

О. И. Евдошенко, аспирант, Астраханский государственный университет, goronet@list.ru

*А. Г. Кравец, докт. техн. наук, Волгоградский государственный
технический университет, gk@gde.ru*

*И. Ю. Петрова, докт. техн. наук, профессор,
Астраханский инженерно-строительный институт, vtempus2@gmail.com*

Разработка онтологии и базы данных для эффективного поиска научно-технической документации¹

В статье приведена динамика роста числа научных документов. Разработана инфологическая модель научно-технического документа. Поэтапно описаны этапы проектирования базы данных, хранящей сведения о документах. Сформирована классификация классов, подклассов, выявлены характеристики, описывающие данные понятия, и разработана понятийная структура онтологии для эффективного поиска информации в базе данных, которые реализованы с помощью редактора *Fluent Editor 2014*.

Ключевые слова: прием совершенствования эксплуатационной характеристики, научно-техническая документация, база данных, СУБД MySQL, инфологическая модель документа, онтология, редактор онтологий *Fluent Editor 2014*.

Введение

Любое техническое устройство представляет собой сложную целостную систему, характеризующуюся набором взаимосвязанных параметров (характеристик). Значение каждой из этих характеристик может быть улучшено для расширения области применения устройства и повышения его конкурентоспособности по сравнению с аналогами. Для улучшения значения этих характеристик (параметров) технического устройства используются специальные приемы и методы — направленное изменение конструкции, схемы, новые материалы и другие способы, с помощью которых в техническом решении получен положительный эффект по сравнению с прототипом.

Для выявления обобщенных приемов улучшения эксплуатационных характери-

стик инженером-конструктором проводится анализ массивов научно-технических документов в определенной предметной области, к которым можно отнести: патенты, авторские свидетельства, научные статьи, учебные материалы, информацию из Интернета и др.

В [4; 5; 7] предложена методика анализа патентной документации определенного класса/подкласса/группы/подгруппы Международной патентной классификации (МПК) в целях выявления обобщенных приемов, включающая:

- подбор описаний изобретений из нужного класса/подкласса/группы/подгруппы МПК;
- изучение принципа действия и конструктивной реализации изобретения;
- изучение принципа действия и конструкций прототипов выбранных изобретений в целях выявления усовершенствованных узлов и деталей;
- составление параметрической структурной схемы принципа действия устрой-

¹ Работа выполнена при финансовой поддержке РФФИ, грант № 15-37-50292 мол_нр.

ства [12] и ее модификаций для улучшения эксплуатационных характеристик;

- определение эффективности принимаемых технических решений в сравнении с таковыми для прототипов по уравнениям или с помощью качественного сравнения показателей эффективности.

Обзор динамики роста объема научно-технической документации

Патенты, авторские свидетельства

В настоящее время ежегодно к патентным фондам добавляется около 1 млн документов. По данным IP5 (группа из пяти крупнейших ведомств интеллектуальной собственности в мире <http://www.fiveipoffices.org/>), в 2013 г. зарегистрировано 957 000 патентов (на 4% больше, чем в 2012 г.), подано 2,1 млн заявок на регистрацию патентов (на 11% больше, чем в 2012 г.) [11].

В табл. 1 продемонстрирована динамика роста выдачи патентов в Российской Федерации (по данным Федерального института промышленной собственности).

Научные статьи, учебные материалы, информация из Интернета и др.

Динамика роста числа публикаций (научных статей) за 2014–2015 гг. по данным электронной библиотеки eLibrary.ru представлена в табл. 2.

Для эксперимента был проведен патентный анализ по базам данных Всемирной организации интеллектуальной собственности (<http://patentscope.wipo.int/>). В результате по запросу «*биморфные пьезоэлектрические устройства*» было найдено 1216 патентов за период 2004–2014 гг. При аналогичном запросе в базе данных служб международного патентного поиска (www.espacenet.com) — 2495 патентов.

Ориентироваться в таком огромном объеме документов, используемых для анализа в целях выявления приемов улучшения эксплуатационных характеристик инженеру-конструктору сложно, а поиск и анализ каждого документа может быть продолжительным. Это обуславливает острую потребность в создании базы данных, хранящей сведения о научно-технических документах в структурированном виде, которую можно *пополнять, своевременно актуализировать и корректно использовать*.

Цель данной статьи — разработать проект БД, хранящей сведения о научно-технической документации в структурированном виде, и организовать эффективный поиск в этой базе данных с использованием онтологий.

Основным условием создания такой базы данных должна быть возможность ее интеграции с системой «Интеллект» [2; 3] и ба-

Таблица 1. Динамика выдачи патентов в РФ

Table 1. Dynamics of the issue of patents in the Russian Federation

Показатели	2010	2011	2012	2013	2014	2014 в % к 2013
Выдано патентов, всего из них	30 322	29 999	32 880	31 638	33 950	107,31
Российским заявителям	21 627	20 339	22 481	21 378	23 065	107,89
Иностранцами заявителям	8 695	9 660	10 399	10 260	10 885	106,09

Таблица 2. Динамика роста числа научных публикаций

Table 2. Dynamics of growth of number of scientific publications

Год	На 1 января	На 1 апреля	На 1 июля	На 1 октября
2014	18341 516	18643 731	18928 686	19223 528
2015	19877 921	20244 155	20686 361	21 118 033

зой знаний по приемам улучшения эксплуатационных характеристик [6].

Разработка базы данных научно-технических документов

Процесс создания базы данных включает в себя три этапа: концептуальное, логическое и физическое проектирование. Результаты, достигнутые на каждом этапе, представлены ниже.

Концептуальное проектирование базы данных

Концептуальное (инфологическое) проектирование — построение инфологической модели предметной области, т. е. информационной модели наиболее высокого уровня абстракции.

Любой документ обладает структурой с набором полей: название, дата опубликования, авторы и т. д. Полный список полей указан в разработанной инфологической модели документа, которая представлена в следующем виде:

$$D = \{T, H, HЗ, ОР, ДО, A_i^K, МПК_j^M, Ц_z^L, P, I_a^N, ФТЭ_{Num}^R\},$$

где T — тип документа (патент, авторское свидетельство, научная статья);
 H — номер документа <текст>;
 $HЗ$ — название документа <текст>;
 $ОР$ — название организации <текст>;
 $ДО$ — дата опубликования <дата>;
 A_i^K — сведения об авторах: фамилия, имя, отчество, SPIN, Scopus Author ID, RESEARCHERID, РИНЦ Author ID (K — общее количество авторов, i — номер автора в общем перечне эксплуатационных характеристик) <структура данных>;
 $МПК_j^M$ — сведения о МПК: раздел, класс, подкласс, номер, полное название (M — общее количество, j — номер в общем перечне) <структура данных>;
 $Ц_z^L$ — цели, достигаемые в приеме, который описан в документе (L — общее количество, z — номер цели в общем перечне) <массив данных>;
 P — изображение конструкции <рисунок>;

I_a^N — сведения об источнике документа: номер, название, вид источника (статья, электронный ресурс), номера страниц, год опубликования (N — общее количество источников, a — номер источника в общем перечне) <структура данных>;

$ФТЭ_{Num}^R$ — множество физико-технических эффектов, описанных в документе (Num — номер эффекта в БД системы «Интеллект», R — общее количество эффектов) <массив данных>.

Логическое проектирование базы данных

Логическое (даталогическое) проектирование — создание схемы базы данных на основе конкретной модели данных.

На основании выявленных инфологических моделей были определены базовые отношения (табл. 3).

Таблица 3. Базовые отношения и атрибуты отношений

Table 3. Basic relations and attributes of the relations

Отношение	Имена атрибутов отношения
Документ	Тип, номер, название, дата опубликования, цели, источник документа, авторы, МПК, организация, эскиз
Авторы	Фамилия, имя, отчество, SPIN, Scopus Author ID, RESEARCHERID, РИНЦ Author ID
МПК	Раздел, класс, подкласс, номер, полное название
Источник	Номер, название, вид источника, номера страниц, год опубликования, ISSN, ISBN, учредитель

На основе набора схем отношений была построена схема базы данных реляционной модели, изображенная на рис. 1.

Физическое проектирование базы данных

Физическое проектирование — создание схемы базы данных для конкретной СУБД. База данных разработана с использованием СУБД MySQL, обладающей следующими показателями:

— большое разнообразие встроенных функций;

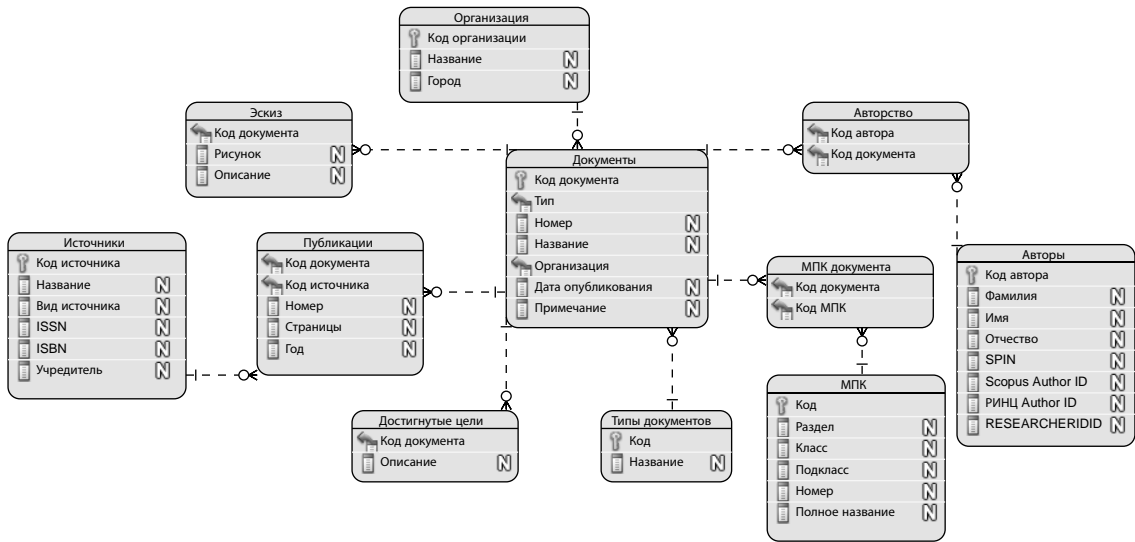


Рис. 1. Схема базы данных
Fig. 1. Database scheme

- наличие средств безопасности как на системном уровне, так и на уровне приложений;
- возможность интеграции с различными языками программирования;
- возможность интеграции с системой «Интеллект» и базой знаний по приемам улучшения эксплуатационных характеристик.

В процессе эксплуатации разработанной базы данных были выявлены следующие недостатки, описанные в работе [1]:

- медленный доступ к данным;
- результатом запроса к базе данных является информация, хранимая в самой БД, отсутствие логического вывода;
- при использовании баз данных проблематично моделировать сложные связи в отличие от онтологического подхода.

Преодолеть перечисленные недостатки баз данных позволяют онтологии, являющиеся современной формой представления знаний и позволяющие производить автоматизированную обработку семантики информации с целью ее эффективного использования (представления, преобразования, поиска).

Построение онтологии для быстрого поиска документов

Онтология — система фундаментальных понятий, которая предоставляет разработчику возможность смоделировать и представлять определенную область мира посредством очевидных определений и таксономических структур [9; 10].

К основным достоинствам применения онтологий можно отнести следующее:

- обеспечение простоты развития системы;
- получение логического вывода на основе хранимых данных;
- возможность применения согласованной (разделяемой всеми) терминологии с точно определенной семантикой, позволяющего интегрировать и использовать данные (информацию) из разных источников;
- возможность динамического изменения данных.

Формальным определением онтологий может служить следующая модель:

$$O = \{C, E, At, R, A\},$$

где *C* — понятия (классы) онтологии;

Таблица 4. Обозначение классов, экземпляров классов и отношений между ними
Table 4. The description of classes, of instances of classes and of relation between them

Классы (C) и экземпляры класса (E)		Отношения между понятиями (R)	
Название класса	Обозначение экземпляра класса	Название отношения	Описание отношения между экземплярами класса
Document (документ)	D-N (N- номер документа)	Type	$E_D = E_{TD}$
Author (автор)	A-N (N — номер автора)	Authorship	$E_D = E_A$
Organization (организация)	Or-N (N- номер организации)	Owner	$E_D = E_{Or}$
Typedoc (тип документа)	Td-N (N-номер документа)	Publication	$E_D = E_S$
Source (источник документа)	S-N (N-номер документа)		

E — экземпляры онтологии;
At — атрибуты понятий и экземпляров онтологии;
R — отношения между понятиями;
A — аксиомы онтологии.

Создание онтологий — сложный и интерактивный процесс. В нем участвуют эксперты конкретных предметных областей и специалисты инженерии знаний. На сегодняшний день разработаны подходы, позволяющие до определенной степени автоматизировать этот процесс. Однако подавляющее большинство существующих онтологий разработано «вручную» с применением специальных технических средств — редакторов онтологий.

Для построения онтологии использовался редактор *Fluent Editor 2014 (Cognitum)*, который обладает характеристиками, описанными в [8].

На основании модели документа необходимо создать классы онтологии и основные отношения между понятиями в редакторе *Fluent Editor 2014*, представленные в табл. 4.

Фрагмент описания онтологии на *Controlled English* для добавления нового документа представлен в листинге 1.

Листинг 1. Описание онтологии в *Fluent Editor*

Listing 1. The description of ontology in *Fluent Editor*

```
A-1 is an author.
D-1 is a document.
Or-1 is an organization.
S-1 is a source.
Td-1 is a typedoc.
A-1 has-authorships D-1.
D-1 includes Pr-9 and numbers equal-to '2472253'.
Fte-10 has-opisans D-1.
Or-1 has-owners D-1.
S-1 has-publications D-1.
Td-1 types D-1.
A-1 has-fios equal-to 'Головнин В. А.'.
D-1 has-datas equal-to '17.08.2011'.
D-1 has-mpks equal-to 'H01L41/08'.
D-1 names equal-to 'Пьезоэлектрический прибор и способ его изготовления'.
Or-1 names equal-to 'Тверской государственный университет (RU) '.
S-1 names equal-to 'Официальный бюллетень «Изобретения. Полезные модели» №1/2013'.
Td-1 has-typedocnames equal-to 'патент'.
```

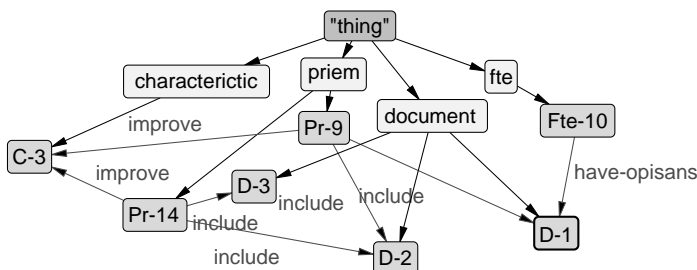


Рис. 2. Фрагмент онтографа для поиска научно-технической документации:

characteristic — класс Характеристики; *document* — класс Документы; *priem* — класс Приемы; *fte* — класс ФТЭ; *Pr-9*, *Pr-14* — экземпляры класса Приемы; *Fte-10* — экземпляр класса ФТЭ; *C-3* — экземпляр класса Характеристики; *D-1*, *D-2*, *D-3* — экземпляры класса Документы

Fig. 2. The fragment of OntoGraf for search of scientific and technical documentation:

characteristic — class of Characteristics; *document* — class of Documents; *priem* — class of Receptions; *fte* — class of PTE; *Pr-9*, *Pr-14* — instances of classes of Receptions; *Fte-10* — instances of classes of PTE; *C-3* — instances of classes of Characteristics; *D-1*, *D-2*, *D-3* — instances of classes of Documents

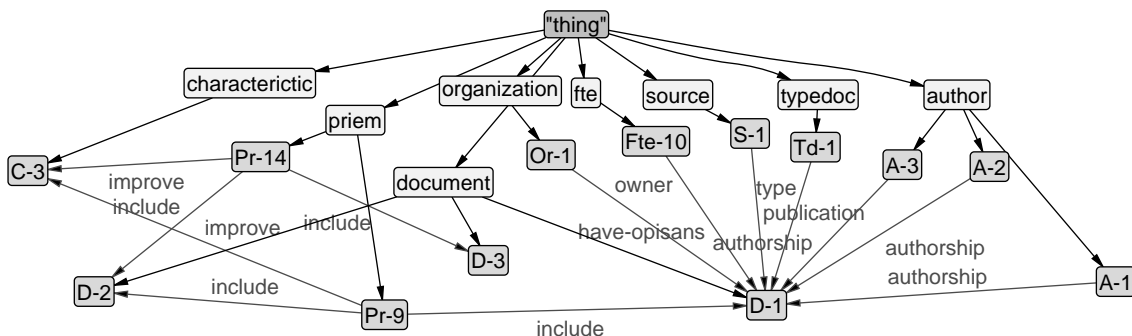


Рис. 3. Фрагмент онтографа после интеграции онтологий по приемам улучшения эксплуатационных характеристик и научно-технической документации:

organization — класс Организации; *typedoc* — класс Типы документов; *author* — класс Авторы; *A-1*, *A-2*, *A-3* — экземпляры класса Авторы; *Or-1* — экземпляр класса Организации; *S-1* — экземпляр класса источники; *Td-1* — экземпляр класса Типы документов

Fig. 3. The fragment of OntoGraf after integration of ontologies on methods of improvement of operational characteristics and scientific and technical documentation:

organization — class of Organizations; *typedoc* — class of types of Documents; *author* — class of Authors; *A-1*, *A-2*, *A-3* — instances of classes of Authors; *Or-1* — instances of classes of Organizations; *S-1* — instances of classes of source; *Td-1* — instances of classes of types of documents

Materialized Graph - result					Result
Who-or-What improves C-3 ?					
Прием	Описание	Название документа	Дата	Номер	МПК
Pr-9	Центральная прокладка изготавливается из бериллиевой бронзы	Пьезоэлектрический прибор и способ его изготовления	17.08.2011	2472253	H01L41/08
Pr-14	Ассиметричный биморф, состоящий из металлической пластины с приклеенным к ней поляризованным по толщине плоским пьезоэлементом	Комбинированный акустический приемник	27.01.2005	2245604	H04R17/00

Рис. 4. Пример запроса к онтологии и отображение результата

Fig. 4. A query example to the ontology and display of result

Аналогично добавлены все проанализированные документы, и в результате была создана онтология, которая на данный момент содержит сведения более чем о ста научно-технических документах.

Интеграция онтологий по приемам улучшения эксплуатационных характеристик и научно-технической документации

Разработанная онтология была интегрирована с онтологией по приемам улучшения эксплуатационных характеристик. Фрагменты графов онтологии по приемам улучшения эксплуатационных характеристик до и после интеграции с онтологией по научно-технической документации представлены на рис. 2 и 3.

На рис. 4 отображен результат запроса к онтологии в среде *Fluent Editor* на выбор приемов улучшения заданной эксплуатационной характеристики. По результатам запроса найдено два приема улучшения механической прочности биморфной пластины и представлены подробные сведения о документах.

Таким образом, пользователь может получить не только информацию о приеме улучшения характеристики (название, ссылка на документ), но и подробную информацию о самом документе.

Заключение

Авторами проведен обзор динамики роста объема научно-технической документации по данным научной электронной библиотеки и патентных ведомств. В результате проведенного исследования на концептуальном этапе проектирования была сформулирована инфологическая модель документа, на основании которой были разработаны онтология в среде *Fluent Editor* и база данных в СУБД MySQL, позволяющие оперативно проводить информационный поиск научно-технических документов, определять взаимосвязь между ними по авторам, МПК, типам и источникам. Созданная онтология интегрирована с онтологией по приемам

улучшения эксплуатационных характеристик и системой «Интеллект». Применение онтологии облегчает информационный поиск экспертам и инженерам-конструкторам при создании нового технического устройства и оценке его патентоспособности (по новизне и изобретательскому уровню).

Список литературы

1. Гончар А. Д. Сравнительный анализ баз данных и баз знаний (онтологий) применимо к моделированию сложных процессов // Современные научные исследования и инновации. 2014. № 5. URL: <http://web.snauka.ru/issues/2014/05/34325> (дата обращения: 25.09.2015).
2. Зарипова В. М., Лунев А. П., Петрова И. Ю. Научить инновационному мышлению — задача университета // Инновации. 2012. № 11 (69). С. 62–69.
3. Зарипова В. М., Цырульников Е. С., Киселев А. А. «Интеллект» для развития навыков инженерного творчества // Alma Mater (Вестник высшей школы). 2012 (1). Изд-во РУДН.
4. Петрова И. Ю. Микроэлементы систем управления с распределенными параметрами различной физической природы. М.: Наука, 1979. — 111 с.
5. Петрова И. Ю., Гурская Т. Г. Приемы усовершенствования электрокинетических преобразователей // Датчики и системы. 2007. № 10. С. 18–21.
6. Петрова И. Ю., Евдошенко О. И., Лежнина Ю. А. Концептуальная модель подсистемы выбора приемов улучшения эксплуатационных характеристик технических устройств // Вестник Оренбургского государственного университета. 2015. № 3 (178). С. 249–258.
7. Петрова И. Ю., Евдошенко О. И., Зарипова В. М., Гурская Т. Г. Приемы совершенствования эксплуатационных характеристик биморфных сенсоров и актюаторов // Прикаспийский журнал: управление и высокие технологии. 2014. Вып. 4 (28). С. 213–226.
8. *Cognitum. Fluent Editor*. URL: <http://www.cognitum.eu/semantics/FluentEditor/> (дата обращения: 05.09.2015).
9. Guarino N, Garetta P. Towards very large knowledge bases — knowledge building and knowledge sharing. IOS Press, 1995. P. 25–32 (Chap. Ontologies and knowledge bases — towards a terminological clarification).
10. Huhns M, Singh M. Ontologies for agents. Internet Computing, IEEE. 1997. 1 (6):81–3.
11. IP5 Statistics Report 2013 Edition <http://www.fivepofices.org/statistics/statisticsreports/2013edition.html> (дата обращения: 08.09.2015 г.).
12. Zariyova V., Petrova I. System of Conceptual Design Based on Energy-Informational Model // PROGRESS IN SYSTEMS ENGINEERING, Proceedings of the the 23rd International Conference on Systems Engineering, August, 2014, Las Vegas, NV, Series: Advances in Intelligent Systems and Computing. 2015. Vol. 1089. P. 365–373.

References

- Gonchar A. D. *Comparative analysis of databases and knowledge bases (ontologies) applicable to modeling complicated process*. Available at: <http://web.snauka.ru/issues/2014/05/34325> (accessed: 25.09.2015) (in Russian).
- Zaripova V. M., Lounev A. P., Petrova I. Yu. The university task is to train in innovative thinking. *Innovatsii — Innovations*, 2012, no. 11 (69), pp. 62–69 (in Russian).
- Zaripova V. M., Tsyurul'nikov E. S., Kiselev A. A. «Intellekt» dlya razvitiya navykov inzhenerenogo tvorchestva [«Intelligence» for development of skills of engineering creativity]. *Alma Mater (Vestnik vysshei shkoly — Bulletin of the higher school)*, 2012 (1), RUDN Publ.
- Petrova I. Yu. *Mikroelementy sistem upravleniya s raspredelennymi parametrami razlichnoi fizicheskoi prirody* [Microcells of control systems with the distributed parameters of various physical nature]. M., Nauka Publ., 1979. 111 p.
- Petrova I. Yu., Gurskaya T. G. The ways to improve electrokinetic transducers. *Datchiki i sistemy — Sensors and systems*, 2007, no. 10, pp. 18–21 (in Russian).
- Petrova I. Yu., Yevdoshenko O. I., Lezhnina Y. A. Conceptual model of the subsystem of the choice of methods of improvement of operational characteristics of technical devices. *Vestnik orenburgskogo gosudarstvennogo universiteta*, 2015, no. 3 (178), pp. 249–258 (in Russian).
- Petrova I. Yu., Yevdoshenko O. I., Zaripova V. M., Gurskaya T. G. Methods improvement of operational characteristics of bimorph sensors and actuators. *Prikladnyy zhurnal: upravlenie i vysokie tekhnologii*, 2014, no. 4 (28), pp. 213–226 (in Russian).
- Cognitum. Fluent Editor*. URL: <http://www.cognitum.eu/semantics/FluentEditor/> (accessed: 05.09.2015).
- Guarino N, Giarretta P. *Towards very large knowledge bases — knowledge building and knowledge sharing*. IOS Press, 1995, pp. 25–32 (Chap. Ontologies and knowledge bases — towards a terminological clarification).
- Huhns M, Singh M. *Ontologies for agents. Internet Computing*. IEEE, 1997, 1 (6):81–3.
- IP5 Statistics Report 2013 Edition*. Available: <http://www.fiveipoffices.org/statistics/statisticsreports/2013edition.html> (accessed: 08.09.2015).
- Zaripova V., Petrova I. System of Conceptual Design Based on Energy-Informational Model //PROGRESS IN SYSTEMS ENGINEERING, Proceedings of the the 23rd International Conference on Systems Engineering, August, 2014, Las Vegas, NV, Series: Advances in Intelligent Systems and Computing, 2015, vol. 1089, pp. 365–373.

O. Evdoshenko, Astrakhan State University, Astrakhan, Russia, goronet@list.ru

A. Kravets, Volgograd State Technical University, Volgograd, Russia, agk@gde.ru

I. Petrova, Astrakhan Civil Engineering Institute, Astrakhan, Russia, vtempus2@gmail.com

Development of ontology and database for the effective search of scientific and technical documentation

Engineers have to use special methods for improvement of value of operational characteristics (parameters) of the technical device. The design engineer is to carry out the analysis of arrays of scientific and technical documents in a certain subject subarea for search of such methods. To be guided in the huge volume of the documents used for the purpose of search of methods to the expert is difficult, and search and the analysis of each document can be long. The purpose of work is to create the databases project for storing data of scientific and technical documentation in the structured view and to organize effective search in this database with use of ontologies. Authors have developed the logical model of the document and have described stages of database design. Authors created the ontology for the automated processing of semantics of information and its effective use. Classification of classes, subclasses is formed. Characteristics are identified that describe these concepts. Conceptual structure of the ontology is developed and designed in Ontorion Fluent Editor. Process of creation of classes of ontology, instances of classes and relations between instances is described. The created ontology is integrated with ontology «generalized methods of improvement of operational characteristics, system «Intelligence» and allows the user to receive not only information on reception of improvement of the characteristic (the name, the link to the document), but also detailed information on the most scientific and technical document.

Keywords: improving performance, scientific and technical documentation, database, MySQL Workbench, logical models of document, ontology, ontology editor Fluent Editor 2014.

About authors: O. Evdoshenko, Postgraduate; A. Kravets, Dr of Technique; I. Petrova, Dr of Technique, Professor

For citation: Evdoshenko O., Kravets A., Petrova I. Development of ontology and database for the effective search of scientific and technical documentation. *Prikladnaya Informatika — Journal of Applied Informatics*, 2015, vol. 10, no. 5 (59), pp. 85–92 (in Russian).