

Картографическое представление текстового контента

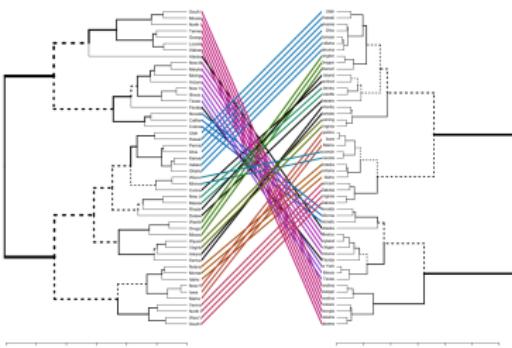
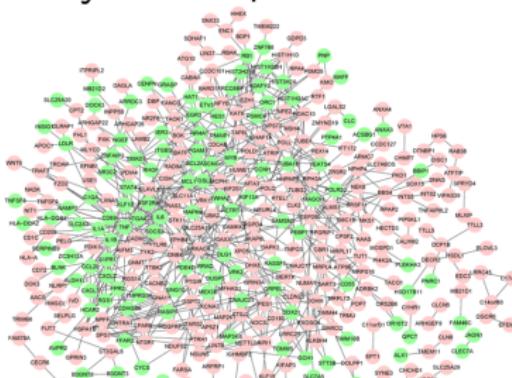
Д.В.Лемтюжникова

МФТИ, МАИ

Москва, 2019

Сложность

Сложность существующих конструкций для визуализации больших блоков данных



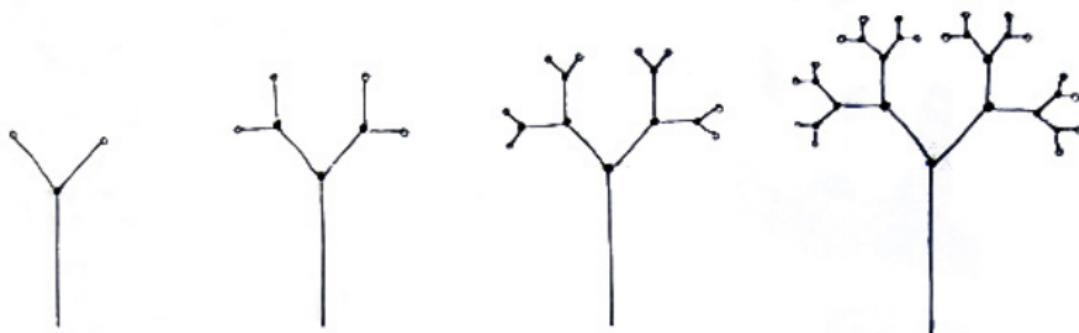
Простота

Простые формы малоинформационны



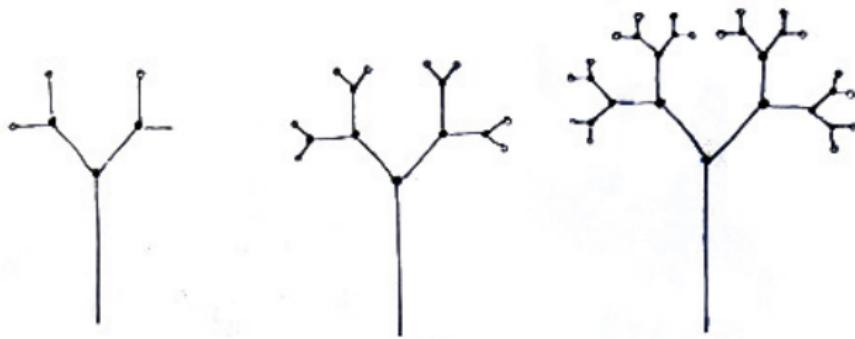
Фракталы – способ представления иерархии знаний

Знание иерархично, обладает нетривиальной структурой на разных масштабах → фрактальна

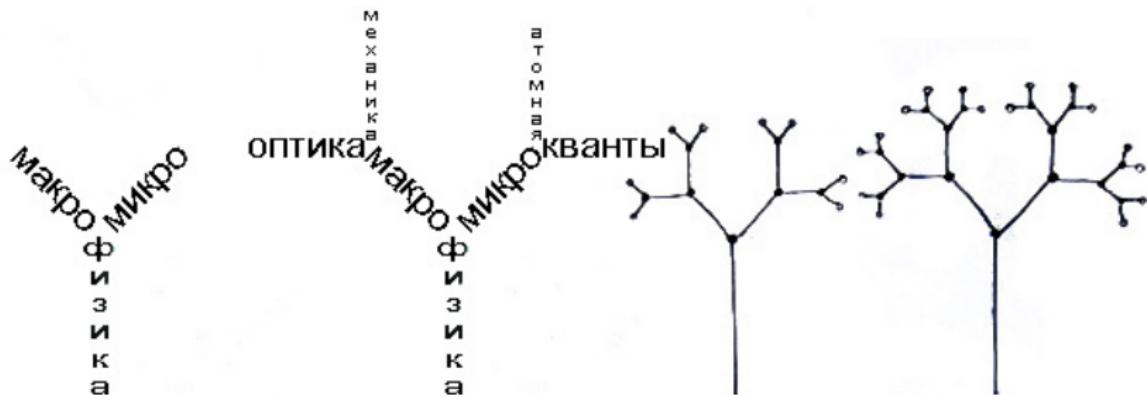


Области знаний

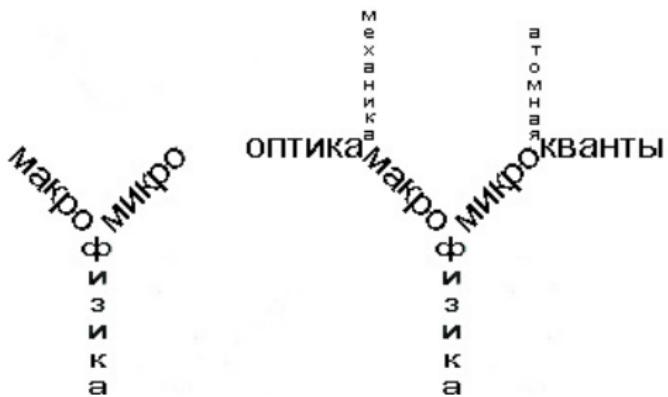
макро
микро
физика



Вершины фракталов – коллекции документов



Дружественный интерфейс и фракталы

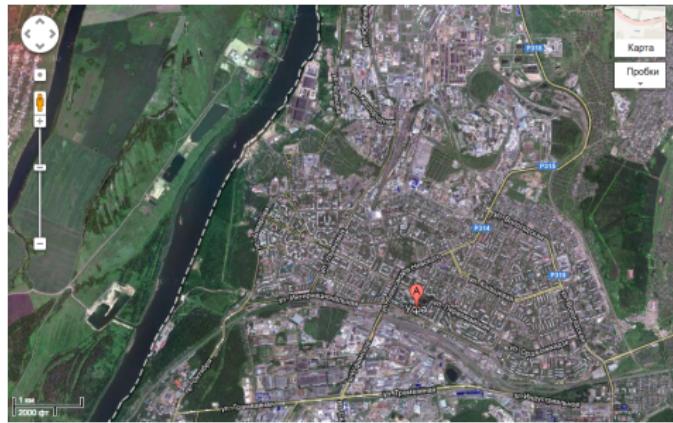


Знакомая метафора

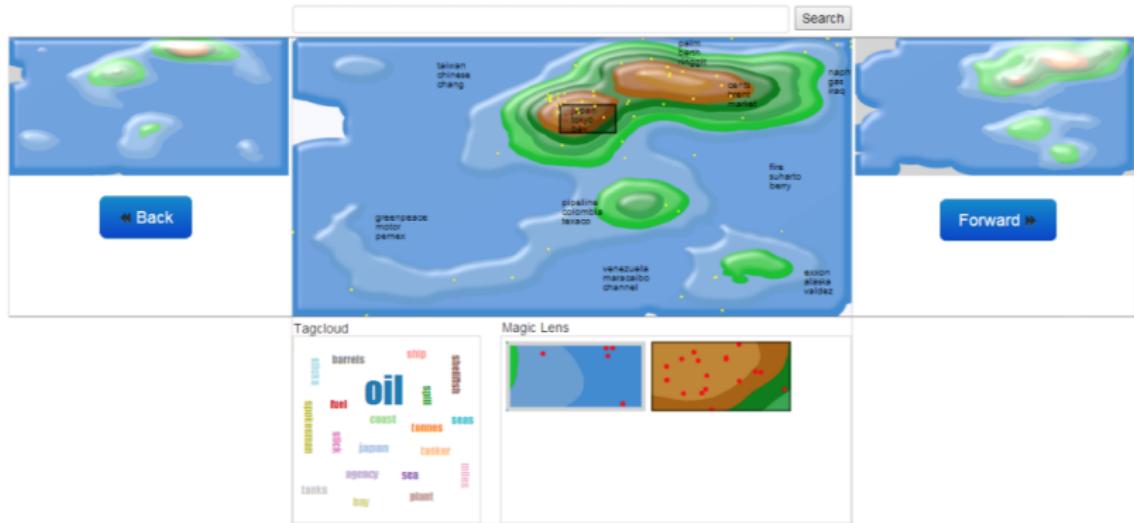
- Знание, как и незнание, постоянно растёт
- Знание необходимо упорядочить



Привычный интерфейс



Пример работы с картами



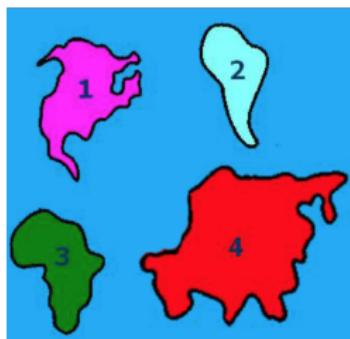
Как строить карту?



Знание и незнание

знание — «материки»

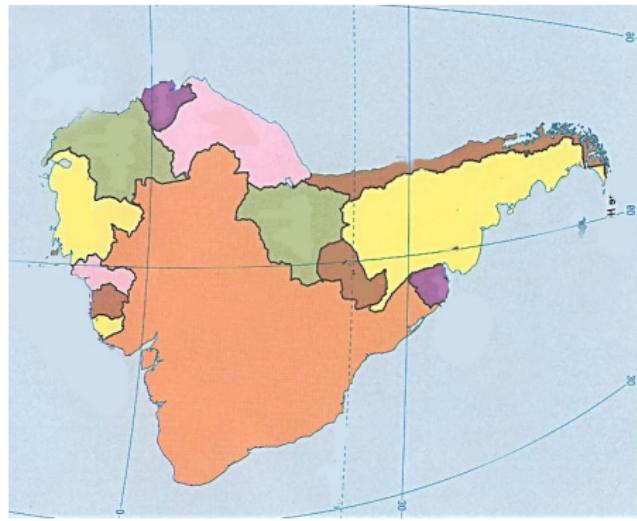
незнание — «вода»



- 1) о живой природе
- 2) естественное
- 3) техническое
- 4) социально-гуманитарное

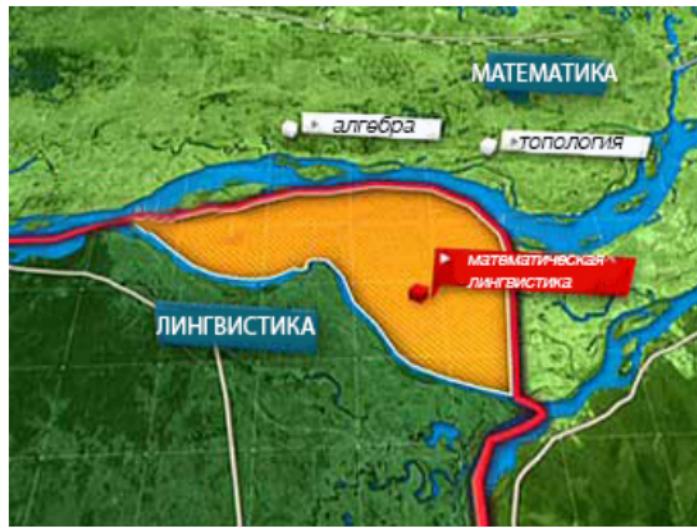
Страны

Каждый «материк» состоит из соответствующих наук — «стран материка»



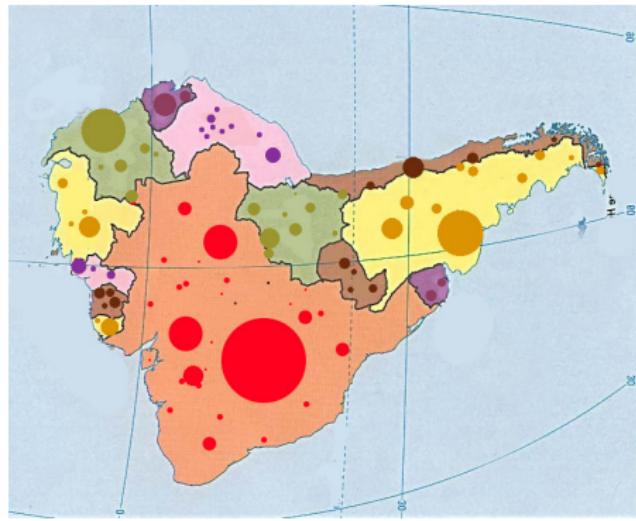
Неоднозначные выводы

Спорные территории между «странами» — междисциплинарные науки



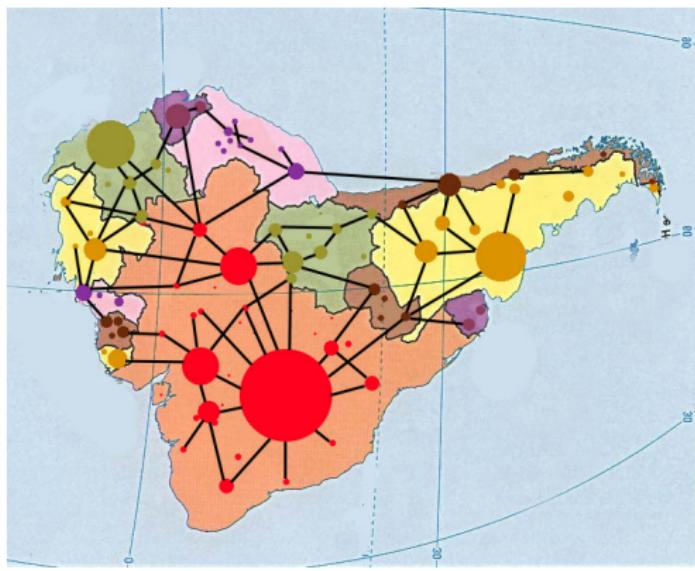
Города и тематическое моделирование

Каждая «страна» содержит «населённые пункты» — изучаемые темы



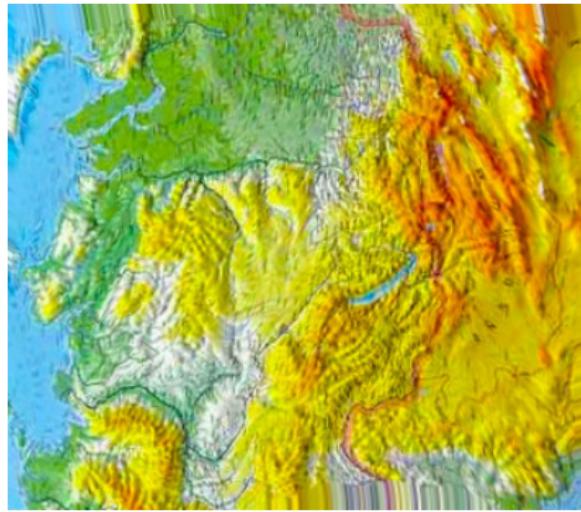
Дороги

Многие темы связаны друг с другом



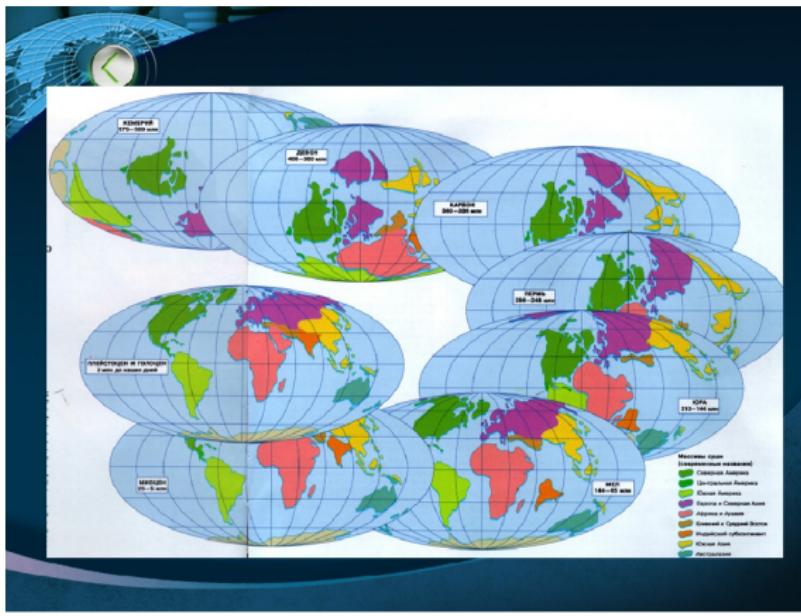
Уровень разработанности

«Карта высот» — степень глубины знания:
дальность от «уровня моря» — незнания



История развития

Как меняется наше знание



Наукоёмкость

Как отделять научное знание от ненаучного?

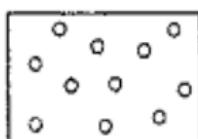


От знаний к документам

«населённые пункты» — темы
«дома» — документы

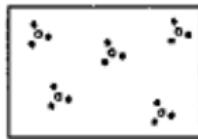


Обозначения для научных источников знания



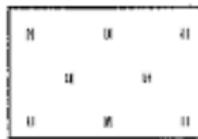
Статьи

≡ Монография



Диссертации

▽ Информационный тупик



Тезисы

! Спорный источник

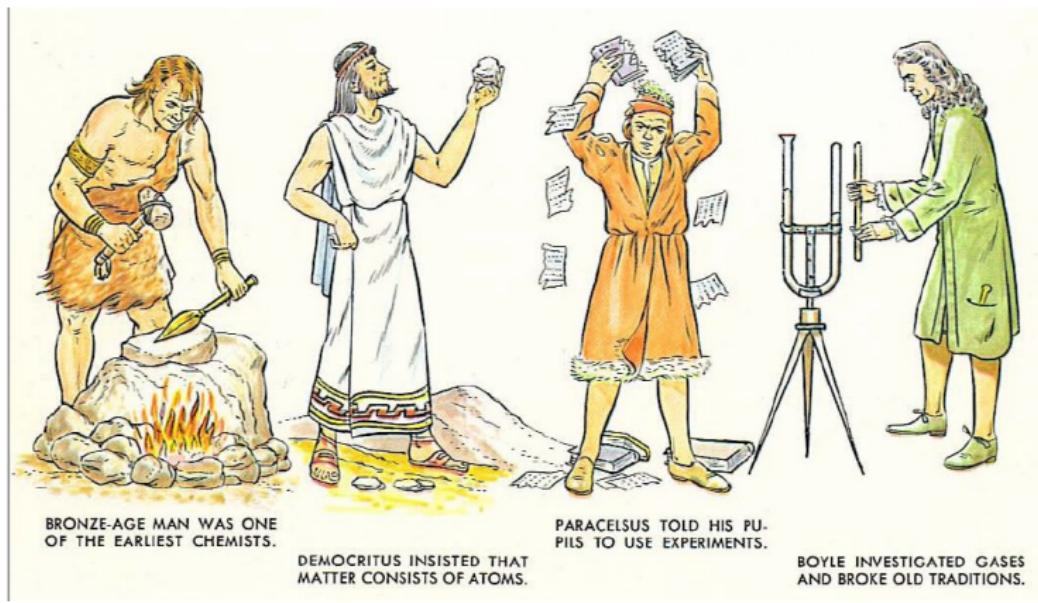
↗ Научная школа

Позиционирование документа в теме

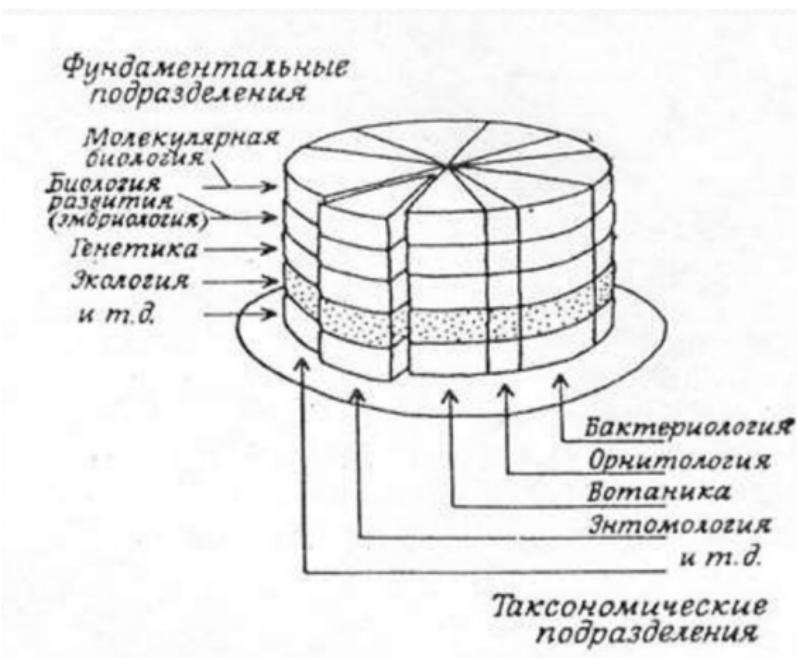
- 👑документ, принципиально важный для понимания темы, обзор
- 👑сильный научный результат в данной теме
- 👑документ открыл новую ветку исследований
- ⚡оригинальный ход
- ⚡прямолинейное развитие темы
- ⚡небольшое развитие темы

Плоская карта (антиУДК)

Эмпирическое определение принципов знания



Пирог биологических наук



Задача информационного поиска



Проблема перегруженности статистики поиска
(стоп-слова в NLP)
Проблема «найди то-не- знаю-что»

Информационный поиск — процесс извлечения документов из массива



Анализ новостных потоков

- иерархия
-
- эмоциональная окраска
- регуляция

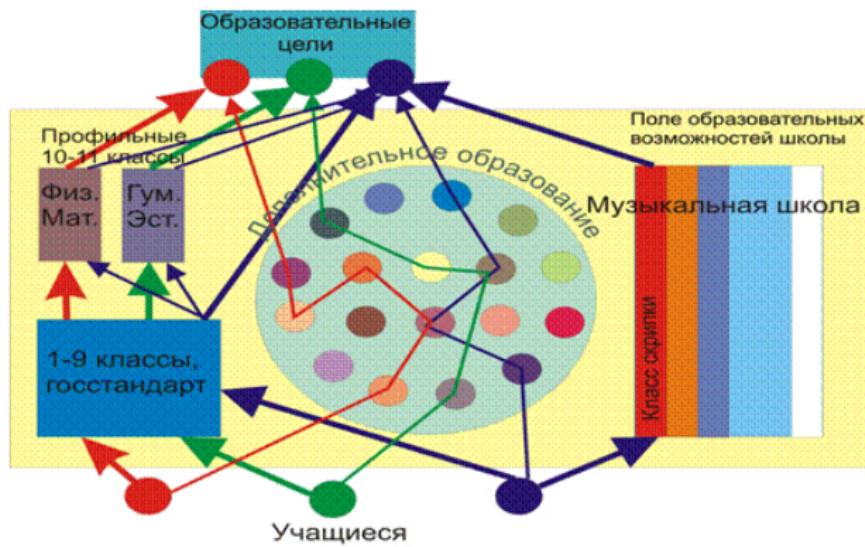


Интерфейс пользователя как игра

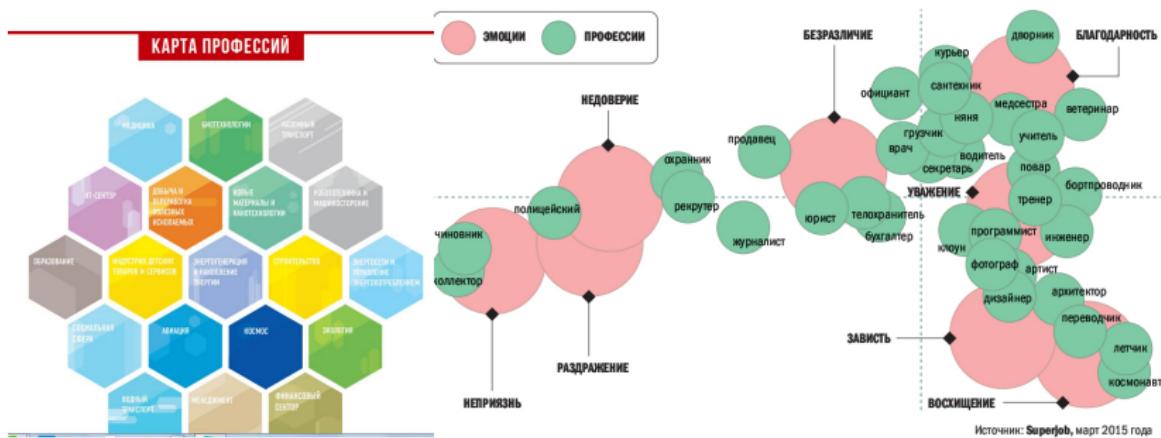
Поиск информации на карте знаний



Траектория индивидуального развития пользователя



Сейсмическая карта профессий



Спасибо за внимание!
Дарьяна Лемтюжникова darabtt@gmail.com