

Министерство образования Республики Беларусь
Учреждение образования
Белорусский государственный университет
информатики и радиоэлектроники

УДК 004.6 + 004.8

Чеушев
Кирилл Васильевич

Модели и алгоритмы семантической навигации
в больших массивах данных

АВТОРЕФЕРАТ

на соискание степени магистра технических наук

по специальности 1-40 80 05 – Математическое и программное обеспечение
вычислительных машин, комплексов и компьютерных сетей

Научный руководитель
Глухова Л.А.
к.т.н., доцент

Минск 2018

КРАТКОЕ ВВЕДЕНИЕ

С момента создания первых средств вычислительной техники развивается идея наделить их способностью не только быстро выполнять рутинные операции, но и реализовывать функции, которые относят к семантическим уровням высшей нервной деятельности, интеллекту человека. Сегодня можно говорить, что завершился первый полувековой цикл развития этого научного и практического направления – интеллектуализации компьютерных систем – и набирает силу вторая волна популярности исследований, разработок и практического применения их результатов.

Интеллектуальные информационные системы крайне востребованы практически во всех областях жизнедеятельности. Они сейчас обеспечены несопоставимо более мощными ресурсами. И несмотря на то, что многие апробированные в «первой волне» подходы оказались ошибочными, в целом верной остаётся главная идея – реализовывать информационные системы «по образу и подобию» нервной системы животных и человека, воплощая в аппаратуре и программах современные знания об архитектонике и когнитивной функции мозга.

Прагматичная практика «второй волны» стремится решать актуальные задачи в самых различных областях и для этого строит интеллектуальные информационные системы, соединяющие в себе множество удачных решений различных направлений исследований, стремясь получить от них не только частные эффекты, но и эффекты кумулятивные, системные.

Этому же подходу стремится следовать и автор в разработке и реализации подхода «семантическая навигация в больших информационных массивах».

Источниками заимствования идей служат нейробиология в целом и особенно Системная психофизиология, развиваемая на базе Теории функциональных систем, физиология сенсорных систем, синергетика. Идея состоит не в том, чтобы попытаться воспроизвести в модели какую-то архитектуру или алгоритм работы живого мозга, а в том, чтобы следовать более общим принципам, воспроизводя их «по образу и подобию», пользуясь аналогиями, метафорами, а главное – стремясь уловить и воспроизвести в технической реализации системное единство.

Центральная исходная посылка заключается в том, что на основе (а) анализа и обобщения накопленного опыта и (б) комплекса «заимствования идей» из нейробиологии и системной психофизиологии можно достичь такого уровня унификации решений и создать такой набор «заготовок» на всех необходимых уровнях, что в практике создания ИИС появится возможность не решать всякий раз творческую задачу разработки того или иного алгоритма из класса операций семантической навигации, а применять решения данного комплекса, настраивая их на конкретные условия.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Главная **цель** диссертационной работы состоит в разработке целостного системного комплекса методических и алгоритмических решений, составляющих необходимую основу для реализации в современных интеллектуальных информационных системах широкого класса функциональных возможностей, кратко именуемых «семантической навигацией».

Для достижения поставленной цели в работе решаются следующие **основные задачи**:

- проводится анализ предметной области с опорой на потребности современных интеллектуальных информационных систем и акцентом на современные знания биологического характера в области, выбранной автором в качестве теоретической основы для «биологических заимствований» (Глава 1);
- разрабатывается общая концепция, модель и методика семантической навигации в информационных массивах (Глава 2);
- разрабатываются составляющие элементы операциональной триады структурирование-экстрагирование-поиск (Главы 2 и 3)
- разрабатываются необходимые метрики и алгоритмы (Главы 2 и 3);
- разрабатываются и экспериментально исследуются технологические цепочки из созданных элементов для решения практических задач (Глава 3).

Тема диссертации **актуальна**, а её результаты **востребованы практикой** проектирования современных информационных систем, поскольку широкий класс функциональных возможностей, которые в данной работе для краткости обобщаются как «семантическая навигация», сегодня не только составляет ключевую функциональность многих ИС, которые уместно классифицировать как Интеллектуальные информационные системы, но и широко применяются в самых разнообразных информационных системах различного назначения. Например, это компоненты обеспечения «дружественного пользовательского интерфейса», такие как именуемые в англоязычной литературе функции поддержки действий пользователя «as you type» и «instant search», компоненты подбора аналогичных товаров в системах анализа продаж конкурентов, компоненты персонализации рекламных и других предложений, рекомендательные функции, функции навигации по контенту в сервисах книг, фильмов и других ИР и многие другие функции, для реализации которых необходимо в больших массивах данных (а) подобрать и ранжировать так называемые «семантические детерминанты», а именно характерные и/или специфичные в заданном контексте информационные единицы (например, теги) и/или (б) подобрать семантически подобные объекты (скажем, книги, фильмы, товары и т.д.).

Главная гипотеза работы состоит в том, что:

(а) множество функций современных интеллектуальных информационных систем может быть реализована путем построения иерархической системы из трёх операциональных элементов:

- экстрагирования семантических детерминантов,
- структурирования дисперсной среды ИР,
- поиска и подбора подобных сущностей,

причём так, что на более высоком иерархическом уровне элемент триады представляет собой композицию двух других элементов предыдущего уровня;

(б) для этих функциональных операций могут быть найдены типовые решения – модели, методики синтеза алгоритмов, основные метрики, – допускающие тиражирование в типовом виде для получения работающих в первом приближении компонентов семантической навигации с последующей настройкой на основе обратной связи.

Эта ключевая гипотеза базируется на ряде заимствований автором идей из Теории функциональных систем, Системной психофизиологии, нейробиологии, синергетики. Её проработка, описываемая в диссертационной работе, комплекс экспериментальных исследований и практическая апробация в коммерческом продукте показали правильность исходных посылок, продуктивность выбранного пути достижения поставленной цели и практическую значимость полученных результатов.

Все новые результаты, излагаемые в диссертационной работе, получены автором самостоятельно.

Основные результаты исследования докладывались на следующих конференциях:

52-й научная конференция аспирантов, магистрантов и студентов. (Минск, 25 - 30 апреля 2016 года);

X Международная научно-методическая конференция (Минск, 7–8 декабря 2017 года). Дистанционное обучение – образовательная среда XXI века.

Структура и объём диссертации. Диссертационная работа состоит из введения, общей характеристики работы, трёх глав, заключения, списка использованных источников, приложения. Диссертация изложена на 122 страницах машинописного текста и содержит 9 рисунков, расположенных на 8 страницах, 3 таблицы, занимающие 3 страницы, список литературы из 103 наименований, размещённый на 7 страницах, 2-х приложений, занимающих 23 страницы.

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **введении** определена предметная область работы, отражены ключевые тенденции её сегодняшнего развития, показана актуальность диссертационной работы, обозначена её практическая ценность.

В **первой главе** проводится анализ состояния и тенденций развития интеллектуальных информационных технологий и систем как предметной области работы. Отмечается, что операции семантической навигации применяются сегодня не только в ряде систем, где такие функции составляют основу функциональности системы в целом, но и во множестве информационных систем самого различного назначения.

Примерами операций семантической навигации в современных ИС могут служить: подбор ключевых слов, фраз, тегов к различным информационным блокам (текстам, тематическим рубрикам, сюжетам и т.д.), кластеризация, классификация и предшествующее её обучение, построение дайджестов, в том числе кластеризованных, построение аннотаций и так называемых сниппетов (контекстных аннотаций), машинное обучение как процесс формирования профилей тех или иных сущностей, поиск похожих текстов, рубрик, сюжетов, тегов, персонализация рекламных предложений, персональные рекомендации в различных интернет-сервисах и многое другое, что можно обобщённо определить как семантически направляемые (на основе анализа больших массивов данных) последовательные действия, ведущие к определённым целям и результатам, в том числе априори неизвестным или слабо определённым, и корректируемым непосредственно в процессе навигации.

Вместе с тем, анализ литературы не выявил источников, в которых столь схожие по внешним целям и внутренней сущности операции рассматривались бы в целостном комплексе, с обобщениями, едиными подходами.

Вторая глава посвящена разработке **модели семантической навигации** в целом – в виде так называемой *операциональной триады: экстрагирования семантических детерминантов, структурирования дисперсной среды информационного ресурса, поиска подобных сущностей* – и моделей компонентов этой триады.

На множестве примеров практически значимых цепочек семантической навигации показано, что все рассмотренные сложные схемы достижения цели посредством семантической навигации сводимы к иерархически-последовательному развёртыванию во времени и по уровням абстрагирования перечисленных операций триады, причём так, что последовательность пары из них порождает третий элемент триады на более высоком иерархическом уровне, и он становится первым элементом последовательности уже этого более высокого уровня.

Данный результат создаёт гипотезу и основу для её проработки, что коль скоро множество внешне отличающихся по конкретике целей и достигаемому результату операций семантической навигации внутренне сводимо к трём основным типам, то для этих типов уместен и поиск типовых решений, которые как минимум создадут базис для первоначальной реализации с тем, чтобы развивать её, воспользовавшись обратной связью.

Такой путь синтеза решений для практической разработки (первоначальная сборка из типовых заготовок и последующее эволюционное развитие с обратной связью и планируемым результатом в качестве системообразующего фактора) в целом соответствует концепции системогенеза, как она проработана в современном развитии Теории функциональных систем.

Следующим шагом от сформулированных обобщений и гипотезы стала методика синтеза алгоритмов семантической навигации.

Разработана и реализована в ряде практических задач **методика синтеза алгоритмов семантической навигации**, «биологически инспирированная» метафорой «дивергентно-конвергентной нейрональной импульсации» в мозгу млекопитающего (млекопитающего потому, что только у этого класса живых существ в эволюции появилась так называемая новая кора головного мозга, имеющая регулярную структуру и отвечающая за высшие мыслительные функции, в том числе научение; биологические заимствования проводятся в основном из результатов исследования учёными этой «новой коры» – неокортекса).

Ключевая идея методики состоит в том, что при необходимости разработать новый алгоритм для некоторой требуемой операции класса семантической навигации:

– потребность в результате, который должен выдавать алгоритм, формулируется из потребности схемы семантической навигации более высокого уровня (той схемы, в которую должна вписаться проектируемая операция) непосредственно в метафорических терминах «дивергентных и конвергентных волн активации данных»; во многих случаях (за исключением, может быть, широко известных операций семантической навигации наподобие классификации, чьи смысловые формулировки уже известны из многолетнего опыта) такую формулировку получить намного быстрее и проще, чем объяснительную формулировку классического вида;

– практическая реализация (допустим, что алгоритм реализуется SQL-процедурой в БД) метафор «волн активации» проста: метафора дивергентной волны, при которой один сигнал активирует множество других (дивергенция), реализуется в SQL-запросе конструкциями <join> или <... where Id in(...)...>; метафора конвергентной волны – частью SQL-запроса с конструкцией <group by>, которая сводит активированные сигналы в заданные группы, резко уменьшая их число (конвергенция), вычисляет характеристики, необходимые метри-

кам, ранжирует посредством вычислений метрик список полученных таким образом результатов (например, ключевых слов рубрики) и с помощью порогов отбирает в следующий каскад навигационной схемы только ту часть кандидатов, которая признана годной.

Практика проектирования большого числа цепей и процессов семантической навигации показала, что все без исключения (из числа тех нескольких десятков, которые потребовалось реализовать в проектируемых системах исходя из их целей и задач) можно синтезировать с помощью описываемой методики, причём проще, быстрее и эффективнее, чем способом «изобретения решения под конкретную ситуацию». В некоторых случаях алгоритм операции, которую требовала цепочка семантической навигации более высокого уровня, удавалось синтезировать именно этой методикой, и только уже после реализации и наблюдения результатов сформулировать «классифицирующий термин» тому, какой результат получает синтезированный алгоритм, тогда как конструктивное описание схемы алгоритма при следовании предлагаемой методике содержалось уже в формулировке требований к его результату.

На основе данной методики разработаны все модели семантической навигации, необходимые для реализации функциональности практической системы, выбранной в качестве экспериментальной платформы для апробации и исследования разрабатываемых в диссертационной работе решений.

Следующим шагом на пути к цели проектировщика – созданию работающих в конкретных системах функциональных элементов семантической навигации – стала гипотеза о возможности нахождения типовых структур основных метрик для элементов операциональной триады. Сформирован набор типовых структур метрик и самих метрик для операций семантической навигации.

Источниками послужили:

– наблюдение результатов прошлого этапа работы, которые показали, что число различных типов каскадов в упоминаемых схемах операций оказалось невелико, несмотря на большое число самих схем;

– сформированное из осмысления ряда биологических данных предположение о том, что количество принципиально разных схем принятия решения в мозгу (разумеется, имея в виду только рассматриваемую здесь узкую область) не может быть слишком велико и, вполне возможно, мозг в задачах семантической навигации задействует небольшое число типовых «эволюционных заготовок», разнообразя их иерархически-последовательным комбинированием и системной специализацией (финальной «настройкой») изначально преспециализированных структур; также к числу биологических заимствований следует отнести использование логарифмических зависимостей в качестве базовых элементов метрик (основанное на представлениях о том, что интенсивность реакции нейронов, как минимум сенсорных систем, зависит от интенсивности вход-

ной стимуляции логарифмически) и представления о доминанте, о котором подробнее говорится тремя абзацами ниже в более точном контексте;

– прежний опыт в реализации подобных метрик, в том числе и неразрешённых затруднений, затормозивших выход на следующий качественный уровень и уровень сложности; в частности:

– сделан вывод о непродуктивности с позиций последующего развития системы в целом придавать метрикам такой вид, при котором их значения отражают тот или иной «физический смысл», точнее не смысл как таковой, а смысл специальный для конкретной ситуации (например, значение метрики релевантности текста запросу в системе новостной аналитики было приблизительно пропорционально числу среагировавших на запрос термов текста; это свойство метрики было плюсом для локальных целей, но стало большим минусом для целей развития системы); посылка заменена на ту, что возможность простого осмысления значения метрики является несомненным достоинством и крайне важно для метода эволюционного приближения к результату, но отражаемый метрикой смысл должен быть более общим и служить целям системы в целом, а не локального участка;

– одним из негативных следствий наделения частных метрик свойством, описанным в предыдущем абзаце, стала трудность выстраивания из таких частных метрик общей интегрированной метрики в ситуациях, когда решение нужно принимать по группе источников информации, а не по единственному источнику, и эти источники могут иметь различающиеся по диапазонам значений метрики;

– по наблюдениям экспериментальных данных, большим недостатком традиционных линейного вида метрик, применяемых в алгоритмах операций семантической навигации, является в той или иной форме усреднение сигналов, приводящее к тому, что сильный сигнал будет «заглушен» совокупностью слабых, особенно если число последних велико; в итоге, например, при подборе семантических детерминантов терм, у которого сработало большое число слабых сигналов («откровенно слабый» по смыслу), может получить в списке кандидатов существенно более высокий ранг, чем отлично семантически подходящий терм, для которого в метрике несколько сигналов были сильны, но остальные слабы (последнее может быть вызвано в том числе и техническими причинами, скажем, шумом или недостатком данных). В преодолении этого затруднения большую роль оказало «биологическое заимствование» представлений из учения А.А. Ухтомского о *доминанте*; принцип нелинейного усиления доминантного сигнала в мозгу остальными нейроимпульсациями стал основой построения нелинейного вида метрик для компонентов семантической навигации.

На основе изложенных соображений внесён ряд инноваций в метрики, используемые в операциях семантической навигации, в результате чего они приобрели более универсальный характер и структурный вид, а описанные выше недостатки были устранены.

Для интеграции информации от разных источников в общий рейтинг предложена **нелинейная метрика с доминантным сигналом** (или несколькими сигналами) и её модулированием значениями остальных сигналов.

Изначально метрика предполагалась как дополнение к традиционному способу интегрирования сигналов (взвешенная сумма и т.п.) в тех случаях, когда требуется выделить доминирующий сигнал. Однако экспериментальные исследования показали, что новая метрика показывает лучшие результаты во всех случаях, поэтому в конечном счёте было принято решение отказаться от использования линейных метрик, а предложенная структура выбрана в качестве базовой для практически всех случаев применения.

Третья глава посвящена разработке конкретных алгоритмов и метрик операций семантической навигации на примере выбранной для экспериментальных исследований практической системы – рекомендательной системы для авторов книг, желающих опубликовать их на интернет-сервисах.

Уточнена до конкретной реализации метрика интегрирования информации от разных источников, общая структура которой предложена в главе 2.

Разработана **метрика релевантности** и её компоненты.

Для операций семантического экстрагирования разработаны *две частные метрики*:

– так называемый «фактор попадания», по смыслу подобный Байесовскому критерию, но в логарифмической форме;

– так называемый «фактор специфичности», который по целям можно соотнести с классическим критерием TF-IDF;

и *объединяющая их интегральная* «метрика релевантности», в которой фактор специфичности нелинейно «модулирует» фактор попадания, усиливая или ослабляя его в зависимости от своего значения и регулировочного порога.

В процессе проведения экспериментальных исследований результаты, получаемые с помощью указанных метрик, сопоставлялись с результатами, полученными с помощью Байесовского критерия, классической метрики TF-IDF и их комбинациями. Эксперименты показали, что в большинстве случаев предлагаемые в работе метрики дают гораздо более адекватные потребности результаты и лишь в некоторых случаях комбинация традиционных критериев показала результаты, в целом схожие с результатами, полученными предлагаемой метрикой (это случай подбора ключевых термов тематическим категориям, где условия подбора благоприятные, в других условиях, скажем, при отборе устойчивых и семантически значимых в заданном контексте фраз ряд апробированных ком-

бинаций линейных метрик оказался бессильным уловить множество самых интересных и значимых словосочетаний, а предлагаемая метрика их не только находит, но и позиционирует на первые места ранжированного списка).

Важным для применения в практике проектирования свойством указанных метрик является их «автонормирование» (фактор попадания и метрика релевантности принимают значения в диапазоне $[0, 1]$, фактор специфичности автонормирован относительно 1) и «осмысленность значений»:

– фактор попадания и метрика релевантности отражают свой смысл ростом от 0 до 1, причём согласно «бионической модели», т.е. осмысление их человеком с психофизиологических позиций (по логарифмическому закону) представляется более адекватным, чем восприятие линейных частотных метрик;

– фактор специфичности имеет значение 1 при «нейтральной специфичности» измеряемой характеристики относительно контекста, значение, меньшее 1, интерпретируется как «объект неспецифичен для данного контекста», большее 1 – «объект специфичен для данного контекста», тем более специфичен, чем большее значение принимает эта метрика.

Экспериментальные исследования разработанной метрики релевантности показали, что она обеспечила существенно лучшие результаты, чем специализированные метрики, применявшиеся до этого в конкретных операциях, и может использоваться как основа для всех операций экстрагирования семантических детерминантов. Её несомненным преимуществом является отличный (и регулируемый) баланс подбора как высокоспецифичных для заданного контекста семантических детерминантов, так и высокохарактерных, что (соблюдение данного баланса) обычно является слабым местом традиционных метрик, вынуждая проектировщиков тратить большие усилия на подбор регулировочных параметров в каждой ситуации применения.

Несколько неожиданным явился экспериментально установленный факт, что метрика релевантности, предложенная для операций подбора семантических детерминантов, пригодна и отлично работает в операциях подбора семантически подобных сущностей, то есть в другом функциональном элементе операциональной триады. А если подбор подобных сущностей осуществляется по нескольким цепочкам семантической навигации, то применима разработанная в главе 2 метрика интегрирования нескольких сигналов.

С помощью этих двух метрик построены, в частности, так называемые «ассоциативные облака» всех информационных сущностей рекомендательной системы – тематических категорий, ключевых слов, поисковых фраз, книг.

Этот результат (нахождение метрик, работающих в широком диапазоне условий применения и более того – в операциях разного типа):

– даёт довод в пользу вывода о правильности выбранного пути исследования – обобщения широкого спектра результатов в области семантической навигации и поиска универсальных конструктивных элементов;

– имеет большую практическую значимость, поскольку существенно сокращает сроки и трудоёмкость проектирования компонентов семантической навигации, избавляя от необходимости скрупулёзного подбора метрик, адекватных каждой ситуации применения.

На основе разработанных в главе 2 методических решений и моделей семантической навигации, предложенных метрик создано несколько десятков алгоритмов и реализующих их SQL-процедур, полностью покрывающих потребности рекомендательной системы и реализующих функциональность рекомендаций (а) тематических категорий, (б) поисковых термов, (в) ключевых слов аннотаций книг, динамически формируемых в процессе и в темпе побуквенного ввода пользователем информации о своей книге (т.е. в реальном времени).

При внедрении разработанных алгоритмов и метрик в проектируемую систему проводились анализ поведения предложенных решений в различных ситуациях применения, а также сопоставительный анализ с известными метриками. Частично результаты этих сопоставлений проиллюстрированы таблицами Приложения А с их комментариями, помещёнными в соответствующих разделах Главы 3.

Приложение Б иллюстрирует на конкретных примерах действий пользователя рекомендательной системы работу разработанных алгоритмов семантической навигации.

ЗАКЛЮЧЕНИЕ

Главный результат диссертационной работы состоит в разработке целостного системного комплекса методических и алгоритмических решений, составляющих необходимую основу для реализации в современных интеллектуальных информационных системах широкого класса функциональных возможностей, кратко именуемых «семантической навигацией».

1. Разработана общая **модель семантической навигации**. На основе анализа и обобщения большого множества операций семантической навигации как с точки зрения потребностей в них и функциональных возможностей, так и с точки зрения технической реализации, выделена так называемая «**операциональная триада**» семантической навигации, а именно операции:

- *структурирования* дисперсной среды ИР;
- *экстрагирования семантических детерминантов*;
- собственно семантической навигации, в большинстве случаев сводимой к *поиску (и/или подбору) семантически похожих сущностей*.

Показано, что все рассмотренные сложные схемы достижения цели посредством семантической навигации сводимы к иерархически-последовательному развёртыванию во времени и по уровням абстрагирования операций перечисленных типов.

2. Разработана и реализована в ряде практических задач **методика синтеза алгоритмов семантической навигации**, «биологически инспирированная» метафорой «дивергентно-конвергентной нейрональной импульсации» в нервной системе. По данной методике разработаны частные модели цепей семантической навигации, необходимые для реализации функциональности практической системы, выбранной в качестве экспериментальной платформы для апробации и исследования разрабатываемых в диссертационной работе решений.

3. Сформирован набор типовых структур метрик и самих метрик для операций семантической навигации.

3.1. Для операций семантического экстрагирования разработаны *две частные метрики* и *объединяющая их интегральная* «метрика релевантности».

По результатам экспериментов и сопоставления с традиционными аналогами, предлагаемая метрика позволила получить более качественные результаты в случаях, когда условия подбора семантических детерминантов благоприятны, и оказалась работоспособной в случаях, когда применение традиционных метрик не обеспечивало приемлемых результатов.

3.2. Для интеграции информации от разных источников в общий рейтинг предложена нелинейная метрика с доминантным сигналом (или несколькими сигналами) и его модулированием значениями остальных сигналов.

Эксперименты показали, что во всех рассмотренных случаях применения предложенная метрика даёт лучшие результаты, чем традиционные линейные метрики (усреднение, взвешенная сумма и другие подобные способы агрегирования нескольких сигналов).

3.3. Экспериментально установлено, что метрика релевантности, предложенная для операций подбора семантических детерминантов, пригодна в операциях подбора семантически подобных сущностей, то есть в другом функциональном элементе операциональной триады.

Этот результат (нахождение метрик, работающих в широком диапазоне условий применения и более того – в операциях разного типа):

– даёт довод в пользу правильности выбранного пути исследования – обобщения широкого спектра результатов в области семантической навигации и поиска универсальных конструктивных элементов;

– имеет большую практическую значимость, поскольку существенно сокращает сроки и трудоёмкость проектирования компонентов семантической навигации, избавляя от необходимости скрупулёзного подбора метрик, адекватных каждой ситуации применения.

4. На основе разработанных методических решений, моделей и схем семантической навигации, предложенных метрик создано несколько десятков алгоритмов и реализующих их SQL-процедур, полностью покрывающих потребности выбранной для апробации рекомендательной системы и реализующих функциональность рекомендаций (а) тематических категорий, (б) поисковых термов, (в) ключевых слов аннотаций книг, динамически формируемых в процессе и в темпе побуквенного ввода пользователем информации о своей книге (т.е. в реальном времени).

Помимо экспериментального исследования предложенных решений проводилась оценка их качества коллективом, проектирующим коммерческую систему с применением этих решений. Главные выводы руководства проекта заключаются в следующем:

- по сравнению с применявшимися ранее решениями (как широко известными, так и собственными разработками) предлагаемые в работе решения оказались по качеству результатов либо близкими (но не худшими), либо лучшими, вплоть до того, что они обеспечивали выход на качественно новый уровень;

- помимо общего повышения качества итоговых результатов, т.е. эксплуатационных характеристик создаваемой системы, большое значение имеют эффекты, достигаемые на основе предложенных в работе обобщенных методик и унифицированных решений. Например, применение предложенной методики синтеза алгоритмов операций семантической навигации позволило существенно упростить и ускорить проектирование алгоритмов и реализующих их программных процедур в ситуации, когда потребность в таких операциях исчисляется десятками. Другой пример: предложенные соискателем типовые метрики показали результаты, по меньшей мере не худшие, чем применявшиеся до того специализированные метрики, а это позволило заменить множество специальных решений на однотипные, что также упростило и ускорило разработку и повысило управляемость проектируемой системой;

- в целом созданные соискателем решения можно квалифицировать как имеющие существенную практическую значимость; их применение возможно в широком спектре современных информационных систем самого различного назначения, поскольку наличие в современных ИС функциональности, так или иначе подпадающей под определение «семантической навигации в информационном пространстве», скорее правило, чем редкость или специальная функциональность; в подобных условиях (т.е. когда компоненты семантической навигации встраиваются в программный продукт, но не являются его основным функциональным звеном, а, следовательно, над продуктом работают и специалисты иного профиля) унифицированность решений приобретает тем более важное значение.

Описанные в работе результаты были реализованы в рекомендательной системе для авторов книг в составе интернет-сервиса помощи авторам в публикации их произведений. После внедрения в сервис разработанной рекомендательной системы его популярность среди авторов резко возросла, и за год (с осени 2016-го до лета 2017-го) с помощью сервиса было опубликовано более 10 000 книг. О достигнутом высоком уровне потребительского качества свидетельствует и тот факт, что вскоре после внедрения в продукт разработанной системы стартап проекта был куплен издательским гигантом Macmillan, и в настоящее время система находится в собственности этой компании.

Результаты, полученные при апробации, подтверждают верность выдвинутых гипотез и направления разработки и позволяют сделать вывод о практической применимости и высокой эффективности разработанных решений.

СПИСОК ОПУБЛИКОВАННЫХ РАБОТ

1. Чеушев, К. В. Инструменты семантической навигации для формирования ресурсов дистанционного образования / К. В. Чеушев, Л. А. Глухова // Дистанционное обучение – образовательная среда XXI века : материалы X международной научно-методической конференции (Минск, 7 - 8 декабря 2017 года). – Минск : БГУИР, 2017. – С. 152 - 153.

2. Чеушев, К. В. Семантическая навигация в больших информационных массивах / К. В. Чеушев // Компьютерные системы и сети : материалы 52-й научной конференции аспирантов, магистрантов и студентов. (Минск, 25 - 30 апреля 2016 года). – Минск : БГУИР, 2016. – С. 89 - 91.