

КОМПЛЕКСНОСТЬ ИНФОРМАЦИОННОГО ПОИСКА

Игорь Наумович Розенберг, профессор, д-р техн. наук,

генеральный директор,

e-mail: ig.rozenb2012@yandex.ru,

ОАО «Научно-исследовательский и проектно-конструкторский институт информатизации, автоматизации и связи на железнодорожном транспорте»

(ОАО «НИИАС»),

http://www.vniias.ru,

Академик академии транспорта, академик международной академии наук Евразии, заведующий кафедрой «Геодезия, геоинформатика и навигация»,

Московский государственный университет путей сообщения Императора Николая II

http://www.miit.ru

Статья раскрывает содержание современных технологий информационного поиска. Статья вводит новые понятия оценки результатов информационного поиска. Проведена дифференциация понятия релевантность. Статья вводит новые характеристики оценки результатов поиска, включая разнообразные оценки релевантности. Статья обосновывает необходимость информационного моделирования при организации информационного поиска. Статья доказывает, что информационное моделирование при информационном поиске должно быть системным и включать дескриптивные и прескриптивные модели.

Ключевые слова: информационные технологии; информационный поиск; паттерн; концепт; информационные единицы.

DOI: 10.21777/2312-5500-2017-1-41-49

Введение

В реальной практике лицу, принимающему решение, или исследователю приходится работать в условиях информационной неопределенности [1] или информационной асимметрии [2]. Иногда информационную асимметрию трактуют как семантический разрыв. Но и в этом случае инструментом преодоления всех перечисленных факторов является информационный поиск (*Information Retrieval, IR*) [3]. В процессе информационного поиска проводят качественное и смысловое оценивание информационного объекта. То есть по существу информационный поиск является набором методов когнитологии когнитивного анализа [4]. Информационный поиск является одной из технологий сбора информации, необходимой для поддержки принятия решений. Информационный поиск применяется для удовлетворения информационной потребности, продиктованной информационной ситуацией [5]. Чаще всего его применяют с использованием специализированных информационно-поисковых систем (ИПС). Информационный поиск решает также онтологические задачи [6].

Материалы и методы. В качестве материалов использованы существующие описания и статистика информационного поиска. В качестве методов использовался системный анализ и информационное моделирование.

Принципы информационного поиска – поиск неструктурированной информации, единицей представления которой является документ произвольного формата. Информационный поиск представляет собой процесс отыскания необходимой информации I_n (знаний или данных) в некотором информационном множестве в соответствии с заданным критерием поиска. Предметом поиска выступает информационная потребность пользователя, неформально выраженная в поисковом запросе. Следует различать информационный поиск и «поиск данных». Критерий информационного поиска и его результаты не детерминированы. Этим информационный поиск отличается от «поиска



И.Н. Розенберг

данных», который оперирует набором формально заданных предикатов, имеет дело со структурированной информацией и чей результат всегда детерминирован. Поиск может осуществляться циклически, рекурсивно, селективно, многоаспектно [7].

Теория информационного поиска изучает все составляющие процесса поиска, а именно: предварительную обработку текста (индексирование), обработку и исполнение запроса, ранжирование, пользовательский интерфейс и обратную связь. При информационном поиске следует учитывать следующие факторы:

1. Информационная потребность в поиске необходимой информации In , или I . *Информационная потребность* – необходимость в получении информационных продуктов и услуг для поддержания жизнедеятельности и развития человеческой личности, социальной группы, общества в целом; внутренний побудитель активности.

2. *Информационный язык* – искусственный язык, используемый в информационно-поисковых технологиях с целью формализации информации, фактов и сведений.

3. *Поисковое предписание* (query, запрос) – в общем случае шаблон текста с искомым образцом, составленный с использованием информационного языка. В работе [4] его называют паттерном.

4. *Дескриптор* (лат. descriptor – описывающий), лексическая единица (слово, словосочетание) информационно-поискового языка, выражающая основное смысловое содержание какого-либо текста. Дескриптор может быть составлен и на естественном языке. Используется при информационном поиске документов в информационно-поисковых системах. В работе [4] его называют концептом.

5. *Индексирование* (indexing, индексация) – процесс составления или приписывания указателя (*индекса*) – служебной структуры данных, необходимой для последующего поиска. Выражение главного содержания текста какого-либо документа в терминах только языка информационно-поисковой системы. Применяется для упрощения поиска нужного документа среди множества других.

6. *Лемматизация* (lemmatization, нормализация) – приведение формы слова к словарному виду, то есть лемме [8].

7. Технология поиска информации и стратегия поиска.

8. Критерий оценки соответствия найденной информации I_p , или X , поисковому запросу.

9. Оценка эффективности поисковой системы или эффективности технологий поиска.

10. Информационное множество (*Data set*), в котором осуществляется поиск. Это множество, элементам которого ставятся во взаимно однозначное соответствие так называемые ключи (*идентификаторы*) – информационные элементы без внутренней структуры. Введение ключа означает установление соответствия между атрибутами информационных объектов и неким новым информационным элементом.

Алгоритмы и методы поиска. Известны различные поисковые алгоритмы. Прямым называется поиск, алгоритм которого основан на последовательном просмотре документов.

Прямой поиск – поиск непосредственно по тексту документов [9], без предварительной обработки (без *индексирования*). Прямой поиск текста заключается в просмотре строки (слева направо) и последовательном сравнении каждой позиции с искомой подстрокой. Для этого сравнивают все символы.

Другие алгоритмы требуют «индексирования», предварительной обработки документов, при котором создается вспомогательный файл, «индекс», призванный упростить и ускорить поиск. Это алгоритмы *инвертированных файлов, суффиксных деревьев, сигнатур*.

Широко используют поисковый алгоритм, основанный на получении инвертированного файла – упорядоченного исчерпывающего списка искомых слов из одного текста. Инвертированный файл (inverted file, инверсный файл, инвертированный индекс,

инвертированный список) – индекс поисковой системы, в котором перечислены слова коллекции документов, а для каждого слова перечислены все места, в которых оно встретилось. Поисковый алгоритм состоит в отыскании нужного слова и загрузке в память уже развернутого списка позиций

Как правило, инвертированный файл сжимают, например указывают только номер документа и число употреблений этого слова в нем. Именно такая упрощенная структура считается основной в классической теории информационного поиска. Второй способ сжатия инвертированного файла предполагает упорядочить позиции для каждого слова по возрастанию адресов и для каждой позиции хранить не полный ее адрес, а разницу от предыдущего. В результате применения описанных методов размер инвертированного файла, как правило, составляет от 7 до 30 процентов от размера исходного текста в зависимости от подробности адресации.

Используют другие, отличные от инвертированного и прямого поиска алгоритмы и структуры данных. Это суффиксные деревья, а также сигнатуры. Метод сигнатур – представляет собой преобразование документа к поблочным таблицам *хеш-значений* его слов, что называют «сигнатурой» и последовательному просмотру «сигнатур» во время поиска.

Сигнатура (signature, подпись) – множество *хеш-значений* слов некоторого блока текста. При поиске по методу сигнатур все сигнатуры всех блоков коллекции просматриваются последовательно в поисках совпадений с *хеш-значениями* слов запроса.

Метод суффиксных деревьев (*suffix trees, suffix arrays, PAT-arrays*) основан на представлении всех значимых суффиксов текста в структуре данных, известной в теории иерархических моделей как лес (trie). Суффиксом в этом индексе называют любую «подстроку», начинающуюся с некоторой позиции текста (текст рассматривается как одна непрерывная строка) и продолжающуюся до его конца. В реальных приложениях длина суффиксов ограничена, а индексируются только значимые позиции – например, начала слов. Этот индекс позволяет выполнять более сложные запросы, чем индекс, построенный на инвертированных файлах

При информационном поиске используют понятие модели поиска. *Модель поиска* – это модель, на основании которой создается оценочная формула, позволяющая ИПС принять решение: какой документ считать найденным и как его ранжировать. Модели информационного поиска принято делить на три вида: теоретико-множественные (булевская, нечетких множеств, расширенная булевская); алгебраические (векторная, обобщенная векторная, латентно-семантическая, нейросетевая) и вероятностные.

Булевская модель (boolean, булева, двоичная) – модель поиска, опирающаяся на операции пересечения, объединения и вычитания множеств. В этой модели наличие искомого слова означает «Да» (документ считается найденным) отсутствие искомого слова означает «Нет» – документ не найден. Недостаток булевской модели состоит в ее непригодности для ранжирования.

Векторная модель – модель информационного поиска, рассматривающая документы и запросы как векторы в пространстве слов, а *релевантность* – как расстояние между ними. Ранжирование в этой модели основано на статистическом наблюдении. По результатам наблюдения получают оценку того, что чем больше частота термина в документе (*TF*) и больше *обратная встречаемость* термина в коллекции (*IDF*), тем выше вес данного документа по отношению к термину. Обозначение $TF*IDF$ широко используется как синоним векторной модели. $TF*IDF$ – численная мера соответствия слова и документа в векторной модели; она тем больше, чем относительно чаще слово встретилось в документе и относительно реже в коллекции

Вероятностная модель – модель информационного поиска, рассматривающая релевантность как вероятность соответствия данного документа запросу на основании вероятностей соответствия слов данного документа идеальному ответу. Релевантность в этой модели рассматривается как вероятность того, что данный документ может ока-

заться нужным пользователю. При этом подразумевается наличие уже существующего первоначального набора релевантных документов, выбранных пользователем или полученных автоматически при каком-нибудь упрощенном предположении.

Вероятность оказаться релевантным для каждого следующего документа рассчитывается на основании соотношения встречаемости терминов в релевантном наборе и в остальной, «нерелевантной» части коллекции [8]. Вероятностные модели обладают некоторым преимуществом, поскольку они ранжируют результаты поиска – располагают документы в порядке убывания «вероятности оказаться релевантным»

Существуют модели поиска, анализирующие смысл текста, например модель латентно-семантического индексирования (выявления скрытого смысла) [10]. Эта алгебраическая модель основана на сингулярном разложении прямоугольной матрицы, ассоциирующей слова с документами. Элементом матрицы является частотная характеристика, отражающая степень связи слова и документа, например $TF*IDF$ [11].

Существует класс алгоритмов, относимых к лингвистическим. Точную границу между статистическими и лингвистическими методами провести сложно [8]. Условно к лингвистическим методам относят методы, опирающиеся на словари (морфологические, синтаксические, семантические), созданные человеком. Многие алгоритмы обработки текстов требуют хотя бы минимального уровня лингвистической обработки.

Лингвистическими методами решаются следующие задачи:

- автоматическое определение языка документа;
- токенизация (графематический анализ): выделение слов, границ предложений;
- исключение неинформативных слов (стоп-слов);
- лемматизация (нормализация, стемминг): приведение словоизменительных форм к «словарной». В том числе и для слов, не входящих в словарь системы;
- разделение сложных слов (компаундов) для некоторых языков (например, немецкого);
- выделение именных групп.

Еще реже в исследованиях и на практике можно встретить алгоритмы словообразовательного, синтаксического и даже семантического анализа. При этом под семантическим анализом чаще подразумевают какой-нибудь статистический алгоритм (LSI, нейронные сети), а если толково-комбинаторные или семантические словари и используются, то в крайне узких предметных областях.

Оценка качества поиска. В результате поиска выявляется некая информация – I_p , которая может в большей или меньшей степени удовлетворять исследователя. В зависимости от соответствия между целью поиска I_n и результатом поиска I_p возможны разные ситуации, характеризующиеся понятием релевантности (рис. 1). Возможны три типа соответствия между результатом поиска и запросом: формальная релевантность, онтологическая релевантность, пертинентность [9, 12].

Формальной релевантностью называют соответствие поискового образа поисковому предписанию по формальным признакам. По этим признакам осуществляется отбор запросов и выдача результатов поиска. Формальная релевантность, как правило, далека от того, что хочет получить исследователь. *Онтологическая релевантность* – соответствие поискового образа поисковому предписанию по смысловым признакам. Она предусматривает сравнение запроса I_n и результата поиска I_p на семантическом (смысловом) уровне. В частности, при документальном поиске сравнение происходит на естественном языке. Смысловая релевантность в меньшей степени ориентирована на формальные признаки, а больше основана на когнитивном анализе [12].

Информация, имеющая смысловую релевантность, формируется на основе семантической операции, поэтому она несет в себе долю неопределенности. Критерий смыслового соответствия формируется человеком и устанавливает соответствие между выдаваемой информацией и смыслом запроса.

Это ставит перед исследователем дополнительную задачу – точно определить, релевантен или не релевантен результат поиска. Такая задача решается когнитивными методами [13–15]. Формальная и смысловая релевантность могут не удовлетворять информационные потребности пользователя. Наиболее полным критерием соответствия результатов поиска запросу считается пертинентность. *Пертинентностью* называют полное соответствие найденных знаний или данных информационным потребностям, устанавливаемое при информационном поиске.

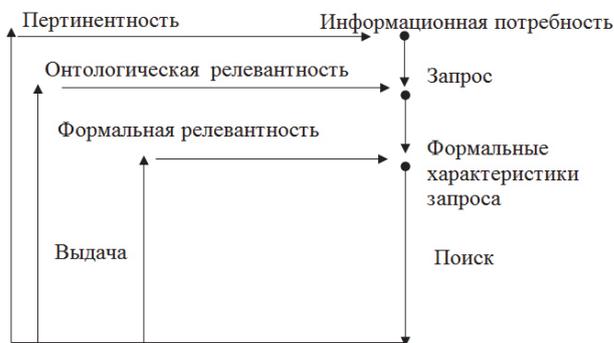


Рис. 1. Соотношение между пертинентностью, онтологической и формальной релевантностью

Следует отметить, что результат поиска определяется не только правильно построенным запросом, но наличием информации о том, что необходимо искать. Как правило, в результате поиска выдается большой объем информации, которая обладает формальной релевантностью. Эта информация анализируется и, если необходимо, в ней проводится уточняющий поиск. Этот анализ и дополнительный поиск уменьшают объем первоначально

полученных данных и создают поле смысловой релевантности. Таким образом парадигматическая [16] цепочка:

Исходные данные → формально релевантные данные → данные с онтологической релевантностью → пертинентные данные.

Реализуется в технологической последовательности

ИПС → оператор → эксперт.

Результат информационного поиска целесообразно оценивать для того, чтобы сравнивать разные поисковые технологии и системы (рис. 2).

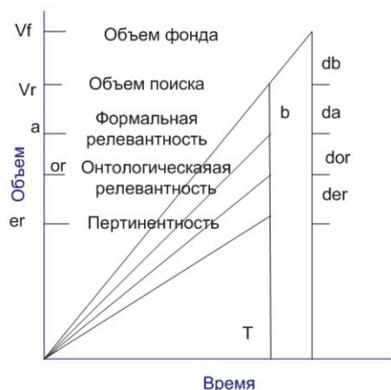


Рис. 2. Характеристики поиска

На рис. 2 имеются следующие обозначения. T – общее время поиска. $Vr = a + b$ – объем поиска. $Vf = a + da + b + db$ – объем фонда, в котором выполнен поиск. $Vf = Vr + dV$ – объем фонда. $dV = da + db$ – часть объема фонда, не использованная при поиске. Объем a – объем формально релевантной информации. Объем or – объем онтологически релевантной информации. Объем er – объем пертинентной информации в результатах поиска. Свод этих характеристик приводится в табл. 1.

Таблица 1

Состав документов, имеющихся в информационном массиве и выданных в результате поиска

Выдача	Формально релевантные	Онтологически релевантные	Пертинентные	Нерелевантные	Всего
Выдано	a			b	$a + b$
		or			
			er		
Не выдано	da	dor	der	db	
Всего	$a + da$	$or + dor$	$er + der$	$b + db$	$a + b + da + db$

Приращения da ; dor ; der ; db – остатки соответствующих характеристик в фонде,

которые не попали в результат поиска. Можно ввести следующие оценки результатов информационного поиска:

Полнота поиска по формальной релевантности: $Pf = a/(a + da)$.

Полнота поиска по онтологической релевантности: $Po = a/(a + da)$.

Полнота поиска по пертинентности: $Pp = a/(a + da)$.

Коэффициент релевантности поиска: $Kpn = a/(a + b)$.

Коэффициент релевантности фонда: $Kpf = (a + da)/Vf$.

Коэффициент онтологичности поиска: $Kon = or/Vr$.

Коэффициент пертинентности поиска: $Knn = er/Vr$.

Скорость поиска за время T : $V_T = Vr/T$.

Эффективность поиска за время T : $Эn = Vr/Vf$.

Эффективность релевантности поиска: $Эpn = a/(a + da)$.

Эти показатели дают возможность оценивать динамические характеристики поиска и качественно оценивать результаты поиска. В отличие от многих оценок информационного поиска по одной релевантности, в данной схеме выделено три типа релевантности, которые дифференцированно позволяют оценить результаты поиска.

Кроме того, введенные характеристики дают возможность оценивать эффективность разных ИПС и проводить сопоставительный анализ различных поисковых систем.

Эффективность информационного поиска измеряется совокупностью разных показателей, в том числе технической и экономической эффективностью. Техническая эффективность информационно-поисковой системы или технологии определяется как мера выполнения функции поиска. Экономическая эффективность поиска оценивается по стоимости выполнения этих функций. Стоимостные факторы могут изменяться с течением времени и регулироваться самим потребителем.

Техническая эффективность зависит от ряда факторов, которые могут быть сведены к двум группам:

1. Объемно-временные характеристики: они включают объем фонда информационных массивов, объем выдачи, время поиска.
2. Группа оценки полноты и точности поиска включает коэффициенты полноты (P), релевантности (K) и эффективности ($Э$).

Анализ является необходимым компонентом информационного поиска, поскольку на его основе принимается решение о завершении или продолжении поиска. Существуют специальные задачи информационного поиска, решение которых позволяет расширять процесс поиска.

Информационный поиск может иметь многоаспектное представление и не сводится к простому просмотру и анализу содержимого какого-то массива. Например, аудит (включая информационный аудит) является разновидностью информационного поиска, в ходе которого осуществляется *информационный поиск соответствий* и *несоответствий* нормативным документам. Технология поиска включает сравнение двух множеств. Множества нормативной документации и информационного множества, описывающего реализацию некой технологии или совокупности практических действий, которое должно соответствовать этой нормативной документации.

Другой пример. Желание найти оптимальное решение можно рассматривать как информационный поиск альтернативы на информационном множестве альтернатив в соответствии с заданным критерием оптимальности.

Еще один пример. Возникает новая задача (технологическая, экономическая, организационная, управленческая, математическая, проектная). Для решения новой задачи можно *создать решение* или *найти известные решения* применительно к новой задаче на множестве известных решений известных задач. Технология поиска включает сравнение двух множеств: множеств условий решения известных задач и множества

условий решения новой задачи, критерием поиска является максимально возможное совпадение условий решения новой задачи условиям решения известной задачи.

Еще один пример. Теоретико-игровой подход позволяет рассматривать задачу информационного поиска как устранение конфликтной ситуации. Пусть Y – информационная потребность, первоначальные данные – X_0 , результат поиска информации – X .

Конфликтные ситуации можно моделировать по-разному. Например, конфликтная ситуация возникает как конфликт между необходимыми данными Y (информационная потребность) и первоначальными данными X_0 , имеющимися в распоряжении пользователя. При отсутствии необходимых данных конфликтная ситуация в теоретико-множественном формализме отображается как $X_0 \cap Y = \emptyset$. Здесь \emptyset – пустое множество.

После завершения поиска также могут возникать конфликтные ситуации как ситуации несоответствия результата поиска информационным потребностям. При формальной и смысловой релевантности или при наличии ошибок в запросе результат поиска X не будет соответствовать в полной мере Y . Это в формальном виде отразится как $X_0 \subseteq Y$.

Технология поиска включает поиск такого множества X , для которого конфликт между X и Y был бы минимальным.

Другой конфликт обусловлен возможными ошибками. Поскольку информационный поиск проводится в некоем ограниченном объеме данных, то необходимо учитывать ошибки двух видов: ошибки выборки и ошибки наблюдения. Чем меньше выборка, тем меньше может быть ошибка наблюдения, но больше ошибка выборки (неадекватность выборки рассматриваемой совокупности). Чем больше выборка, тем больше нагрузка на лицо, анализирующее информацию, и тем больше ошибка наблюдения.

Кроме того, поиск как технологическая операция требует затрат. Чем длительнее поиск, тем больше вероятность нахождения нужной информации. Но чем длительнее поиск, тем больше затраты. Конфликт обусловлен противоречием между затратами на поиск и необходимостью полного анализа информации. Такая ситуация называется транзакционными издержками поиска [17]. Для минимизации издержек необходим анализ рисков. Ценность информации при ее поиске может определяться разными факторами, в том числе и затратами на ее получение, однако с течением времени цена информации как товара может меняться скачком.

Моделирование при организации информационного поиска. Обобщенные модели информационного поиска имеют вид информационных конструкций [18, 19]. Информационные конструкции описывают паттерн и концепт, а также позволяют осуществлять интерпретацию в информационном поле [20]. Сложность моделирования при информационном поиске в том, что приходится моделировать качественно разные сущности: процессы, объекты и ситуации [21]. Процесс информационного моделирования при информационном поиске имеет двойственность. С одной стороны, для оптимизации результата поиска надо моделировать паттерны как описательные модели. С другой стороны, для оптимизации процесса поиска надо моделировать процессы поиска. Это приводит к необходимости построения процессуальных моделей. Таким образом, моделирование в информационном поиске требует применения дескриптивных и прескриптивных моделей [22]. Разнообразие моделей по качественному и количественному признакам ставит задачу их систематизации [23]. Основой построения современных моделей и информационных конструкций являются информационные единицы [24, 25]. При этом в области информационного поиска применяют специальные поисковые информационные единицы [26].

Заключение. Современные технологии информационного поиска являются комплексными технологиями, включающими информационное и когнитивное моделирование, а также применение информационных единиц. Современные технологии информационного поиска включают многоуровневую оценку релевантности, что усложняет

оценку результатов поиска. Современные технологии информационного поиска требуют формирования прескриптивных и дескриптивных моделей в единой системе моделей. Технологии информационного поиска решают задачи снятия информационной неопределенности, уменьшения информационной асимметрии и преодоления семантического разрыва. По этой причине они интегрированы в систему информационных технологий, связанную с обработкой информации и управлением.

Литература

1. *Цветков В. Я.* Информационная неопределенность и определенность в науках об информации // Информационные технологии. 2015. № 1. С. 3–7.
2. *Васютинская С. Ю.* Информационная асимметрия в образовательных технологиях // Образовательные ресурсы и технологии. 2016. № 4 (16). С. 14–20.
3. *Ion A. L.* Algorithms for reducing the semantic gap in image retrieval systems // 2009 2nd Conference on Human System Interactions. – IEEE, 2009. С. 97–102.
4. *Tsvetkov V. Ya.* Cognitive Science of Information Retrieval // European Journal of Psychological Studies. 2015. Vol. 5. Iss. 1. P. 37–44.
5. *Ожерельева Т. А.* Информационная ситуация как инструмент управления // Славянский форум. 2016. № 4 (14). С. 176–181.
6. *Розенберг И. Н.* Онтологический подход в геоинформатике // Образовательные ресурсы и технологии. 2016. № 5 (17). С. 86–95.
7. *Романов В. П.* Теоретические основы информатики. Информационные структуры и фактографический поиск информации. – М., РЭА им. Г. В. Плеханова, 1996. 190 с.
8. *Болбаков Р. Г.* Развитие и применение когнитивно-семантических методов и алгоритмов в мультимедийных образовательных порталных системах: Дис. ... канд. техн. наук. – М.: МИРЭА, 2013. 184 с.
9. *Поляков А. А., Цветков В. Я.* Прикладная информатика / под общ. ред. А.Н. Тихонова. Т. 1. – М.: МАКС Пресс, 2008. 788 с.
10. *Куликова А. А.* Причинность в моделях латентно-структурного анализа и структурных уравнений // Социология: методология, методы, математическое моделирование. 2009. № 29. С. 30–45.
11. *Хенингер М.* Эффективные стратегии поиска в Интернет / пер. с англ. – М.: Центр-пресс, 1998. 160 с.
12. *Шемакин Ю. И.* Теоретическая информатика / под общей ред. К. И. Курбакова. – М.: Рос. экон. акад., 1998. 132 с.
13. *Номоконов И. Б.* Когнитивные методы при лучевой диагностике: монография. – М.: МАКС Пресс, 2016. 60 с.
14. *Цветков В. Я.* Когнитивные аспекты построения виртуальных образовательных моделей // Перспективы науки и образования. 2013. № 3. С. 38–46.
15. *Болбаков Р. Г.* Когнитивные методы оценки качества образования // Образовательные ресурсы и технологии. 2016. № 1 (13). С. 34–39.
16. *Цветков В. Я.* Триада как интерпретирующая система // Перспективы науки и образования. 2015. № 6. С. 18–23.
17. *Розенберг И. Н., Цветков В. Я.* Информационные транзакционные затраты // Международный журнал прикладных и фундаментальных исследований. 2010. № 12. С. 160–161.
18. *Tsvetkov V. Ya.* Information Constructions // European Journal of Technology and Design. 2014. Vol. 5. Iss. 3. P. 147–152.
19. *Дешко И. П.* Информационное конструирование: монография. – М.: МАКС Пресс, 2016. 64 с.
20. *Чехарин Е. Е.* Интерпретация информационных конструкций // Перспективы науки и образования. 2014. № 6. С. 37–40.
21. *Цветков В. Я.* Информационные модели объектов, процессов и ситуаций // Дистанционное и виртуальное обучение. 2014. № 5. С. 4–11.
22. *Цветков В. Я.* Дескриптивные и прескриптивные информационные модели // Дистанционное и виртуальное обучение. 2015. № 7. С. 48–54.
23. *Стоева Д. Р.* Систематизация информационных моделей // Перспективы науки и образования. 2015. № 4. С. 13–18.
24. *Павлов А. И.* Информационные модели и информационные единицы // Перспективы

науки и образования. 2015. № 6. С. 12–17.

25. Rozenberg I. N. Information Construction and Information Units in the Management of Transport Systems // European Journal of Technology and Design. 2016. Vol. 12. Iss. 2. P. 54–62.

26. Tajima K., Hatano K., Matsukura T., Sano R., Tanaka K. Discovery and Retrieval of Logistical Information Units in Web // WOWS. 1999. August. P. 13–23.

Complex information retrieval

Igor' Naumovich Rozenberg, Professor, Doctor of Technical Sciences, Deputy general director of the Research Institute of automated systems in railway transport JSC NIIAS – HEAD OFFICE

The article reveals the contents modern technologies of information retrieval. The article introduces a new concept of evaluation of information retrieval. Article performs differentiation concept of relevance. The article introduces the new features evaluation of search results, including the evaluation of a variety of relevance. The article substantiates the need for information modeling in the organization of information retrieval. The article argues that information modeling for information retrieval should be systematic and include descriptive and prescriptive models.

Keywords: information technology, information retrieval, pattern, concept, information units.

УДК 612.833, 001.6

НЕЯВНЫЕ ЗНАНИЯ В МЕДИЦИНСКОЙ ДИАГНОСТИКЕ

Ольга Юрьевна Номоконова, врач терапевт пульмонологического отделения;
врач аллерголог-иммунолог консультативно-диагностической поликлиники,
e-mail: nomokol@bk.ru,

ГБУЗ Иркутская областная клиническая больница,
<http://www.crbirk.ru/employees>

Исследованы неявные знания как объективный фактор при медицинской диагностике. Статья раскрывает содержание современной медицинской диагностики. Статья доказывает, что в одних случаях неявные знания препятствуют диагностике и требуют их устранения. В других случаях неявные знания являются дополнительным ресурсом медицинской диагностики. Описан процесс диагностики как сложная динамическая система. Раскрыто содержание двух качественных типов диагностических задач.

Ключевые слова: медицинская диагностика; неявные знания; информационные технологии; репродуктивные диагностические задачи; научные диагностические задачи.

Введение

В разнообразных областях деятельности существуют задачи, в которых по исходным данным о некоторой системе необходимо определить: (1) комплекс происшедших в ней изменений, (2) выявить причины изменений, (3) выявить направленность этих изменений [4], (4) разработать механизм снижения отрицательных факторов, обусловленных этими изменениями. Задачи (1) называют фактоустанавливающими [1]. Задачи (1) и (2) называют задачами импакт-анализа [2, 3]. Задачи (1)–(3) называют диагностическими [4–6]. Задачи (4) называют терапевтическими. Диагностика может быть разной: технической, компьютерной, связанной с информационной безопасностью, медицинской. В медицине связанные задачи (1)–(4) определяют диагностику и терапию. Для задач (1) характерно наличие неявного знания и информационной асимметрии [7] между первичной информацией и необходимой для постановки диагноза. Такая информационная асимметрия преодолевается в ходе информационного взаимодействия [8] с объектом диагностики. Информационная асимметрия также преодолевается за счет извлечения неявных знаний [9] и трансформации их в явные знания [10]. Такая информационная асимметрия также преодолевается применением аналитических ис-



О.Ю. Номоконова