

Система статистического, морфологического, кластерного и онтологического анализа данных для ЖКХ

Студент: Желепов А.С.

Группа: ПИмд-11

Кластерный анализ

Задача: разделить массив слов (словосочетаний) на категории «является термином» и «не является термином»

Способ достижения цели: применение FCM-алгоритма

Выбор центров кластеров: набор максимальных и минимальных характеристик

Встречаемость слова	Loglikelihood	Mutual Information	Tscore
5	3.3	7.4	0.98
7	4.5	9.8	0.65
15	2.3	13.5	0.34
1	1.4	8.1	0.68



Процедура сортировки

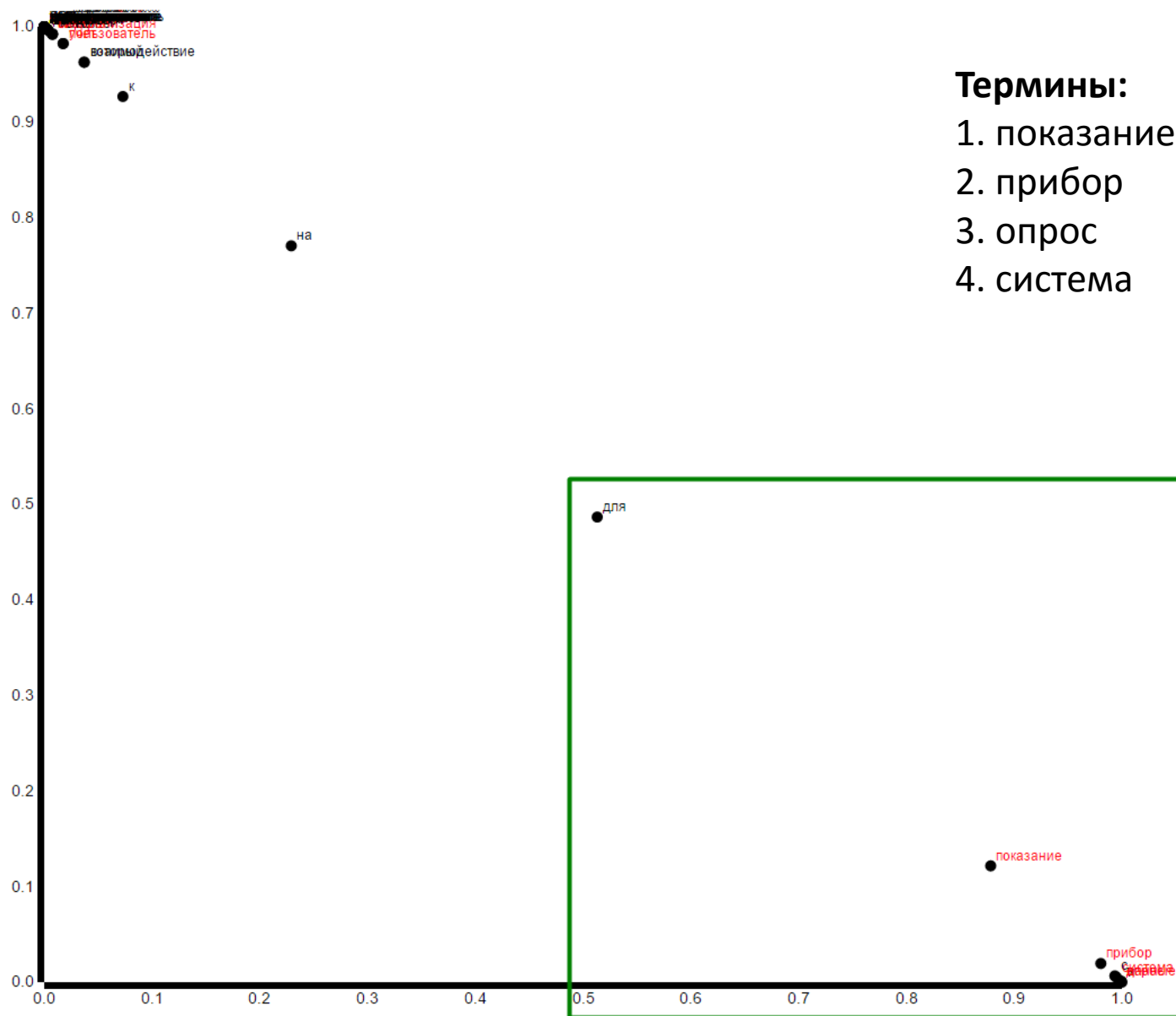
Встречаемость слова	Loglikelihood	Mutual Information	Tscore
1	1.4	7.4	0.34
5	2.3	8.1	0.65
7	3.3	9.8	0.68
15	4.5	13.5	0.95

«Нетермин»

↓ «Термин»

Слова

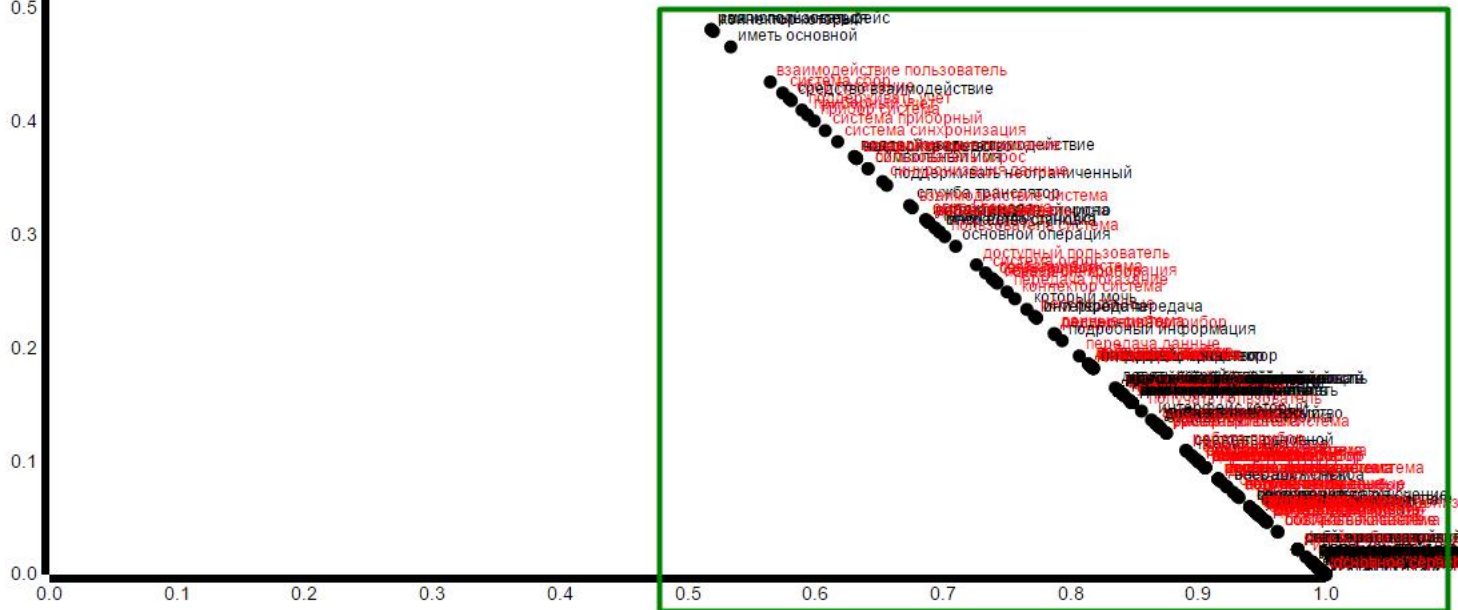
1



Термины:

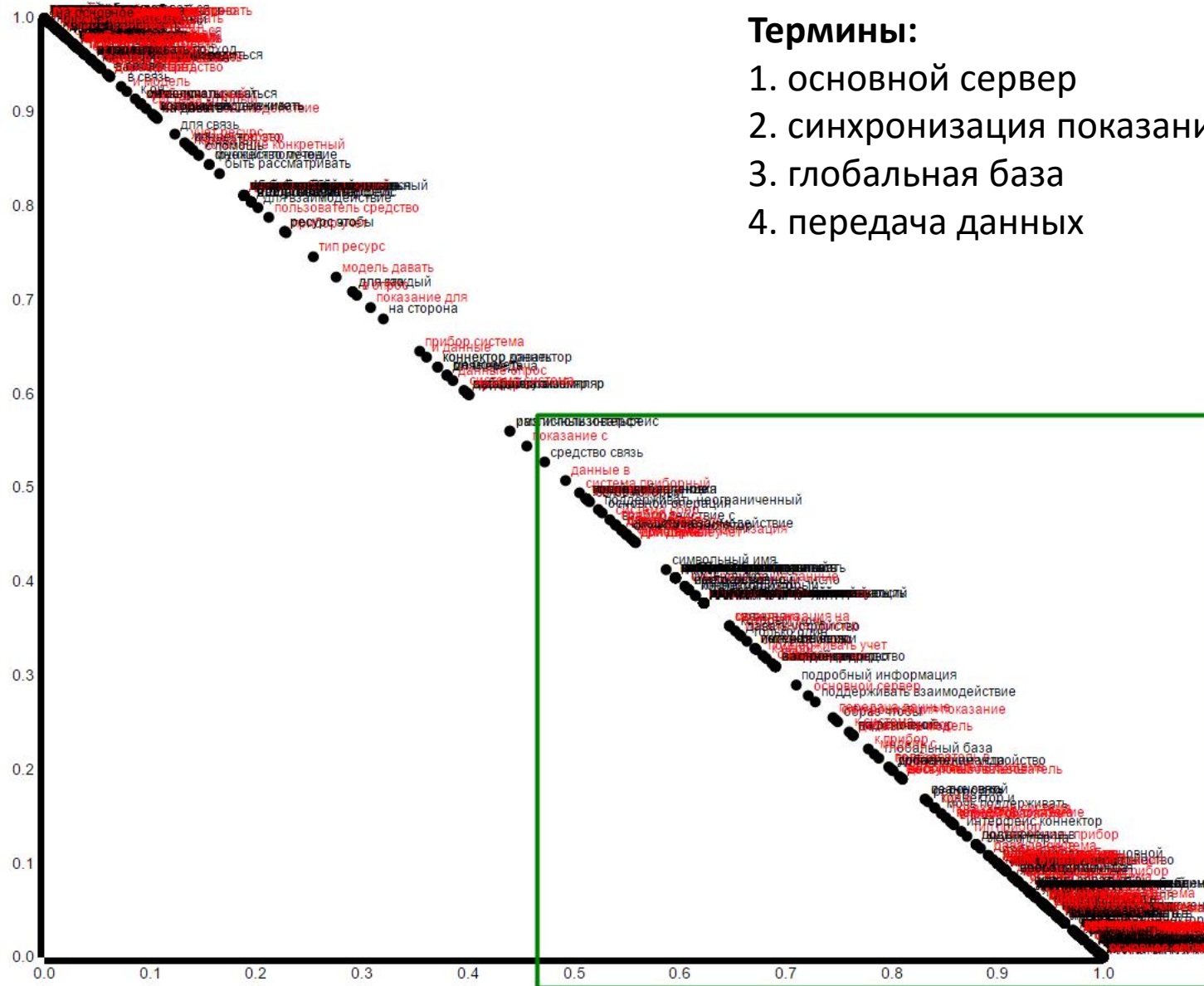
1. показание
2. прибор
3. опрос
4. система

2



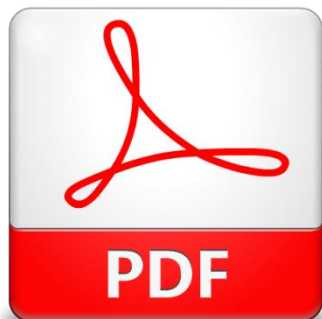
Словосочетания с предлогами

3



Дополнительные возможности

1



Обработка...



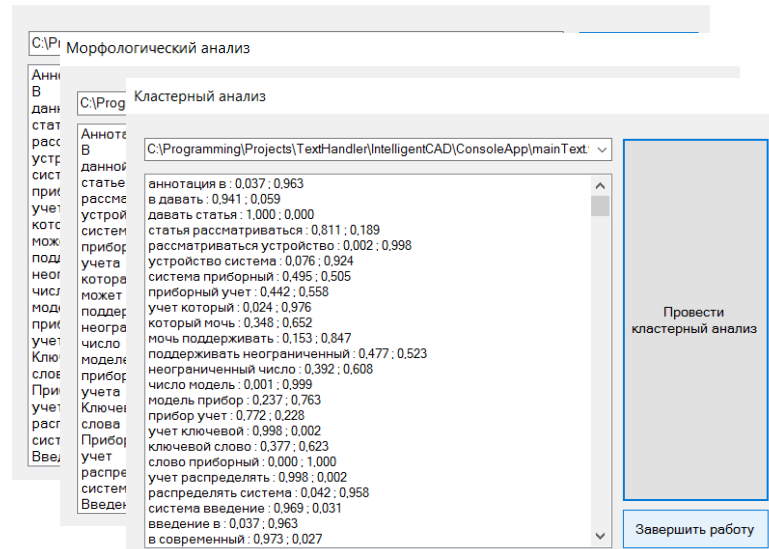
Сохранение...



2

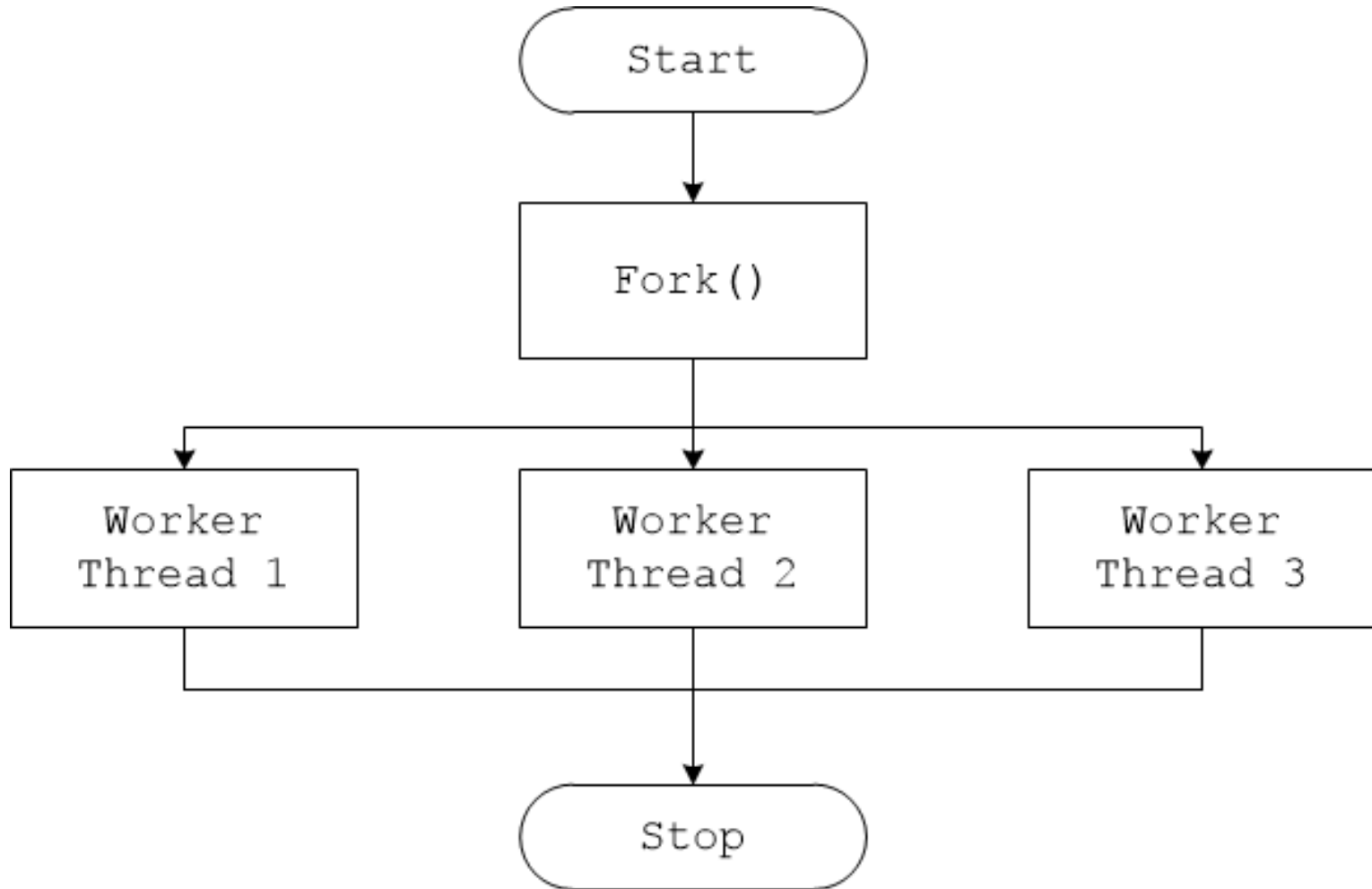


Обработка mystem.exe



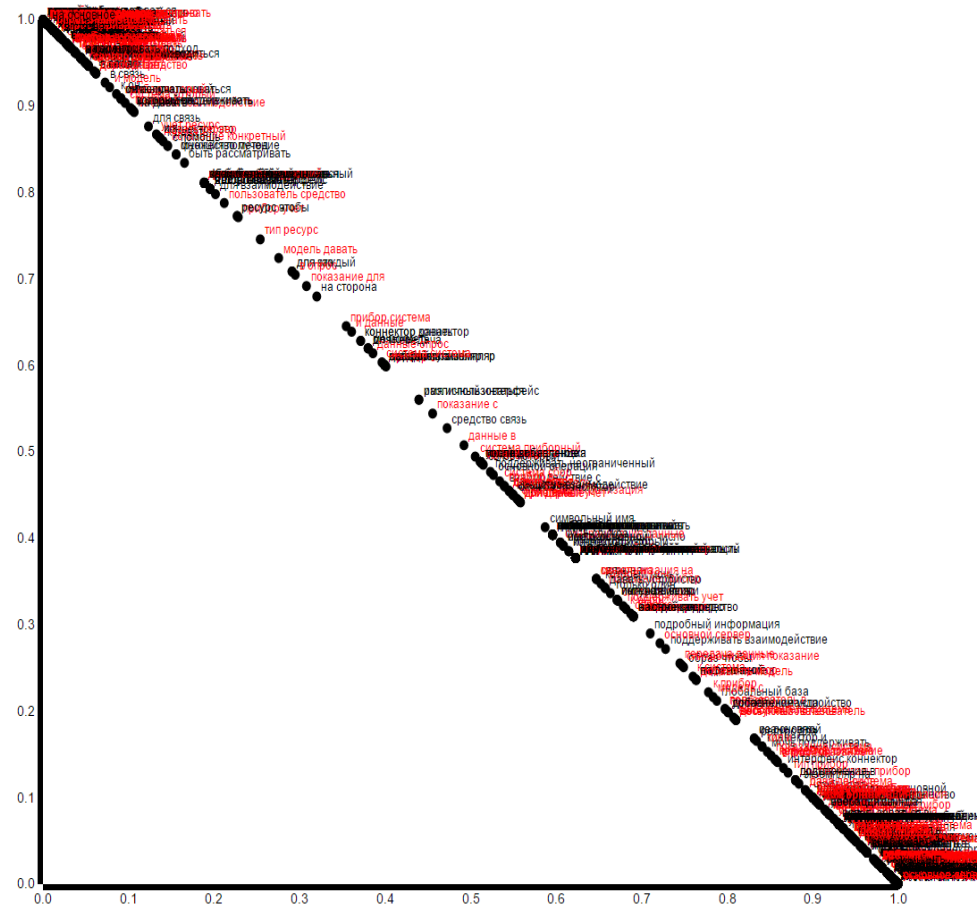
Дополнительные возможности

3



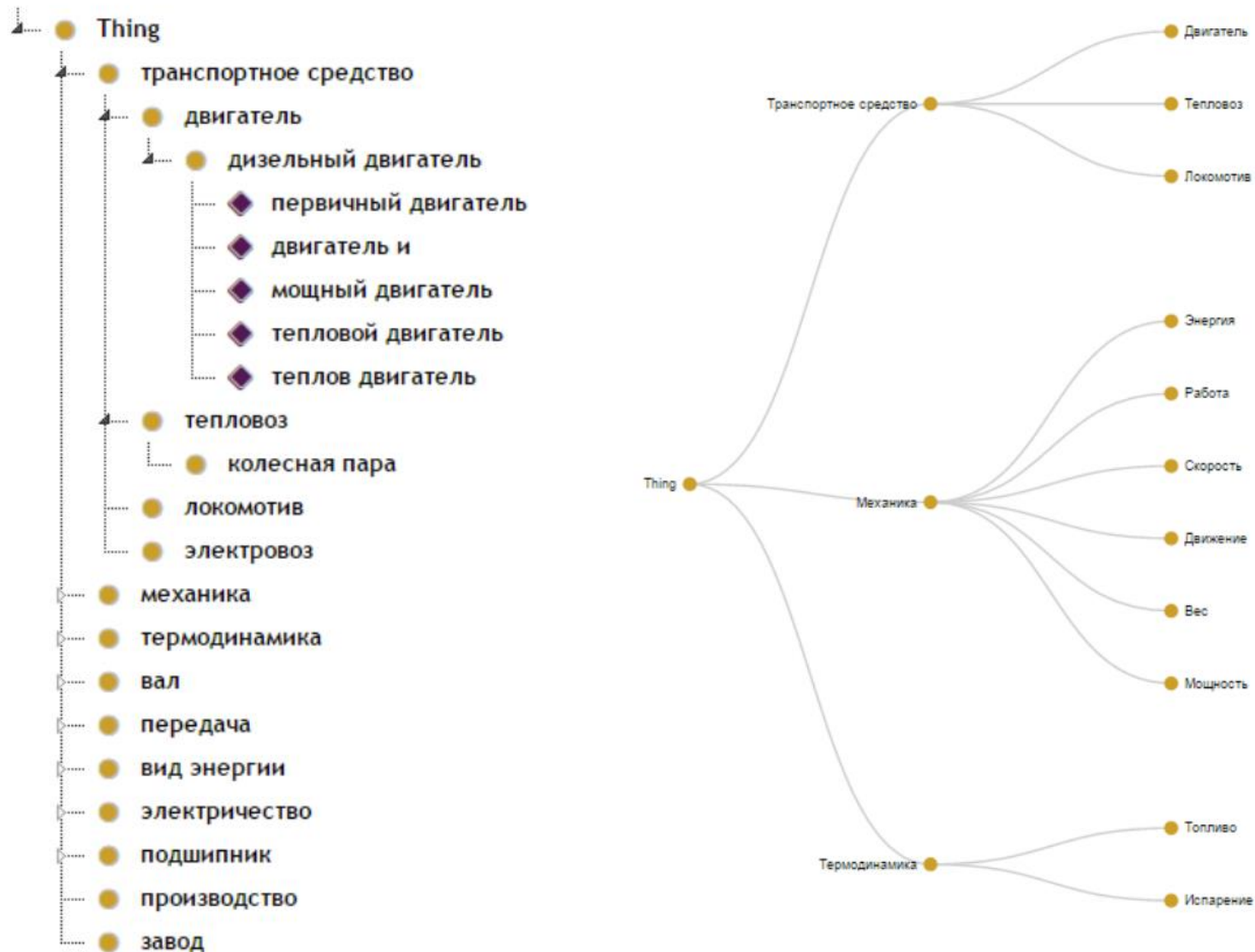
4

Перетащите файлы сюда...



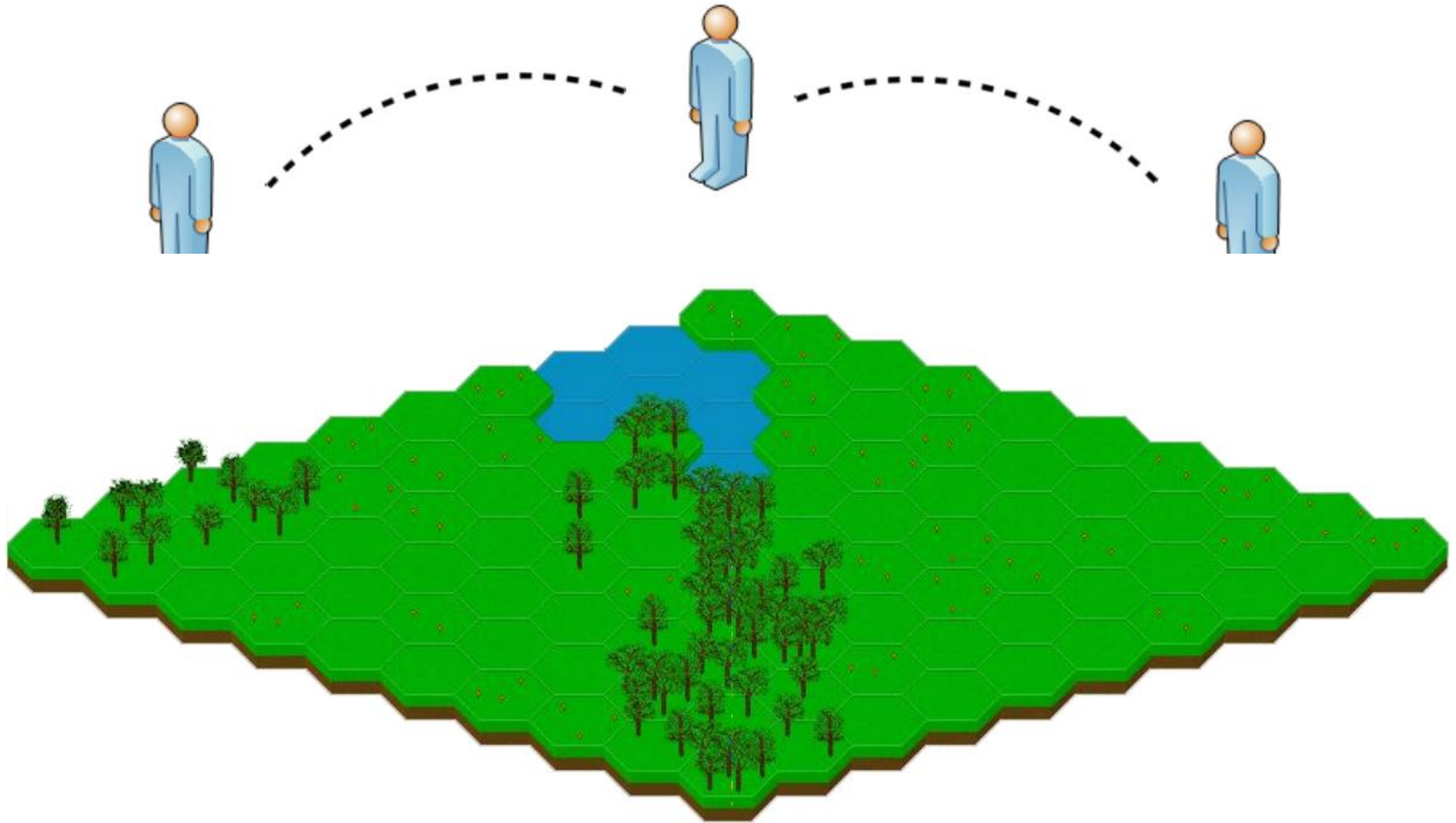
Научно-исследовательская работа

1



AI-модель сбора информации об окружающем мире

2



Спасибо за внимание!