

# Повышение качества поиска в больших объёмах текстовых документов с использованием генетического алгоритма как способ поддержки научных исследований

---

Плешкова Ирина Юрьевна

Пермь, ПГНИУ

[elf\\_irina@mail.ru](mailto:elf_irina@mail.ru)

# Обо мне

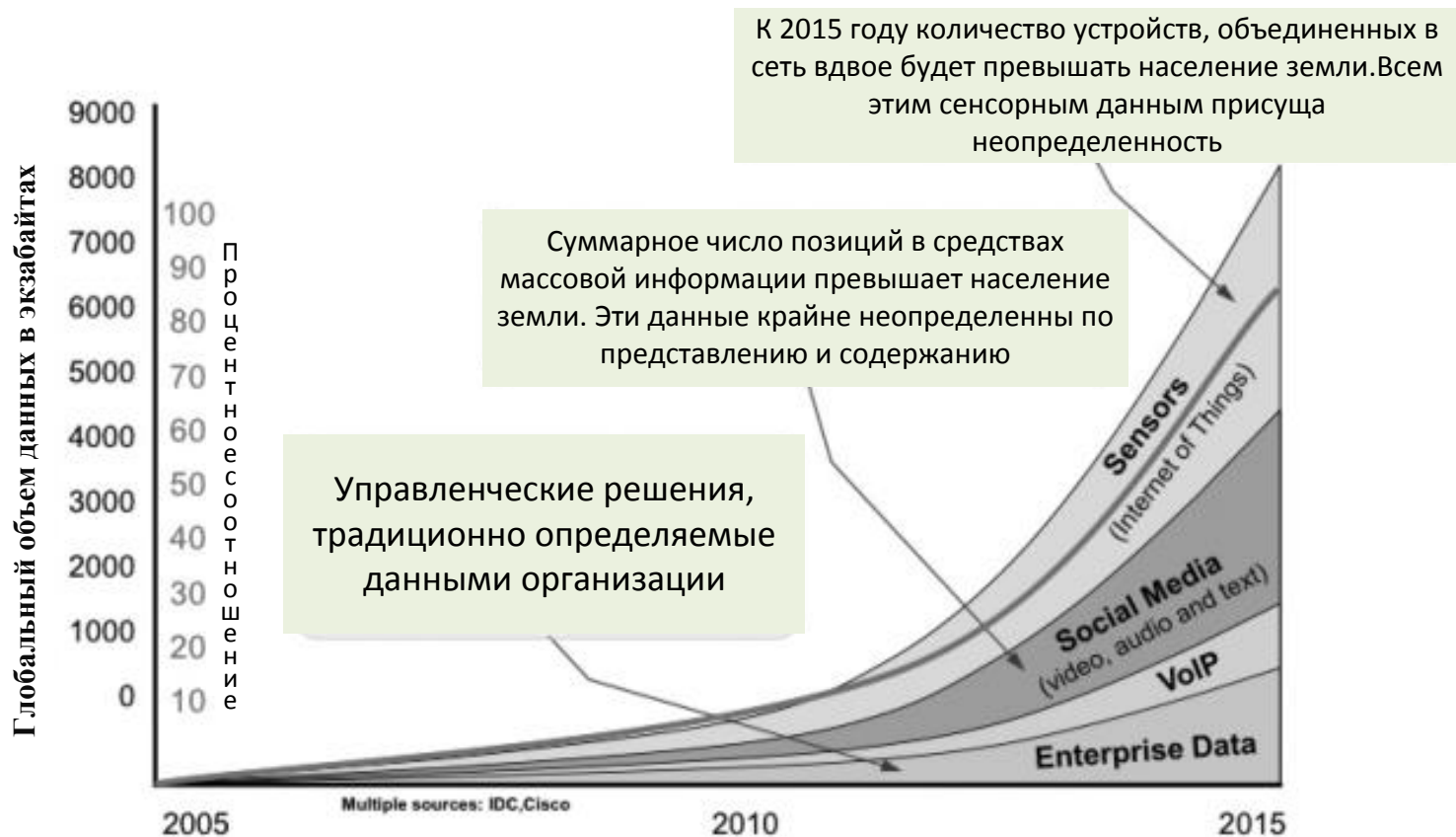
---

Магистрант 2-го курса кафедры математического обеспечения вычислительных систем Пермского государственного университета.

Ассистент кафедры МОВС, ПГНИУ.

Разработчик в малом инновационном предприятии «Кнова».

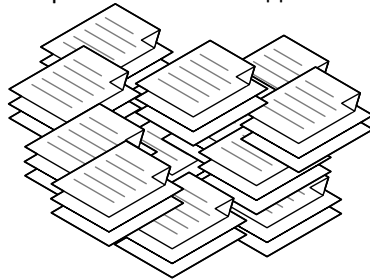
# Прогноз роста данных до 2015 года



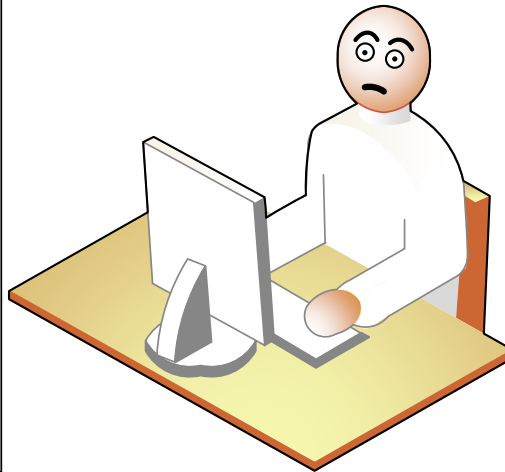
# Бизнес

---

Корпоративное хранилище – сотни терабайт текстовых данных



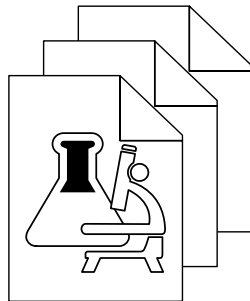
Нормативные акты, отчёты, соглашения с поставщиками,...



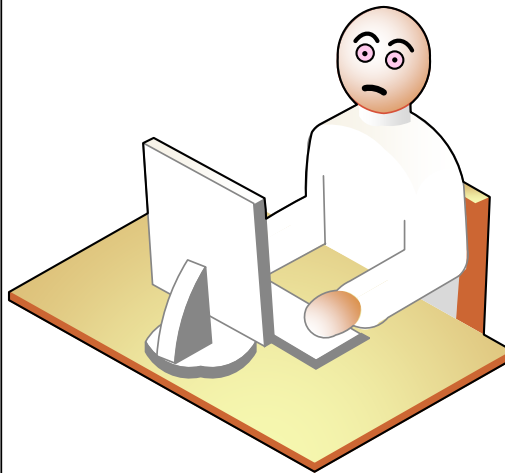
# Наука

---

Исследовательские организации:

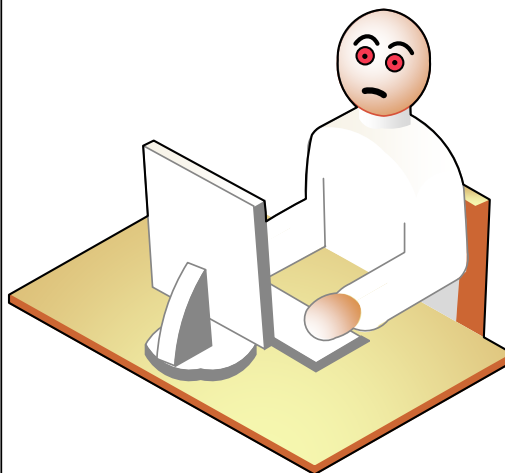


Отчёты о ранее проводимых исследованиях, сборники публикаций,...

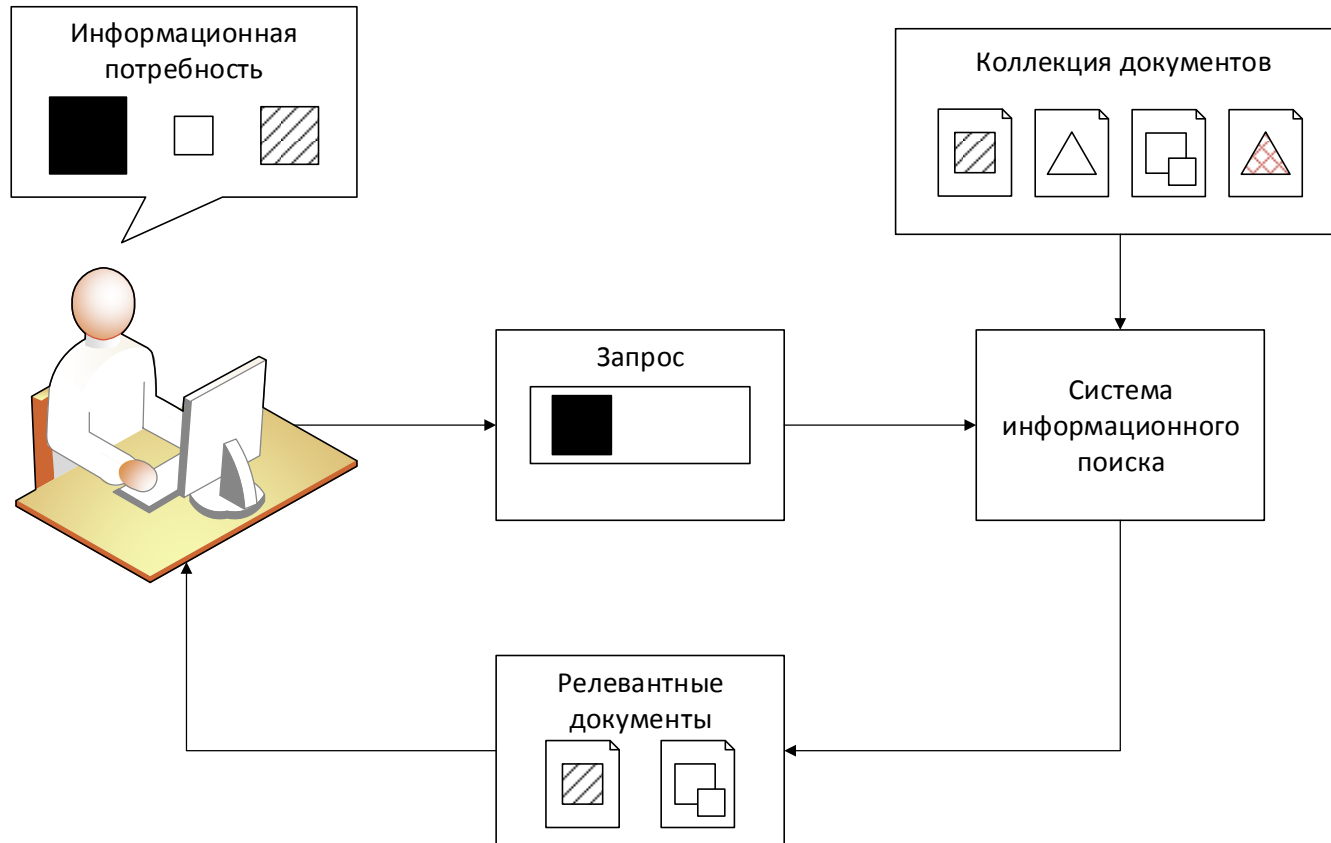


# Библиотеки

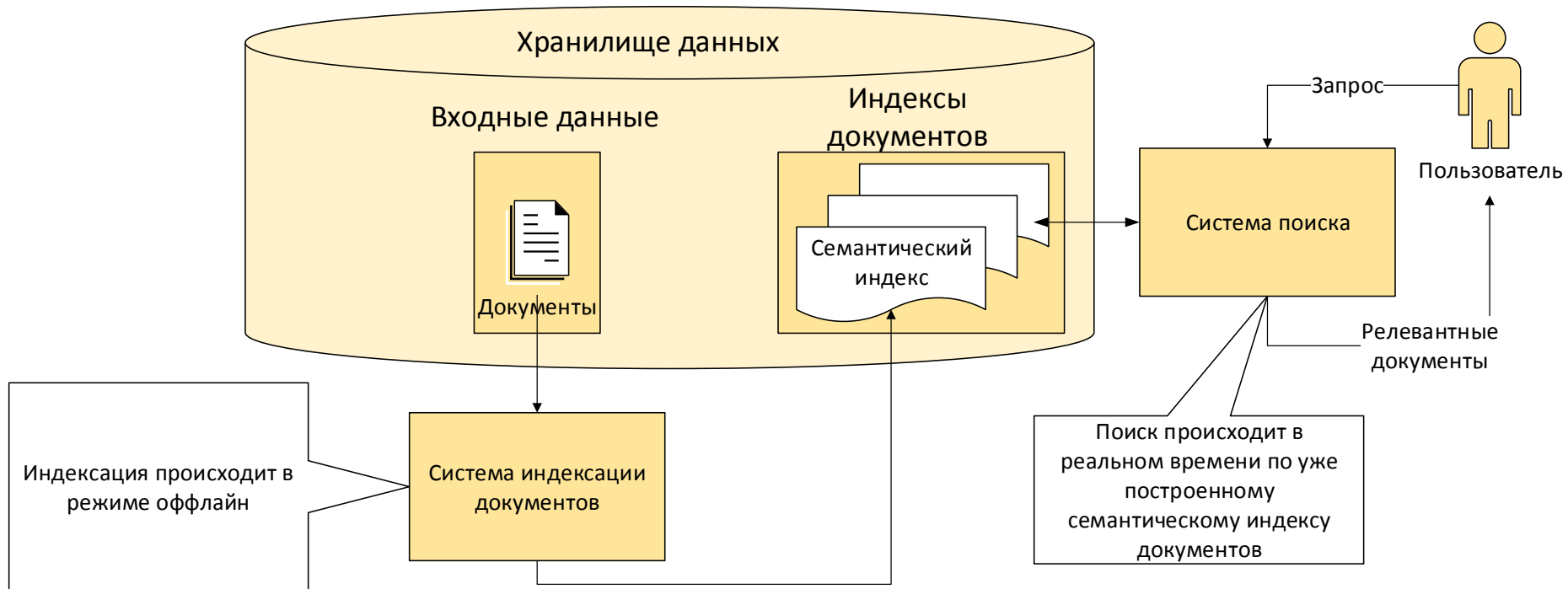
---



# Обобщённый пример



# Архитектура поисковой системы





# Используемые компоненты

---

Были выбраны следующие технологии:

- NGDATA Lily
- Apache Solr
- Apache Spark
- Scala

# Научная статья

УДК 681.7

Г. Савлен  
Томский университет в Томске, Томск

## Интеллектуальная система поддержки решений в нечёткой среде

В работе предложены основные понятия и методы построения интеллектуальной системы поддержки принятия решений, основанные на методах нечёткой логики. Показано, почему такая система является эффективной и удобной для использования, как в промышленности (автоматическое управление в процессе принятия решений в сложных, неструктурированных системах), так и в повседневной жизни. Приведены результаты практической реализации и принципы работы нечёткой системы для решения задач стратегического управления.

### Введение

Глобализация международного рынка, необходимость активно конкурентно реагировать на все стадии перед современным предприятием всё больше требуются. Для эффективного управления производственной системой в настоящее время необходимы систематизация методов и инструментов поддержки решений на всех уровнях её функционирования. В последние десятилетия в области решения сложных неструктурированных производственных систем (таких, как СМТ-Сборные Интерфейсы Машинистов) появились новые направления, которые характеризуются использованием в них методов и технологий искусственного интеллекта для поддержки решений и взаимодействия в управлении производством, а также для мониторинга, диагностики и анализа технологических процессов в системах реального времени. В научной литературе это новое прогрессивное направление получило название интеллектуального производства (от англ. Intelligent Manufacturing). Концепция интеллектуального производства заключается в создании интеллектуальной системы поддержки решений на всех этапах функционирования производства, способной и адаптирующей новые знания, обучение в результате анализа накопленных знаний и опыта, а также адаптации во все динамически изменяющейся внешней среде и текущего состояния всех составляющих элементов производственной системы. Необходимость разработки такой системы искусственного интеллекта объясняется, прежде всего, возросшей сложностью международной информации и процессом принятия решений в производстве, а также тем, что решение многих проблем и управление современным производством связано с обработкой большой, нечёткой или противоречивой информации. Решения часто принимаются в условиях, когда знания трудно или вообще невозможно сформулировать, а основные критерии и значения часто или вообще невозможно определить. Для решения таких задач необходимо применять неструктурированные процедуры принятия решений, включающие системы знания, экспертные и искусственные нейроны.

Заголовок

Аннотация,  
ключевые слова

Основной текст

# Научная статья: пример

---

## Интеллектуальная система поддержки решений в нечёткой среде

В работе представлены основные положения и методика проектирования интеллектуальных систем поддержки принятия решений, основанных на нечётких нейронных сетях. Нечёткие нейронные сети позволяют моделировать и обрабатывать как количественную, так и качественную (лингвистическую) информацию в процессе принятия решений в сложных, многомерных системах. Представлены результаты практической имплементации и применения нечёткой нейронной системы для решения задач стратегического управления.

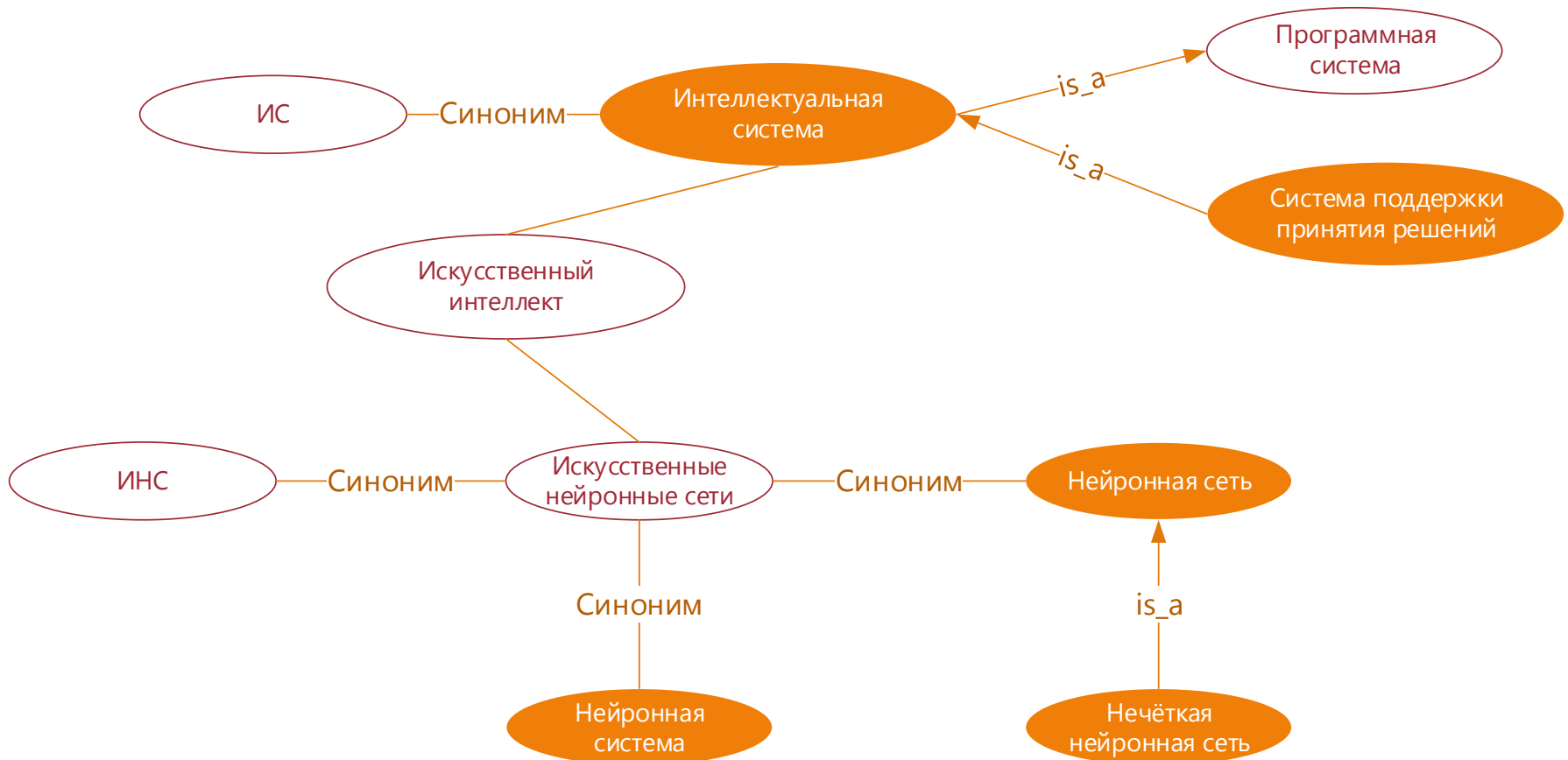
# Научная статья: пример

---

## Интеллектуальная система поддержки решений в нечёткой среде

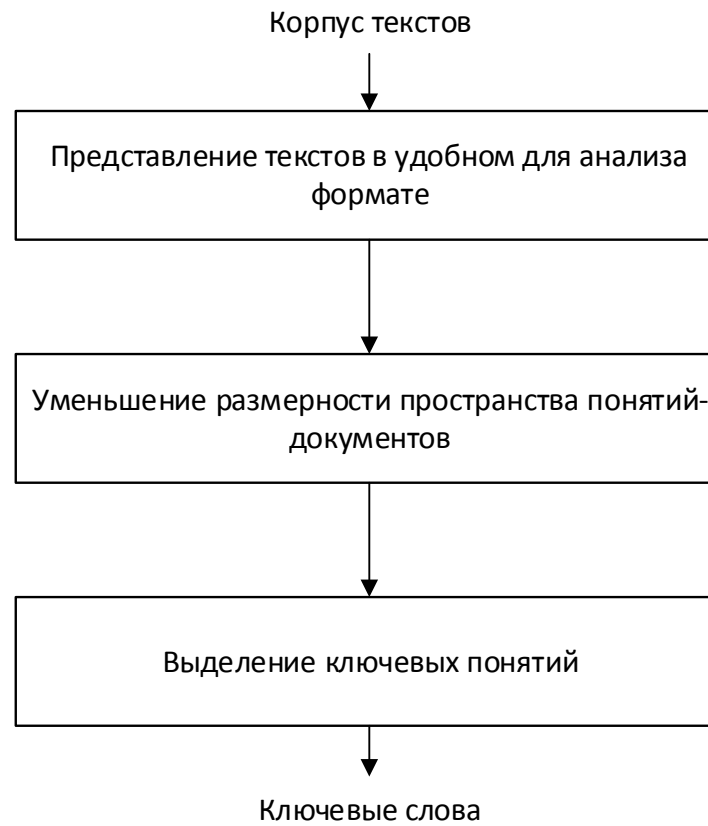
В работе представлены основные положения и методика проектирования интеллектуальных систем поддержки принятия решений, основанных на нечётких нейронных сетях. Нечёткие нейронные сети позволяют моделировать и обрабатывать как количественную, так и качественную (лингвистическую) информацию в процессе принятия решений в сложных, многомерных системах. Представлены результаты практической имплементации и применения нечёткой нейронной системы для решения задач стратегического управления.

# Научная статья: пример



# Индексация: как это происходит обычно

---



# Пример

---

Слово	Вес (количество употреблений)
Нечёткий	57
Система	48
Решение	37
Сеть	37
Нейронный	35
Алгоритм	38
Функция	26
Вид	22
Задача	20
Метод	20

# Индексация: как это происходит обычно

---

Мы можем индексировать только те понятия, которые есть в документе!

В большинстве традиционных схем анализа текстов идёт работа со словами, а не понятиями!

Традиционные способы индексации не учитывают семантические связи между понятиями!



# Предлагаемый подход

---

Использовать онтологию для выделения понятий и связей между ними.

При оценке веса учитывать множество критериев:

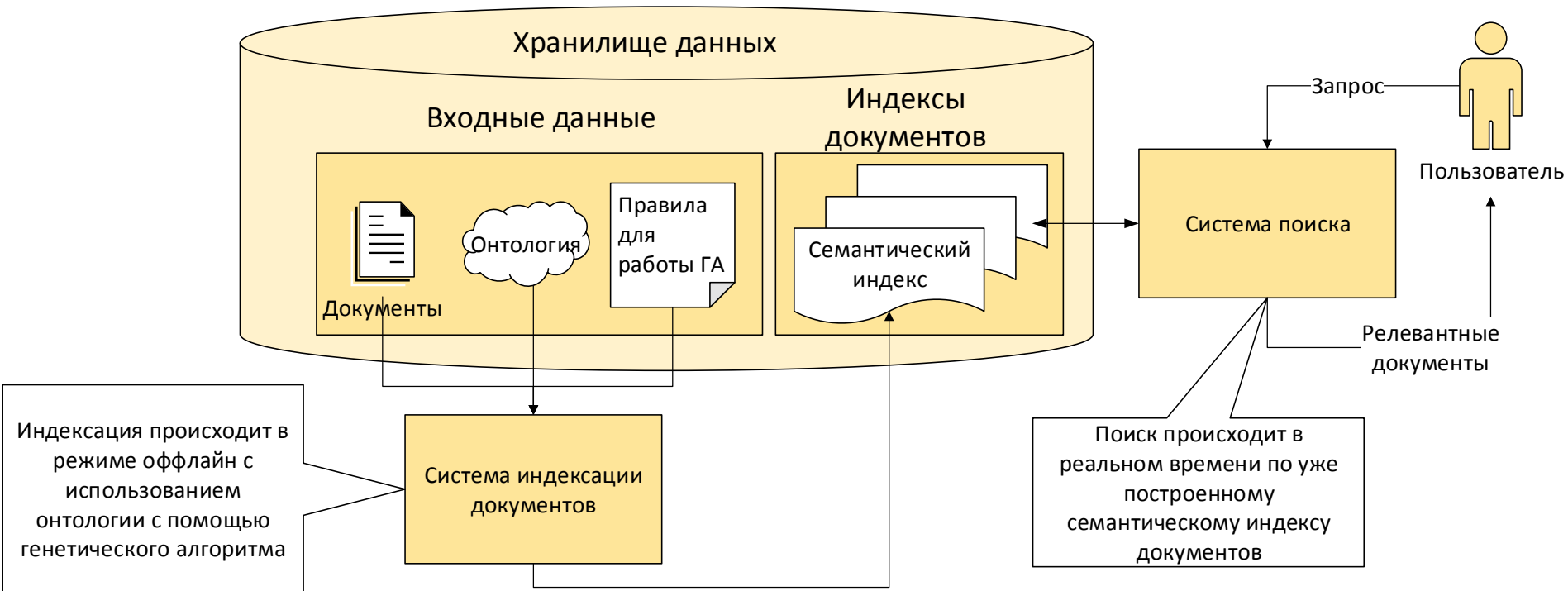
- Статистические
- Семантические
- Структуру текста

# Пример

---

Понятие	Вес
Нейронная сеть	83
Нечёткая нейронная сеть	74
Интеллектуальная система	52
Нейронная система	47
Искусственный интеллект	31
Программная система	15
Искусственная нейронная сеть	13
Система поддержки принятия решений	5

# Предлагаемый подход: Архитектура



# Аналоги

---

<b>Критерий сравнения</b>	<b>Системы синтаксического поиска (Apache Solr, Sphinx, Lucene и др.)</b>	<b>Системы семантического поиска (Nigma, SHOE, TRUST, Inquirus2)</b>
<b>Учёт контекста запроса</b>	-	+
<b>Учёт связанных по смыслу понятий</b>	+/-	+
<b>Скорость работы</b>	+	-

# Преимущества

---

Обработка больших объёмов текстовых документов различных форматов.

Семантический поиск – по смыслу, а не по ключевым словам.

Быстрая скорость работы.

# Спасибо за внимание!

---

Плешкова Ирина Юрьевна

Пермь, ПГНИУ

[elf\\_irina@mail.ru](mailto:elf_irina@mail.ru)

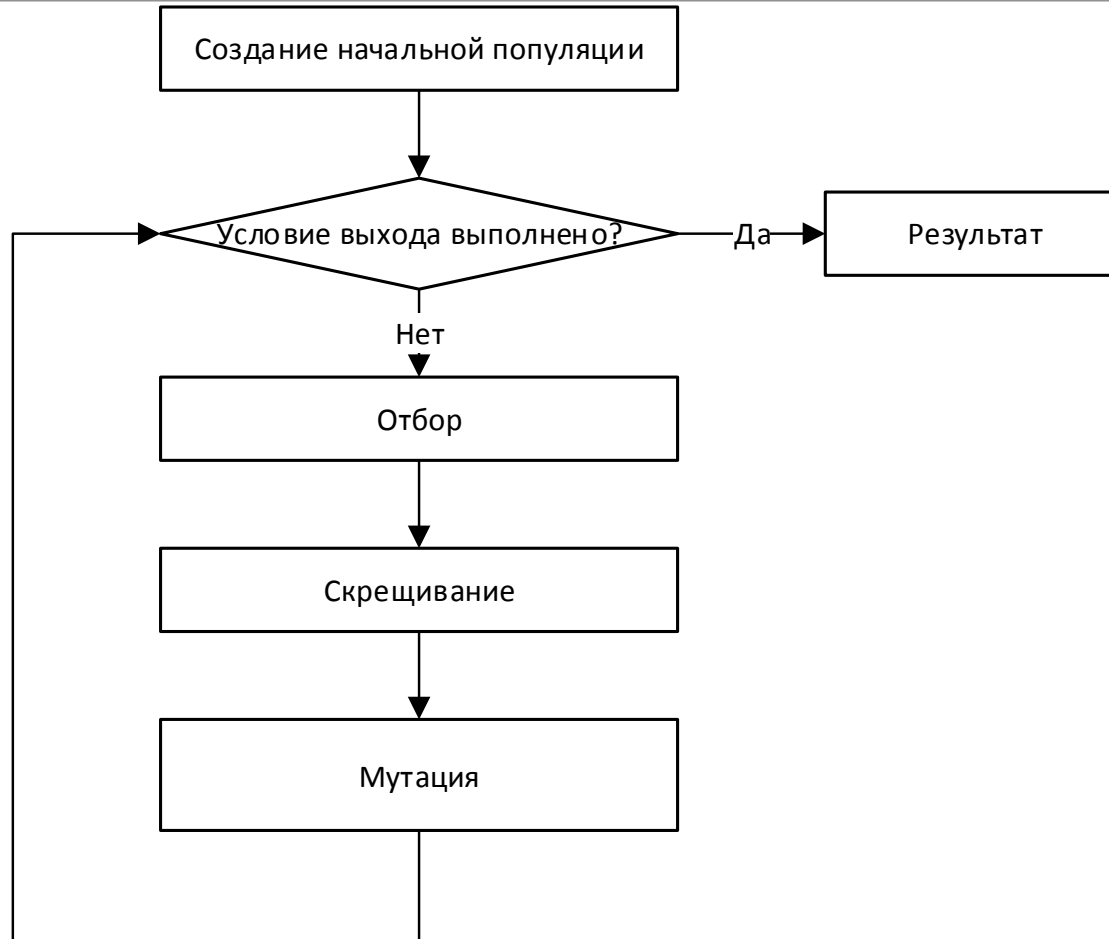
# Генетические алгоритмы

---

Генетические алгоритмы (ГА) – это процедуры поиска, основанные на механизмах естественного отбора и наследования. В них используется эволюционный принцип выживания наиболее приспособленных особей. Они отличаются от традиционных методов оптимизации несколькими базовыми элементами. В частности, генетические алгоритмы:

- Обрабатывают не значения параметров самой задачи, а их закодированную форму.
- Осуществляют поиск решения исходя не из единственной точки, а из их некоторой популяции.
- Используют только целевую функцию, а не ее производные либо иную дополнительную информацию.
- Применяют вероятностные, а не детерминированные правила выбора.

# Схема работы ГА





# Hortonworks Data Platform 2.1

