

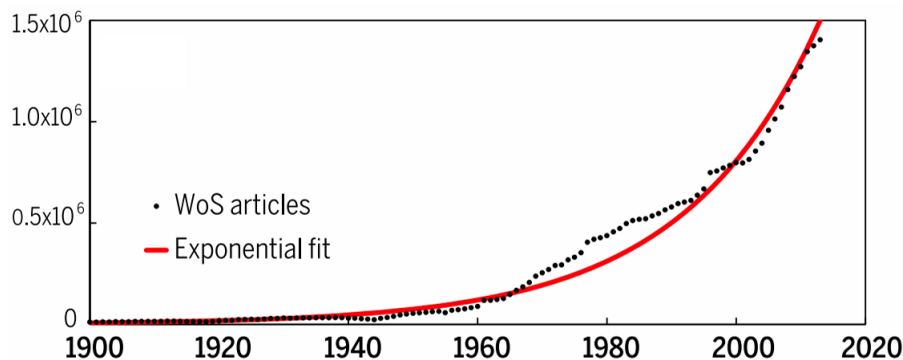
МОДЕЛИРОВАНИЕ И АНАЛИЗ НАУЧНЫХ ОБЪЕКТОВ В ИНФОРМАЦИОННОЙ СИСТЕМЕ АНАЛИЗА НАУЧНОЙ ДЕЯТЕЛЬНОСТИ

Губанов Дмитрий Алексеевич
д.т.н., зав. лабораторией ИПУ РАН,
профессор МФТИ, ВМК МГУ

НТИ'2025, 25 сентября 2025 г.

Информационная система анализа научной деятельности

- Спрос на разработку информационных систем, автоматизирующих работу с большими массивами публикаций
- Любая такая система должна содержать базу публикаций
- Состав аналитических инструментов работы с публикациями в разных системах различен и определяется целями, которые ставят перед собой создатели таких систем



В 2022 году в мире опубликовано
3,3 млн научных статей
(1,6 млн в открытом доступе)

*Ежегодное издание научных статей,
индексируемых в базе данных WoS*

Базы научных публикаций (1)

- РИНЦ
- Web of Science
- Scopus
- OpenAlex
- Google Scholar
- ResearchGate,
- Academia
- ...

Главная цель – анализ цитируемости публикаций, на основе которого вычисляются наукометрические оценки публикаций и их авторов (индекс Хирша), а также научных журналов (импакт-фактор)

Базы научных публикаций (2)

- Semantic Scholar
- Aminer
- Dimensions
- iFORA ...

*Задачи анализа
содержания научных
текстов*

Методы структурирования тематического пространства (универсальные классификаторы)

- УДК
- OECD
- Классификатор РНФ
- ГРНТИ
- ...

00	ОБЩЕСТВЕННЫЕ НАУКИ В ЦЕЛОМ
02	ФИЛОСОФИЯ
03	ИСТОРИЯ. ИСТОРИЧЕСКИЕ НАУКИ
04	СОЦИОЛОГИЯ
05	ДЕМОГРАФИЯ
06	ЭКОНОМИКА И ЭКОНОМИЧЕСКИЕ НАУКИ
10	ГОСУДАРСТВО И ПРАВО. ЮРИДИЧЕСКИЕ НАУКИ
11	ПОЛИТИКА И ПОЛИТИЧЕСКИЕ НАУКИ
12	НАУКОВЕДЕНИЕ
13	КУЛЬТУРА. КУЛЬТУРОЛОГИЯ

*Слишком крупное членение научных направлений и,
соответственно, слишком слабая дифференциация разделов
этих направлений*

Информационная система анализа научной деятельности (ИСАНД)

Ключевая задача — позиционировать научные объекты в едином тематическом пространстве теории управления

Разработка ИСАНД мотивирована запросами следующих потенциальных пользователей:

- исследователи
- редакции научных журналов и организаторы научных конференций
- руководители научных и образовательных организаций, подразделений и коллективов
- организаторы науки

Потенциальные пользователи ИСАНД

Исследователи	<ul style="list-style-type: none">• Анализ актуальных научных направлений• Изучение ключевых понятий• Идентификация влиятельных агентов (исследователей, журналов, конференций, организаций) и научных публикаций
Редакции научных журналов и организаторы научных конференций	<ul style="list-style-type: none">• Соответствие представленных материалов журналу либо конференции• Поиск квалифицированных рецензентов• Поиск потенциальных участников конференции

Потенциальные пользователи ИСАНД

Руководители научных и образовательных организаций, подразделений и коллективов	<ul style="list-style-type: none">• Поиск новых сотрудников и участников проектов• Анализ актуальности научных направлений внутри научной или образовательной организации
Организаторы науки	<ul style="list-style-type: none">• Вопросы, связанные с организационными структурами (организации, подразделения, научные коллективы и исследователи)• Прогноз и оценка перспективности направлений научных исследований

Онтология ИСАНД

Онтология – формальная спецификация согласованного описания (концептуализации) предметной области, разрабатываемая группой экспертов и интерпретируемая как машиной, так и человеком (OWL)

```
Individual: :a1
  Types: JournalArticle
  Facts:
    citesAsDataSource :d1,
    holdsStatusInTime :s1

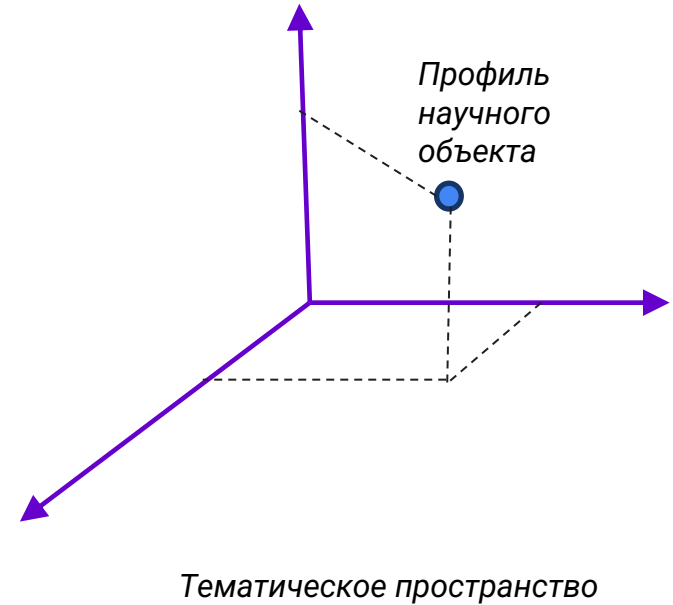
Class: :HighlyCitedArticle
  EquivalentTo:
    JournalArticle
    and (isCitedBy min 100 Work)
```

Онтология ИСАНД ориентирована на представление и сбор знаний в области теории управления:

- как **научного знания**
(онтология теории управления),
- так и знаний, связанных с **научной деятельностью** агентов в данной области

Онтология научного знания (теория управления)

- Предлагаемый классификатор является «системой координат» тематического пространства
- Он имеет четырехуровневую структуру, уровни которой (кроме нижнего) представляют собой дерево
- Уровни нумеруются числами от 0 до 3



Онтология научного знания: метафакторы

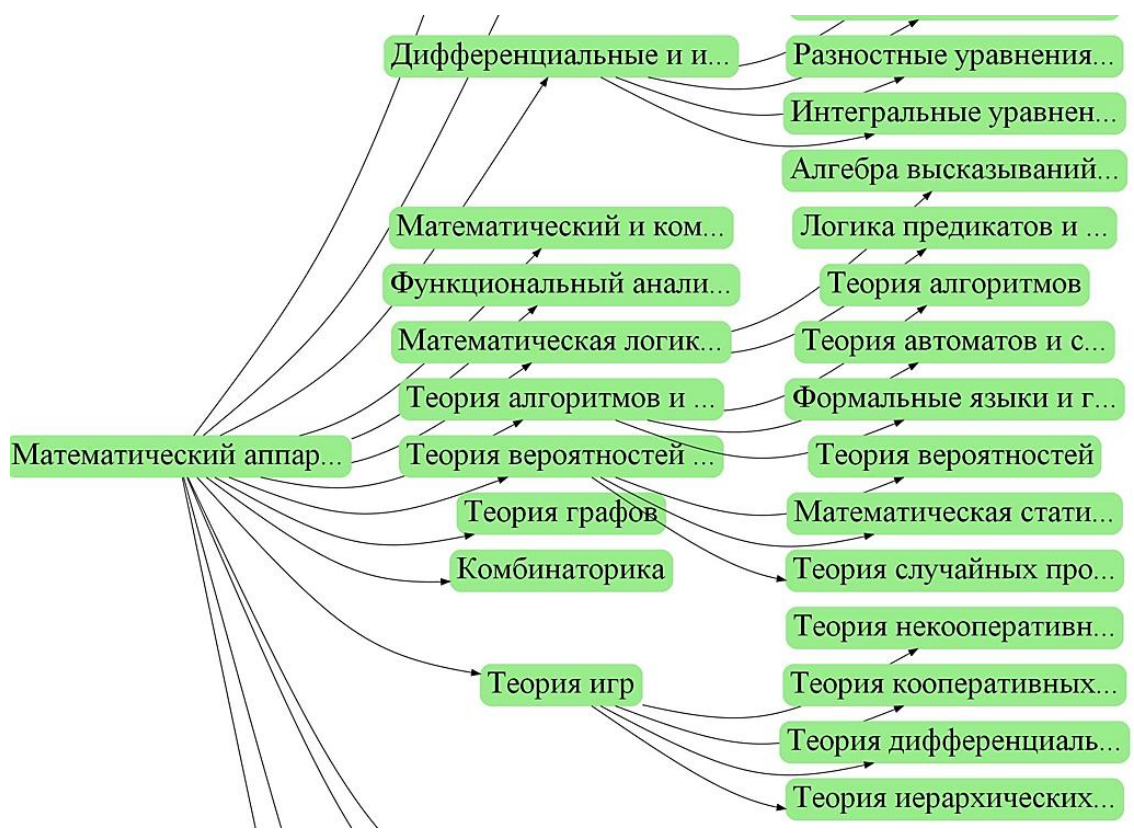
Нулевой уровень содержит 3 фиксированные вершины

- «Математический аппарат»
- «Предметная область»
- «Сфера применения»

Онтология научного знания: факторы и подфакторы

- «Математический аппарат»
 - «Теория игр»
 - «Теория некооперативных игр»
- «Предметная область»
 - «Теория управления в организационных системах»
 - «Механизмы планирования»
- «Сфера применения»
 - «Энергетика»
 - «Атомная энергетика»

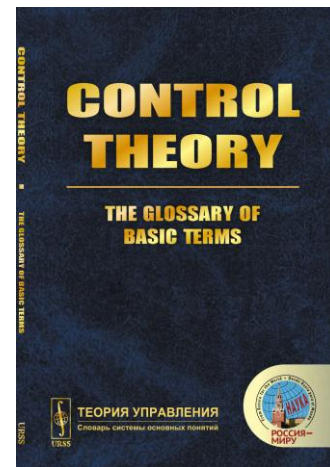
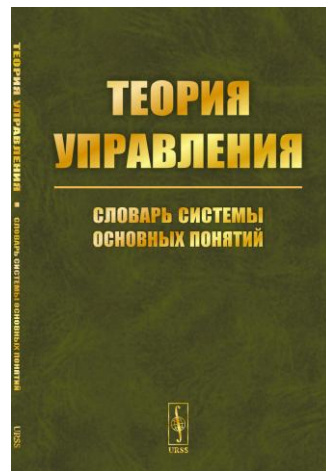
Фрагмент онтологии научного знания (граф классификатора)



Онтология научного знания

Классификатор построен экспертами ИПУ РАН для теории управления, на данный момент он включает в себя

- 3 фактора нулевого уровня
- 53 фактора первого уровня
- 161 фактор второго уровня
- >3000 терминов



Онтология научной деятельности

Онтология научной деятельности предназначена для описания *научных объектов* (агентов и результатов их действий):

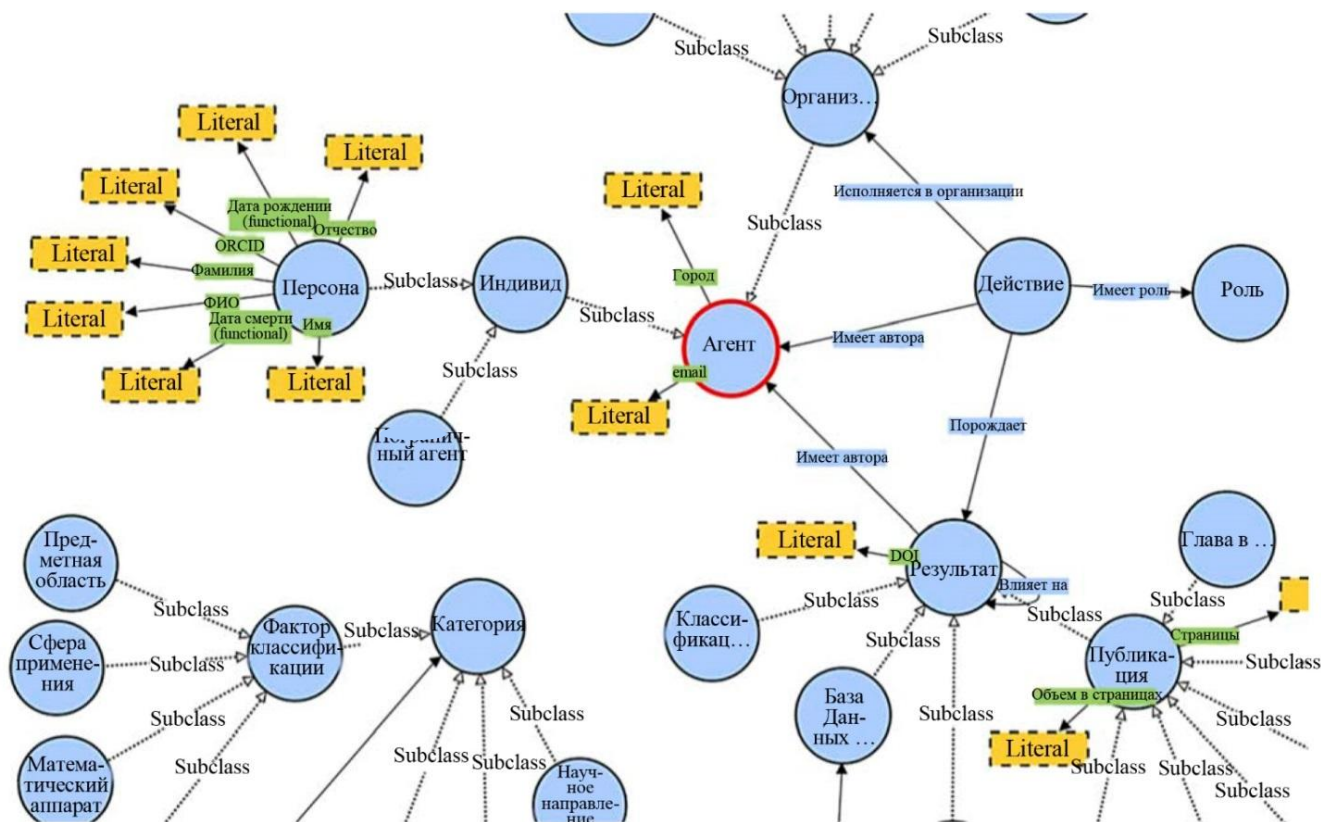
- Автор
- Организация
- Журнал
- Конференция
- Публикация

Текущая версия онтологии (OWL*) включает в общей сложности **50** классов в таксономии (например, «Публикация»), **25** объектных свойства (например, «*влияет на*»), **40** простых свойств (например, «*название*»)

Онтология включает в себя девять классов верхнего уровня, в том числе классы «Агент», «Действие», «Результат», «Категория», «Роль» и «Профиль»

*The Web Ontology Language (OWL) is a **family of knowledge representation languages** for authoring ontologies.

Фрагмент онтологии научной деятельности



Профиль публикации

- $V = \{v_1, \dots, v_n\}$ – множество вершин 1-го уровня (факторов)
- $V_i = \{v_{i1}, \dots, v_{in_i}\}$ – множество вершин 2-го уровня (подфакторов) для i -й вершины 1-го уровня, $m = \sum_{i \in N} n_i$ - общее число подфакторов
- $Q_{ij} = \{1, \dots, q_{ij}\}$ – множество вершин-терминов, характеризующих ij -й подфактор
- L – множество публикаций
- Δ_{lij} – сумма числа вхождений в l -ю публикацию базовых терминов из ij -го подфактора
- Профиль второго уровня публикации l :

$$x_l = (x_{l1}, \dots, x_{lij}, \dots, x_{lm}), \quad \text{где } x_{lij} = \frac{\Delta_{lij}}{\sum_{i \in N} \sum_{j \in N_i} \Delta_{lij}},$$
$$l \in L, \quad j \in N_i, \quad i \in N.$$

- Профиль первого уровня публикации l :

$$X_l = (X_{l1}, \dots, X_{li}, \dots, X_{ln}), \quad \text{где } X_{li} = \sum_{j \in N_i} x_{lij}, \quad l \in L, \quad i \in N.$$

Профиль ученого

Обозначим

- K – множество ученых
- $r(l)$ – количество авторов l -ой публикации
- $\omega(k, l) = \begin{cases} 1, & \text{если } k\text{-й ученый является автором } l\text{-ой публикации;} \\ 0, & \text{в противном случае;} \end{cases}$

Профили второго и первого уровня k -го ученого считаются на основе его публикаций

$$y_{ij}^k = \frac{\sum_{l \in L} \omega(k, l) \frac{x_{lij}}{r(l)}}{\sum_{i \in N} \sum_{j \in N_i} \sum_{l \in L} \omega(k, l) \frac{x_{lij}}{r(l)}}, \quad k \in K, \quad j \in N_i, \quad i \in N$$

$$Y_i^k = \sum_{j \in N_i} y_{ij}^k, \quad k \in K, \quad i \in N$$

Расстояние между профилями

Предлагается применять следующее расстояние между двумя профилями, задаваемыми стохастическими векторами $\alpha = (\alpha_1, \dots, \alpha_n)$ и $\beta = (\beta_1, \dots, \beta_n)$:

$$d(\alpha, \beta) = 1 - \sum_{j=1}^n \min(\alpha_j, \beta_j) = \frac{1}{2} \sum_{j=1}^n |\alpha_j - \beta_j|$$

Замечание В данной метрике расстояние между профилями первого уровня всегда не превосходит расстояние между профилями 2 уровня тех же объектов

Примеры профилей 1-го уровня



Профили ученых ИПУ 1-го уровня (52 фактора) в двумерном пространстве

Новиков Д.А.

1-й уровень	Значение
Теория игр	0,032
Теория графов	0,016
Теория вероятностей и мате	0,014
Теория множеств и отноше	0,008
Комбинаторика	0,005
Теория управления в органи	0,408
Теория выбора и принятия р	0,111
Кибернетика и системный а	0,064
Исследование операций	0,055
Информационная безопаснос	0,035
Образование	0,070
Социальные системы	0,027
Социально-экономические с	0,011
Робототехника	0,011
Военное дело	0,008

Кузнецов О.П.

1-й уровень	Значение
Теория графов	0,085
Математическая логика	0,038
Дифференциальные и интег	0,017
Функциональный анализ	0,012
Теория алгоритмов и форма	0,012
Искусственный интеллект и	0,319
Теория управления в органи	0,133
Информационная безопасно	0,104
Теория выбора и принятия р	0,087
Теория автоматического уп	0,032
Социальные системы	0,050
Робототехника	0,011
Образование	0,008
Технологические процессы	0,005
Авиация	0,000

Поляк Б.Т.

1-й уровень	Значение
Теория оптимизации	0,113
Информатика и теория инфо	0,079
Дифференциальные и интег	0,054
Алгебра и теория чисел	0,012
Математическая логика	0,007
Теория автоматического уп	0,491
Кибернетика и системный а	0,058
Навигация и управление дви	0,042
Мехатроника	0,020
Теория выбора и принятия р	0,020
Образование	0,012
Робототехника	0,004
Космос	0,004
Социальные системы	0,002
Военное дело	0,000

Профиль ИПУ РАН (первые 10 компонент по убыванию)

№	Метафактор	Фактор	Вес
1	Предметная область	Теория автоматического управления	0.102
2	Предметная область	Кибернетика и системный анализ	0.086
3	Математический аппарат	Физика	0.063
4	Предметная область	Искусственный интеллект и интеллектуальное управление	0.048
5	Математический аппарат	Теория вероятностей и математическая статистика	0.045
6	Предметная область	Навигация и управление движением	0.035
7	Предметная область	Теория управления в организационных системах	0.035
8	Математический аппарат	Теория оптимизации	0.033
9	Математический аппарат	Дифференциальные и интегральные уравнения	0.032
10	Сфера применения	Авиация	0.028

Тематически ближайшие к ИПУ РАН конференции

Тип ▲	Сущность ♦	Расстояние ♦	Сходство ♦	Источник ♦
Конференция	Всероссийское совещание по проблемам управления (ВСПУ, Москва)	0,159	Теория автоматического управления • Кибернетика и системный анализ • Физика	БД ИПУ РАН
Конференция	Международная конференция "Управление развитием крупномасштабных систем" (MLSD)	0,207	Кибернетика и системный анализ • Теория автоматического управления • Искусственный интеллект и интеллектуальное управление	БД ИПУ РАН
Конференция	IFAC Conference on Networked Systems NECSYS	0,211	Теория автоматического управления • Кибернетика и системный анализ • Физика	БД ИПУ РАН

Научные объекты в произвольном базисе

$\{\mathbf{e}_j = (e_{j1}, \dots, e_{jm})\}_{j=1, \dots, k}$ – набор базисных векторов
(вообще говоря, не являющийся естественным)

Тогда научный объект $\mathbf{x} = \sum_{j=1}^k x_j \mathbf{e}_j$ в естественном базисе характеризуется m -компонентным стохастическим вектором $(\sum_{j=1}^k x_j e_{j1}, \dots, \sum_{j=1}^k x_j e_{jm})$.

Расстояние между объектами в произвольном базисе

Расстояние между научными объектами \mathbf{x} и \mathbf{y} :

$$d(\mathbf{x}, \mathbf{y}) = \frac{1}{2} \sum_{l=1}^m \left| \sum_{j=1}^k e_{jl}(x_j - y_j) \right|$$

Информационная система анализа научной деятельности (ИСАНД) в области теории управления

- База научных публикаций по теории управления
- Другие научные объекты, связанные с публикациями:
 - Авторы
 - Организации
 - Журналы
 - Конференции
- Формальные модели представления научных объектов
- Набор инструментов поиска научных объектов
- Набор аналитических инструментов

ИСАНД в других научных областях?

Данные
о публикациях и
других научных
объектах

Онтология
научного знания
(классификатор)

ИСАНД
в научной области
NNN

Спасибо за внимание!

Губанов Дмитрий Алексеевич
gubanov@ipu.ru