

Значение критерия оценки прозрачности  $O$  постоянно для ПС, т.е. не зависит от конкретной решаемой вычислительной проблемы.

Предлагаемая система критериев оценки эффективности ПС организации РВ может применяться для анализа эффективности различных ПС организации РВ, а также выбора ПС с более высокими значениями приоритетных критериев для решаемой распределенной вычислительной задачи.

Список использованных источников:

1. Coulouris, G. Distributed Systems: Concepts and Design (5th Edition). Coulouris G. [and others]. – Boston: Addison-Wesley, 2011. – 1008 p.
2. Черемсинов, Д.И., Проектирование и анализ параллелизма в процессах и программах / Д.И. Черемсинов. – Минск: Беларус. навука, 2011. – 300 с.
3. 5200.28-STD Trusted Computer System Evaluation Criteria // National Security Institute [Электронный ресурс]. – 2015. – Mode of access: <http://csrc.nist.gov/publications/history/dod85.pdf>. – Date of access: 18.02.2016.
4. Farooqi, K. ISO Reference Model of Open Distributed Processing (RM-ODP) // Institute For Enterprise Architecture Developments [Электронный ресурс]. – 2015. – Режим доступа: <http://www.enterprise-architecture.info/Images/Document-s/RM-ODP2.pdf>. – Дата доступа: 15.01.2016.

## СЕМАНТИЧЕСКАЯ НАВИГАЦИЯ В БОЛЬШИХ ИНФОРМАЦИОННЫХ МАССИВАХ

*Белорусский государственный университет информатики и радиоэлектроники  
г. Минск, Республика Беларусь*

*Чеушев К. В.*

*Глухова Л. А. – канд. техн. наук, доцент*

Предложено системное решение для актуальной задачи навигации в больших информационных массивах, основанное на применении комплекса аналитических функций, создающего возможность целенаправленного движения пользователя в пространстве информационного ресурса, направляемого не априори установленными ссылками, а его текущими потребностями и интересами. Выделена обеспечивающая семантическую навигацию операциональная триада «структурирование – экстрагирование – поиск подобного», элементы которой собираются в цепи и каскады, реализующие все необходимые аналитические функции, такие как сложные высокоточные формы поиска, классификация, кластеризация, автоматическое выделение ключевых термов, построение сниппетов, аннотаций, дайджестов и т.д., позволяющих пользователю удовлетворить сложные информационные потребности, итерационно двигаясь с помощью аналитической системы к искомому результату.

Экспоненциальный рост доступных в сетях информационных ресурсов (ИР) создаёт заманчивую перспективу их использования в профессиональных, образовательных и других целях. Гигантские ИР сами по себе ещё не достаточное условие доступности и качества знаний, с ростом масштабов ИР увеличивается и усложняется проблема нахождения элементов ИР, удовлетворяющих конкретной информационной потребности пользователя (ИПП), критериям полноты, точности и качества. Острая ещё пару десятилетий назад проблема информационного голода трансформировалась в информационное перенасыщение, когда мы практически уверены в том, что в пределах нашей доступности нужная информация есть, но слабо представляем, как её получить, отобрать, сократить объём подлежащего «ручному» анализу до приемлемого. Раньше основные надежды возлагались на то, что информация перед помещением в состав ИР будет переработана человеком, должным образом структурирована и снабжена семантическими и другими метками, но сегодня всё чаще на это рассчитывать уже нельзя – объёмы новых ИР велики настолько, что качественно их обработать и всесторонне структурировать люди уже не в состоянии, но и в любом случае структурирование по одним критериям может не коррелировать с критериями конкретной ИПП. В текстовых ресурсах разметка для обеспечения гиперссылочной навигации уже перестала достаточно полно обеспечивать возможности доступа к искомой информации, а полнотекстовый поиск зачастую порождает огромный «шум», что вынуждает либо проводить очень ёмкую последующую обработку найденного, либо ограничиться использованием малого количества найденных документов, сомневаясь в их качестве и полноте удовлетворения ИПП. Велика и возрастает актуальность и востребованность так называемой семантической навигации – развитого способа ориентирования в огромных информационных ресурсах с опорой не на априори расставленные гиперссылки, а на интеллектуальные автоматические инструменты.

Под семантической навигацией понимается процесс движения аналитика или просто пользователя в информационном пространстве ИР, при котором он имеет возможность итерационно корректировать свои цели и интересы и переходить в соответствии с ними к другим подобластям ИР, стартуя с какой-то позиции, возможно семантически даже отдалённой от искомого результата (т.е. допускается нечёткое начальное представление цели). Отличие от навигации в классическом понимании (по априори установленным ссылкам) состоит в том, что направление движение в пространстве ИР не определено заранее, оно обеспечивается функциональными аналитическими механизмами непосредственно в момент взаимодействия пользователя с информационно-аналитической системой.

Исследования и эксперименты показывают, что для осуществления всех главных поисковых и аналитических функций необходимо и достаточно три функциональных (операциональных) механизма:

- структурирование дисперсной среды ИР;
- экстрагирование семантических детерминантов;
- собственно семантическая навигация как переход от одних элементов ИР к другим.

Эта тройка, будучи собранной в последовательности и в иерархию, обеспечивает все нужные смысловые операции разных уровней. При этом в каждую из них на более высоком уровне рассмотрения обычно входят две другие с предыдущего иерархического уровня. Так, операцию кластеризации как структурирование довольно высокого уровня можно выразить цепью операций экстрагирования и семантической навигации более низкого уровня.

Структурирование здесь – это выделение из дисперсной среды ресурса групп (кластеров) элементов, “откликающихся” на одну или несколько информационных потребностей. Это может быть поиск, и тогда результатом станет один кластер найденных текстов (а потребность сформулирована запросом), или параллельная группа поисков по разным основаниям, и тогда результат – группа кластеров. Если основания (так называемые профили классов) заданы априори, то это задача классификации, а если априори смысловые основания структурирования не заданы и их нужно определять в процессе, то это кластеризация. В отличие от часто понимаемых классификации и кластеризации как “раскладывания по полочкам” обязательно всех элементов ресурса здесь такая цель не ставится – в этом нет нужды, а нужно выделить из ИР только те группы элементов, которые потенциально удовлетворяют ИПП и только им. И второе – элемент ИР будет помещаться в такое количество кластеров, на какое у него будет достаточный “отклик”. Кластеризация в этой модели не только самостоятельная аналитическая задача, но и нередко решается как промежуточная в другом процессе. Например, построение кластеризованного дайджеста – по сути экстрагирование – содержит в себе кластеризацию как служебную задачу.

Экстрагирование семантических детерминантов – самый интеллектуально ёмкий элемент триады, во многом определяющий успех аналитических цепочек. Прежде всего это ранжирование элементов ресурса и выделение верхней части ранжированного списка, селекция тех элементов – семантических детерминантов, – которые наиболее полезны для получения планируемого результата, удовлетворения заданной информационной потребности. В более сложных случаях за подбором значимых элементов следует ещё и придание им структуры. Роль экстрагирования, как и других операций рассматриваемой триады, двойственна: оно может иметь самостоятельное значение (выделение ключевых термов, построение контекстных (так называемые сниппеты) и бесконтекстных аннотаций, дайджестов), а может входить в состав других операций. Так, кластеризация существенно основана на экстрагировании групп ключевых термов, которые могут составлять “смысловое ядро” кластера.

Семантическая навигация как элемент операциональной триады предполагает переход от одних элементов ИР к другим. И если структурирование и экстрагирование сужают рассматриваемое пространство ИР при решении аналитической задачи, то переход расширяет его. Ключевой конструктив этой операции – поиск (или отбор) подобного. Так, если взять текст, то смысловой переход это поиск текстов, подобных на заданный. Операция семантической навигации на уровне поиска подобных включает в себя пару других операциональных элементов: чтобы найти подобные тексты, система экстрагирует из “эталона” ключевые термы и выполняет по ним операцию структурирования (поиск, а затем, возможно, классификацию или кластеризацию найденного).

Навигация поиском подобных становится точнее, если взять не один текст, а кластер похожих текстов: тогда проще отобрать подходящие ключевые термы, отражающее как раз то, что в текстах кластера общее. Так формируется “экстракт кластера” в форме ранжированного списка ключевых термов. Полезна ещё одна операция – получение своего рода “экстракта кластера” уже в другой форме – в форме той части текстов кластера, которые сильнее других “откликаются” на выделенное множество ключевых термов, т.е. текстов, максимально релевантных тематике кластера и концентрированно отражающих эту тематику (тогда как у пользователя нигде не возникла необходимость формулировать её в явном виде, и это одно из наиболее ярких достоинств подхода).

Определённая таким образом семантическая навигация может быть сколь угодно долгой, поскольку комплекс функциональных механизмов обеспечивает как сужение пространства анализа, так и его расширение. Сужение пространства анализа достигается применением того или иного способа структурирования (классификации, кластеризации) с последующим переходом к одному из кластеров. В таком кластере также возможно сужение (повторная кластеризация или поиск по словарю, запросу, фрагменту текста и т.д.) или выделение “экстракта” в виде наиболее выразительных текстов, а возможно и расширение путем поиска похожих текстов или поиска по ключевым словам или дайджестам, которые система предложит для этого кластера.

Семантическая навигация должна стать действенным механизмом решения поисково-аналитических задач, в том числе априори неизвестных. Действительно, если создать необходимый набор интеллектуальных инструментов – функций для всех элементов рассмотренной триады, собрать их в более высокоуровневые цепочки, связать их в единое целое с интерфейсными решениями, то можно получить очень мощный комплекс, работающий по принципу взаимодействия пользователя и аналитической системы достижению полезного результата, с положительной обратной связью, реализующий цепь “синергических транзакций” взаимодействия человек-машина. Начав с посылки даже самого малого смыслового сигнала, хоть как-то отражающего информационную потребность, пользователь получит отклик системы (например, кластер текстов со сниппетами, подсветкой автоматически выявленных ключевых термов, дайджестом), с которым он может провести действия, уточняющие потребность (отобрать наиболее релевантные и представительные

тексты или предложения, поднять веса некоторым термам и т.п.) и запустить следующую аналитическую цепь, которая даст уже более полный и точный результат – и так далее до достижения нужного качества и достаточной полноты решения поставленной задачи.

На основе предложенного и проверенного эксперимента комплекса решений планируется создание небольшой информационно-аналитической системы, на которой будут отработаны все предлагаемые решения. В качестве ИР по ряду мотиваций планируется использовать документальный ресурс Википедии. С одной стороны, этот ресурс доступен и понятен, с другой – реализованная система будет полезна на практике, реализуя ряд важных востребованных аналитических функций. И, наконец, разметка данного ресурса позволит в некоторых случаях применить автоматизированные приёмы анализа результатов работы компонентов системы, что даст возможность быстрее и качественнее подобрать параметры алгоритмов и метрик.

## ХАРАКТЕРИСТИКИ КАЧЕСТВА СПЕЦИФИКАЦИЙ ТРЕБОВАНИЙ

*Белорусский государственный университет информатики и радиоэлектроники  
г. Минск, Республика Беларусь*

*Чиркова А. Ю.*

*Бахтизин В. В. – к. техн. наук, доцент*

Качество программного обеспечения можно определить как соответствие явно установленным функциональным и эксплуатационным требованиям и явно указанным стандартам разработки. Такое определение качества программного обеспечения подчеркивает, что требования к программному обеспечению – основа, относительно которой определяется качество ПО.

Обеспечение качества программного обеспечения – непрерывный процесс в течение всего жизненного цикла ПО, который охватывает: методы и средства анализа, проектирования и кодирования; методику многоуровневого тестирования; контроль программной документации и внесенных в нее изменений; процедуры обеспечения соответствия стандартам в области разработки ПО, соответствие которым определено в задании на разработку данного ПО.

Разработка любого программного средства (ПС) начинается с определения требований, которые предъявляются к нему. Качество конечного продукта напрямую зависит от того, насколько качественно сформулированы требования [1]. По оценкам Д. Леффингуэлла и Д. Уидринга, стоимость исправления ошибок на финальных этапах разработки ПС возрастает на порядок [2]. Именно поэтому важно обеспечивать качество разрабатываемого продукта еще на этапе создания спецификации требований.

Обеспечение качества спецификации невозможно без его оценки, что в свою очередь не возможно без соответствующих методов оценки качества, основанных на модели качества. Современные стандарты, такие как, например, IEEE 830 [3], ISO/IEC/IEEE 29148 [4], ISO/IEC 25030 [5] содержат только рекомендации и методики составления спецификаций требований, а также характеристики качества требований. Однако они не описывают ни модели качества, ни метрики(меры), по которым можно было бы оценить качество требований.

Модель качества ПС в соответствии с ISO/IEC 25010 [6] представлена в иерархическом виде, где на первом уровне отображены характеристики (показатели) качества, а на втором уровне – подхарактеристики. Всего модель содержит восемь характеристик качества программного продукта: функциональное соответствие, эффективность функционирования, совместимость, практичность, надежность, защищенность, сопровождаемость, мобильность. Однако данная модель качества может быть применена только к программному продукту, или к компьютерной системе, которая включает в себя программное обеспечение, так как большинство подхарактеристик имеют отношение к ПС и системам. Поэтому модель качества спецификаций требований должна содержать характеристики и подхарактеристики качества, оценивающие не только на качество требований, но и влияющие на качество ПС в конечном итоге.

Качество данных является ключевым компонентом качества и полезности информации, полученной из этих данных, и большинство бизнес-процессов зависит от качества получаемых данных. Таким образом, спецификация требований также должна отвечать критериям качества данных.

Согласно стандарту ISO/IEC 25012 [7] модель качества данных содержит следующие характеристики: правильность, полнота, согласованность, достоверность, актуальность, доступность, соответствие, конфиденциальность, эффективность, точность, трассируемость, понятность, пригодность, переносимость, восстанавливаемость.

Основываясь на моделях качества, предложенных стандартами ISO/IEC 25010 и ISO/IEC 25012, а также на требованиях и рекомендациях стандарта ISO/IEC 25030, были выделены характеристики качества требований, а также подхарактеристики.

Спецификацию требований к ПО следует рассматривать как множество отдельно взятых требований, которые должны отвечать характеристикам качества, предъявляемым требованиям, а также как систему требований, к которой должна соответствовать своему набору характеристик качества.

Каждое требование к ПО должно обладать следующими свойствами. Каждое требование определяет важность, характеристики ПС и ограничения.

Требования должны быть актуальными и не устаревать с течением времени. Они не должны зависеть