

Системные основы программной инженерии

Задание к РГР

Собрать материал и провести анализ с позиций системной инженерии одного из крупных проектов (существующего или гипотетического), оценить в нем IT-составляющую. Провести анализ предметной области, оценку параметров системы, сделать набросок бизнес-плана.

Варианты заданий:

1. Разработка и производство интегральных схем
2. Большой адронный коллайдер
3. Стратегический дата-центр
4. Космический стартовый комплекс
5. Скоростная железная дорога в экстремальных условиях (высокогорье, крайний север, пустыня)
6. Система управления войсками в современных условиях
7. Морская буровая платформа
8. Система управления здравоохранением
9. Стратегическая подводная лодка
10. Магистральный пассажирский самолет

Вариант. Всемирная база знаний

Гипотетический проект: собрание артефактов, подтвержденных общепризнанных знаний и фактов по всем сторонам деятельности человеческого общества.

Ключевые вопросы

В современном состоянии поисковые системы Интернета являются “информационной помойкой”. Проблемы поиска:

- при поиске не учитывается структура знаний той области, где осуществляется поиск
- отсутствует оценка достоверности, общепризнанности, степени доверия источника
- “замусоривание” поиска по специализированной проблеме рекламной, общеизвестной информацией, связанной с ключевыми словами
- появление в новостях (особенно науки) недостоверных, непроверенных, фейковых и провокационных фактов
- зачастую сами копипастеры, популяризаторы вносят элемент сенсационности, недостоверности

Основная проблема: **оценка достоверности “факта”**. Ключевой момент - система должна содержать “факты” во взаимосвязи с источниками достоверности, обоснованности, признанности со стороны других источников.

Метафоры:

- демократия в ядерной физике

- вопросы религии и веры, науки и лженауки, политики и пропаганды - оставить одно “правильное мнение” или “демократическое” многообразие заблуждений

Области применения

- наука
- образование, уровни вхождения в
- технологии (инжиниринг)
- артефакты - факты, события, документы, места

Экологическая ниша

- с одной стороны - неструктурированные поисковые системы
- с другой стороны - специализированные источники и БД

Индуктивный подход. Пример и сценарии: как есть и как надо

Новость науки

Новость: ученые научились получать водород из воды при помощи обычных батареек.

Поиск: новость, фамилии, университет, найдена статья.

Факты: вода практически не разлагается на водород и кислород (электролиз). Обычный электролиз солей. Разработаны наноповерхности (или что-то в этом роде), которые являются катализатором этого процесса, так что напряжения обычной батарейки достаточно для электролиза.

Резюме: сам факт использования батарейки никакого технологического смысла не имеет.

Новость науки

Новость: проект TrueNorth, “ученые разгадали/воспроизвели” работу человеческого мозга, отчасти с подачи разработчиков.

Поиск: достаточно прямолинейно на IBMResearch и статьи по проекту

Факты: кристалл $5 \cdot 10^9$ вентиляей - 4096 ядер, каждое моделирует 256 импульсных нейронов + коммутации, тактовая частота 1 кГц, последовательная обработка нейронов и сигналов синапсов. Каждый нейрон - сумматор с пороговым сравнением, веса синапсов. Архитектура **ни в какое сравнение с биологическим нейроном не идет.** Факты практического применения сторонними разработчиками не обнаружены. Методы решения задач нейронными сетями - традиционные. Для импульсных нейронных сетей

Резюме: решение проблем за счет количественного роста элементов с достаточно низким быстродействием

Прототипы

Специализированные “общественные” системы для хранения знаний - Wikipedia, WikiData

Wikipedia.ru -текущее состояние

статей - 1300000

страниц - 4768000

файлов - 189000

Wikipedia.ru - история - статей (тыс)

2004 - 10
2005 - 50
2006 - 100
2007 - 200
2008 - 300
2009 - 350
2010 - 500
2011 - 700

- Один из первых сборов средств был проведен с [18 февраля](#) до [1 марта](#) 2005 года и собрал \$94 000, превзойдя ожидания на \$21 000.
- [6 января](#) 2006 года закончился сбор Q4 2005, собрав в общей сложности более \$390 000.
- В июне 2007 года было объявлено, что Немецкая Википедия будет финансироваться государством.

Оценка: линейный рост, национальные варианты не связаны, нет API и семантической структуры материала. Финансирование спонсорское, государственное, порядок 10^5 \$, сообщества экспертов

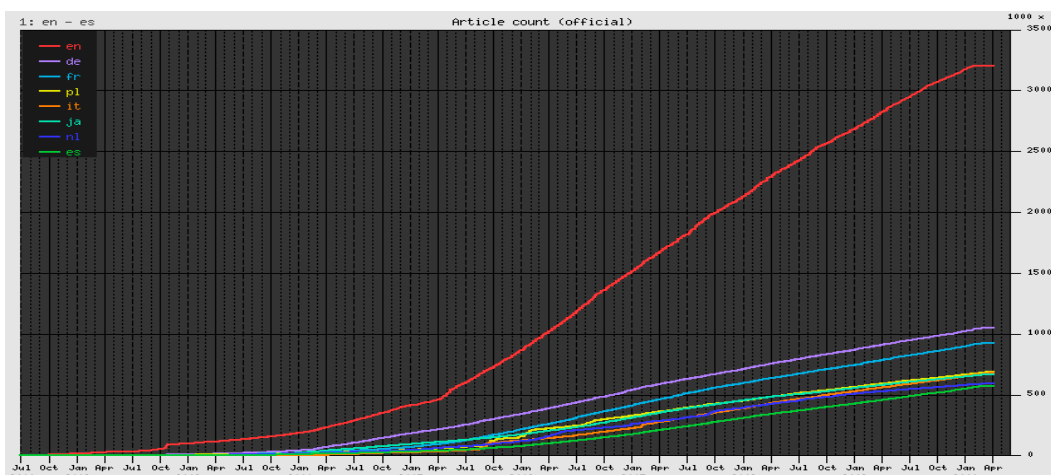
Wikidata - <https://ru.wikipedia.org/wiki/Викиданные>

«**Викиданные**» (англ. *Wikidata*) — совместно редактируемая база знаний, созданная Фондом Викимедиа. Она используется для обеспечения централизованного хранения **данных**, которые могут содержаться в статьях **Википедии** — например, **интервики-ссылок**, значков статусных статей и списков или статистической информации: дат рождения, численности населения и т. п.^[1] Содержимое Викиданных распространяется по лицензии **Creative Commons CC0**.^[2]

Проект официально запущен 30 октября 2012 года, разработка проекта ведётся под руководством **Wikimedia Deutschland**. «Викиданные» стали первым новым проектом **Фонда Викимедиа** с 2006 года^[3]. Создание проекта было осуществлено за счёт пожертвований **Allen Institute for Artificial Intelligence**, **Gordon and Betty Moore Foundation** и **Google** на общую сумму 1,3 миллиона **евро**^{[4][5]}.

22.09.2015 - $21 \cdot 10^6$ фактов

03.03.2016 - $23 \cdot 10^6$ фактов



Оценка: сравнительно низкая скорость роста (10% за полгода)

Смежные области, технологии

- BigData
- базы знаний

Оценка объемов

<http://www.protown.ru/information/hide/3598.html> - численность ученых по странам
<http://rf-smi.ru/cul/31091-kolichestvo-uchenyh-i-lzheuchenyh-v-mire-pochti-sravnyalos.html> - количество ученых и лжеученых (6 млн.)

Задачи (процессы)

- поиск необходимой информации пользователями
- систематизация (классификация), группировка, оценивание
- ввод артефактов, экспертная оценка, уровень однозначности
- взаимодействие с другими наукометрическими БД и технологическими БД (мониторинг)
- взаимодействие с поисковыми системами
- API для поиска и для пополнения

Научные проблемы

Представление знаний (фактов) с учетом их достоверности, общепризнанности, множественной интерпретации, доказательности (формальное доказательство, общепризнанный “житейский” факт, догма (вера)). Организация поиска, группировка, систематизация и т.п..

<http://www.swsys.ru/index.php?id=2555&page=article> - Расширение базы знаний с учетом доверия к новому знанию

Автоматизация

- реферирование и систематизация научных статей (обнаружение фейков, оценка значимости)

БД научных и технологических артефактов

- наукометрические БД (Scopus 2009 - $38 \cdot 10^6$, РИНЦ)
- БД по нейроинформатике <http://www.bioinformatix.ru/bioinformatika/bazyi-dannyih-po-bioinformatike.html>
- Protein Data Bank (10^5) https://ru.wikipedia.org/wiki/Protein_Data_Bank
- БД исторических артефактов - на уровне идей, отсутствие инфраструктуры и согласия
- БД литературных артефактов
- БД медицинских артефактов
- Свободное ПО (репозитории)

Риски, проблемы

- классификация - оригинальная в науке (академические, гуманитарные, прикладные), медицина, религия
- социальные аспекты - противодействие: наука, псевдонаука, религия, политика, пропаганда
- альтернативные точки зрения (общепризнанность факта) “Волга впадает в Каспийское море”, интерпретации, толкования. При жестких рамках - ограничение **заинтересованных пользователей**, при широких - противоречия, взаимоисключающие суждения
- правовые аспекты - авторское право, “клевета”
- языки, проблемы перевода, интернационализация и локализация, глобальная и национальные базы знаний
- метрика (количественная, качественная) - сопоставление артефактов

Ресурсы

- оценка объемов - необходимость собственного датацентра
- наука - базы знаний, ориентированные на суждения, системы логического вывода и поиска в таких базах знаний
- разработка как таковая
- сопровождение технологическое
- эксперты - штатные, внештатные
- финансирование

Ссылки

<https://habrahabr.ru/post/195650/> - базы знаний. Введение