

УДК 378.4:004

## СТРУКТУРИРОВАНИЕ КОНТЕНТА ИНФОРМАЦИОННОГО ПРОСТРАНСТВА ТЕХНИЧЕСКОГО УНИВЕРСИТЕТА С ИСПОЛЬЗОВАНИЕМ ПРОЦЕССНОГО ПОДХОДА И СЕМАНТИЧЕСКОЙ ИДЕНТИФИКАЦИИ

Г. Г. Куликов<sup>1</sup>, М. А. Шилина<sup>2</sup>, Г. В. Старцев<sup>3</sup>, А. А. Бармин<sup>4</sup>

<sup>1</sup>gennadyg\_98@yahoo.com, <sup>2</sup>maria.shilina@gmail.com, <sup>3</sup>startsev@gmail.com, <sup>4</sup>abarmin@outlook.com

ФГБОУ ВПО «Уфимский государственный авиационный технический университет» (УГАТУ)

*Поступило в редакцию 1 ноября 2013 г.*

**Аннотация.** Рассматривается процесс обеспечения идентифицируемости и прослеживаемости слабо-структурированных информационных ресурсов информационного пространства технического университета. Приводится методика идентификации информационных ресурсов на основе системных моделей процессов предметной области. Приводится пример реализации предложенной методики с использованием информационно-поисковых систем на базе веб-портала выпускающей кафедры.

**Ключевые слова:** бизнес-процесс; системная модель; теория категория; информационно-поисковые системы.

### ВВЕДЕНИЕ

Опыт российских и зарубежных университетов показывает, что использование современных информационных технологий в деятельности вуза в настоящее время является необходимостью. Требования к автоматизации отражены в стандарте ISO 9001:2011, директивах ENQA, а также в образовательных стандартах третьего поколения и новой редакции Федерального закона «Об образовании». Отмечается, что необходимым условием развития современных университетов является создание и поддержание единого информационного пространства для обеспечения информационной прозрачности и мониторинга, в том числе со стороны государства, повышения качества образования, а также устранения территориальных барьеров (то есть академической мобильности, в том числе виртуальной). В связи с присоединением России к Болонскому соглашению, потребовалось создание новых образовательных стандартов, ориентированных на поддержку концепции «образование в течение всей жизни».

Создание и поддержание единого информационного пространства подразумевает использование банков и баз знаний, технологий их сопровождения и использования, а также информационных телекоммуникационных систем, обеспечивающих взаимодействие и удовлетво-

рение информационных потребностей пользователей.

В процессе функционирования университет накапливает значительный объем данных. Использование этих данных может быть эффективным, если к ним обеспечивается оперативный доступ. Таким образом, актуально решение проблемы структурирования информационного пространства образовательного учреждения на основе систематизации процессов сбора, обработки и анализа данных об учебном процессе, а также обеспечения их идентификации и прослеживаемости.

Свойства идентификации и прослеживаемости являются необходимыми требованиями для адекватности построения сложных систем. В формализованных (структурированных) областях деятельности (область конструирования, производства и т.п.) эти требования составляют основу соответствующих стандартов (CAD / CAM / CAE).

Проблема прослеживаемости – это прежде всего документирование идентифицированных объектов в различной документации предметной области, а следовательно, и в едином информационном пространстве, являющимся ее отображением.

Как показывает опыт, информационное пространство современного вуза является гетерогенным. В этом случае данные, характеризую-

щие учебный процесс и отдельные его объекты, оказываются распределенными по различным репозиториям и имеют разный формат представления. Кроме того, в процессе создания информационных объектов принимают участие несколько пользователей, поэтому ни один из них не может обладать знаниями обо всех его характеристиках. Таким образом, эффективное удовлетворение информационных потребностей пользователей возможно только с использованием развитых информационно-поисковых систем.

В статье рассматриваются вопросы формирования и структурирования единого информационного пространства вуза с использованием CALS-технологий, методика идентификации информационных объектов на основе их семантической аннотации, и обеспечение идентификации и прослеживаемости информационных объектов с использованием информационно-поисковых систем. Также приводится пример создания такой системы на примере выпускающей кафедры технического вуза.

#### НЕОБХОДИМОСТЬ ФОРМИРОВАНИЯ ЕДИНОГО ИНФОРМАЦИОННОГО ПРОСТРАНСТВА ВУЗА С ПРИМЕНЕНИЕМ CALS- ТЕХНОЛОГИЙ

Университеты представляют собой консервативные социально-экономические системы. Тем не менее переход к информационному обществу, а следовательно, изменение способов, сроков обработки информации, привел к значительному увеличению ее объемов, но вместе с тем – к расширению доступа к этой информации, существенному снижению межгосударственных и внутрироссийских барьеров, что требует от университетов непрерывной адаптации к новым условиям. Немаловажным фактором является также постоянное изменение потребностей рынка труда, а также новации в глобальной информационной инфраструктуре [1].

Необходимым условием выполнения требований стандартов серии ГОСТ Р ИСО 9000 и ENQA является высокая степень автоматизации бизнес-процессов и соответствующих процессов управления с целью обеспечения информацией, необходимой для поддержки процессов, жизненного цикла продукции и процессов, составляющих основу создания, поддержания и постоянного улучшения системы качества организации, а также их непрерывный мониторинг.

Структура информационного пространства большинства вузов характеризуется большим

количеством информационных систем, автоматизирующих различные задачи и характеризующиеся разрозненным хранением данных, представленных в различных форматах (рис. 1).

Информационные системы (ИС) для управления финансово-хозяйственной деятельностью вуза: 1С:Предприятие, Галактика, Парус, SAPR3 и т.д.	Системы веб-конференций, видеоконференций, вебинаров на платформе AdobeConnect, OpenMeetings, Mirapolis, CiscoWebExMeetingsCenter и т.д.
ИС, обеспечивающие учет, хранение, обработку и анализ информации об основных процессах вуза: 1С:Университет, Галактика Управление Вузом, а также собственные разработки вузов и т.д.	Системы поддержки учебного процесса для организации электронного и дистанционного обучения, т.н. LMS, CMS (Blackboard, Moodle, Прометей, 1С:Электронное обучение, ILIAS и т.д.)
Системы документооборота: 1С:Документооборот, DocVision, Логика СЭД, Directum, Дело, Ефрат-Документооборот и т.д.	Системы для проведения аттестации, компьютерного тестирования, анкетирования (ФЭПО, VeralTest, eTest и т.д.)
Иные ИС (в том числе системы мониторинга, ВРMS, системы управления проектами и т.д.)	Веб-порталы университета, его филиалов и подразделений.

Рис. 1. Основные виды информационных систем в информационном пространстве вуза

Для использования этих данных в качестве обратной связи для информационной поддержки принятия решений требуется решить задачу сбора, централизованного хранения, применения специализированных процедур обработки, а также решение проблемы дублирования и логического несоответствия. В результате исходные данные превращаются в ценную аналитическую информацию, являющуюся надежной базой для принятия управленческих решений.

Доступ к актуальным первичным данным позволяет оперативно получать новую информацию и в условиях изменяющихся бизнес-процессов – новые знания. Однако многократно увеличивающиеся объемы обрабатываемой информации и сложность такой обработки и интерпретации данных требуют использования принципиально новых решений, методов и технологий, использования хранилищ данных и средств многомерного анализа, а также развитых поисковых инструментов. При этом разработка поисковой системы требует создания и реализации соответствующих математических моделей и методик.

Любая система может быть представлена как целостный объект или как совокупность связанных и взаимодействующих друг с другом составных частей — объектов меньшего масштаба. Информационное отображение физических объектов или процессов называют *информационным объектом* (ИО). Совокупность информационных объектов, отражающих свойства

системы и протекающих в ней процессов, называют *информационным пространством*.

Понятие единого информационного пространства или интегрированной информационной среды является ключевым понятием концепции *CALS (Continuous Acquisition and Life – Cycle Support)*, успешно применяемой многими зарубежными и российскими промышленными предприятиями.

При создании единого информационного пространства вуза необходимо руководствоваться *базовыми принципами CALS*, к которым относят:

- системную информационную поддержку жизненного цикла образовательного процесса, основанную на использовании информационных систем, формирующих единое информационное пространство;
- безбумажное представление информации, основанное на электронном обмене данными между участниками процессов жизненного цикла;
- информационную интеграцию, которая может быть достигнута за счет стандартизации описаний объекта управления;
- реинжиниринг бизнес-процессов;
- стандартизацию структур данных и интерфейсов доступа к ним [2].

Кроме того, реализация принципа стандартизации структур данных и интерфейсов доступа к ним накладывают определенные ограничения на разрабатываемые модели. Например, характеристики информационного объекта "Студент" не должны противоречить спецификации *IMS LIP (Learner Information Package)*, признанной стандартом за рубежом и поддерживаемой многими *E-learning*-системами, что является необходимым условием для обеспечения информационной интеграции и взаимодействия с внешними потребителями информации, например, в процессе академической мобильности (в т. ч. виртуальной), и соответствует принципам *CALS*.

Характерным свойством информационного пространства является его *структурированность*. Это означает, что выделены его элементы, установлены связи между ними, введены обозначения, элементы и связи упорядочены. Свойство структурированности в разных видах информационных пространств может быть выражено в разной степени. Высокий уровень структурированности обеспечивает возможность представления информации в виде документов и манипулирования данными с помо-

щью программно-технических средств информационных систем.

Благодаря формированию единого информационного пространства появляется возможность не просто хранить зафиксированные в электронном виде результаты учебного процесса (оценки и т. п.), а гораздо более адекватные модели знаний и процессов, что позволит обеспечить более высокое качество информационной поддержки учебного процесса, прослеживаемость показателей и, как следствие, повышение эффективности управления.

#### **МЕТОДИКА ИДЕНТИФИКАЦИИ СЛАБОСТРУКТУРИРОВАННЫХ ИНФОРМАЦИОННЫХ ОБЪЕКТОВ НА ОСНОВЕ СЕМАНТИЧЕСКОЙ АННОТАЦИИ**

Информационные ресурсы являются элементами информационного пространства (совокупности результатов семантической деятельности человека), которое представляется в виде совокупности:

- банков и баз знаний;
- технологий их сопровождения и использования;
- информационных телекоммуникационных систем, функционирующих на основе общих принципов и обеспечивающих информационное взаимодействие организаций и граждан, а также удовлетворение их информационных потребностей.

Таким образом, основными элементами информационного пространства являются:

- информационные ресурсы;
- средства информационного взаимодействия;
- информационная инфраструктура.

Информационные ресурсы, представленные в различных хранилищах, отличаются по степени структурированности. Так, ресурсы, представленные в виде реляционных или нереляционных баз данных, являются наиболее структурированными, так как определены модели данных, а следовательно, все атрибуты данных. Информационные ресурсы, представленные в форме текстовых файлов, являются неструктурированными и уникально идентифицируются только именем файла. Современные файловые системы позволяют создавать не только иерархические, но и фасетные классификации файлов.

Также существует представление информационных ресурсов в форме баз знаний. Наи-

большее распространение получили базы знаний на основе гипертекстовых технологий. В данном случае информационные ресурсы идентифицируются с помощью уникального имени, связи между ними устанавливаются с помощью гиперссылок.

Таким образом, в каждом из описанных представлений:

- существует класс объектов  $Ob_c$ , обладающих общими свойствами;

- между объектами этого класса существуют отношения – реляционные связи в базе данных, гипертекстовые связи в базе знаний, отношения категоризации в файловой системе. Эти отношения имеют различную природу, но в общем случае они отображают одни элементы одного класса в другие элементы того же или другого класса, то есть являются функциями. Таким образом, можно сделать вывод о наличии морфизмов  $Hom_c(A, B)$  между объектами  $(A, B) \in Ob_c$ ;

- для каждой пары морфизмов  $f \in Hom(A, B)$  и  $g \in Hom(B, C)$  определена композиция  $f \circ g \in Hom(A, C)$  – объекты информационного пространства связаны и возможен переход от одного объекта к другому, например, с помощью гиперссылок, в этом случае операция композиции ассоциативна [3].

На основе изложенных выше положений можно сделать вывод о существовании категории над множеством объектов и связей информационного пространства (объекты и связи информационного пространства – граф информационного пространства).

Таблица 1

## Полнота системной модели

Модель/аспект описания системы	ФМ, %	ИМ, %	ДМ, %	Итого, %
Функциональное содержание системы	60–80	10	10	80–90
Информационное содержание системы	10	60–70	10	80–90
Описание поведения системы во времени в пространстве	10	10	60–70	80–90

Бизнес-процесс (деловой процесс) – множество из одной или нескольких связанных операций или процедур, в совокупности реализующих некоторую цепь производственной деятельности, осуществляемой обычно в рамках заранее определенной организационной струк-

туры, которая описывает функциональные роли участников и отношения между ними.

При исследовании, проектировании и создании реальной системы необходимо определить элементы и структуру их взаимодействия во времени и пространстве с помощью различных мер. При этом существует диалектическая пара между реальной системой и ее информационным отображением – информационной системой.

В прикладном аспекте проблемы создания автоматизированных информационных систем тесно связаны с методами CASE-технологий (Computer-Aided Software Engineering), объединяющих множество формализованных методов проектирования в едином информационном пространстве. Структурное представление основных категорий и положение диалектики позволяет сформулировать следующий четырехмерный базис построения информационной системы, как отражения реальной системы:

- система элементов и отношений (функциональное содержание системы);
- система сущностей и связей (информационное содержание);
- причинно-следственные связи (динамика);
- семантика (количественная неопределенность).

Степень полноты интегрированного, формализованного описания свойств реальной системы, ее функциональной, информационной, динамической моделей можно оценить показателями, приведенными в табл. 1 [4].

Анализ итоговых показателей, приведенных в табл. 1, показывает, что если полное адекватное описание системы может составлять 300%, то в модели 10–20% всех знаний и данных о реальной системе остается неопределенным.

Методология SADT поддержана рядом CASE-средств, построенных на стандартах IDEF0, IDEF1, IDEF1X, IDEF/CPN – ERWinDataModeler, ERWinProcessModeler, BusinessStudio, Ramusi другие.

В стандарте IDEF первоначально обеспечивается единая информационная среда, и система формализованных правил увязывает в единый комплекс модели: семантическую, функциональную, информационную и динамическую, поэтому дальнейшее рассмотрение ведется в аспекте стандарта SADT.

Множество информационных объектов вуза представим в виде подмножеств структуриро-

ванных, слабоструктурированных и неструктурированных информационных объектов:

$$D = D_S \cup D_H \cup D_U, \quad (1)$$

где  $D$  – множество информационных объектов информационного пространства,  $D_S$  – подмножество структурированных информационных объектов,  $D_H$  – подмножество слабоструктурированных информационных объектов,  $D_U$  – подмножество неструктурированных информационных объектов.

Подмножество структурированных информационных объектов можно представить в виде модели в нотации IDEF1X. В этом случае информационные объекты являются сущностями, а отношения между ними – связями. Сущности имеют уникальное имя и содержат ключевые и неключевые атрибуты. Ключевые атрибуты однозначно идентифицируют каждый экземпляр сущности в рамках информационного пространства, что позволяет отследить его положение в бизнес-процессе.

Модель структурированного информационного пространства можно представить в виде ориентированного графа, где узлами выступают сущности, а дугами – отношения между ними:

$$D_S = (E_d, R_d), \quad (2)$$

где  $E_d$  – множество сущностей предметной области,  $R_d$  – множество отношений между сущностями.

Каждая сущность в структурированном информационном пространстве представляется в виде уникального имени и совокупности атрибутов:

$$E_d^i = (n, A_k, A_d), \text{ где } A = A_d \cup A_k, \quad (3)$$

где  $n$  – наименование сущности,  $A$  – атрибуты сущности:

$A_d$  – неключевые атрибуты сущности,

$A_k$  – ключевые атрибуты сущности.

Слабоструктурированные информационные объекты содержат в себе структурированную и неструктурированную части. Структурированная часть информационного объекта может быть представлена в виде информационной модели, неструктурированная часть – в виде совокупности терминов некоторой онтологии, характерной для предметной области.

$$D_H = (D_S \cup T_O), \quad (4)$$

где  $D_S$  – структурированная часть информационного ресурса,  $T_O$  – множество терминов онтологии  $O$ .

Модель онтологии имеет вид:

$$O = (T, R, F), \quad (5)$$

где  $O$  – онтология,  $T$  – термины предметной области, которые определяют онтологию  $O$ ,  $R$  – конечное множество отношений между терминами предметной области,  $F$  – конечное множество функций интерпретации заданных на терминах или отношениях онтологии  $O$ .

Неструктурированные информационные объекты представлены только совокупностью терминов некоторой онтологии.

Для представления неструктурированных информационных объектов с помощью онтологии необходимо определить термины предметной области. Термины можно выделить путем индексирования всех информационных ресурсов или из тезауруса моделей бизнес-процессов, выполняющихся в информационном пространстве.

Тезаурус представляет собой множество упорядоченных пар терминов предметной области  $t_i \in T$  и их описаний  $d_i \in Desc, i = 1..N$ :

$$Thes = \{(t_i, d_i)\}. \quad (6)$$

В тезаурус модели бизнес-процесса входят термины всех входных и выходных данных, нормативных документов и механизмов, участвующих в бизнес-процессе. На основе терминов предметной области формируется многомерный классификатор информационного пространства, в рамках которого могут быть идентифицированы неструктурированные ресурсы.

Классификацию и идентификацию неструктурированных информационных объектов можно провести на основе терминов, присутствующих в них, то есть представить информационный ресурс в виде вектора терминов:

$$R = \{t_1, t_2, \dots, t_n\}. \quad (7)$$

Представим модель бизнес-процесса также в виде подмножеств терминов онтологии предметной области:

$$M = (I, C, O, M) = (T_i, T_c, T_o, T_m), \quad (8)$$

где  $I$  – входные данные,  $T_i$  – термины, связанные со входными данными,  $C$  – нормативные документы,  $T_c$  – термины, связанные с нормативными документами,  $O$  – выходные данные,  $T_o$  – термины, связанные с выходными данными,  $M$  – механизмы и исполнители бизнес-процесса,  $T_m$  – термины, связанные с механизмами и исполнителями.

Таким образом, каждую функцию модели бизнес-процесса можно описать в виде множества терминов:

$$F_j = (T_i^j, T_c^j, T_o^j, T_m^j), (T_i^j, T_c^j, T_o^j, T_m^j) \in T^j. \quad (9)$$

Используя (9) и (7), можно идентифицировать объекты, относящиеся к каждой конкретной функции бизнес-процесса, то есть обеспечить их прослеживаемость в рамках бизнес-процесса.

Использование семантической идентификации не позволяет однозначно идентифицировать информационный объект, соответствующий запросу, но позволяет выделить группу объектов, удовлетворяющих запросу в определенной степени. Степень соответствия между информационным объектом можно определить на основе семантического расстояния между ними.

Запрос к информационному пространству представляется в виде совокупности множеств терминов  $T_q$  и булевых операций между ними  $R_q$  и может быть представлен в дизъюнктивной нормальной форме  $q_{dnf}$ :

$$q = (T_q, R_q) = q_{dnf} = \bigvee_{i=1..N} q_i, \quad (10)$$

где  $q_i$ — $i$ -я конъюнктивная компонента запроса  $q$ .

Пусть  $i$  — индекс термина  $t_i$  из таксономии  $O$  ( $i=1, \dots, M$ ),  $r^j$  — информационный объект, принадлежащий информационному пространству  $D$ . Информационный объект будем рассматривать как вектор  $r^j = (w_1^j, w_2^j, \dots, w_M^j)$ , где  $w_i^j \geq 0$  — вес, ассоциированный с парой  $(t_i, r^j)$ .

Для каждого термина  $t_i$ , который не входит во множество терминов, связанных с информационным объектом  $r^j$ , его вес равен нулю:  $w_i^j = 0$ .

Введем также инверсную функцию  $g$ , соответствующую индексу термина  $t_i$ , которая определяется следующим образом:  $g_i(r^j) = w_i^j$ . Данная функция показывает вес  $i$ -го термина в  $j$ -м информационном объекте.

Мера близости информационного ресурса  $r^j$  и запроса  $q$  —  $sim(r^j, q)$  определяется выражением:

$$sim(r^j, q) = \begin{cases} 1, & \text{если } \exists q_i: (q_i \in q_{dnf}) \wedge (\forall k: g_k(q_i) = g_k(r^j)), \\ 0, & \text{иначе.} \end{cases} \quad (11)$$

То есть  $sim(r^j, q)$  принимает значение 1, если существует такая конъюнктивная компонента  $q_i$ , входящая в дизъюнктивную нормальную форму  $q_{dnf}$ , что инверсная каждого термина  $t_i$  данной конъюнктивной компоненты совпадает с этой же инверсной функцией для инфор-

мационного объекта  $r^j$ . В противном случае  $sim(r^j, q)$  оказывается равной 0.

Таким образом, если  $sim(r^j, q) = 1$ , то информационный объект  $r^j$  является релевантным запросу  $q$ .

Определим граф со следующими объектами и ребрами:

$n$  — неструктурированное информационное пространство,

$i$  — семантически идентифицированное пространство,

$ident: n \rightarrow i$  — семантическая идентификация информационного пространства,

$m$  — множество моделей процессов информационного пространства,

$modeling: n \rightarrow m$  — формализация бизнес-процессов информационного пространства в виде множества моделей,

$s$  — семантически аннотированные модели бизнес-процессов информационного пространства,

$annotation: m \rightarrow s$  — семантическое аннотирование моделей бизнес-процессов на основе терминов онтологии и тезауруса моделей,

$r$  — результат идентификации информационных ресурсов на основе моделей бизнес-процессов,

$search: (s, i) \rightarrow r$  — семантический поиск в аспекте модели бизнес-процесса.

Матрица смежности для полученного графа представлена в табл. 2.

Таблица 2

**Матрица смежности графа информационного поиска**

	$n$	$i$	$m$	$s$	$r$
$n$		ident	modeling		
$i$					search
$m$				annotation	
$s$					search
$r$					

**ОБЕСПЕЧЕНИЕ ИДЕНТИФИЦИРУЕМОСТИ И ПРОСЛЕЖИВАЕМОСТИ ИНФОРМАЦИОННЫХ ОБЪЕКТОВ С ИСПОЛЬЗОВАНИЕМ ИНФОРМАЦИОННО-ПОИСКОВЫХ СИСТЕМ**

Предложенная выше модель аннотирования и идентификации может быть реализована с использованием информационно-поисковых систем. Информационно-поисковая система — это программно-аппаратный комплекс, предостав-

ляющий возможность индексирования и поиска информации.

Информационный поиск – это процесс поиска в большой коллекции (хранящейся, как правило, в памяти компьютера) некоторого неструктурированного материала (обычно документа в виде файла), удовлетворяющего информационные потребности пользователя.

Под неструктурированными данными понимаются данные, которые не имеют ясной, семантически очевидной и легко реализуемой на компьютере структуры. В реальности полностью неструктурированных данных не существует, даже обычные текстовые документы имеют структуру, они разделены на разделы, которые оформлены заголовками, текст разбит на абзацы, абзацы на предложения. Все эти структуры очевидны для естественного языка, но не всегда пригодны для автоматизированного поиска.

Под документами в данном контексте понимаются любые объекты, на основе которых решено построить систему информационного поиска. Группа документов, по которым выполняется поиск, называется коллекцией, термин – единица индексирования.

Простейший вариант поиска документальной информации в такой коллекции предполагает последовательный просмотр всех документов – прямой поиск. Данный подход обладает существенным недостатком – при каждом пользовательском запросе приходится просматривать все пространство, поэтому поиск занимает значительное время.

Во избежание последовательного просмотра всех документов, заранее составляется индекс документов. В этом случае для каждого документа создается запись, в которой указано, содержит ли он конкретный термин предметной области. В результате возникает бинарная инцидентность «термин-документ», которая может быть представлена в форме матрицы. Теперь в зависимости от порядка просмотра этой матрицы (по строкам или по столбцам) можно получить вектор термина, в котором указано, в каких документах он встречается, либо вектор документа, в котором указано, какие термины в нем употребляются.

Для большой коллекции документов матрица может занимать значительный объем памяти и не помещаться целиком в память одного компьютера. Кроме того, матрица будет чрезвычайно разреженной, то есть содержать лишь небольшое количество ненулевых элементов. Для решения данной проблемы используется техно-

логия создания инвертированного индекса, алгоритм создания:

1. Записываем в память словарь терминов (dictionary) (иногда его называют словником (vocabulary) или лексиконом (lexicon)).

2. Для каждого термина создается список документов, содержащих данный термин.

3. Каждый элемент такого списка получает дополнительную информацию о координате термина в документе – словопозицию (posting).

Получившийся в результате список (postinglist) и является инвертированным индексом.

Цель использования информационно-поисковых систем – поиск документов из коллекции, которые являются наиболее релевантными по отношению к произвольным информационным потребностям, сообщаемым системе при помощи однократных, инициированных пользователем запросов. Документ является релевантным с точки зрения пользователя, если он содержит ценную информацию, удовлетворяющую его информационную потребность.

Для простейших моделей поиска поисковый запрос представляется в форме булевых выражений вида:

«Образование» И «Высшее»

Обработка такого запроса выполняется по следующему алгоритму:

1. Обнаруживаем термин «Образование» в словаре.
2. Находим список его словопозиций.
3. Обнаруживаем термин «Высшее» в словаре.
4. Находим список его словопозиций.
5. Находим пересечение списков словопозиций [5].

Словарь терминов информационно-поисковой системы может формироваться автоматизированно на основе терминов документов предметной области, также он может быть сформирован пользователем самостоятельно на основе тезауруса моделей предметной области.

Использование моделей бизнес-процессов позволяет сузить область поиска за счет добавления в поисковый запрос параметров, характерных для каждого бизнес-процесса. Например, для бизнес-процесса «Учет успеваемости» поисковый запрос может быть сформирован в виде:

«Иванов» И «Философия» И «Журнал успеваемости»

Таким образом, использование информационно-поисковых систем позволяет идентифицировать и проследить состояние документа в рамках бизнес-процесса.

#### **РЕАЛИЗАЦИЯ ИНФОРМАЦИОННОГО ПОИСКА В КОНТЕНТЕ КОРПОРАТИВНОГО ВЕБ-ПОРТАЛА ВЫПУСКАЮЩЕЙ КАФЕДРЫ ТЕХНИЧЕСКОГО ВУЗА**

Автоматизированная информационная система образовательного учреждения в процессе своей деятельности накапливает существенный объем данных. Эти данные находятся в различных хранилищах и представлены в разнообразных форматах. Использование порталных технологий на основе сервис-ориентированной архитектуры в образовательном учреждении позволяет обеспечить централизованный доступ к информационному пространству с использованием тонкого клиента – интернет-браузера. Использование информационно-поисковых систем в качестве сервиса информационного пространства обеспечивает оперативный поиск информации.

Обмен данными между сервисами корпоративного веб-портала выполняется с использованием веб-сервисов, для обеспечения взаимодействия которых используются следующие стандарты:

- XML – расширяемый язык разметки, используемый для хранения и передачи структурированных данных;
- SOAP – протокол обмена сообщениями на базе XML;
- WSDL – язык описания внешних интерфейсов веб-службы на базе XML.

Информационно-поисковая система является сервисом, выполняющим две функции:

1. Создание и обслуживание поискового индекса объектов информационного пространства.
2. Выдача ссылок на объекты информационного пространства, удовлетворяющие поисковому запросу.

Каждый бизнес-процесс представляет собой отдельную задачу веб-портала, доступ к которой регламентируется с помощью системы прав доступа. Создание индекса объектов информационного пространства требует доступа к задачам веб-портала с правами супер-пользователя – все объекты информационного пространства должны попасть в поисковый индекс.

Использование объектно-ориентированных технологий при разработке веб-портала позволяет логически группировать классы, относящиеся к одному бизнес-процессу в один пакет. Это позволяет разрабатывать и использовать задачи независимо. Тем не менее наличие большого количества унаследованных технологических решений создает дополнительные сложности при создании сервис-провайдера (serviceprovider).

Для решения возникшей задачи был использован принцип инверсии управления, который используется для уменьшения связанности в объектно-ориентированных приложениях. Использование данного принципа предполагает создание требуемых объектов во время выполнения приложения, что позволяет использовать метаданные для гибкой настройки приложения.

Фрагмент информационной модели, описывающий метаданные о сущностях предметной области приведен на рис. 2.

Каждая сущность объектно-ориентированного информационного пространства представлена в виде класса. Некоторые сущности должны иметь возможность долговременного хранения своего состояния в базе данных, поэтому перед сохранением и обновлением данных должна проверяться согласованность данных на уровне логики бизнес-процессов. В системе реализованы типовые валидаторы для проверки данных на пустоту, выбор из списка и соответствии требуемому формату данных.

Описание объектов предметной области с использованием метаданных позволяет легко адаптировать отображение данных под требования пользователя – в системе реализована поддержка перевода интерфейса CRUD-операций на несколько языков.

На основе метаданных экземпляры объектов предметной области сохраняются в XML-документы, которые затем отправляются в поисковый сервис.

Получение результатов поиска также происходит путем обмена XML-сообщениями между сервисом веб-портала и сервисом информационного поиска. XML-сообщения со ссылками на объекты предметной области в соответствии с метаданными, хранящимися в базах данных, преобразуются в доступный для пользователя вид.

Динамическая модель взаимодействия пользователя, веб-портала и сервиса информационного поиска представлена на рис. 3.

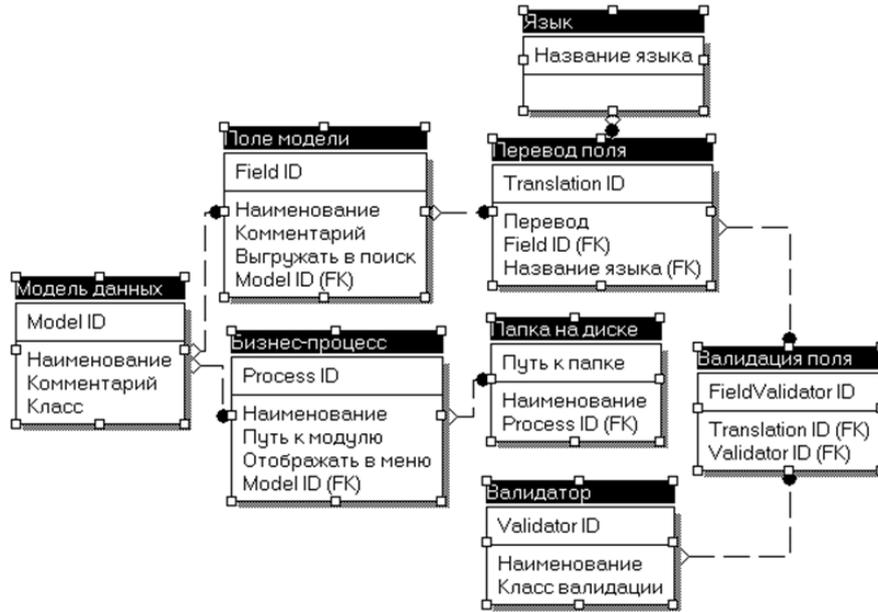


Рис. 2. Фрагмент информационной модели метаданных предметной области

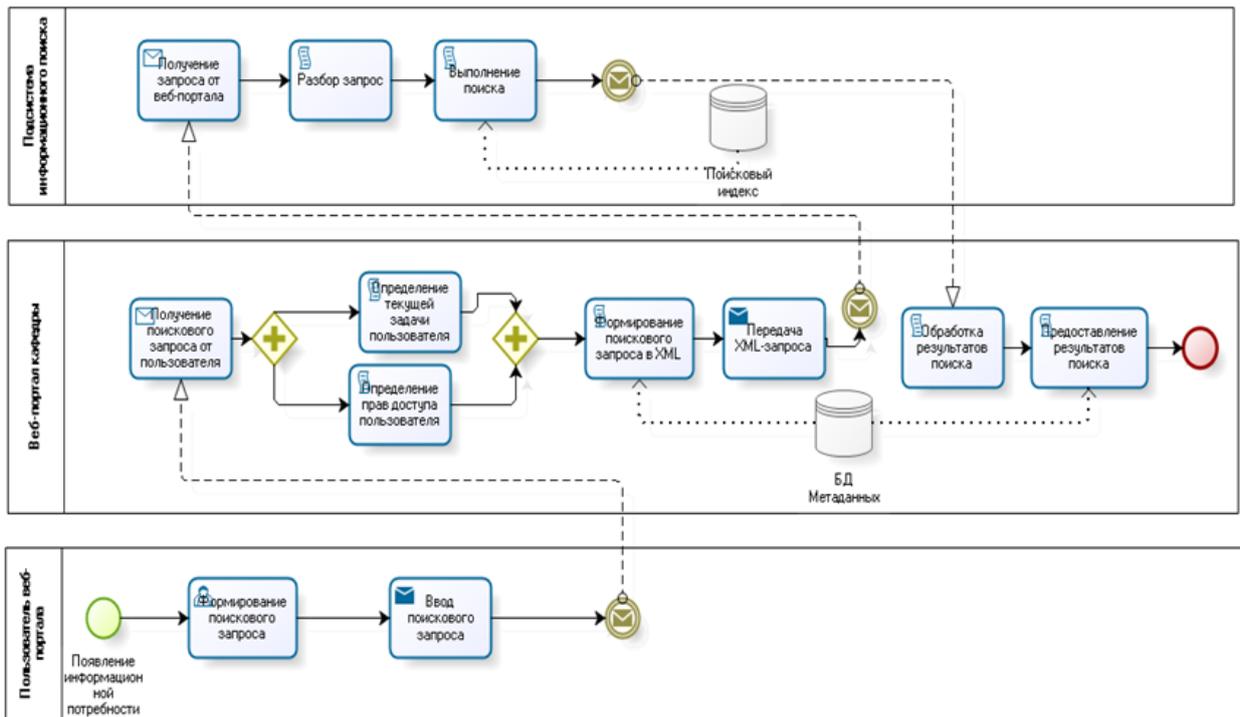


Рис. 3. Динамическая модель взаимодействия с сервисом поиска

Описанное выше решение реализовано с использованием объектно-ориентированного языка программирования PHP 5, системы управления базами данных MySQL, сервера приложений Apache. В качестве сервиса информационного поиска выступает типовое решение ApacheSolr.

### ЗАКЛЮЧЕНИЕ

Свойства идентификации и прослеживаемости являются необходимыми требованиями для адекватности построения сложных систем. В формализованных (структурированных) областях деятельности (область конструирования, производства и т.п.) эти требования составляют

основу соответствующих стандартов (CAD / CAM / CAE). Единое информационное пространство является ключевым компонентом концепции CALS. Данные, содержащиеся в едином информационном пространстве, относятся, в том числе, и к системе менеджмента качества. Таким образом, можно сделать вывод, что структурирование единого информационного пространства, систематизация сбора, обработки данных ведут к повышению качества и оперативности информационной поддержки управления, а следовательно, повышению эффективности управляемых процессов.

Семантическое аннотирование слабоструктурированных объектов информационного пространства позволяет идентифицировать их в рамках бизнес-процессов образовательного учреждения, организовать информационную поддержку жизненного цикла образовательного процесса на всех его этапах жизненного цикла. Таким образом, устанавливается соответствие между моделями бизнес-процессов и их экземплярами в информационном пространстве вуза.

Использование информационно-поисковых систем позволяет в автоматизированном режиме производить идентификацию объектов информационного пространства, выполнять их поиск и навигацию.

Применение информационно-поисковой системы в качестве сервиса корпоративного портала позволяет выполнять оперативный поиск информации в различных источниках. Доступ к корпоративному portalу осуществляется с помощью тонкого клиента – веб-браузера, доступного практически на любых устройствах. Таким образом, можно получать оперативную информацию об объектах информационного пространства.

#### СПИСОК ЛИТЕРАТУРЫ

1. Тихомирова Н. В. Образовательный процесс в электронном университете: условия и направления трансформации // Высшее образование в России. 2011. № 2. С. 3–11. [ N.V. Tihomirova, "Education in electronic university," (in Russian), *High education in Russia*, no. 2, pp. 3-11, 2011. ]
2. Агапов Р. Н., Куликов Г. Г., Старцев Г. В., Шилина М. А. Формирование иерархического интегрированного информационного пространства образовательного учреждения с использованием интранет-технологий // Открытое образование, 2011. № 4 (87). 2011. С. 29–37. [ R. N. Agarov, G. G. Kylikov, G. V. Startsev, M. A. Shilina, "Forming hierachical integrated information space of high school using intranet technologies," (in Russian), *Open Education*, no. 4 (87), pp. 29-37, 2011. ]
3. Кук Д., Бейз Г. Компьютерная математика. М.: Наука, 1990. 384 с. [ D. Kук, G. Beiz, *Computer math*, (in Russian). Moscow: Nauka, 1990. ]

4. Куликов Г. Г., Набатов А. Н., Речкалов А. В. Автоматизированное проектирование информационно-управляющих систем. Уфа: УГАТУ, 1998. 104 с. [ G. G. Kylikov, A. N. Nabatov, A.V. Rechkalov, *Automated design of automated management systems*, (in Russian). Ufa: USATU, 1998. ]

5. Ландэ Д. В., Снарский А.А., Безсуднов И. В. Интернетика: Навигация в сложных сетях: модели и алгоритмы. М.: Либроком, 2009. 264 с. [ D.V. Lande, A.A. Snarsky, I. V. Bezsudnov, *Internetica: navigation in networks*, (in Russian). Moscow: Librocom, 2009. ]

#### ОБ АВТОРАХ

**КУЛИКОВ Геннадий Григорьевич**, проф., зав. АСУ. Дипл. инж. по автом. машиностроения (УАИ, 1971). Д-р техн. наук по сист. анализу, авт. управлению и тепл. двигателям (УАИ, 1989). Иссл. в обл. АСУ и упр. сил. установками ЛА.

**ШИЛИНА Мария Анатольевна**, доц. той же каф. Дипл. информатик-экономист (УГАТУ, 2006). Канд. тех. наук по АСУ (УГАТУ, 2009). Иссл. в обл. упр. учеб. процессом с исп. инф. технологий.

**СТАРЦЕВ Геннадий Владимирович**, доц. той же каф. М-р техн. и технол. по инф.-упр. системам (УГАТУ, 2003). Канд. техн. наук по АСУ (УГАТУ, 2006). Иссл. в обл. проектир. инф. систем на основе веб-технологий.

**БАРМИН Александр Александрович**, асп. каф. АСУ. Дипл. информатик-экономист (УГАТУ, 2011). Готовит дис. в обл. инф.-поисковых систем.

#### METADATA

**Title:** Structuring of subject area of technical university using process approach and semantic identification.

**Authors:** G. G. Kylikov<sup>1</sup>, M. A. Shilina<sup>2</sup>, G. V. Startsev<sup>3</sup>, A. A. Barmin<sup>4</sup>

**Affiliation:**

Ufa State Aviation Technical University (UGATU), Russia.

**Email:** <sup>1</sup>gennadyg\_98@yahoo.com, <sup>2</sup>maria.shilina@gmail.com, <sup>3</sup>startsev@gmail.com, <sup>4</sup>abarmin@outlook.com

**Language:** Russian.

**Source:** Vestnik UGATU (scientific journal of Ufa State Aviation Technical University), vol. 18, no. 4 (65), pp. 115-124, 2014. ISSN 2225-2789 (Online), ISSN 1992-6502 (Print).

**Abstract:** Article present process of identification and tracking of information resources of information area of technical university. Presenting methodic of identification of information resources based on system models, supplement example with information retrieval system.

**Key words:** Business process; category theory; system models; subject area.

**About authors:**

**KULIKOV, Gennady Grivorievich**, Prof., Dept. of Automated Systems. Dipl. Ingeneer automator (UAI, 1987). Dr. of Tech. Sci. (UGATU, 1989).

**SHILINA, Maria Anatolievna**, Dept. of Automated Systems. Informatics and economist (USATU, 2006), PhD (USATU, 2009).

**STARTSEV, Gennady Vladimirovich**, Dept. of Automated Systems. Master of technique and technology (USATU, 2003). PhD (USATU, 2006).

**BARMIN, Aleksandr Aleksandrovich**, Postgrad. (PhD) Student, Dept. of Automated Systems. Informatics and economist (USATU, 2011).