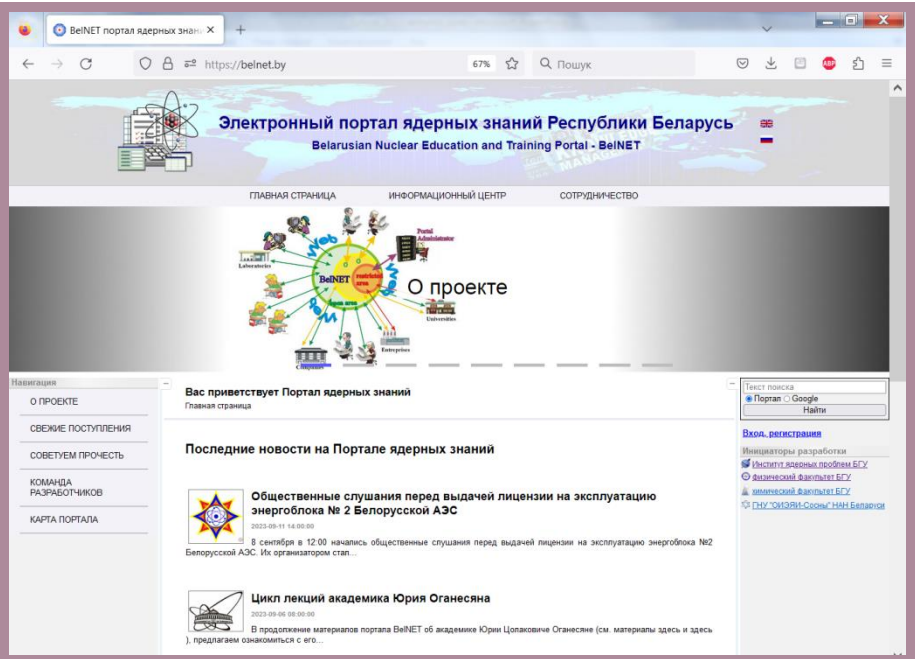


Портал ядерных знаний BelNET – семантический портал

С. Н. Сытова,

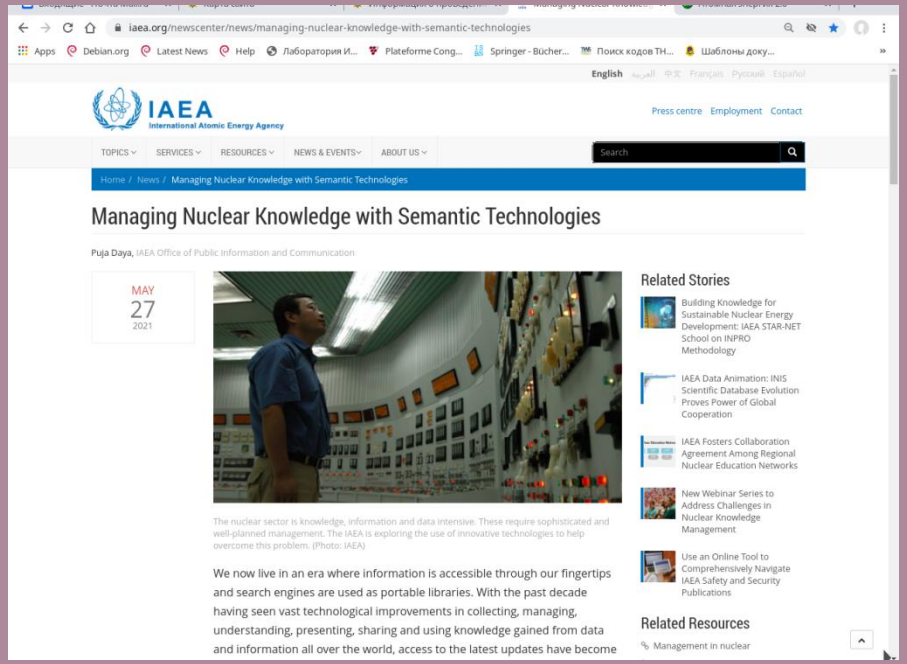
В.В.Гавриловец, А.П. Дунец,

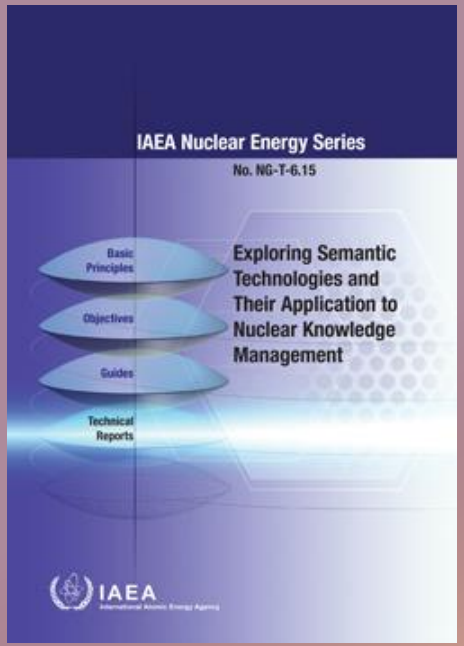
А.Н. Коваленко, С.В.Черепица

sytova@inp.bsu.by

https://belnet.by/

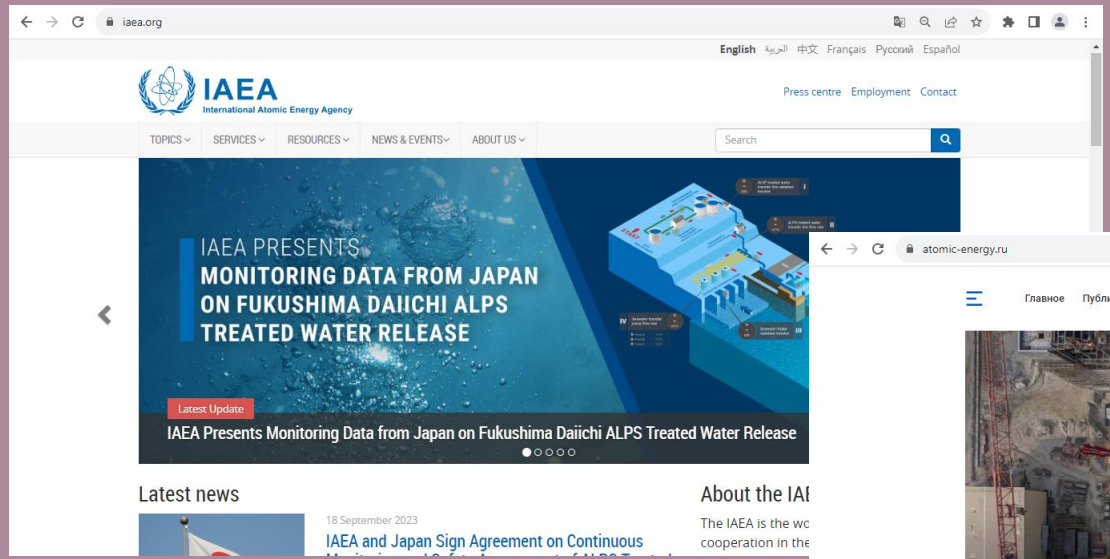
Использование семантических технологий в области ядерных знаний





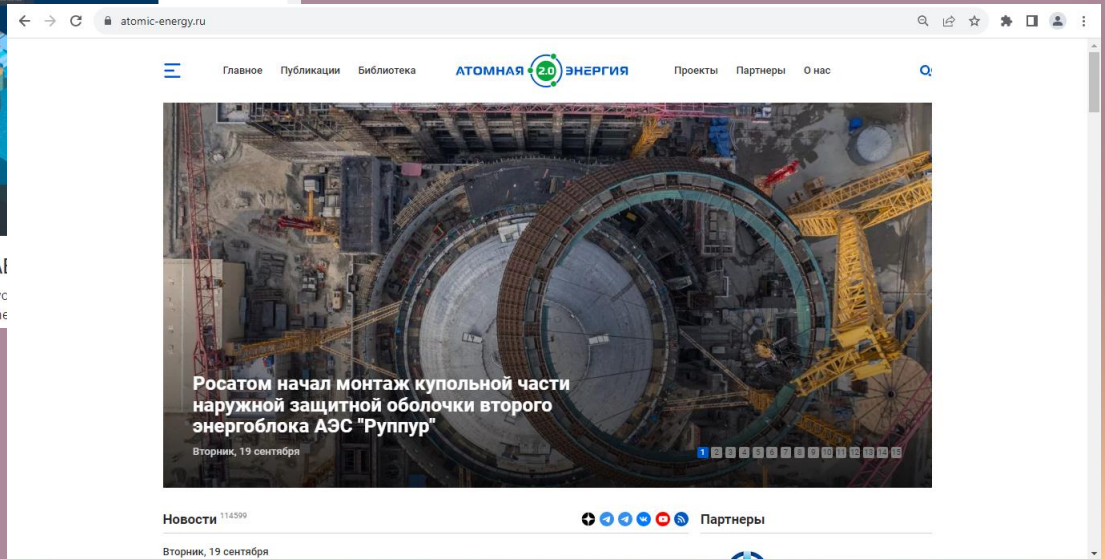
https://www.iaea.org/newscenter/news/managing-nuclear-knowledge-with-semantic-technologies

Примеры семантических порталов в области ядерных знаний



https://www.iaea.org/

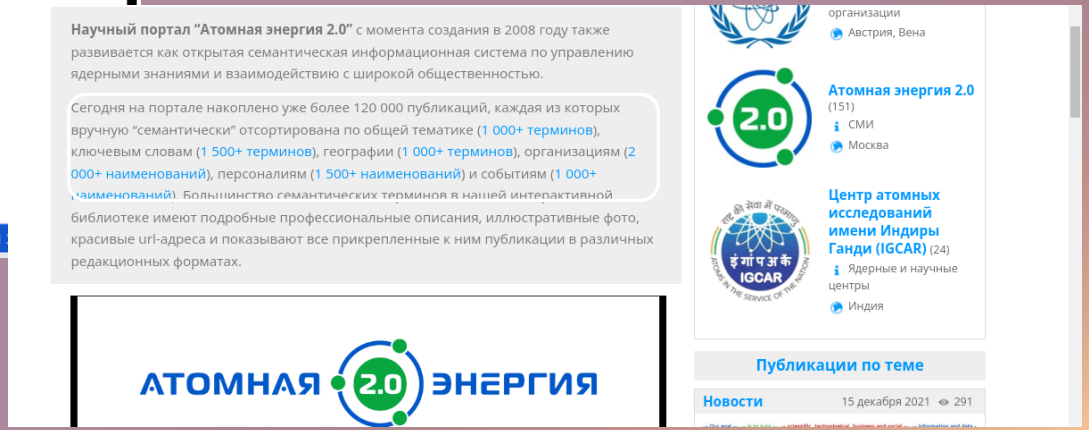
https://www.atomic-energy.ru/



Использование семантических технологий в области ядерных знаний

https://www.atomic-energy.ru/news/2021/11/16/119447





Семантические технологии. Основные понятия

Онтология в информатике – это попытка формализации некоторой области знаний с помощью концептуальной схемы, состоящей из структуры данных, содержащей все релевантные классы объектов, их связей и правил (теоремы, ограничения), принятых в этой области. На принципах онтологий создаются инструментов удобной навигации по ресурсам портала и эффективный поиск по данным портала.

Глоссарий (лат. glossarium «собрание глосс») — словарь узкоспециализированных терминов в какой-либо отрасли знаний с толкованием, иногда переводом на другой язык, комментариями и примерами. Глоссарий без дополнительных связей между терминами может рассматриваться как онтология с пустым множеством отношений.

Тезаурус (от греч. «сокровище») – собрание сведений (словарь с дополнительными отношениями), полномерно охватывающий понятия, определения и термины специальной области знаний или сферы деятельности. Тезаурусы – это особая разновидность словарей с семантическими отношениями между терминами (синонимы, антонимы, паронимы, гипонимы, гиперонимы и т. п.).

Данные и онтология с правилами построения и вывода представляют собой ***базу знаний***.

Для онлайн-поиска в поисковых запросах могут использоваться ***контролируемые термины (дескрипторы)*** – ключевые слова, произвольные текстовые слова или их комбинация, которые должны входить в тезаурус или глоссарии. Для выбора релевантных ссылок из результатов поиска очень полезными элементами являются реферат, заголовок и дескрипторы.

Таксономия – иерархическая структура портала может быть построена на основании семантических технологий, в частности, на основании одного или нескольких тезаурусов.

Белорусский портал ядерных знаний BelNET:
*Bel*arusian *N*uclear *E*ducation and *T*raining

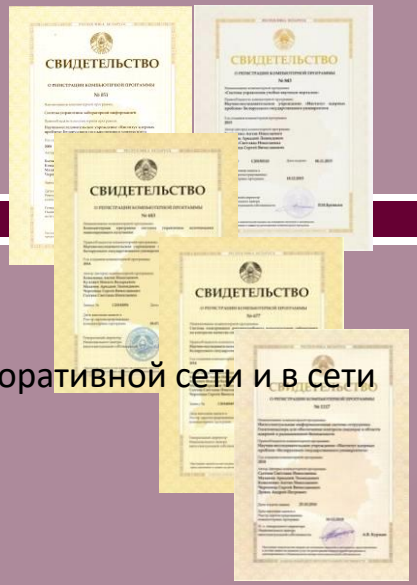
https://belnet.bsu.by/

https://belnet.by/

https://net.inpnet.net/

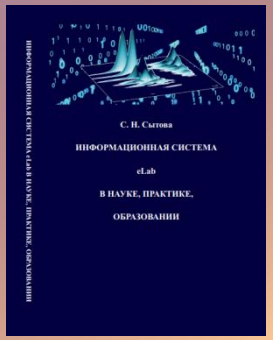


Фреймворк **eLab** на основе свободного программного обеспечения



Принципы организации и особенности системы

- Работа под Windows и Linux в многопользовательском режиме с работой внутри корпоративной сети и в сети интернет либо на отдельно выделенном компьютере;
  - с вводом данных через заполнение веб-форм в режиме on-line;
  - с разделением прав доступа разных категорий пользователей;
  - через Web-интерфейс посредством широко распространённых браузеров.
- Работает на защищенном сервере, на стороне клиента ничего не устанавливается, достаточно браузера.
- Организация удаленного доступа через VPN, обеспечение протокола HTTPS.
- ПО eLab является открытым для модификаций пользователями, возможность пользователю вносить изменения в шаблоны итоговых документов.
- Позволяет в рамках одного установленного экземпляра продукта поддерживать одновременно документооборот многих лабораторий и организаций, причём разного профиля.
- Повышенные требования к системе защиты информации.



Система управления контентом научно-образовательного портала **eLab-Science**





Реализованы все необходимые функции портала, включая возможность удаленной правки структуры портала и занесения документов, разнообразной сортировки и фильтрации, а также два уровня доступа к документам в зависимости от прав пользователей, создание сложных материалов с формулами, рисунками и т.д., оригинальный механизм тестирования при выполнении лабораторных работ.

https://belnet.by/

Схема процессного системного подхода в eLab-Science

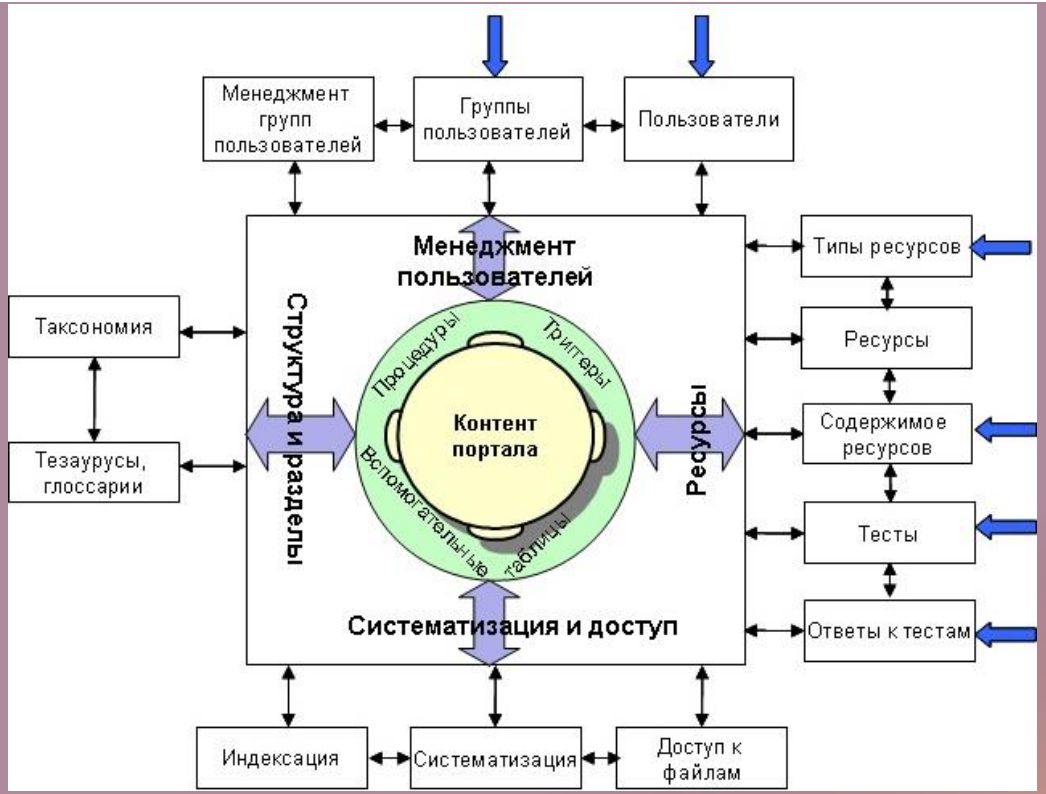
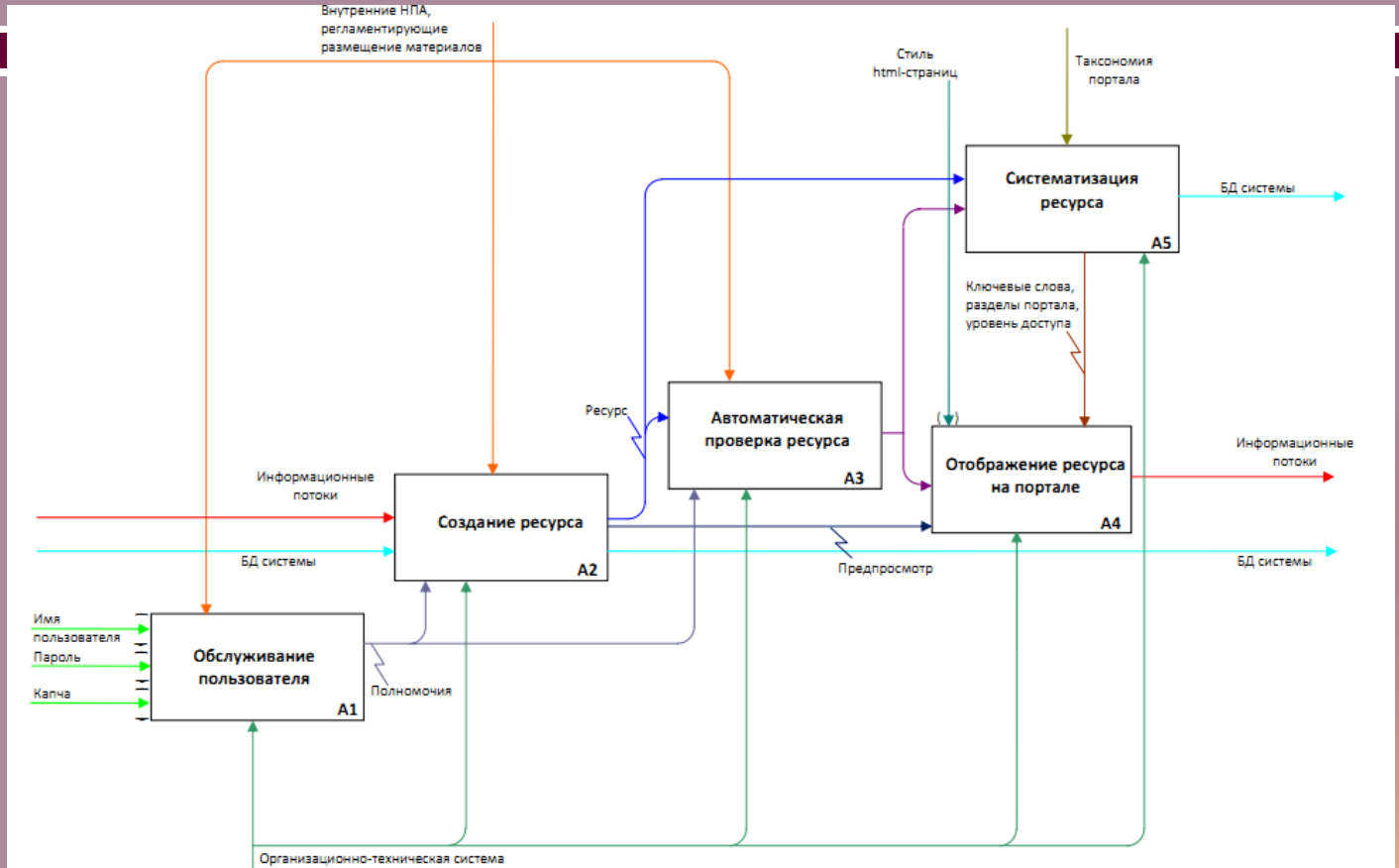
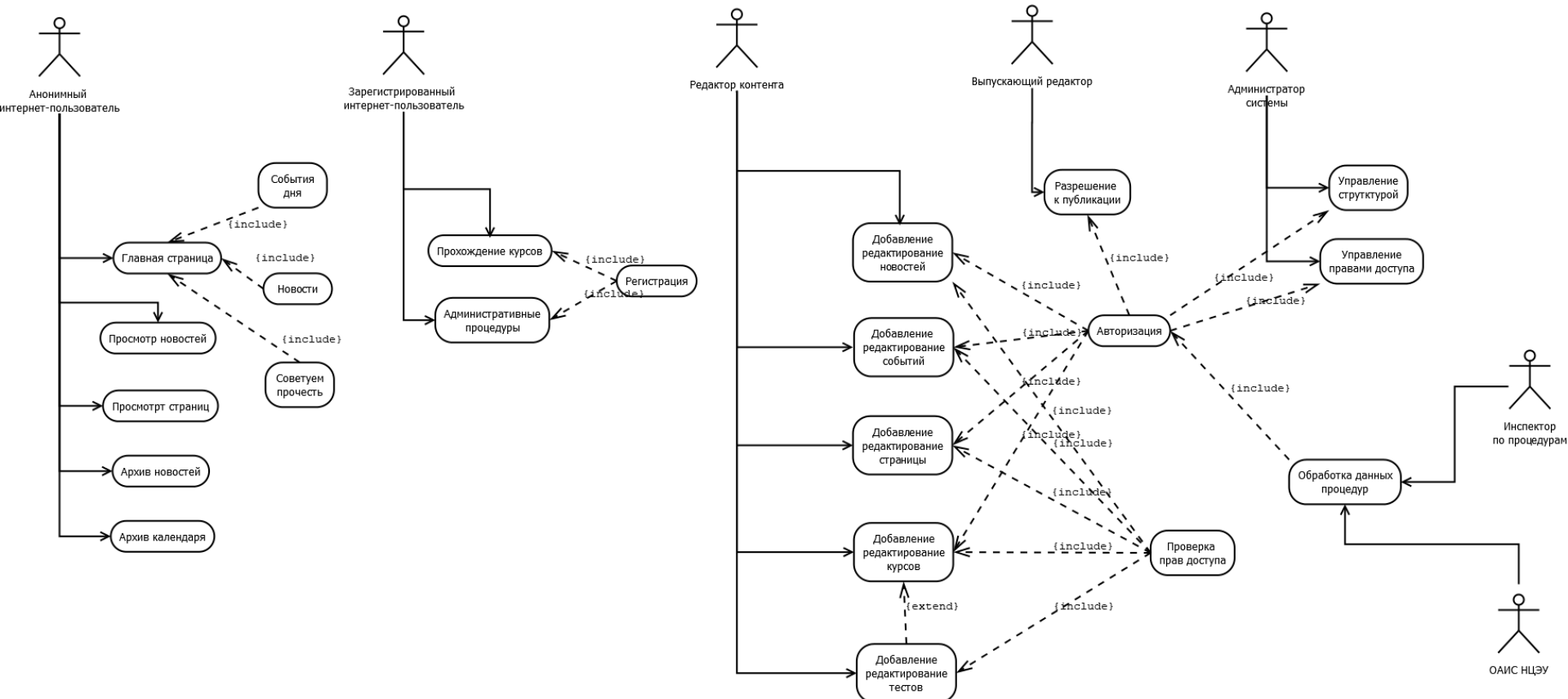


Схема функциональной структуры портала BelNET в обозначениях IDEF1



UML диаграммы прецедентов (сценариев использования) для портала ядерных знаний



Полнотекстовый поиск в eLab



- Извлечение текста из PDF, DOC, DOCX, RTF посредством **Apache Tika** https://tika.apache.org/
- Получение лемм - стемминг
  - по словарю **phpMorphy** http://phpmorphy.sourceforge.net/
  - по алгоритму **Snowball** http://snowballstem.org/

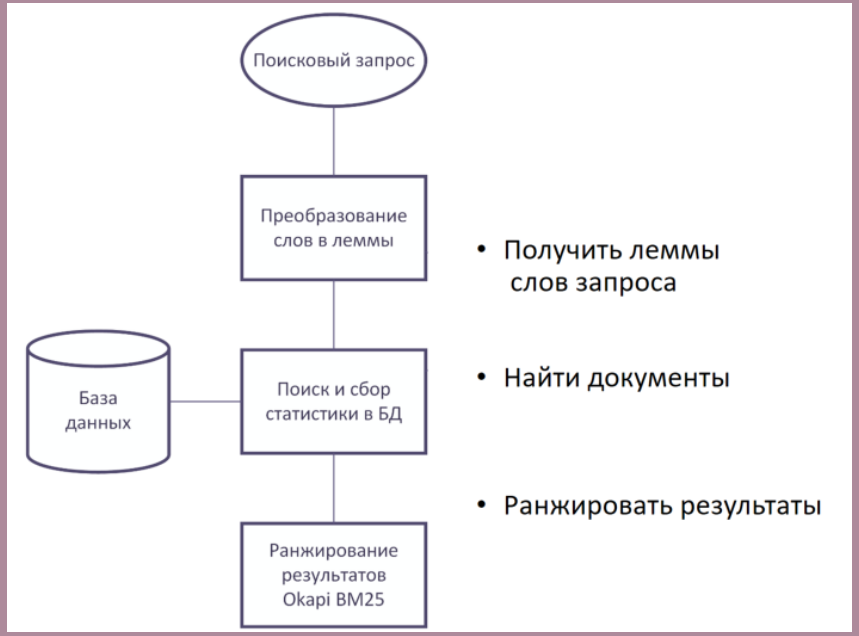
Ворон к ворону летит,
Ворон ворону кричит…

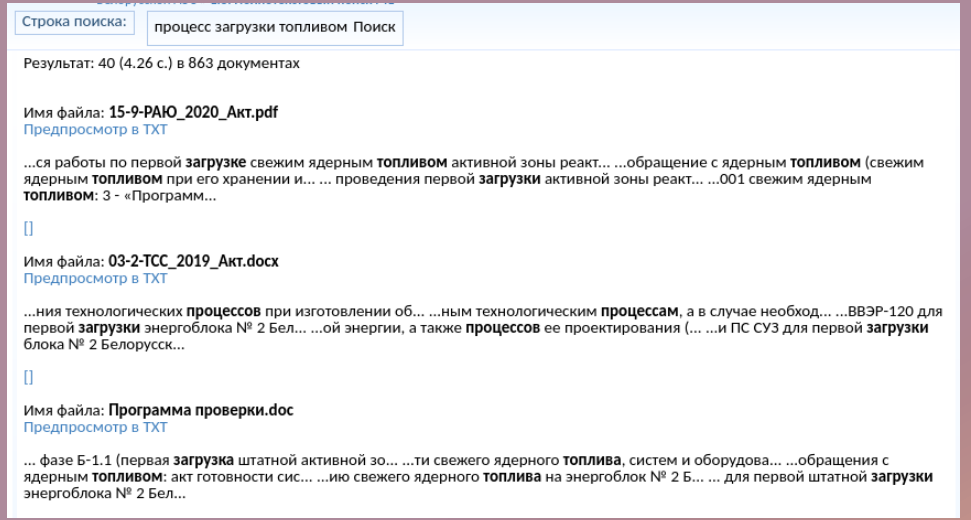


Ворон – 4 вхождения
Лететь – 1 вхождение
Кричать – 1 вхождение

Слова преобразуются в основные формы – леммы

Полнотекстовый поиск в eLab




Строка поиска: процесс загрузки топливом Поиск

Результат: 40 (4.26 с.) в 863 документах

Имя файла: **15-9-РАЮ_2020_Акт.pdf**
Предпросмотр в TXT

...ся работы по первой **загрузке** свежим ядерным **топливом** активной зоны реакт... ...обращение с ядерным **топливом** (свежим ядерным **топливом** при его хранении и... ... проведения первой **загрузки** активной зоны реакт... ...001 свежим ядерным **топливом**: 3 - «Программ...

[]

Имя файла: **03-2-ТСС_2019_Акт.docx**
Предпросмотр в TXT

...ния технологических **процессов** при изготовлении об... ...ным технологическим **процессам**, а в случае необход... ...ВВЭР-120 для первой **загрузки** энергоблока № 2 Бел... ...ой энергии, а также **процессов** ее проектирования (... ...и ПС СУЗ для первой **загрузки** блока № 2 Белоруссk...

[]

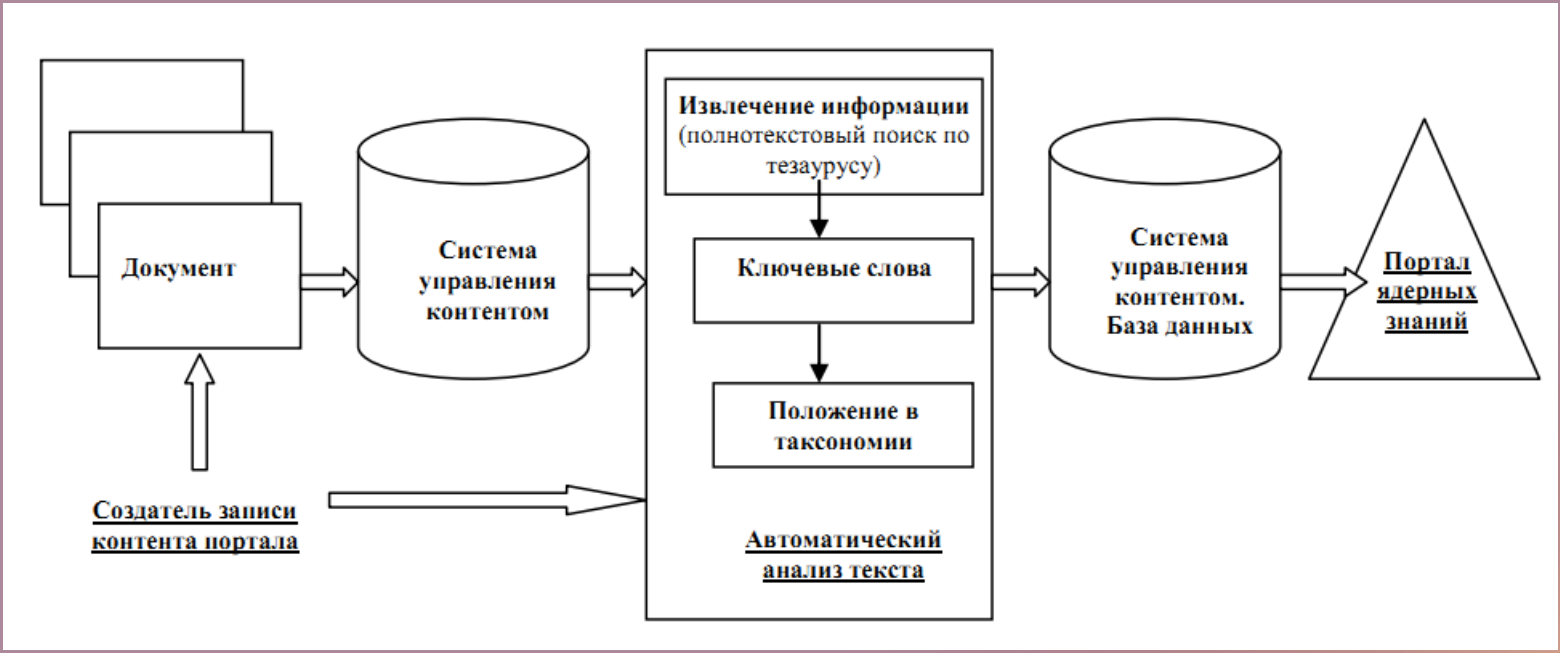
Имя файла: **Программа проверки.doc**
Предпросмотр в TXT

... фазе Б-1.1 (первая **загрузка** штатной активной зо... ...ти свежего ядерного **топлива**, систем и оборудова... ...обращения с ядерным **топливом**: акт готовности сис... ...ию свежего ядерного **топлива** на энергоблок № 2 Б... ... для первой штатной **загрузки** энергоблока № 2 Бел...

Стемминг* – разные словоформы приводятся к одному виду с визуализацией ключевых слов во фрагментах текста.

Lovins, J. B. Development of a Stemming Algorithm. Mechanical Translation and Computational Linguistics (1968), 11, 22–31

Алгоритм размещения информационного ресурса на портале ядерных знаний



Алгоритм размещения информационного ресурса на портале ядерных знаний

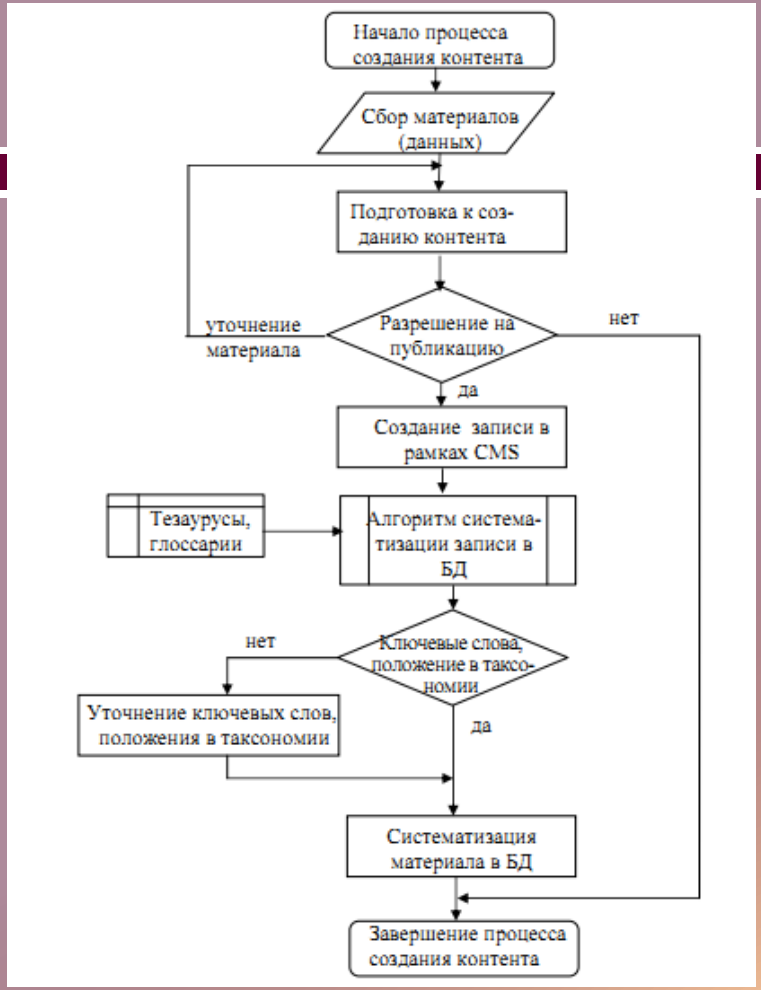
Синоним
- Леммы

Понятие
- Наименование
- Определения
- Подробное описание

Источник
- Тэг
- Библиографическое описание
- Ссылка

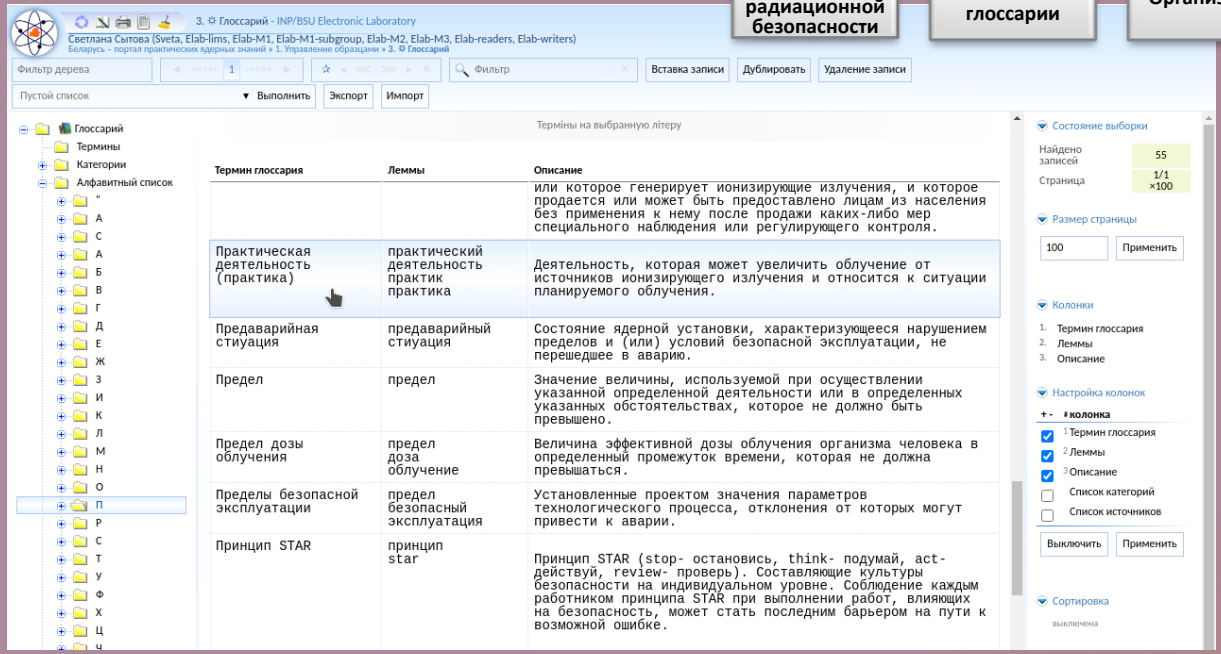
Категория
- Тэг
- Наименование

Концептуальная модель БД глоссария



Тезаурус и глоссарий по ядерной и радиационной безопасности алгоритм





Тестирование семантического алгоритма

Атом

- Конечное состояние
- Процесс
- Работа
- Система
- УЕ
- Элементы

1. Атом (начальные сведения)

Издревле ученых, прежде всего химиков, волновали вопросы: как устроено вещество, можно ли бесконечно дробить его на все более мелкие части? Идея о том, что этот процесс не бесконечен, возникла у древнегреческих и древнеиндийских философов. Но лишь в 17-18 веках химикам удалось экспериментально доказать, что вещество не может быть подвергнуто дальнейшему расщеплению на составляющие элементы с помощью химических методов, а состоит из атомов (от др.-греч. ἄτομος – неделимый, не разрезаемый).

В конце 19-го и начале 20-го веков были открыты частицы, намного меньшие чем атом (субатомные частицы). Стало проясняться, что реальная частица, которой было присвоено имя атома, в действительности не является неделимой, причем имеет собственную особую структуру.

Таким образом было установлено, что ***атом*** представляет собой мельчайшую частицу химического элемента, например, железа или меди, обладающую его химическими свойствами. Мельчайшие частицы сложных веществ, например, воды (H_2O), представляют собой *молекулы*, которые состоят из двух и более атомов.

В 1911 году Э. Резерфордом было доказано, что любой атом состоит из ***ядра*** и ***электронов***, образующих электронную оболочку вокруг ядра. Отрицательно заряженные электроны удерживаются положительно заряженным ядром благодаря кулоновским (электрическим) силам притяжения. Этим



В глоссарии BelNET:

«Процесс – последовательность действий или операций, в особенности ряд последовательных стадий изготовления продукта или некоторых других операций. Ряд взаимосвязанных или взаимодействующих операций, которые преобразуют вкладываемые ресурсы в конечные результаты».

В тексте «Атом» есть фраза: «Издревле ученых, прежде всего химиков, волновали вопросы: как устроено вещество, можно ли бесконечно дробить его на все более мелкие части? Идея о том, что этот процесс не бесконечен, возникла у древнегреческих и древнеиндийских философов».

То есть семантика в тексте и семантика в определении для слова «процесс» различаются.

В тексте статьи «УЕ» – это условная единица. В глоссарии BelNET – «УЕ» – это учетная единица.

Тестирование семантического алгоритма

Ядро

- Активность
- Атомная электростанция (АЭС)
- Деление
- Излучение
- Модель
- Процесс
- Работа
- Радиоактивность
- Радиоактивный
- Синтез
- Система
- УЕ
- Фон - семантика
- Элементы
- Ядерный реактор

4. Ядро. Символические обозначения ядер

В 1932 году В. Гейзенбергом, Д. Иваненко было доказано, что *ядра* всех элементов, исключая водород, состоят из частиц двух сортов: протонов и нейтронов (их общее название – ***нуклоны***, от лат. *nucleus* «ядро»). В этом же году Дж. Чедвик экспериментально открыл нейтрон. ***Нейтрон*** (от лат. *neuter* — ни тот, ни другой) имеет массу около 1838 электронных масс (примерно на две массы электрона больше, чем у протона), но не имеет электрического заряда.

Нуклоны в ядре притягиваются друг к другу мощными силами притяжения, которые называются ***ядерными силами***. Ядерное (или *сильное*) взаимодействие компенсирует кулоновское отталкивание положительно заряженных протонов и обеспечивает устойчивость большинства ядер.

В последние десятилетия выяснилось, что нуклоны имеют достаточно сложную структуру, однако в практических задачах их по-прежнему можно считать элементарными частицами.

Для описания ядер используют три важных числа. Число протонов в ядре ***Z*** одновременно определяет и число электронов в атоме, а значит и порядковый номер элемента периодической системе. Число нейтронов обозначается ***N***, в сумме с Z они дают число нуклонов в ядре, или ***массовое число A***:

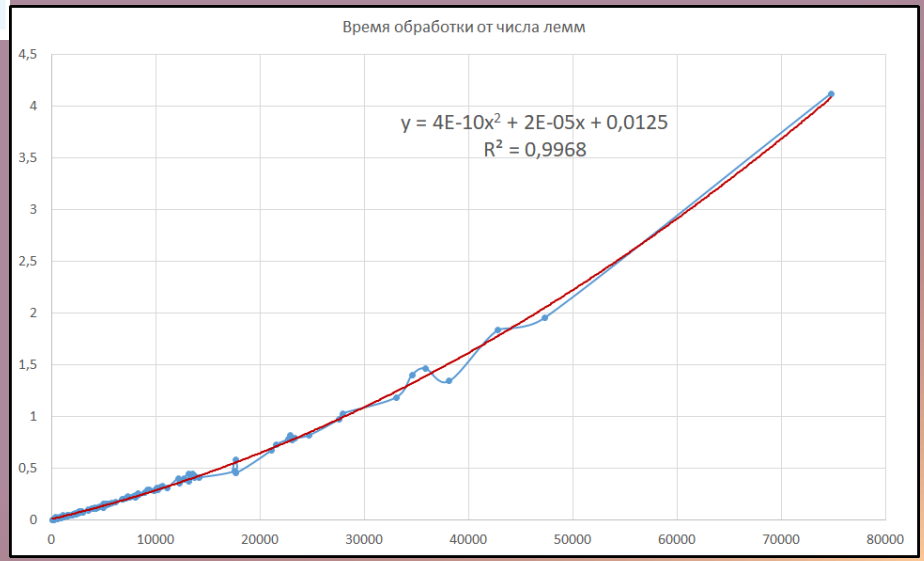
$$A = Z + N.$$

Здесь получено нарушение семантики для термина «Фон»: «их взаимное электрическое отталкивание даже на фоне сильного ядерного притяжения».

Результаты испытаний работы системы под нагрузкой

Число символов	Число лемм в тексте	Число найденных терминов	Время обработки, сек
15 438	1 208	6	0.03
24 921	1 951	15	0.05
182 764	17 669	43	0.45
4 870 803	165 478	113	45.50

Для тестирования использовался процессор Intel i5-7200U с тактовой частотой 2,5 ГГц. Объем оперативной памяти практически не влиял на результат, так как требуемый объем ОЗУ не превысил 50 Мб.



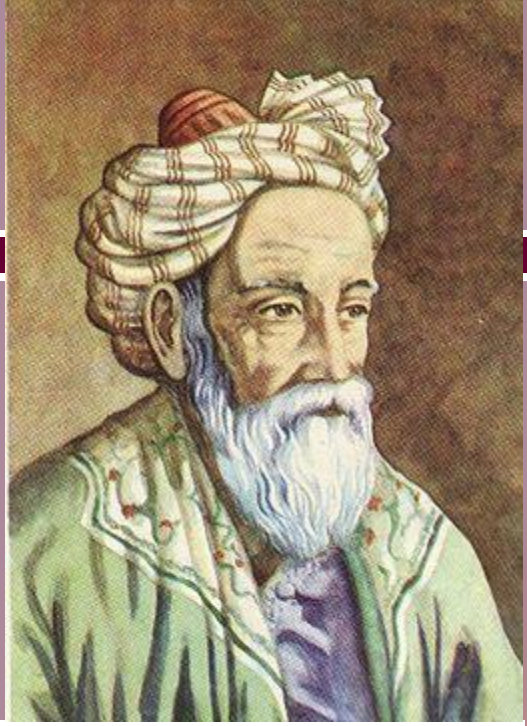
Оказалось, что даже для очень больших документов (самый большой документ содержал около 300 страниц и почти 5 миллионов символов) время работы находится в приемлемых рамках – 45 с. Типичные документы – статьи и т.п. объемом несколько страниц обрабатываются менее секунды.

Поскольку от работы данных алгоритмов не требуется мгновенного результата, работа м. б. выполнена в фоновом режиме. Размещение одной записью огромных документов не оптимально с точки зрения читабельности и наглядности материала. Лучше использовать разбиение материала на главы и разделы.

Таким образом, можно констатировать, что ***результаты реализации алгоритма являются очень хорошими***.

Много лет размышлял я над жизнью земной.
Непонятного нет для меня под луной.
Мне известно, что мне ничего не известно! —
Вот последняя правда, открытая мной.

Омар Хайям



Спасибо за внимание!

sytova@inp.bsu.by