

*А.А. Барчук^{2,3}, В.С. Гайдуков¹, А.М. Беляев^{2,3}, А.И. Арсеньев², В.И. Кузнецов¹,
Ю.И. Комаров², В.И. Сатиков¹, С.А. Тараканов¹, С.В. Канаев², А.В. Нефедова², А.В. Черная²*

Применение онтологии при скрининге онкологических заболеваний

¹ФГАОУ ВО «Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики»

²ФГБУ «НИИ онкологии им. Н.Н. Петрова» Минздрава России

³ГБОУ ВПО «СЗГМУ им. И.И. Мечникова» Минздрава России,
Санкт-Петербург

Одной из значимых проблем современной медицины, препятствующей в частности эффективному внедрению новых диагностических методик, таких как популяционный скрининг, является постоянное увеличение объемов важной медицинской информации, а также недостаточное внимание к анализу динамики состояния пациентов. Указанные проблемы могут быть решены за счет обеспечения информационной поддержки медицинского специалиста, как в процессе проведения исследований, так и при формировании рекомендаций по дальнейшему ведению пациентов. В статье рассматриваются возможные пути решения этих задач на основе разработки программных инструментов создания и работы с базами знаний, содержащих рекомендации по наблюдению и лечению различных видов заболеваний, а также интеллектуальной поддержки принятия решений на примере онкологических заболеваний. Результаты апробации данных решений позволяют говорить об их эффективности и применимости в клинической практике.

Ключевые слова: онкология, база знаний, онтология, семантические сети, поддержка принятия решений, большие данные, популяционный скрининг

Введение

В современной медицине специалистам зачастую приходится иметь дело с большими объемами первичных данных, которые нуждаются в обработке при диагностике различных заболеваний. Серьезной технической проблемой являются всё увеличивающиеся объемы важной для диагностики информации, которые становятся попросту невозможно обрабатывать прежними статистическими методами, учитывая ограниченные резервы человеческой памяти отдельного исследователя. Учитывая эти сложности и чрезвычайно высокую стоимость диагностических

ошибок, поскольку речь идет о человеческой жизни, проблема обеспечения медицинских специалистов современными средствами автоматизации рабочих процессов и поддержки принятия решений становится всё более актуальной.

Кроме того, нередко встречаются случаи как гиподиагностики, когда использование недостаточно эффективных методик не обеспечивает своевременного выявления патологии, так и гипердиагностики, когда используются неоправданно сложные диагностические мероприятия, приводящие к снижению качества жизни, различным осложнениям, а иногда и к проведению излишних лечебных мероприятий.

Во многих европейских странах основные акценты в настоящее время делаются на превентивную медицину, информационную поддержку врача, на создание новой информационной среды его деятельности. Это позволяет повысить качество диагностики, лечения и их доступность. В то же время внедряемые подходы к раннему выявлению заболеваний, среди которых одним из наиболее современных является популяционный скрининг, в свою очередь также являются источником больших массивов диагностических данных. В рамках популяционного скрининга первичные обследования проводятся среди клинически бессимптомных лиц, попадающих по определенным критериям в группы риска по целевому типу заболевания. Тем самым расширяется сфера контроля медицинского персонала, охватывавшая при традиционной системе обследования лишь пациентов с явными симптомами заболеваний, что серьезно увеличивает нагрузку на медицинский персонал или требует увеличения его количества, поэтому внедрение скрининговых программ ограничено во всем мире. Таким образом, в настоящее время все более возрастает потребность в автоматизации сбора, обработки, обмена информацией, отборе участников, контроле качества процессов и поддержки принятия решений медицинских специалистов в связи с ростом масштабов медицинских исследований.

Другой значимой проблемой является то, что недостаточно внимания уделяется динамическому наблюдению за изменениями, предшествующими патологическому процессу и в последующем течении заболевания. Оптимизация диагностического алгоритма путём определения и стандартизации этапов дополнительного обследования участников скрининга в зависимости от полученных предварительных данных позволит медицинским учреждениям выйти на новый этап своего развития. Решение перечисленных проблем может быть обеспечено за счет информационной поддержки медицинского специалиста в процессе, как проведения исследований, так и формирования рекомендаций по дальнейшему ведению пациентов.

Обзор современных решений в области онтологии и баз знаний для врачей

Вообще, **онтология** (др. греч. ὄντος + λόγος – учение о сущем, бытии) это раздел философии изучающий фундаментальные принципы бытия, его наиболее общие сущности и категории, структуру и закономерности, существующий в единстве с теорией познания и логикой.

Онтология в информатике — это попытка всеобъемлющей и подробной формализации некоторой области знаний с помощью концептуальной схемы. Обычно такая схема состоит из структуры данных, содержащей все релевантные (субъективно соответствующие в момент времени) классы объектов, их связи и правила (теоремы, ограничения), принятые в этой области. Онтологии используются в процессе программирования как форма представления знаний о реальном мире или его части. В отличие от философского термина онтология в информатике имеет два существенных отличия: 1) имеет формат, который компьютер сможет обработать; 2) информационные онтологии создаются всегда с конкретными целями и они оцениваются больше с точки зрения *применимости*, чем *полноты*.

В информатике и исследованиях искусственного интеллекта существует понятие база знаний (БЗ) – это особого рода база данных, разработанная для оперирования знаниями (метаданными). Знания представляют собой закономерности предметной области (принципы, связи, законы), полученные в результате практической деятельности и профессионального опыта, позволяющие специалистам ставить и решать задачи в данной области. Именно для представления знаний в компьютерной форме используются онтология – компьютерная форма представления знаний в виде семантической сети, где в качестве главных концептов выступают объекты с их атрибутами, отношения между объектами и сценарии (алгоритмы) функционирования объектов [5, 12, 16].

Создание и использование онтологий позволяет отделить знания и логику от программного кода, предоставляя возможность конечному пользователю без специальных знаний в области программирования пополнять базу знаний.

Применение онтологического подхода играет ключевую роль в процессе принятия решений, поскольку использование онтологий позволяет создавать наиболее точную и максимально приближенную к реальности модель предметной области, содержащую все основные понятия и отношения, которые так или иначе должны учитываться в процессе принятия решений [3]. Эта семантическая сеть может содержать как декларативные, так и процедурные компоненты и включать такие понятия (концепты) как физические или абстрактные объекты, свойства, процессы, отношения, атрибуты. К примеру, семантическая сеть для представления онтологии клинических рекомендаций в области заболеваний будет содержать описание патологий в данной сфере, методов их лечения, диагностики и профилактики, перечень медицинских специалистов, занимающихся лечением, лекарственные препараты, лечебные процедуры. С каждым из перечисленных объектов связано конечное число возможных сценариев действий, отношений, атрибутов. Например, в зависимости от вида поставленного диагноза пациенту может быть назначен тот или иной метод лечения. Симптомы заболевания и течение болезни влияют на используемые методы обследования и клинические рекомендации. Без использования онтологического подхода становится практически невозможным создать эффективную систему поддержки принятия решений при популяционном скрининге и использовать её в клинической практике, так как в этой сфере важен «адресный», т.е. ориентированный на пациента, подход к его наблюдению и лечению.

Традиционно, в информационных системах уровень хранения данных представлен в виде классических объектных и реляционных баз данных, предназначенных для хранения массивов однородных данных, имеющих жёсткую (фиксированную) структуру и набор атрибутов. В отличие от баз данных, рассчитанных на хранение больших по объёму и иерархически структурированных однородных массивов, например, «заболевание – метод лечения», онтологии позволяют работать с совершенно разнородными данными, не имеющими фиксированной структуры (данные о заболеваниях, способах диагностики, медицинских учреждениях и т.д.), структура которых является сетевой и часто дополняется или меняется [4], в частности, с принятием каждого нового закона или реформированием бизнес-процесса предоставления медицинских услуг, изменением методики лечения, появлением нового лекар-

твенного препарата или метода диагностики и др. При этом онтология может играть роль мета-данных, описывающих не только структуру данных (например, данных о заболевании или методике лечения), но и методов доступа к этим данным во внешних базах данных [7, 10].

Ряд проведенных исследований показывает, что использование технологии семантических сетей для построения онтологий актуально на сегодняшний день в различных сферах применения, в частности различных областях медицины [6, 8, 9, 11, 15].

Применение онтологий для создания рекомендаций по наблюдению и лечению онкологических заболеваний

Авторами статьи разработан прототип базы знаний, содержащий клинические рекоменда-

ции по лечению онкологических заболеваний в структурированном виде, пригодном для использования программным обеспечением – в виде онтологии. В качестве примера приведены разработанные клинические рекомендации по диагностике рака легкого (РЛ), в основе которых в свою очередь лежат рекомендации, предложенные и постоянно совершенствуемые организацией National Comprehensive Cancer Network (NCCN); версия 2.2015. от 17.03.2015 г. Доказано, что для выявления раннего периферического РЛ основным и наиболее эффективным с точки зрения популяционного скрининга способом исследований можно считать низкодозную компьютерную томографию (НДКТ), обеспечивающую при относительно небольшой дозе облучения достаточный уровень детектирования узлов в легких [13]. Отправной точкой является первое обследование пациента из группы риска

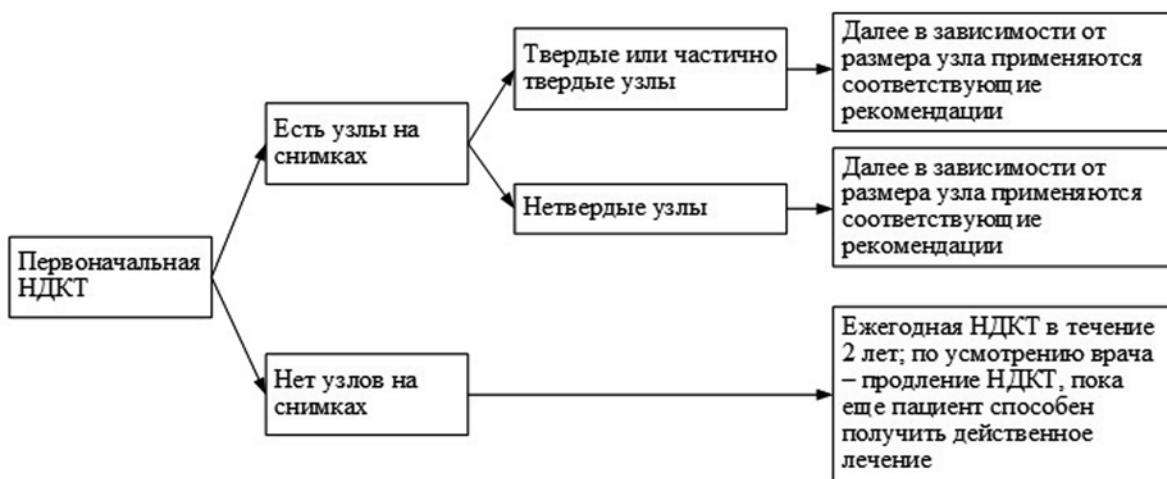


Рис. 1. Отправная точка при скрининге рака легкого



Рис. 2. Рекомендации по неплотным (нетвердым) узлам

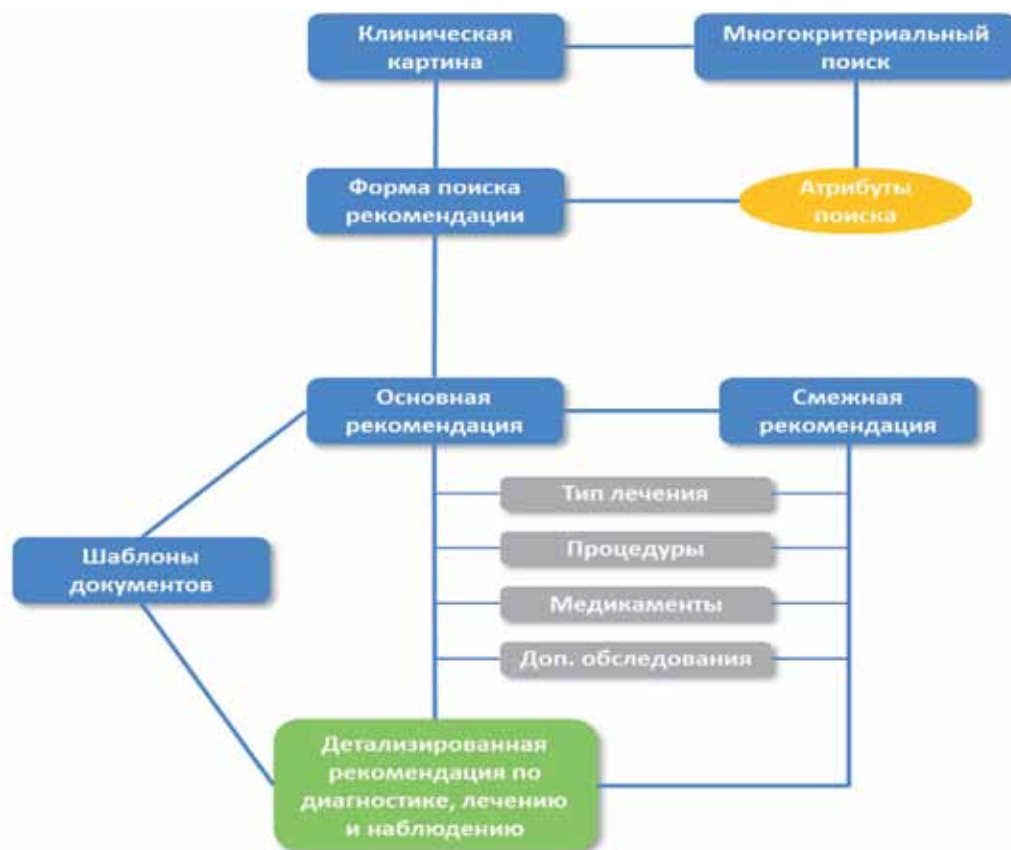


Рис. 3. Семантическая сеть клинических рекомендаций

(рис. 1). В первую очередь рекомендации строятся в зависимости от плотности узла в легком. При этом выделяют три группы на основе значения единиц Хаунсфилда: плотный (твердый) [14, 17], неплотный (нетвердый) узел, или узел по типу чистого матового стекла [14, 18], и частично плотный (полутвердый) узел, или узел по типу смешанного матового стекла.

Другими факторами и условиями, которые влияют на дальнейшие диагностические процедуры, являются размер (диаметр, или объём) узла и динамика его изменения [2]. Так, например, рекомендации по дальнейшему наблюдению в случае обнаружения нетвердых узлов приведены на рис. 2.

В рамках популяционного онкологического скрининга рекомендации заключаются в определении онтологического алгоритма наблюдения участников из группы риска, проведения необходимых диагностических процедур в зависимости от ряда факторов и условий и возможному дальнейшему направлению пациента в специализированный медицинский центр для лечения. В рамках такого подхода решается сопутствующая проблема организации наблюдения и оценки динамики состояния пациентов.

Реализация программных средств для создания онтологий клинических рекомендаций и поддержки принятия решений

Предлагаемый прототип использован с целью разработки программных инструментов для создания баз знаний, работы с ними и содержал рекомендации по наблюдению и лечению заболеваний, а также по поддержке принятия решений в этой области. Указанные программные инструменты разрабатывались применительно к сфере онкологии, однако могут быть с равным успехом применены и для других видов заболеваний. Программные инструменты реализованы в качестве модулей интеллектуальной распределенной системы популяционного скрининга (СПС) онкологических заболеваний [1]. Данная СПС, обеспечивающая автоматизированный сбор, обработку, анализ и управление информацией, а также контроль указанных процессов, рассматривается как основа для запуска программ популяционного скрининга онкологических заболеваний, которые на настоящий момент только внедряются во всем мире.

Модуль онтологии клинических рекомендаций

Фрагмент семантической сети (онтологии) для представления клинических рекомендаций по наблюдению и лечению онкологических заболеваний представлен на рис. 3.

На рис. 3 синим и зелёным цветом выделены сущности базы знаний. По семантической сети видно, что клиническая рекомендация связана с формой поиска. При поиске клинической рекомендации специалистом описывается клиническая картина либо используется многокритериальный поиск по одному или нескольким атрибутам поиска. После обработки поискового запроса, прототип базы знаний выдаёт основную клиническую рекомендацию, смежную клиническую рекомендацию (в случае, если для описанного заболевания такая имеется) и шаблоны документов, которые необходимо будет заполнить для того, чтобы пациент мог продолжить проходить обследование или лечение. Детализированная клиническая рекомендация определяется шаблонами документов, а так же основной и смежной клиническими рекомендациями, которые, в свою очередь, определяются типом лечения, процедурами, медикаментами и дополнительными обследованиями.

Разработанный модуль онтологии клинических рекомендаций предназначен для описания в формализованном виде знаний из методов медицинских обследований, клинических рекомендаций и стандартов оказания медицинской помощи при диагностике и лечении онкологических заболеваний. Применение онтологического подхода позволяет наиболее полно описать медицинские знания, содержащиеся в клинических рекомендациях, пополнять эти знания и использовать их в процессе принятия решений. Модуль обеспечивает следующие возможности: а) создание, редактирование, добавление или удаление шаблонов клинических рекомендаций; б) ввод неограниченного количества шаблонов клинических рекомендаций; в) ввод шаблонов клинических рекомендаций с нормами и стандартами оказания медицинской помощи посредством форм ввода данных, содержащих такие свойства информации как сущность и атрибуты; г) добавление и редактирование структуры онтологии клинических рекомендаций посредством добавления/удаления связей между сущностями посредством соответствующего редактора; д) визуализированное предоставление данных клинических рекомендаций с возможностью их редактирования, добавления или удаления посредством простых действий. Модуль содержит конструктор онтологии, который позволяет создавать и изменять структуру знаний, содержащихся в базе знаний.

При помощи конструктора онтологии можно создавать, изменять и удалять сущности базы знаний. После создания сущности можно сформировать конкретную клиническую рекомендацию для определённого заболевания, содержа-

щую сведения о профилактике, диагностике, лечении заболеваний и реабилитации после выздоровления или достижения ремиссии. Каждую сущность можно связать с другой сущностью. При помощи отношений формируется структура онтологии, например, клиническую рекомендацию посредством отношений можно связать с определённым видом лечения, медикаментами, которые будут использованы в процессе лечения, процедурами и дополнительными обследованиями. Создаваемая клиническая рекомендация содержит атрибуты, каждый из которых может быть использован для поиска рекомендации специалистом.

Модуль интеллектуальной поддержки принятия решений

В рамках модуля реализованы следующие функциональные возможности: а) предоставление информации о способах лечения или способах минимизации риска развития онкологических заболеваний пациента в виде встречной сцены рекомендаций; б) предоставление детальной информации о рекомендуемых пациенту медикаментах, их дозировках и режимах введения; в) предоставление информации о назначаемых пациенту дополнительных обследованиях, которые необходимо будет пройти; г) предоставление информации как в текстовом, так и в графическом виде.

Семантическая сеть для представления онтологии клинических рекомендаций содержит описание клинической картины заболевания и позволяет, на основании предоставленной пользователем информации о симптомах заболевания, выдавать адресную клиническую рекомендацию, содержащую сведения о профилактике, диагностике и лечении заболевания. Доступ к формируемым рекомендациям может получить как медицинский специалист, так и непосредственно наблюдаемый пациент.

Поиск клинической рекомендации для врача-онколога производится на основе описания клинической картины посредством единовременного заполнения данных. Также врачу доступен поиск и сортировка рекомендаций в зависимости от статуса, сферы действия, уровня действия, типа специалиста (например, можно осуществить поиск только по действующим международным рекомендациям в области онкологии, предназначенным для врачей, специализирующихся по онкологическим заболеваниям в области органов грудной полости).

Пациент или медицинский специалист с малым опытом или стажем может осуществить поиск необходимой клинической рекомендации путем работы с интерактивной анкетой — вы-

бирается вид обследования, в зависимости от которого выделяются ключевые данные, по которым будет проводиться анализ. Затем пациент описывает клиническую картину в формате анкетирования, а по окончании всех итераций (повторений) формируется некая клиническая картина, в соответствии с которой выдаётся встречная адресная клиническая рекомендация, содержащая перечень типов назначаемого лечения, видов медикаментов их дозировках и режимах использования, а так же процедур и обследований, которые необходимо пройти.

Результаты апробации полученных решений

Апробация программных инструментов была проведена в составе СПС на медицинских объектах Самарского региона в условиях, максимально приближенных к реальной клинической практике. В рамках модуля онтологии клинических рекомендаций при помощи конструктора онтологии врачами-онкологами было произведено создание и изменение структуры и содержимого базы знаний посредством добавления и редактирования сущностей, атрибутов сущности, связей между сущностями. Результатом стало создание сущности базы данных. В рамках модуля интеллектуальной поддержки принятия решений в формате анкетирования была описана клиническая картина заболевания пациента и в соответствии с этой картиной распределенная система популяционного скрининга выдала встречную адресную клиническую рекомендацию. По результатам апробации был отмечен высокий потенциал СПС как инструмента автоматизации процессов в области медицины и поддержки специалистов в медицинской практике.

Выводы

Описанные в настоящей статье инструменты по созданию баз медицинских знаний и интеллектуальной поддержки принятия решений на основе интерактивной работы с такими базами знаний являются важной составляющей частью решения проблем увеличения объемов медицинских данных и оптимизации нагрузки на высокоспециализированных медицинских специалистов. Особую ценность данные решения представляют для области формируемых сейчас во всем мире программ, направленных на раннее выявление и предупреждение развития заболеваний, так как данная область связана с расширением контингента обследуемых ввиду необходимости выявления скрытых признаков или установления факторов риска заболеваний. Разработанные решения могут найти применение в рамках частных задач, таких как снижение

нагрузки на медицинских специалистов, кроме того, в составе интеллектуальных систем, могут быть направлены на внедрение эффективных программ популяционного скрининга по различным видам заболеваний, в частности онкологическим.

Настоящая работа проводится при финансовой поддержке государства в лице Минобрнауки России в рамках проекта RFMEFI57814X0008

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Barchuk A.A., Podolsky M.D., Gaidukov V.S. et al. Intelligent distributed system of population cancer screening // *Voprosy onkologii.* – 2015. – Vol. 61. – № 4. – P. 517-522.
2. Barchuk A.A., Podolsky M.D., Tarakanov S.A. et al. Automated identification, interpretation and classification of focal changes in the lungs on the images obtained at computed tomography for lung cancer screening // *Voprosy onkologii.* – 2015. – Vol. 61. – № 6. – P. 913-919.
3. Brook R.H., Vaiana M.E. Using the Knowledge Base of Health Services Research to Redefine Health Care Systems // *Journal of General Internal Medicine.* – 2015. – Vol. 30. – № 10. – P. 1547-1556.
4. Budovec J.J., Lam C.A., Kahn C.E.Jr. Informatics in radiology: radiology gamuts ontology: differential diagnosis for the Semantic Web // *Radiographics.* – 2014. – Vol. 34. – № 1. – P. 254-264.
5. Griffon N., Charlet J., Darmoni S. Knowledge representation and management: towards an integration of a semantic web in daily health practice // *Yearbook of medical informatics.* – 2013. – № 8. – P. 155-158.
6. Jafarpour B., Abidi S.R., Abidi S.S. Exploiting Semantic Web Technologies to Develop OWL-Based Clinical Practice Guideline Execution Engines // *IEEE J Biomed Health Inform.* – 2016. – Vol. 20. – № 1. – P. 388-398.
7. Jing X., Kay S., Marley T. et al. Integration of an OWL-DL knowledge base with an EHR prototype and providing customized information // *Journal of Medical Systems.* – 2014. – Vol. 38. – № 9. – P. 1-14.
8. Livingston K.M., Bada M., Baumgartner WA Jr, Hunter LE. KaBOB: ontology-based semantic integration of biomedical databases // *BMC Bioinformatics.* – 2015. – Vol. 16. – № 126. – 21 p.
9. Liyanage H., Krause P., De Lusignan S. Using ontologies to improve semantic interoperability in health data // *Journal of Innovation in Health Informatics.* – 2015. – Vol. 22. – № 2. – P. 309-315.
10. Machado C.M., Rebholz-Schuhmann D., Freitas A.T. et al. The semantic web in translational medicine: current applications and future directions // *Briefings in bioinformatics.* – 2015. – Vol. 16. – № 1. – P. 89-103.
11. Mohammed O., Benlamri R. Developing a semantic web model for medical differential diagnosis recommendation // *Journal of Medical Systems.* – 2014. – Vol. 38. – № 10. – P. 1-18.
12. Panahiazar M., Taslimitehrani V., Jadhav A., et al. Empowering Personalized Medicine with Big Data and Semantic Web Technology: Promises, Challenges, and Use Cases // *Proceedings of the 2014 IEEE International*

- Conference on Big Data, IEEE Big Data 2014, Oct. 2014. – 2015. – P. 790-795.
13. Silverman J.D., Paul N.S., Siewerdsen J.H. Investigation of lung nodule detectability in low-dose 320-slice computed tomography // *Med Phys.* – 2009. – Vol. 36. – № 5. – P. 1700-1710.
 14. Sone S., Hanaoka T., Ogata H. et al. Small peripheral lung carcinomas with five-year post-surgical follow-up: assessment by semi-automated volumetric measurement of tumour size, CT value and growth rate on TSCT // *Eur Radiol.* – 2011. – Vol. 22. – № 1. – P. 104-119.
 15. Wang Z., Sagotsky J., Taylor T., et al. Accelerating cancer systems biology research through Semantic Web technology // *Wiley Interdisciplinary Reviews: Systems Biology and Medicine.* – 2013. – Vol. 5. – № 2. – P. 135-151.
 16. Wu H., Yamaguchi A. Semantic Web technologies for the big data in life sciences // *Bioscience trends.* – 2014. – Vol. 8. – № 4. – P. 192-201.
 17. Xu D.M., van Klaveren R.J., de Bock G.H. et al. Role of baseline nodule density and changes in density and nodule features in the discrimination between benign and malignant solid indeterminate pulmonary nodules // *European Journal of Radiology.* – 2009. – Vol. 70. – № 3. – P. 492-498.
 18. Zhang L., Yankelevitz D.F., Carter D. et al. Internal Growth of Nonsolid Lung Nodules: Radiologic-Pathologic Correlation // *Radiology.* – 2012. – Vol. 263. – № 1. – P. 279-286.

*A.A. Barchuk^{2,3}, V.S. Gaidukov¹, A.M. Belyaev^{2,3},
A.I. Arseniev², V.I. Kuznetsov¹, Yu.I. Komarov²,
V.I. Satikov¹, S.A. Tarakanov¹, S.V. Kanaev²,
A.V. Nefedova², A.V. Chernaya²*

The use of ontology in screening for cancer

¹Saint Petersburg National Research University of Information Technologies, Mechanics and Optics

²N.N. Petrov Research Institute of Oncology

³I.I. Mechnikov North-West State Medical University
St. Petersburg

One of the most important problems of modern medicine, which, in particular, precludes the effective implementation of new diagnostic methods such as population screening, is the steady increase of volumes of important medical data, as well as insufficient attention to the analysis of the dynamics of the patients' condition. These problems can be solved by the information support of medical specialist in the process of research and in the formation of recommendations for further management of patients. In the study, we examined the possible ways of solving these problems through the development of software tools for creation of knowledge bases of recommendations for monitoring and treatment of various diseases, as well as intelligent decision support by the example of cancer. The results of tests of these solutions allow speaking about their effectiveness and applicability in clinical practice.

Key words: oncology, knowledge base, ontology, semantic networks, decision support, big data, population screening

Поступила в редакцию 29.11.2016 г.