

УДК 001(043.3/.5):004.5

**Гогунский В.Д.**, д-р техн. наук, проф.;

**Коляда А.С.**, ассистент;

кафедра управления системами безопасности жизнедеятельности,  
Одесский национальный политехнический университет;

---

## ИНТЕГРАЛЬНАЯ ОЦЕНКА ПУБЛИКАЦИОННОЙ АКТИВНОСТИ В ОСНОВНЫХ НАУКОМЕТРИЧЕСКИХ БАЗАХ

---

*В.Д. Гогунський, А.С. Коляда. Інтегральна оцінка активності публікацій в основних наукометричних базах.* Показана можливість виконання інтегральної оцінки публікаційної активності авторів за допомогою розробленого програмного продукту, як одного з інструментів моніторингу публікаційної активності наукових співробітників. Розв'язана задача витягання і візуалізації метаданих публікацій, які індексуються в основних міжнародних наукометричних базах даних.

**Ключові слова:** публікації, моніторинг, наукометричних баз, візуалізація.

*В.Д. Гогунский, А.С. Коляда. Интегральная оценка публикационной активности в основных наукометрических базах.* Показана возможность выполнения интегральной оценки публикационной активности авторов с помощью разработанного программного продукта, как одного из инструментов мониторинга публикационной активности научных сотрудников. Решена задача извлечения и визуализации метаданных публикаций, которые индексируются в основных международных наукометрических базах данных.

**Ключевые слова:** публикации, мониторинг, наукометрические базы, визуализация.

*V.D. Gogunsky, A.S. Kolyada. Integral assessment activity in major publications scientometric databases.* The possibility of implementation of integrated assessment Publication activity developed by the authors using the software as a tool for monitoring activity Publication researchers. The problem of extracting metadata and visualization of publications indexed in major international scientometric databases.

**Keywords:** publishing, monitoring, scientometric database, visualization.

Задачей данного программного продукта является предоставить список публикаций соискателя, которые индексируются в международных наукометрических базах данных. Одной из первых стадий разработки программного проекта является сбор информации, анализ, спецификация, и проверка требований к программному обеспечению [1 - 11].

Опыт разработки известных информационных технологий показывает, что вопросы, связанные с управлением требованиями, оказывают критически важное влияние на программные проекты, в определенной степени – на сам факт возможности успешного завершения проектов [12 - 17].

Требованиями к функциям данного проекта являются [18 - 26]:

– извлечение информации из веб страниц;

- критерием информации является ФИО автора;
- работа с наиболее распространенными наукометрическими базами данных: Scopus, Web of Science, Copernicus, BASE, Science Index и др.;
- обработка и анализ результатов для определения нерелевантной информации и фильтрации ее;
- предоставление информации пользователю.

#### *Программный проект*

Определения архитектуры, компонентов, интерфейсов и других характеристик системы и ее компонентов называется проектированием. Результат процесса проектирования – дизайн [27 - 32]. Проектирование - инженерная деятельность, в которой надлежащим образом анализируются требования для создания описания внутренней структуры ПО и являющейся основой для его конструирования [24, 33, 34]. Программный дизайн (как результат деятельности проектирования) должен описывать архитектуру программного обеспечения, то есть представлять декомпозицию программной системы в виде организованной структуры компонент и интерфейсов между ними [13]. Важнейшей характеристикой готовности дизайна является тот уровень детализации компонентов, который позволяет заняться их конструированием. Проектирование программных систем можно рассматривать как деятельность, результат которой состоит из двух составных частей:

- архитектурный или высокоуровневый дизайн – описание высокоуровневой структуры и организации компонентов системы;
- детализированный дизайн – описывающий каждый компонент в том объеме, который необходим для конструирования.

Разделяют следующие виды дизайна:

- D-дизайн – декомпозиция структуры программного обеспечения в виде набора фрагментов или компонент;
- FR-дизайн – семейство архитектурных представлений, базирующихся на шаблонах;
- I-дизайн – создание высокоуровневой концепции, видения того, что из себя будет представлять программная система - данный вид дизайна является результатом анализа требований и их трансформации в подходы к реализации.

Проектирование программного обеспечения в понимании программной инженерии подразумевает D- и FR-дизайн. I-дизайн в большей степени относится к работе с программными требованиями.

#### *Высокоуровневый дизайн*

Декомпозиция структуры программного проекта в виде набора компонент представлена на рис 1. Проект представляет собой программный комплекс из нескольких приложений, взаимодействие которых предоставляет сервис поиска и извлечения публикаций указанного автора. Основными компонентами системы являются:

- веб приложение SMD (Science Metric Databases);
- веб сервис Scrapyd;
- дополнительные компоненты, с которыми работают основные это:

- сервер реляційної бд mysql;
- сервер NoSQL БД MongoDB.

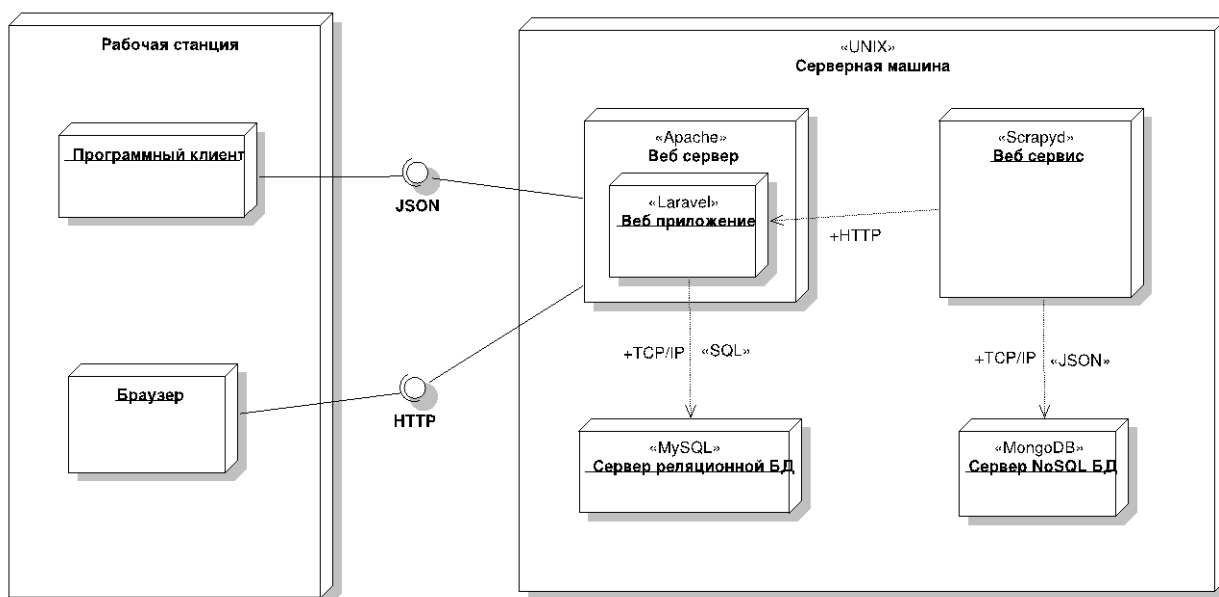


Рисунок 1 - Архитектура программного проекта по извлечению публикаций

Веб приложение SMD представляет собой графический интерфейс пользователя, а также предоставляет программный интерфейс для использования поиска публикаций другими приложениями. Веб сервис Scrapyd представляет сервис по извлечению структурированных данных из НМБД, а также управляет запуском соответствующих программ-пауков отдельной для каждой НМБД. Таким образом, функционал программной системы разделен на отдельные модули — приложения, который работают независимо друг от друга. Веб приложение SMD использует сервис Scrapyd во время для поиска публикаций по запросу пользователя. Эти приложения общаются между собой по HTTP протоколу в JSON формате.

Веб приложение SMD использует реляционную базу данных (MySQL) в качестве хранилища данных, таких как, информация о пользователях, список поддерживаемых НМБД, история результатов поиска публикаций и др. Веб сервис Scrapyd использует документо-ориентированную базу данных (NoSQL) для временного хранения результатов поиска на внешнем диске, таким образом, не увеличивая объем использования оперативной памяти при извлечении большого количества публикаций. Доступ к базам данных предоставляют отдельные приложения — СУБД, с которыми приложения работают по протоколу TCP/IP.

Работа с приложением выполняется с помощью веб браузера. Также есть программный доступ к интерфейсу в формате JSON.

Основные варианты использования включают такие основные варианты использования приложения:

- регистрация пользователей в системе – создание учетной записи пользователя для привязки найденных публикаций к пользователю;
- поиск публикаций – основной вариант использования, который включает в себя несколько этапов и участие двух актеров, которые инициируют действия.

С одной стороны это пользователь, который запускает поиск по заданным параметрам, с другой стороны это сервис поиска (scrapyd), который управляет этим процессом. Основные этапы поиска публикаций это извлечение информации, ее анализ (включая латентно-семантический) и сохранение результатов;

- история поиска публикаций – навигация по истории выполненных поисковых запросов;
- просмотр результатов поиска состоит из двух вариантов использования;
- привязка публикаций к пользователю и отображение статистики по найденным публикациям или публикациям привязанных к пользователю.

#### *Используемые средства и технологии*

В данном проекте используются несколько языков программирования, различные библиотеки и приложения, которые отображены в табл. 1. Колонка «Где используется» показывает, какой из двух основных компонентов использует ту или иную технологию:

- **PHP** (*PHP: Hypertext Preprocessor*) – скриптовый язык программирования общего назначения, интенсивно применяемый для разработки веб-приложений.

*Таблица 1 - Используемые средства и технологии*

Категория	Значение	Где используется
Языки программирования	PHP, Javascript	Веб приложение
	Python	Веб сервис
Фреймворки, библиотеки	Laravel	Веб приложение
	Guzzle	Веб приложение
	Scrapy	Веб сервис
	Selenium WebDriver	Веб сервис
	Gensim NLTK	Веб сервис Веб сервис
Внешние приложения	Apache web server	Веб приложение
	PhantomJS	Веб сервис
	MySQL server	Веб приложение
	MongoDB server	Веб сервис
	Scrapyd	Веб сервис

- **JavaScript** – прототипно-ориентированный сценарный язык программирования. JavaScript обычно используется как встраиваемый язык для программного доступа к объектам приложений. Наиболее широкое применение находит в браузерах как язык сценариев для придания интерактивности веб-страницам.

– **Python** – высокоуровневый язык программирования общего назначения, ориентированный на повышение производительности разработчика и читаемости кода.

– **Laravel** – бесплатный веб-фреймворк с открытым кодом, предназначенный для разработки с использованием архитектурной модели MVC (Model View Controller — модель-представление-контроллер).

– **Guzzle** – библиотека для PHP с помощью которой легко слать HTTP запросы и незатруднительно интегрировать приложение с веб сервисами.

– **Scrapy** это фреймворк для обхода веб-сайтов и извлечения структурированных данных, которые могут быть использованы для широкого приложений.

– **Selenium** – это инструмент для тестирования Web-приложений. **Selenium WebDriver API** используется для доступа к браузеру.

– **Gensim** является библиотекой на языке программирование Python и предназначена для автоматического извлечения семантических тем из документов. Алгоритмы в gensim: латентного семантический анализ, латентный распределения Дирихле.

– **NLTK** (Natural Language Toolkit) — набор библиотек и программ для символической и статистической обработки естественного языка на языке программирования Python.

– Apache HTTP-сервер – свободный веб-сервер.

– **PhantomJS** — скриптовый браузер без графического интерфейса, используемый для автоматизации взаимодействия с веб-страницами.

– **MySQL** – свободная реляционная система управления базами данных.

– **MongoDB** – документо-ориентированная система управления базами данных (СУБД) с открытым исходным кодом, не требующая описания схемы таблиц.

– **Scrapyd** представляет собой приложение для развертывания и запуска SCRAPY пауков. Это позволяет развертывать ваши проекты и контролировать своих пауков с помощью JSON API.

#### *Графический интерфейс пользователя*

Интерфейс пользователя веб приложения представлен в виде веб страниц (html+css+javascript), которые браузер обрабатывает и отображает результат в графическом виде. На рис 3 показана главная или корневая страница приложения, где находится приветствие и список наукометрических баз данных с которыми работает приложение. Панель навигации в шапке общая для всех страниц этого приложения, где можно переключиться на страницу задания параметров поиска или просмотреть историю предыдущих поисковых запросов с их результатами. Так же панель навигации содержит кнопки для входа зарегистрированных пользователей и регистрации новых.

Страница формы поиска публикаций показана на рис. 4. Здесь задается ФИО автора, ключевые слова его публикаций и выбирается список НМБД, по которым будет выполняться поиск. При вводе вариантов (+) ФИО автора можно использовать несколько вариантов, например, на разных языках.

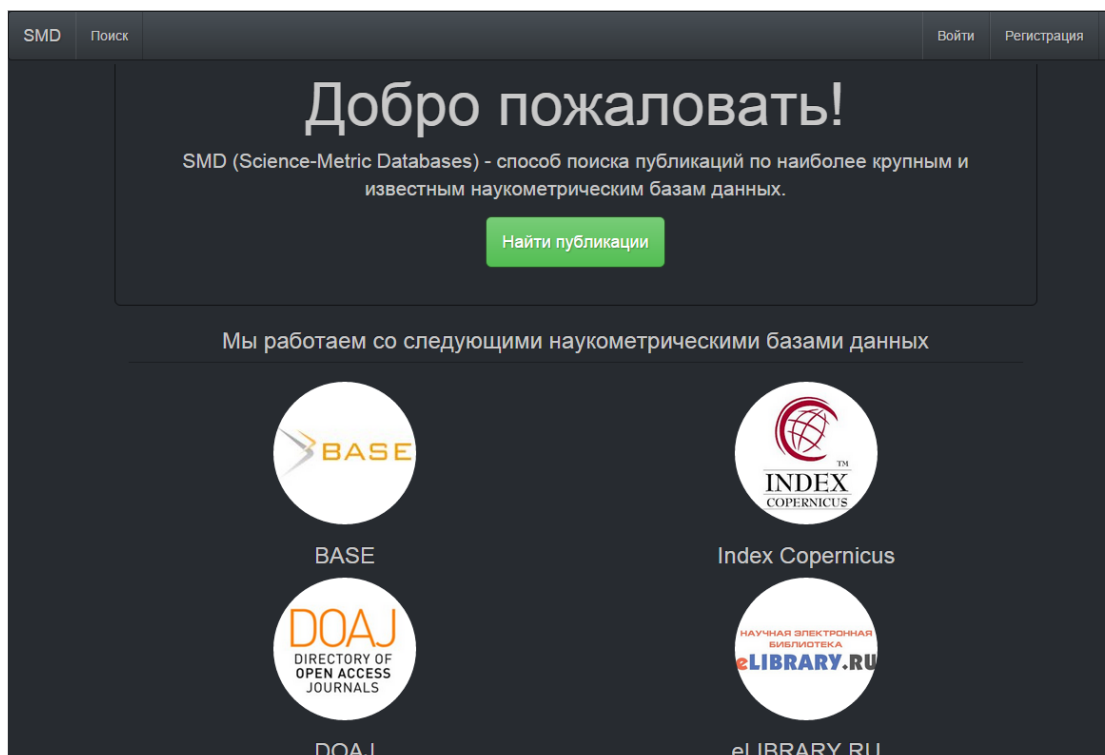


Рисунок 3 – Фрагмент головной формы системы

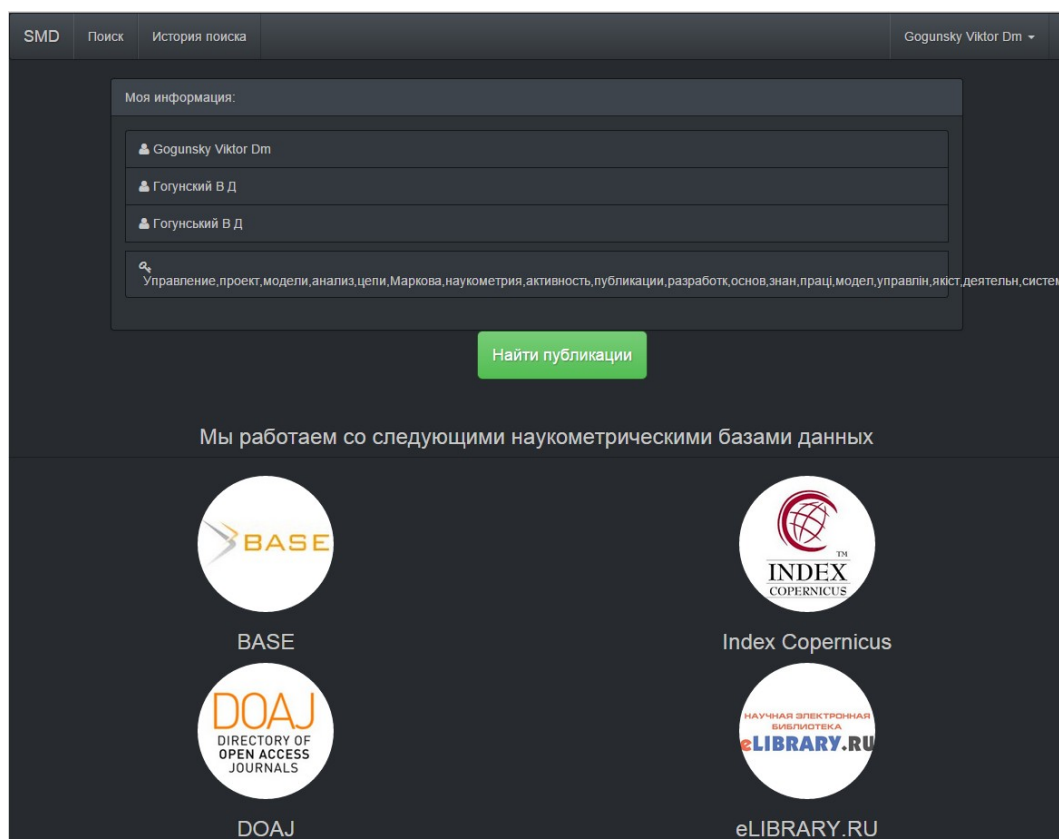


Рисунок 4 – Фрагмент формы поиска

Страница результатов содержит список публикаций с мета-информацией, а также статистику публикаций по НМБД и по дате.

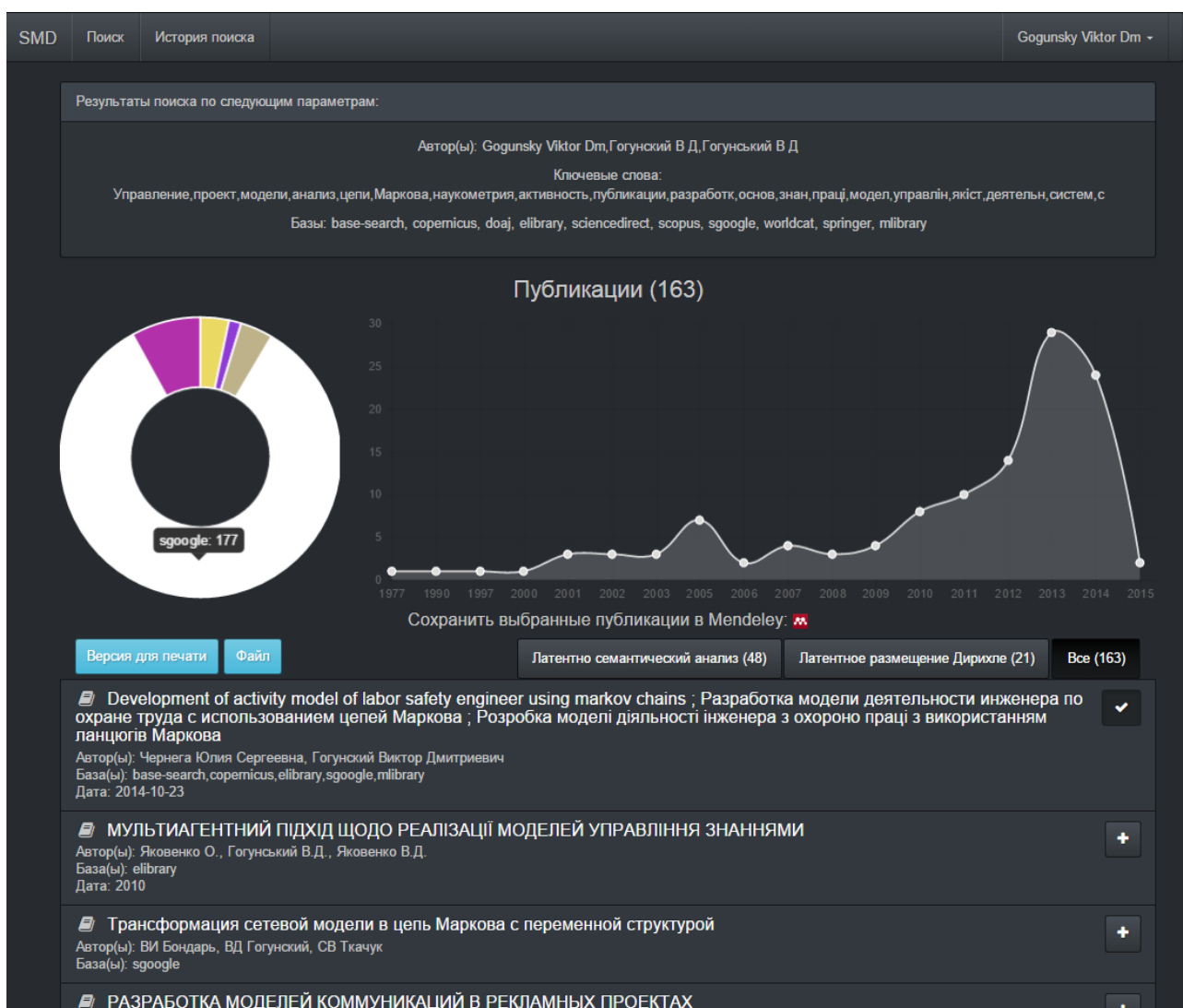


Рисунок 5 – Фрагмент формы отображения результатов поиска

Предусмотрено три варианта отображения данных поиска: результаты, полученные с помощью латентно-семантического анализа; данные размещения Дирихле; все публикации. Выявлено, что результаты латентно-семантического анализа и данные размещения Дирихле в значительной степени зависят от совокупности задаваемых ключевых слов. Поэтому в SMD предусмотрена процедура автоматического формирования набора ключевых слов.

Полученные результаты можно распечатать, записать в файл, а также сохранить полученные данные в системе Mendeley.

*Выводы.* Показана возможность выполнения интегральной оценки публикационной активности авторов с помощью разработанного программного продукта, как одного из инструментов мониторинга публикационной активности научных сотрудников. Решена задача извлечения и визуализации метаданных публикаций, которые индексируются в международных наукометрических базах данных. Программный продукт состоит из веб приложения и веб сервиса, которые взаимодействуют между собой. Веб сервис реализует поиск и извлечение данных, а веб приложение предоставляет графический интерфейс пользователя, который отображает найденные публикации и содержит интерфейс инициализации поиска. Веб приложение

также предоставляет программный интерфейс для возможного автоматизированного использования его.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Костирко, Т. Н. Університети України: приєднання до руху відкритого доступу [Текст] // Вісник ОНУ. – 2011. – Том 16. – Випуск 1/2 (5/6). – С. 283 – 289.
2. Негри, А. А. Концепция проекта агрегирующей аналитической информационной системы для работы с наукометрическими базами данных [Текст] / А. А. Негри, Е. В. Колесникова, Ю.С. Барчанова // Информ. технології в освіті, науці та виробництві: зб. наук. праць. – 2013. - № 4(5). - С. 52 – 56.
3. Коляда, А. С. Извлечение информации из слабо структурированных Веб страниц [Текст] / А. С. Коляда, В. Д. Гогунский // Вост.-Европ. журнал передовых технологий. – 2014. - № 1/9 (67). – С. 51 – 54.
4. Kolesnikova, K.V. The development of the theory of project management: project initiation study law [Text] / K.V.Kolesnikova // Management of development of complex systems. – 2013. - № 17. – С. 24 – 30.
5. Бурков, В. Н. Параметры цитируемости научных публикаций в наукометрических базах данных [Текст] / В. Н. Бурков, А. А. Белощицкий, В. Д. Гогунский // Управління розвитком складних систем. - 2013. - № 15. - С. 134 - 139.
6. Бушуев, С.Д. Наукометричні бази: характеристика, можливості і завдання [Текст] / С.Д. Бушуев, А.О. Білощицький, В.Д. Гогунський // Управління розвитком складних систем. – 2014. - № 18. – С. 145 – 152.
7. Білощицький, А. О. Наукометричні бази та індикатори цитування наукових публікацій [Текст] / А. О. Білощицький, В. Д. Гогунський // Інформаційні технології в освіті, науці та виробництві. – № 4 (5). – О. : АО Бахва, 2013. – С. 198 – 203.
8. Буй, Д.Б. Scopus та інші наукометричні бази: прості питання та нечіткі відповіді [Текст] / Д.Б. Буй, А.О. Білощицький, В.Д. Гогунський // Вища школа. – 2014. - № 4. – С. 37 -40.
9. Новиков, Д. А. Наукометрия и экспертиза в управлении наукой [Текст] / Д. А. Новиков, М. В. Губко // Упр. больш. сист. «Наукометрия и экспертиза в управлении наукой». — М. : ИПУ РАН, 2013. - Спецвыпуск, № 44. - С. 8 - 13.
10. Моніторинг публікаційної активності науковців та кафедр вищих навчальних закладів [Текст] / В.М. Тонконогий, К.В. Колеснікова, К.О. Логінова, А.О. Негрі // Информ. технології в освіті, науці та виробництві. – 2013. - С. 196 -209.
11. Гогунский, В.Д. Наукометрические данные научного издания «Управление развитием сложных систем» [Текст] / В.Д. Гогунский, А.С. Коляда, В.А. Яковенко // Управління розвитком складних систем. – 2014. - №19. – С. 6 – 11.
12. Колесникова, Е.В. Теория проектного управления: закон контроля параметров риска / Е.В. Колесникова // Вісник Одес. нац. морського ун-ту. – 2013. - № 3 (39). – С. 220 – 232
13. Колесникова, Е.В. Управление знаниями в IT-проектах [Текст] / Е.В. Колесникова, А.А. Негри // Вост.-Европ. журн. передовых технологий. – 2013. – № 1/10 (61). – С. 213 – 215.
14. Гогунський, В.Д. Наукометричні бази: характеристика, можливості і завдання [Текст] / В.Д. Гогунський, А.С. Коляда, Г.О. Оборський // Шляхи реалізації кредитно-модульної системи. – 2014. - № 8. – С. 3 – 12.
15. Оборський, Г. О. Scopus: достовірність даних за запитами щодо числа публікацій університетів [Текст] / Г. О. Оборський, В. Д. Гогунський, В. А. Волобоев // Информ. технології в освіті, науці та виробництві : зб. - 2014. – № 2 (7). - С. 179 – 290.
16. Оборський, Г.О. Стандартизація і сертифікація процесів управління якістю освіти у вищому навчальному закладі [Текст] / Г.О. Оборський, В.Д.Гогунський, О.С. Савельєва // Тр. Одес. политехн. ун-та. - 2013. - № 1 (35). - С. 252 – 256.
17. Гогунський, В. Д. Створюємо свій акаунт “GOOGLE Академія” [Текст] / В.Д. Гогунський, О. Є. Колесніков // Вища школа. – 2014. - № 9. – С. 55 - 58.



18. Коляда, А. С. Автоматизация извлечения информации из наукометрических баз данных [Текст] / А. С. Коляда, В. Д. Гогунский // Управління розвитком складних систем. – 2013. - № 16. – С. 96 – 99.
19. Вайсман, В.О. Сучасна концепція проектно-орієнтованого командного управління підприємством [Текст] / В.О. Вайсман, К.В. Колеснікова, В.В. Натальчишин // Сучасні технології в машинобудуванні : зб. наук. праць. – 2013. - № 8. – С. 246 - 253.
20. Оборський, Г.О. Нові тенденції і завдання щодо підготовки науковців вищої кваліфікації [Текст] / Г.О. Оборський, В.Д. Гогунський // Інформ. технології в освіті, науці та виробництві : зб. наук. праць. – Вип. 2. – Одеса : АО Бахва, 2013 - С. 15 – 22.
21. Лизунов, П.П. Проектно-векторное управление высшими учебными заведениями [Текст] / П.П. Лизунов, А.О. Белощицкий, С.В. Белощицкая // Управління розвитком складних систем. – 2011. - № 6. – С. 135 – 139.
22. Бушуев, С. Д. Напрями дисертаційних наукових досліджень зі спеціальності «Управління проектами та програмами» [Текст] / С. Д. Бушуев, В. Д. Гогунський, К. В. Кошкін // Управління розвитком складних систем. – 2012. - № 12. – С. 5 – 7.
23. Колесніков, О.Є. Основні аспекти впровадження дистанційної освіти [Текст] / О.Є. Колесніков, В.Д. Гогунський // Інформаційні технології в освіті, науці та виробництві: зб. – 2012. - № 1. – С. 34 – 41.
24. Белощицкий, А. А. Управление проблемами в методологии проектно-векторного управления образовательными средами [Текст] / А.А. Белощицкий // Управління розвитком складних систем. - 2012. – № 9. – С. 104 – 107.
25. Оборский, Г.А. Актуальность дистанционного обучения [Текст] / Г.А. Оборский, А.Е. Колесников, В.А. Граменицкий // Шляхи реалізації кредитно-модульної системи. – 2013. - № 7. – С. 3 - 8.
26. Коляда, А. С. Применение латентного размещения Дирихле для анализа публикаций из наукометрических баз данных [Текст] / А. С. Коляда, В.А. Яковенко, В. Д. Гогунский // Тр. Одес. политехн. ун-та. - 2014. – 1 (43). – С. 186 – 191.
27. Коляда, А. С. Достоверность идентификации авторства научных публикаций на основе латентно семантического анализа [Текст] / А. С. Коляда, В. Д. Гогунский // Восточно - Европейский журнал передовых технологий. - 2014 – № 3/2 (69). – С. 36 – 40.
28. Гогунський, В.Д. Разработка концепции системы наукометрической базы данных [Текст] / В.Д. Гогунський, В.А. Яковенко А.С. Коляда // Управління розвитком складних систем. – 2014. – № 20. – С. 143 – 147.
29. Гогунський, В. SCOPUS: знайдемо свої публікації [Текст] / В.Д. Гогунський, Д.Б. Буй // Вища школа. – 2014. - № 8. – С. 113 - 115.
30. Гогунський, В.Д. Наукометричні бази: характеристика, можливості і завдання [Текст] / В.Д. Гогунський, А.С. Коляда, Г.О. Оборський // Шляхи реалізації кредитно-модульної системи. – 2014. - № 8. – С. 3 – 12.
31. Vaysman, V.A. The planar graphs closed cycles determination method / V.A. Vaysman, E.V. Kolesnikova, D.V. Lukianov // Тр. Одес. политехн. ун-та. - 2012. - № 1 (38). - С. 222-226.
32. Колеснікова, К. В. Матричная диаграмма и «сильная связность» индикаторов ценности в проектах / К. В. Колеснікова, Т. М. Олех // Электротехнические и компьютерные системы. – № 7(83). – К. : Техніка, 2012. – С. 148 – 153.
33. Колесников, А.Е. [Формирование информационной среды университета для дистанционного обучения](#) / А.Е. Колесников // Управління розвитком складних систем. – 2014. - № 20. – С. 21 – 26.
34. Оборский, Г.А. Наукометрические исследования публикационной активности как составляющая инновационного развития университета [Текст] / Г.А. Оборский, В.М. Тонконогий, В.Д. Гогунский // Високі технології в машинобудуванні : зб. наук. праць. - Харків : НТУ «ХП», 2014. - № 1 (24). – С. 130 – 138.