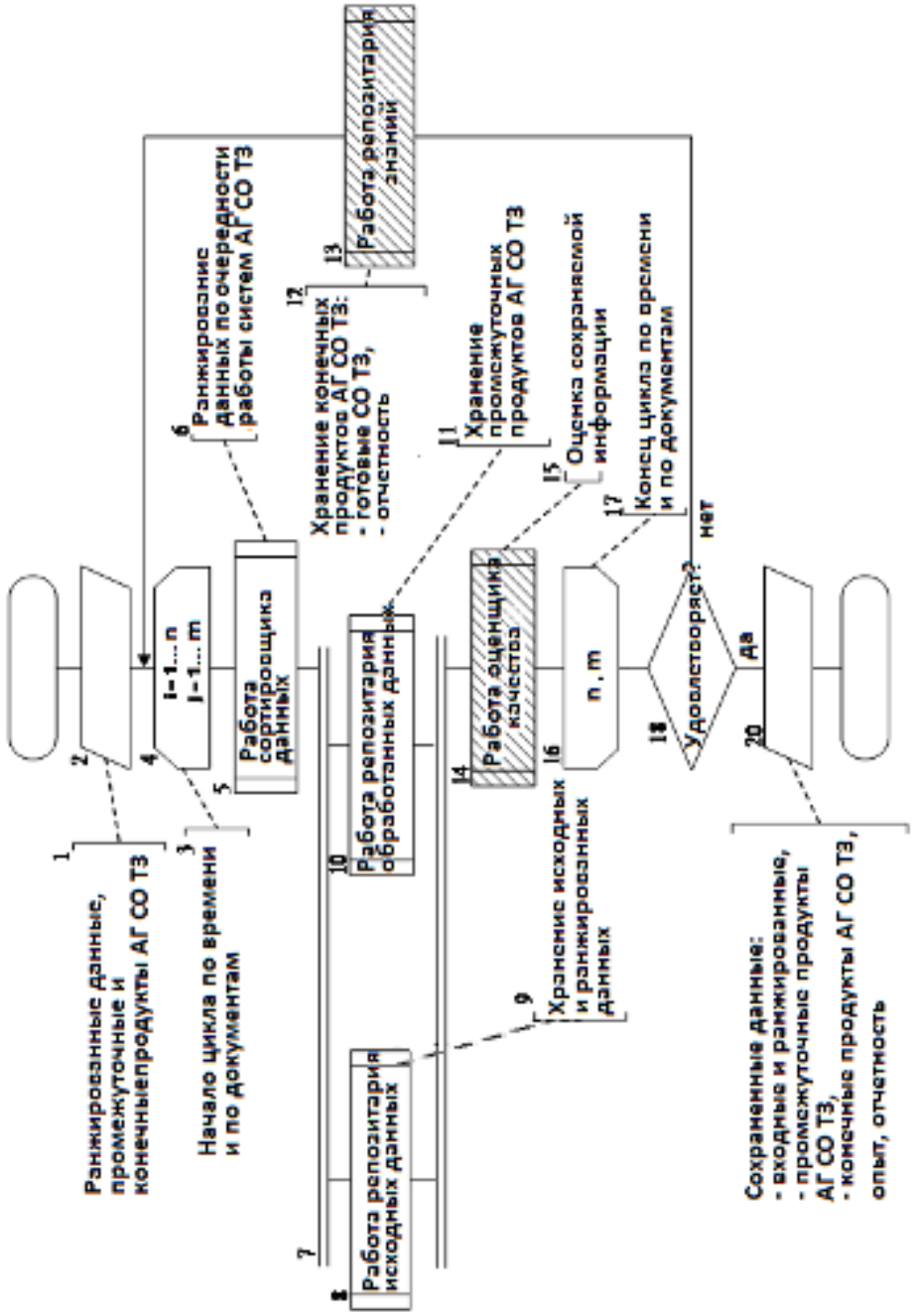


Рис. 3. Алгоритм функционирования системы репозитария данных и знаний

Работа сортировщика данных (рис.4) заключается в ранжировании входных данных по степени обработки системами АГ СО ТЗ и создании в резервных копий документов, которые претерпели изменения.

Работа подсистемы 4.5 формализована в виде алгоритма на рис.5. По списку проблем выполняется локальный поиск решения, если решение проблемы не найдено, то поиск продолжается в интернете. При отсутствии решения – обращение к специалисту в предметной области проблемы.

Работа подсистемы 4.6 состоит в том, что информация различного рода, попадающая в систему репозитария, оценивается по критериям пользователя и отсеивается при неудовлетворительной оценке.

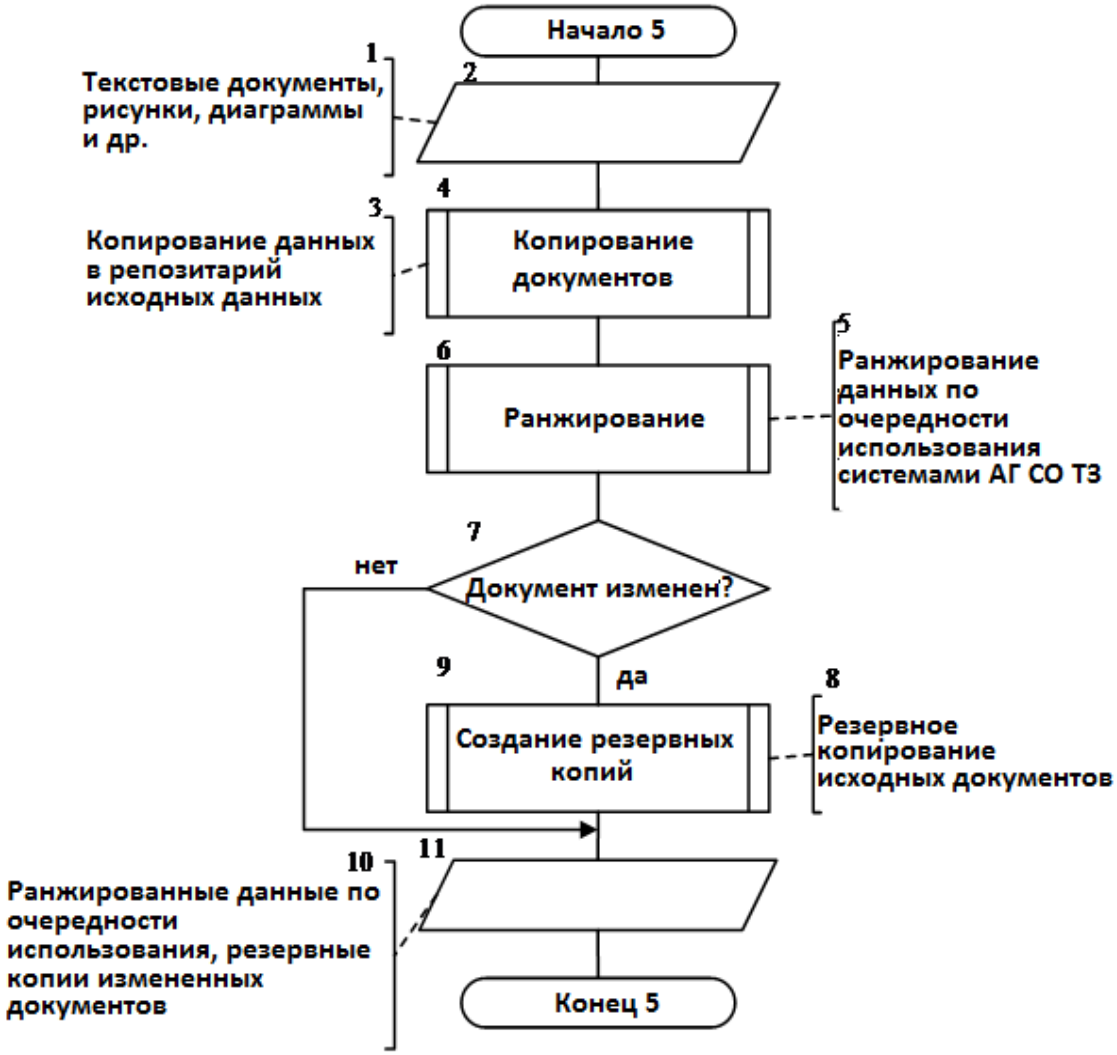


Рис. 4. Алгоритм функционирования блока 5 (подсистема 4.1) сортировщика данных

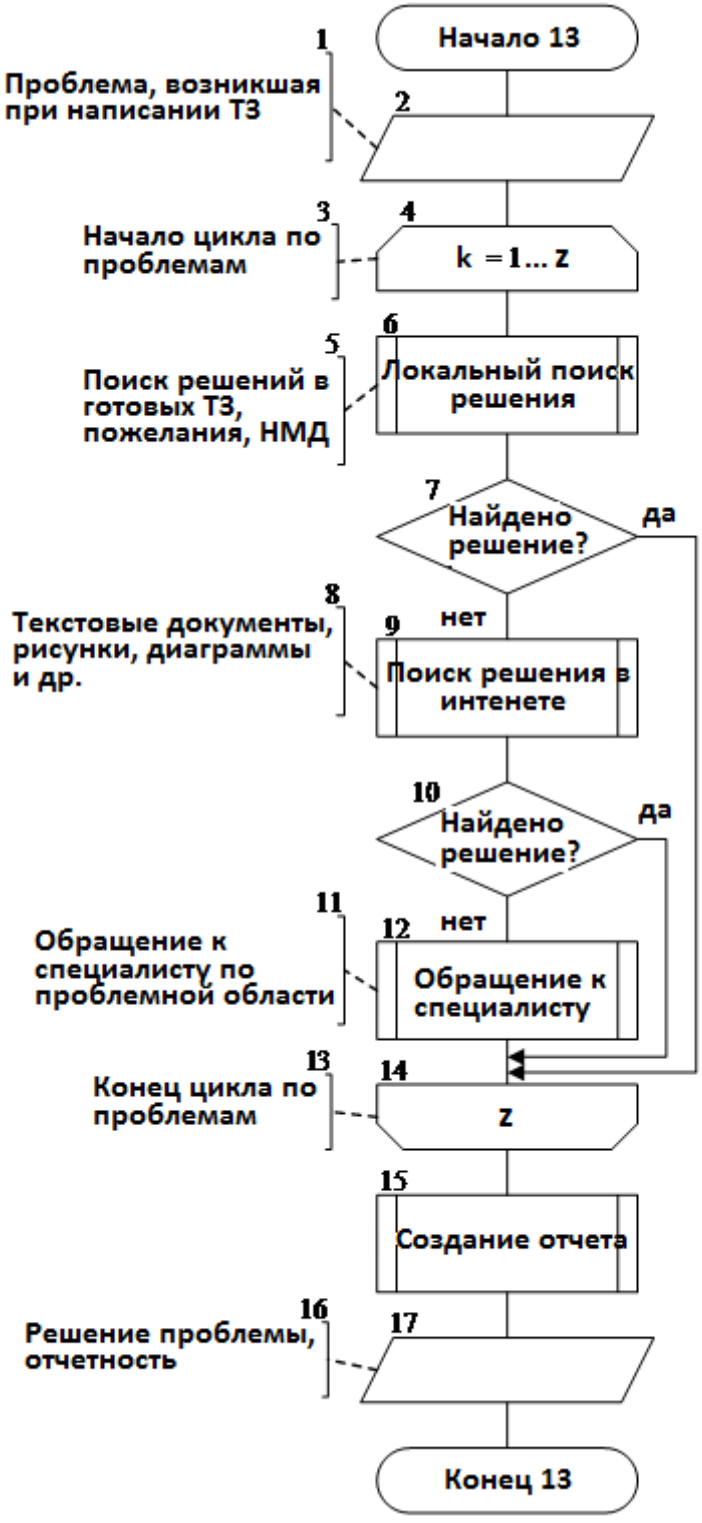


Рис. 5. Алгоритм функционирования блока 13 (подсистема 4.5) репозитария знаний

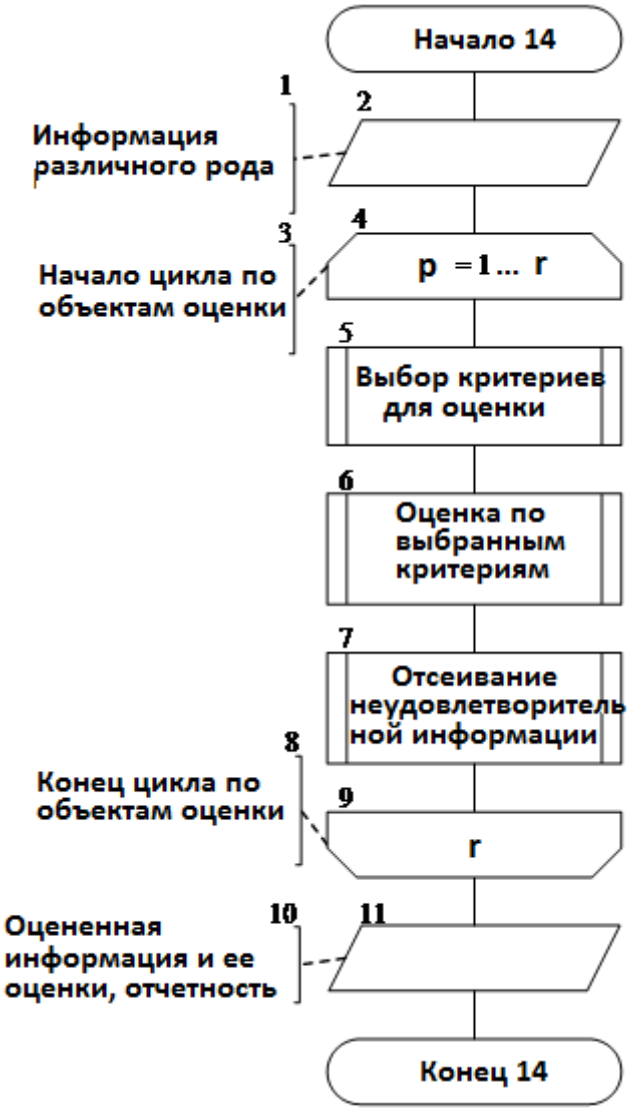


Рис. 6. Алгоритм функционирования блока 14 (подсистема 4.6) оценщика качества

Схема взаимодействия подсистем системы репозитария

Система репозитария представляет собой иерархию каталогов. В корневом каталоге располагаются база данных пользователей и сами проекты СО ТЗ. В базе данных пользователей хранятся списки всех проектов и документов, задействованных в соответствующем проекте, а также - учетные записи пользователей с различными уровнями доступа. Каждый каталог проекта содержит 3 подкаталога: репозитарий исходных данных, репозитарий обработанных данных и репозитарий знаний.

Данные, загруженные в АГ СО ТЗ при помощи СЭВИ [6], хранятся в репозитарии исходных данных, который, в свою очередь, состоит из репозитария исходных данных и репозитария ранжированных данных(здесь хранятся необработанные и распределенные по системам АГ СО ТЗ данные соответственно), а также - из репозитария резервных копий.

Последний хранит версии документов до внесения в них каких-либо изменений. В базе данных ведется учет всей информации, присутствующей в системе репозитария. После обработки данные хранятся в репозитарии 4.3 обработанных данных – это либо промежуточные результаты, либо завершенные проекты. В процессе создания СО ТЗ накапливаются различные знания, опыт - все это хранится в базе знаний в виде математических формул, решающих правил и т.п. с пояснениями и ссылками в интернет на различные информационные ресурсы. Схема приведена на рис.7.

Программная реализация

Программа реализована в среде MS Visual Studio 2013 на языке C#.

Реализованы модули: сортировщика 4.1 данных, репозитария 4.2 исходных данных, репозитария 4.3 обработанных данных, репозитария 4.5 знаний, оценщика 4.6 качества.



Рис. 7. Схема взаимодействия подсистем системы репозитария *АГ СО ТЗ*
(4.2, 4.3, 4.5 – см. рис.2; x, 0, y – данные; f – правила отбора данных для вставки в ТЗ)

Пример экранной формы представлен на рис.8. В левой части организовано иерархическое представление структуры репозитария, в правой – содержимое выделенной папки. Имеется возможность создания, копирования, удаления папок и файлов. Настройка резервного копирования и управление БД осуществляется на вкладках «Резервное копирование» и «БД» соответственно.

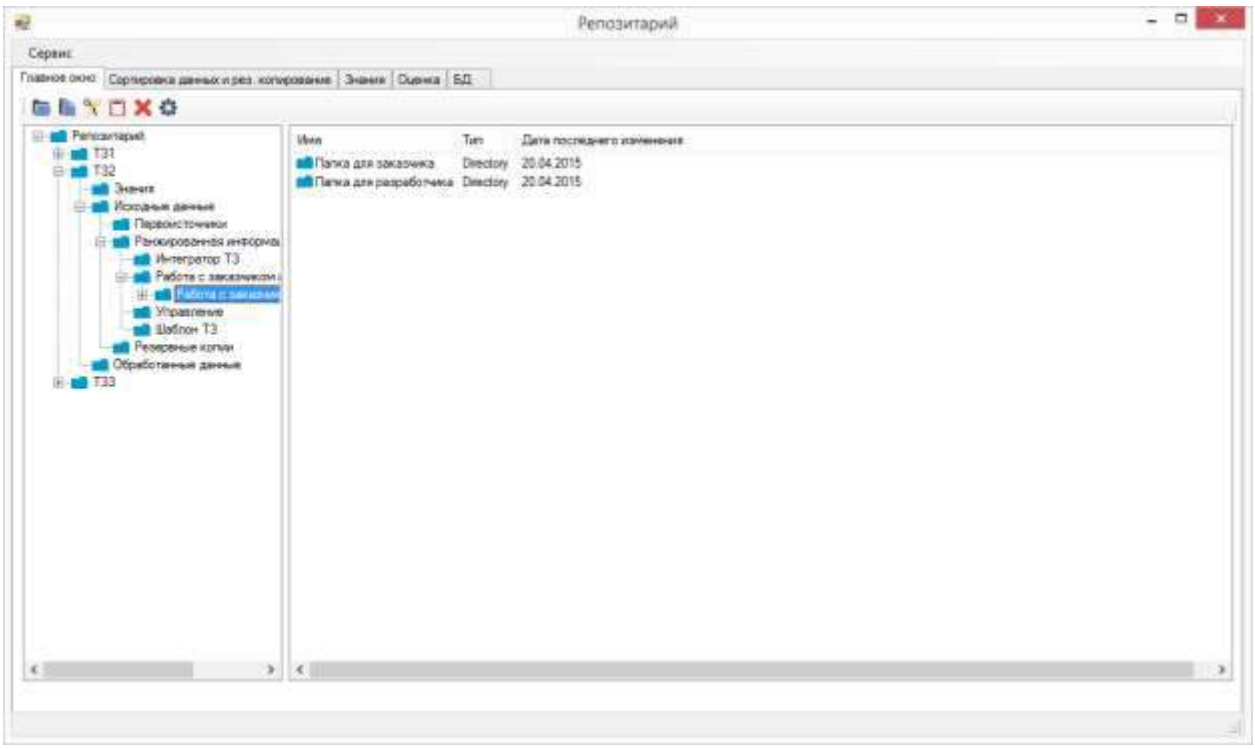


Рис. 8. Пример экранной формы

Результаты и выводы:

– поставлена задача развития системы репозитария;

– найден прототип, дана его критика;

– разработаны и приведены структура и алгоритмы функционирования системы репозитария и ее подсистем;

– проработано устройство хранилища входной информации – репозитария исходных данных, репозитария обработанных данных и репозитария знаний, выбран принцип взаимодействия разделов репозитария между собой.

– на основании разработанных моделей создана исследовательская версия программного обеспечения системы репозитария данных и знаний АГ СО ТЗ.

Детально разработанная система репозитария может быть использована как часть АГ СО ТЗ с целью хранения и ранжирования данных, необходимых для создания технического задания, а также - в качестве самостоятельного инструмента для накопления информации, например, при составлении отчетов о каком-либо виде деятельности или при написании научных трудов, методических пособий, диссертаций и пр.

Список литературы

1. Глушков В.М. Основы безбумажной информатики / В. М. Глушков - М.: Наука. Главная редакция физико-математической литературы, 1982. 552 с. Гольдштейн С.Л. Системная

интеграция бизнеса, интеллекта, компьютера / С.Л. Гольдштейн, - Екатеринбург: ИД «ПироговаЪ», 2006, - 392 с.

2. Гольдштейн С.Л. О функционировании автоматизированного генератора системно обоснованного технического задания на медицинскую информационную систему / С.Л. Гольдштейн, Е.М. Грицюк, Д.А. Леонов // Электронный научный журнал "Системная интеграция в здравоохранении". 2012. № 1. С.20-32. URL: http://www.sys-int.ru/files/2012.1/146/sys_int_130_1_15_2012.pdf

3. Гольдштейн С.Л. О Структуре автоматизированного генератора системно-обоснованного технического задания на информационные системы / С.Л. Гольдштейн, Е.М. Грицюк // Системы управления и информационные технологии, №1, 2012, с.70-74.

4. Zotero / [Официальный сайт zotero]. URL: https://www.zotero.org/.

5. Гольдштейн С.Л. Развитие системы электронизации входной информации / С.Л. Гольдштейн, Е.М. Грицюк, Д.А. Леонов // Электронный научный журнал "Системная интеграция в здравоохранении". 2012. № 2. С.5-18. URL: http://www.sys-int.ru/files/2012.2/152/sys_int_136_2_12_2012.pdf

Грицюк Елена Михайловна, - к.м.н., врач-эпидемиолог ГБУЗ СО ДКБВЛ НПЦ «Бонум», 620149, Екатеринбург, ул. Бардина, 9а, тел: (343)240-42-68 bonum@bonum.info