«ОБЛАЧНЫЕ» ВЫЧИСЛЕНИЯ В РЕШЕНИИ ЗАДАЧ КЛАССИФИКАЦИИ НА ПРИМЕРЕ МУЖСКОГО БЕСПЛОДИЯ: НОВЫЙ ИСТОЧНИК МЕДИЦИНСКИХ ДАННЫХ

ШАМИЛЬ МАНСУРОВИЧ ГИМАДЕЕВ, зав. терапевтическим отделением Сармановской ЦРБ, Сарманово, РТ **АЙРАТ ИЛЬДУСОВИЧ ЛАТЫПОВ**, ведущий программист отдела информатизации ГБОУ ВПО «Казанский государственный медицинский университет Минздравсоцразвития РФ» **ОЛЬГА РАФАИЛОВНА РАДЧЕНКО**, доцент кафедры гигиены, медицины труда с курсом медицинской экологии ФПК и ППС ГБОУ ВПО «Казанский государственный медицинский университет Минздравсоцразвития РФ» **СТАНИСЛАВ ВИТАЛЬЕВИЧ РАДЧЕНКО**, директор ООО «НПФ «Алтын Кэз», Казань, e-mail: info@altinfo.ru

Реферат. В статье рассматриваются методы и предварительные результаты применения гибридной архитектуры прикладного программного обеспечения в здравоохранении, основанной на интеграции систем масштаба ЛПУ с «облачными» сервисами Google Docs. Открытый API сервисов Google Docs позволил автоматически обработать значительные дополнительные объемы анамнестических данных, поступающих от пользователей Интернета, программными средствами собственной разработки. Увеличение объемов анамнестических данных и возможность процедурного доступа к ним существенны для решения задачи классификации идиопатического мужского бесплодия. Мы предполагаем существование адаптивной формы мужского бесплодия, поддающегося неспецифическому лечению и мерам реабилитации. Выделение адаптивной формы мужского бесплодия эффективно с экономической и медицинской точек зрения, но требует обработки значительного объема данных. Практически такие данные могут быть получены из ранее не описанного источника, широкое использование которого возможно только благодаря открытому API сервисов Google Docs.

Ключевые слова: мужское бесплодие, «облачные» вычисления, источники медицинских данных, Google Docs, анамнез, автоматизация сбора данных, гибридная архитектура.

«CLOUD» COMPUTING IN SOLUTION OF CLASSIFICATION TASKS ON THE EXAMPLE OF MALE INFERTILITY: NEW SOURCE OF MEDICAL DATA

SH.M. GYMADEYEV, A.I. LATYPOV, O.R. RADCHENKO, S.V. RADCHENKO

Abstract. The article consider some methods and preliminary results of hybrid software architecture application in public health, based on HIS and Google Docs «cloud» computing services integration. Google Docs open API allowed the automatic processing of Internet users' extended data, which concerns variable anamnestic details, by self-provided software. Significant data extension in couple with procedural accessibility are vital to classification tasks in idiopathic males infertility area. We suggest the adaptive origin of some male infertility forms, potentially curable by non-specific treatment and rehabilitation measures. The allocation of adaptive infertility forms is medically and economically effective, but demands the anamnestic data processing in a great scale. Virtually, we cannot receive such data amounts beside data source, which was not described and classified earlier. Its wide use is possible due to Google Docs services open API only.

Key words: male infertility, «cloud» computing, medical data sources, Google Docs, anamnesis, automation of data collection, hybrid software architecture.

Существующая в стране демографическая проблема усугубляется низкой рождаемостью, связанной в том числе со значительным распространением бесплодия в браке. Мужское бесплодие в общей структуре бесплодия составляет примерно половину случаев [9]. Являясь в целом управляемым фактором, мужское бесплодие, тем не менее, содержит плохо управляемый компонент, снижающий эффективность как медицинских, так и социально-гигиенических мероприятий, направленных на повышение рождаемости. Речь идет о случаях идиопатического бесплодия, т. е. неспособности мужчины зачать ребенка в ситуации, когда известные для этого патофизиологические причины отсутствуют или не выявлены [10].

Мужское бесплодие, классифицируемое как идиопатическое, в действительности представляет собой неоднородную сущность. В определенных условиях доля идиопатического бесплодия может уменьшаться, как это в свое время произошло при установлении роли микроделеций в Y-хромосомах в уничтожении оплодотворяющей способности спермы [12]. С гигиенической и социально-гигиенической точек зрения при наличии предположения о случае идиопатического бесплодия важным оказывается дополнительно классифицировать случай как курабельный либо инкурабельный. В данной задаче классификации ошибка первого рода приведет к существенным безрезультативным затратам различных ресурсов как со стороны системы здравоохранения, так и со стороны пациента и к снижению качества жизни последнего. Ошибка второго рода будет означать предотвратимую потерю репродуктивного элемента для популяции. Таким образом, решение задачи дополнительной классификации идиопатического мужского бесплодия (ИМБ) целесообразно строить по стратегии максимин.

На неоднородность и потенциальную курабельность в группе ИМБ указывают случаи спонтанной реконвалесценции, которые могут наблюдаться в отсутствие лечения через несколько месяцев или лет после обращения к врачу [8]. Ранее в выборках пациентов с ИМБ нами была выявлена подгруппа мужчин, для которых объединяющим признаком оказывается действие антропотехногенного пресса в виде совокупности неблагоприятных факторов, каждый из которых по своей интенсивности не достигает нозогенного порога [4]. Мы предположили, что бесплодие у таких пациентов носит

адаптивный характер и поддается мерам лечебнопрофилактического воздействия, в ряде случаев — нестандартным. Детальное обсуждение данного вопроса выходит за рамки настоящей статьи.

Возможность дифференциации подвидов ИМБ почти исключительно обусловлена возможностью сбора и анализа дополнительных объемов анамнестических данных, получаемых от пациента. Врач, решающий задачу дифференциальной диагностики ИМБ, вынужден при подозрении на его адаптивный характер полагаться исключительно на ретроспективный материал, поскольку обращение пациента к специалисту всегда происходит уже после того, как некоторая предполагаемая совокупность факторов оказала действие на организм пациента, но на момент обращения прямой оценке не подлежит. Это обстоятельство придает источникам анамнестических данных особую важность, значительно превосходящую важность тех же источников в других ситуациях.

Ранее для количественной оценки важности источников медицинских данных (ИМД) нами был произведен анализ 289 историй болезни с изучением частоты использования различных видов медицинской информации, необходимой консультантам для формирования заключения, а также проведено анкетирование 690 сотрудников в 19 ЛПУ РТ и РФ и стран СНГ. Значение информативности определялось как произведение вероятностей обращения к ИМБ и получения информации из него. В иерархии полученных приоритетов по совокупности общетерапевтических случаев данные

анамнеза оказались на предпоследнем месте [2]. В случаях обращений мужчин к врачу по поводу бесплодия в браке данные анамнеза устойчиво занимают по важности первое-второе места. Под анамнестическими данными в ранее проведенных исследованиях всегда подразумевались данные, получаемые специалистом во время непосредственного общения с пациентом на приеме. При этом возможности сбора анамнеза на приеме у врача объективно ограничены как временем приема, так и организационными причинами. Кроме того, на достоверность полученных данных влияют небольшой объем выборки, обусловленный малым количеством пациентов, согласившихся заполнить бумажную анкету, и недостаточная мотивированность пациента, торопящегося заполнить эту анкету в условиях амбулаторного приема у конкретного специалиста. Как правило, достаточное для обработки количество анкет и амбулаторных карт (200—300 шт.) накапливается у одного врача за период в 2—4 года. В указанные сроки состояние здоровья у части пациентов уже успевает значительно измениться в ту или другую сторону, но более существенным оказывается возможное прекращение действия изучаемых факторов давления среды на организм или появление новых.

Проведенное анкетирование персонала позволило сформировать список основных источников данных (ИД), используемых в работе медицинскими и немедицинскими специалистами ЛПУ (табл. 1).

В сформированном списке присутствуют такие элементы, как «источники Интернета» и «данные

Таблица 1

Классификация источников данных в ЛПУ по видам, методам и источникам получения информации
в соответствии с результатами анкетирования персонала

Вид информации	Метод	Источник информации	Источник данных	
Растровые диагностические изображения	Инструментальное исследование визуализации	Оборудование визуали- зации	визуали- Пациент	
Векторные диагностические изображения	Функциональное исследование	Оборудование функцио- нальной диагностики		
Результаты лабораторных исследований	Лабораторное исследование	Лабораторное оборудо- вание		
Результаты осмотра	Физикальное исследование	Медицинский персонал		
Анамнез	Расспрос		Физическое лицо, ассоциированное с пациентом	
Записи о назначениях	Назначения		Пациент	
Записи о предшествующих назначениях и вмешательствах	Выполнение назначений			
Заключение консультанта	Консультация		Очный консультант	
	Дистанционная консультация	Коммуникационное оборудование и ПО	Удаленный консультант	
Содержание внешних медицинских документов	Оформление направления, выписки, справки	Персонал стороннего ЛПУ	Стороннее ЛПУ	
Содержание внешних немедицинских документов	Оформление финансового до- кумента	Персонал стороннего учреждения	Стороннее немедицинское учреждение	
Содержание нормативных документов	Получение и изучение нормативных документов	Персонал вышестоящего учреждения	Вышестоящее учреждение	
Содержание внутренних немедицинских документов	Регистрация платежа	Контрольно-кассовый аппарат	Немедицинский персонал	
Содержание внешних немедицинских документов	Оформление счетов-реестров	Специализированное ПО		
Все виды информации о па- циенте и действиях над ним	Получение и изучение внутренних медицинских документов	Медицинский персонал	Внутренние медицинские документы	
Литературная публикация	Поиск и изучение литературы	Книги, периодические из- дания, сборники и БД	Библиотека*	

^{*}Термин «библиотека» для обозначения источника данных использован в наиболее широком смысле.

литературы». Примеры, приводимые респондентами, позволяют понять, что в Интернете почти всегда рассматриваются некоторые документы, реже какие-то данные, размещенные на заранее не определенном множестве сайтов. Ко всем документам и данным предполагается ручной непроцедурный доступ. Важность, эмпирически определенная для источников Интернета, семантически связана именно с источниками, интерпретируемыми указанным образом, и не распространяется на все способы получения информации с помощью интернет-технологий. Совокупность пользователей Интернета, оказывающихся авторами контента, обрабатываемого непроцедурно, в соответствующей информационной модели выступает в качестве неуправляемого параметра. Отметим, что данные, получаемые из источников, традиционно понимаемых как «источники Интернета», пригодны для извлечения информации только на этапе обработки человеком-оператором. Это обстоятельство в зарубежных источниках рядом авторов именуется «низкая доступность для обработки» и описывается как свойство, изначально присущее Web [14]. Однако развитие ИТ и web-технологий в частности постепенно привело к возможности использования, а следовательно, и необходимости выделения в качестве самостоятельной таксономической единицы такого ранее неклассифицированного ИД, который, обладая внешними признаками традиционных источников Интернета, не совпадает с ними семантически. Причины соответствующего хода развития ИТ и его последствия для деятельности системных аналитиков и разработчиков в медицине и здравоохранении ранее обсуждались нами в работе [1].

Вторичная деперсонализация активности пользователей ИТ, произошедшая в ходе эволюции Web, и рост пропускной способности каналов при одновременном удешевлении передачи каждой единицы данных поставили в особое положение технологии распределенной обработки данных, в которой компьютерные ресурсы и мощности предоставляются пользователю как интернет-сервис. Такие технологии в настоящее время принято называть «облачными» вычислениями («cloud» computing). Интересно, что со структурнофункциональной точки зрения «облачные» вычисления на персональных компьютерах не являются чем-то новым — достаточно вспомнить историю офисного пакета Lotus eSuite, выпуск которого в 1997 г. сопровождался большой шумихой. Пакет был пригоден для использования на компьютерах, на которых установлен лишь web-браузер, бизнес-логика отрабатывалась на сервере, а функциональность на клиентской стороне обеспечивалась загружаемыми по мере необходимости апплетами и сервлетами Java [3]. Однако продукт, значительно опередивший свое время, не имел какоголибо коммерческого успеха по причине психологической неподготовленности пользователей, связанной с недостаточной распространенностью самой сети Интернет и связанных с ней технологий и метафор. Предоставление функциональности программного обеспечения (ПО) через Интернет в виде услуги, в том числе услуги аренды, с одной стороны, вернуло пользователей к знакомой по мэйнфреймам ситуации централизованного выделения вычислительных ресурсов. а с другой стороны, породило термин SaaS (software as a service), подчеркивающий, что, в отличие от прежней системы «терминал-мейнфрейм», предоставляемой

услугой оказывается доступ не к аппаратному, а программному обеспечению. Характерно, что в модели SaaS в качестве сервера используется своего рода виртуальный мэйнфрейм, выглядящий для пользователя как единая сущность, но в действительности представляющий собой динамически распределенный комплекс физических серверов.

На сегодняшний день в медицине и здравоохранении не произошло достаточно глубокого осмысления как роли, так и ограничений SaaS ни в задачах обработки данных, ни в задачах системной интеграции. Свидетельством этому оказываются дискуссии, которые ведутся в сообществе разработчиков и организаторов здравоохранения по поводу адекватности «облачной» парадигмы для построения и модернизации АСУ в здравоохранении [6].

Ранее нами была предложена схема организации гетерогенных информационных ресурсов для решения задачи обеспечения эффективности распределенной математической модели хронической почечной недостаточности [5]. В предложенной схеме уже предполагалась возможность автоматизированной верификации параметров модели по мере пополнения прецедентной базы, происходящей из многих ИД. Однако говорить об использовании принципиально нового ИД при реализации этой схемы не приходится. Источником данных для прецедентной базы продолжают оставаться пользователи локального ПО, а Интернет выполняет в схеме роль среды передачи данных. Ограничениями, свойственными описанному способу web-интеграции ИМД, следует считать трудоемкость невозможного в автоматическом режиме подключения каждого нового ИД, включающего многоэтапную процедуру регистрации каждого пользователя, необходимость развертывания проприетарного ПО *in situ*, даже в случае значительной географической удаленности рабочих мест, и проблему формирования и поддержания мотивации предметных специалистов, призванных участвовать в обеспечении системы дополнительными данными.

Вновь задействованная парадигма ориентирована на существование широкой, априорно мотивированной аудитории в лице озабоченных своим состоянием пользователей Интернета, возможность использования свободного ПО, возможность автоматизированного процедурного доступа к пополняемым массивам данных, интеграцию с унаследованным и (или) не подлежащим замене проприетарным и закрытым ПО без дополнительных угроз для персональных данных, содержащих сведения о состоянии здоровья. На день написания настоящей статьи возможность реализации такой парадигмы целиком обязана существованию единственной технологии, а именно — Google Docs. Значение этой технологии в здравоохранении на постсоветском пространстве в настоящее время остается недооцененным, вероятно, вследствие ее относительной молодости. Во всяком случае, в отечественной литературе нам не удалось найти работ, указывающих на прецеденты системного применения Google Docs в решении лечебно-диагностических проблем, либо задач управления здравоохранением. Зарубежная литература в этом отношении дает основание говорить о широком использовании сервиса для нужд медицины [3, 7, 11, 15]. Относясь к сфере классических «облачных» вычислений, технология Google Docs первоначально позиционировалась как основанная на Web и независимая от аппаратной и программной платформы

альтернатива специализированным коммерческим офисным приложениям, ориентированная на совместную работу. Появление и распространение такого типа ПО было предсказано нами в 2006 г. в статье, посвященной влиянию Web на эволюцию ИТ-отрасли [1]. В дальнейшем развитие технологии привело вендора к беспрецедентному решению, которое обеспечило качественный скачок возможностей системы. Речь идет об открытии АРІ (интерфейса программирования приложений). Открытый API «облачного» ПО позволяет создавать и эксплуатировать распределенные приложения с принципиально новой архитектурой. В контексте решения проблемы мужского бесплодия архитектура распределенной системы, использующей Google Docs API, позволяет на порядок ускорить наполнение прецедентной базы в части ценных анамнестических данных, а также возможности их анализа. В более широком контексте обеспечивается появление и использование ранее неклассифицированного ИМД с качественно новыми свойствами.

В применяемой нами архитектуре распределенной системы, предназначенной для массового сбора анамнеза и обработки анамнестических данных пациентов с предположительным диагнозом ИМБ, все процедуры взаимодействия собственного проприетарного компонента со свободным компонентом, основанным на сервисе Google Docs, полностью скрыты от пользователя. Для врача и исследователя развертывание функциональной части приложения в Web выглядит как тривиальное создание нового документа и не требует каких-либо специальных навыков (рис. 1).

Вид и структура опросной формы также определяются несложными действиями, сводящимися преимущественно к работе «мышью» (рис. 2). Публикация,

ограниченная публикация, либо закрытие формы осуществляется изменением статуса «Совместный доступ» для выбранных документов из списка ранее созданных (рис. 3).

В свою очередь, пользователь Интернета осуществляет привычную для него навигацию на странице сайта с последующим заполнением web-формы (рис. 4). Ссылка на страницу может быть получена пользователем по одному из традиционных каналов либо найдена самостоятельно в результате тематического серфинга.

Оповещение о появлении новых данных, подлежащих ручной обработке, производится в соответствии с пользовательскими настройками, доступными врачу (рис. 5).

Настройки включают опции оповещения по признаку появления новых данных, внесения/изменения данных определенной категории, факта отправки заполненной формы на сервер. Также доступен выбор между немедленным сообщением по факту события и получением сводки за прошедшие сутки (рис. 6).

Автоматическая передача данных из заполненных форм в базу данных задействуется при условии выставления соответствующей разрешающей отметки (рис. 7).

Благодаря открытому API сервис Google Docs предоставляет возможность автоматической обработки данных по запросам со стороны внешнего ПО. Архитектура системы, основанной на «облачном» взаимодействии проприетарного и открытого ПО и пользователей различных категорий, позволяющей говорить об ИМД нового вида, представлена на рис. 8.

Отличительные признаки нового ИМД приведены в *табл.* 2.

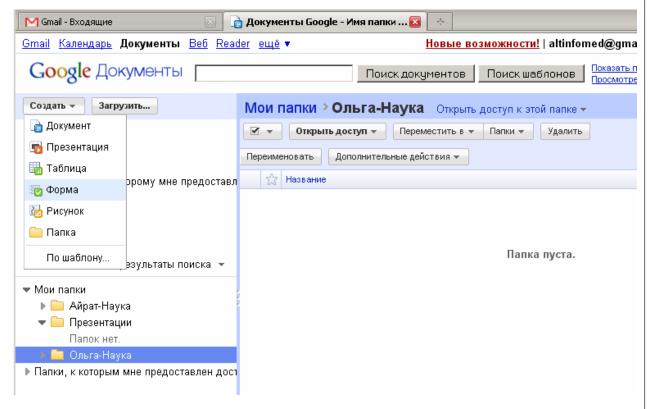


Рис. 1. Ручное создание нового объекта Google Docs на стороне врача

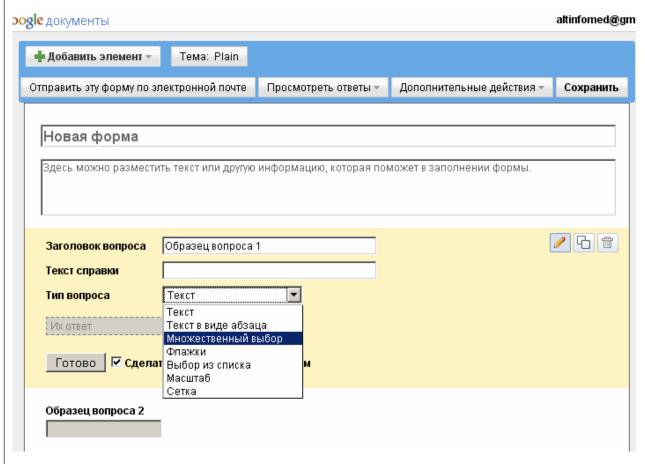


Рис. 2. Стандартные возможности по созданию опросных форм охватывают все известные структурные элементы автоматизированных тестов

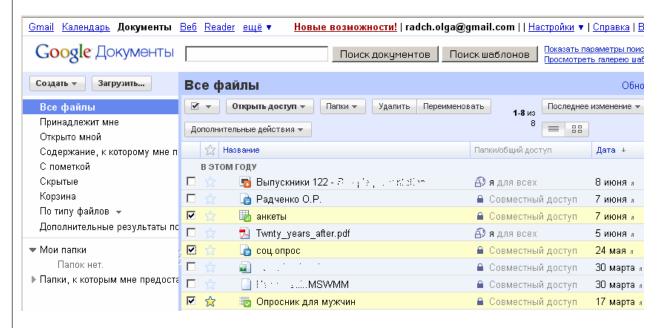


Рис. 3. Для ранее созданных документов Google Docs поддерживается групповое изменение прав доступа

	ее 5 лет назад			
курил раньше, бросил мен	нее 5 лет назад			
С курю				
С Другое:				
5 B ()		,		
Если Вы курите (курили), ук при курении более 5 сигарет в		ния (в года	ax)	
12	-,			
			_	_
Из нижеперечисленных про пищу ежедневно, хотя бы 1				
пишу смедневно, хогл овг г	раз в педелю, т	1 раз в	не чаще 1	совсем не
	ежедневно	т раз в неделю		
молоко коровье	О	•	0	0
сыр	0	_		
	~	•	0	0
творог жирный	•	0	0	0
творог жирный	e	0	0	c
творог жирный яйцо куриное желток	0	o •	0	0
творог жирный яйцо куриное желток масло подсолнечное	© O	o •	0	o o
творог жирный яйцо куриное желток масло подсолнечное соевые продукты	e 0 0	o o	0 0	0 0 0

Рис. 4. Формализованный ввод при сборе анамнеза — вид на стороне пользователя-пациента

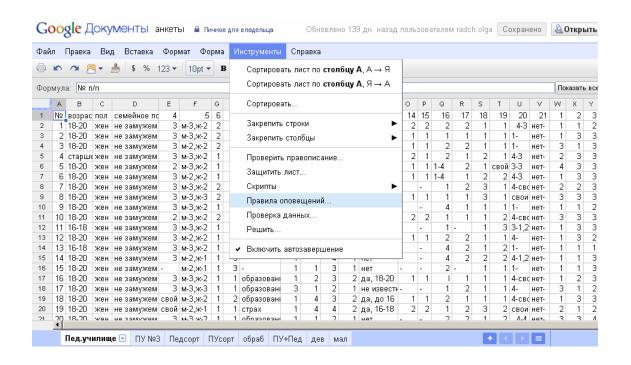


Рис. 5. Настройка правил оповещений доступна из главного меню документа на стороне администратора

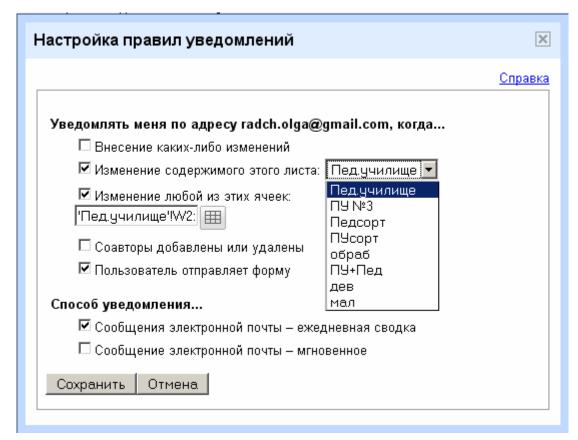


Рис. 6. Уведомления о наступлении системных событий в Google Docs допускают практически произвольные сигнальные признаки

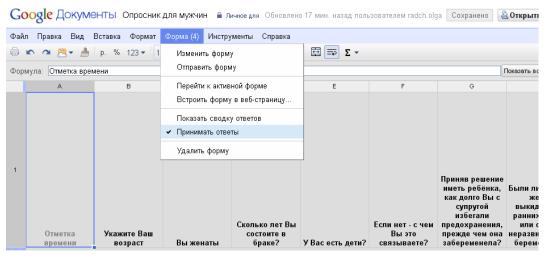


Рис. 7. Выставление разрешающей отметки для приема ответов приводит к автоматическому заполнению БД, доступной через Google API

Таблица 2 Признаки нового вида источника медицинских данных, существенные в решении задач системной интеграции

Показатель	Традиционный интернет- пользователь	Традиционный источник Интернета	Традиционный пациент	Традиционный удаленный пациент	Структурированный «облачный» пациент
Процедурность	Нет	Нет	Нет	Нет	Да
Автоматизация	Нет	Нет	Да	Да	Да
Анонимность	Да	Да	Нет	Нет	Да
Экстерриториальность к ЛПУ	Да	Да	Нет	Да	Да
Полнота/ целостность	Нет	Нет	Да	Нет	Да

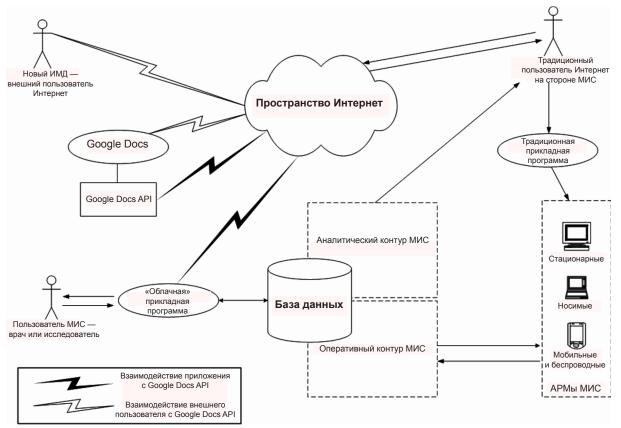


Рис. 8. Схема взаимодействия открытого и проприетарного ПО при использовании ИМД нового вида

Результаты, полученные при обработке данных в разработанной нами распределенной ИС, основанной на интеграции проприетарных модулей локального уровня и модулей масштаба предприятия с гетерогенными «облачными» сервисами, включающими бесплатный компонент, опираются на возможность получения и автоматизированного анализа ранее недоступных объемов анамнестических данных. Они дают основание ввести в практику диагностики и лечения мужского бесплодия дополнительный классификационный признак, состоящий в наличии когорты мужчин с жалобами на невозможность зачатия, подвергающихся воздействию множественных факторов внешней среды субнозологической интенсивности. Адаптивный характер бесплодия в данной классификационной группе прогностически благоприятен в отношении неспецифических методов лечения и реабилитации. С другой стороны, методы системной интеграции, ориентированные на использование широкодоступных «облачных» сервисов, предоставляют возможность эксплуатации значимого ИД нового вида, ранее не учтенного при исследовании медицинских информационных ресурсов.

ЛИТЕРАТУРА

- Гимадеев, Ш.М. Интеграция источников медицинской информации: цели и методология / Ш.М. Гимадеев, А.И. Латыпов, С.В. Радченко // Врач и информационные технологии. — 2006. — № 6. — С.61—67.
- Гимадеев, Ш.М. Роль клинической значимости видов медицинской информации в проектировании телемедицинских систем / Ш.М. Гимадеев, А.И. Латыпов, С.В. Радченко // Гастроэнтерология. — 2004. — № 4, спец. вып. — С.75—76.
- Зырянов, М. Esuite скоро заговорит по-русски / М. Зырянов // Computerworld Росси. — 1998. — № 20.
- Радченко, О.Р. Влияние образа жизни мужчин, состоящих в бесплодном браке, на показатели эякулята

- / О.Р. Радченко // Проблемы репродукции. 2010. № 6. С.94—98.
- Радченко, С.В. Применение распределенной компьютерной модели хронической почечной недостаточности для повышения качества жизни пациентов / С.В. Радченко, Д.Ф. Хазиахметов, О.Р. Шакулова-Радченко // Информационные технологии в здравоохранении.—2002. № 11/12. С.22—25.
- Шеян, И. Облака для здравоохранения / И. Шеян // Computerworld Россия/MedIT. — URL: http://www.osp.ru/ resources/focus-centers/MedIT/news/news_499.html
- Booly: a new data integration platform / L.H. Do, F.F. Esteves, H.J. Karten, E. Bier // BMC Bioinformatics. — 2010. — Vol. 11(1). — P.513.
- Comhaire, F.H. Simple model and empirical method for the estimation of spontaneous pregnancies in couples consulting for infertility / F.H. Comhaire // Int. J. Androl. — 1987. — Vol. 10(5). — P.671—680.
- Dondero, F. Epidemiologic and diagnostic aspects of male sterility / F. Dondero, A. Lenzi // Ann. Ital. Med. Int. — 1986. — Vol. 1(2). — P.152—160.
- Esteves, S.C. Novel concepts in male infertility / S.C. Esteves, A. Agarwal // Int. Braz. J. Urol. — 2011. — Vol. 37(1). — P.5—15.
- Kippenbrock, T. Google docs: a better method than a paper clinical schedule / T. Kippenbrock, E. Holloway, D.D. Moore // Comput. Inform. Nurs. — 2010. — Vol. 28(3). — P.138—140.
- Krausz, C. Y-chromosome microdeletions in «fertile» males / C. Krausz, K. McElreavey // Hum. Reprod. — 2001. — Vol. 16(6). — P.1306—1307.
- Patient web portals to improve diabetes outcomes: a systematic review / C.Y. Osborn, L.S. Mayberry, S.A. Mulvaney, R. Hess // Curr. Diab. Rep. — 2010. — Vol. 10(6). — P.422—435.
- Thiessen, P. WAI-ARIA live regions and channels: ReefChat as a case example / P. Thiessen, E. Russell // Disabil. Rehabil. Assist. Technol. — 2009. — Vol. 4(4). — P.276—287.
- Using web and social media for influenza surveillance / C.D. Corley, D.J. Cook, A.R. Mikler, K.P. Singh // Adv. Exp. Med. Biol. — 2010. — Vol. 680. — P.559—564.