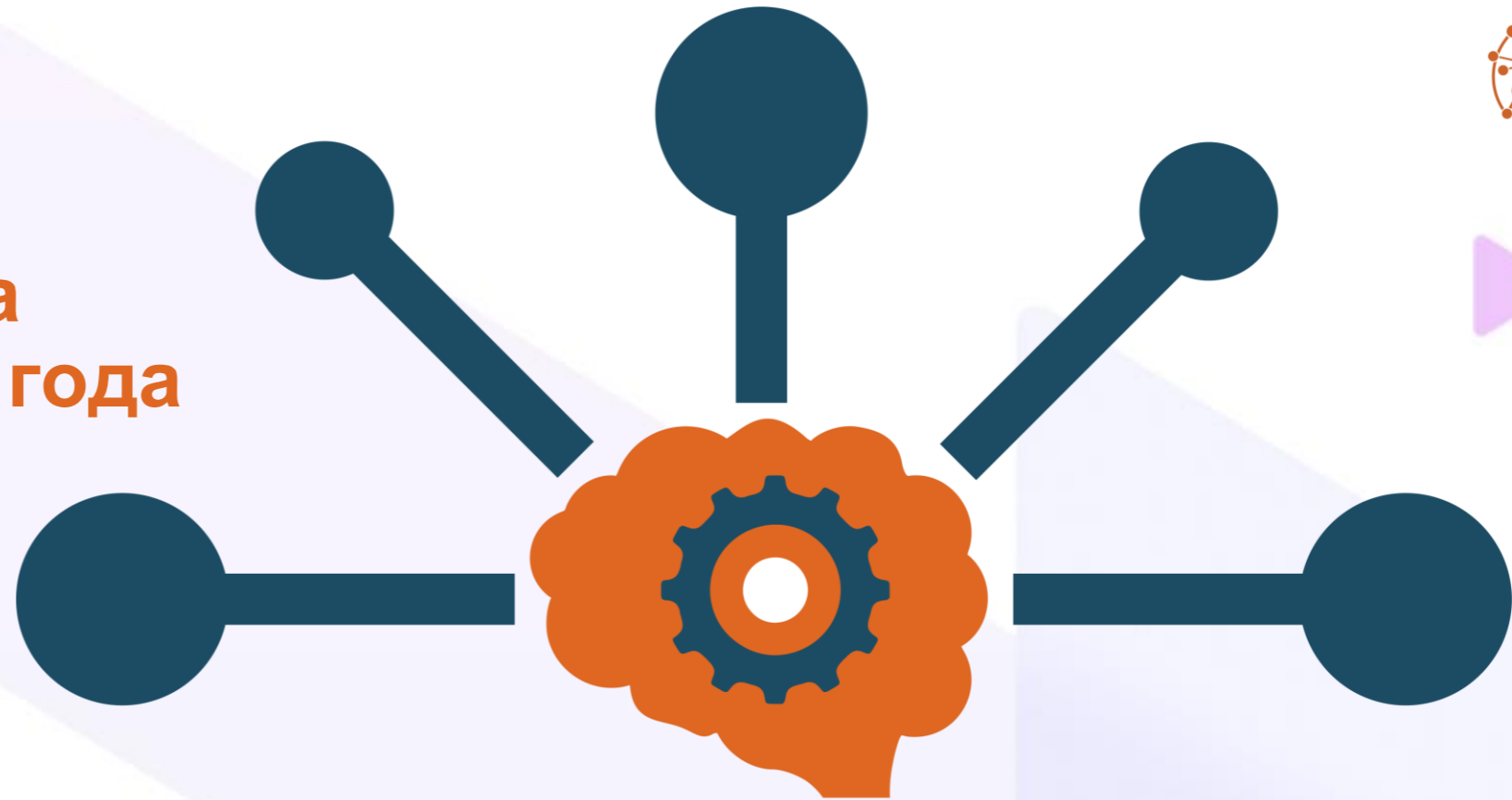




Доклад для сообщества  
AGIRussia, 6 июля 2023 года



# Семантический анализатор Real AI

Построение семантической модели текста  
на базе свойство-ориентированного подхода

Александр Гурьев  
([admin@real-ai.ru](mailto:admin@real-ai.ru))



## О себе



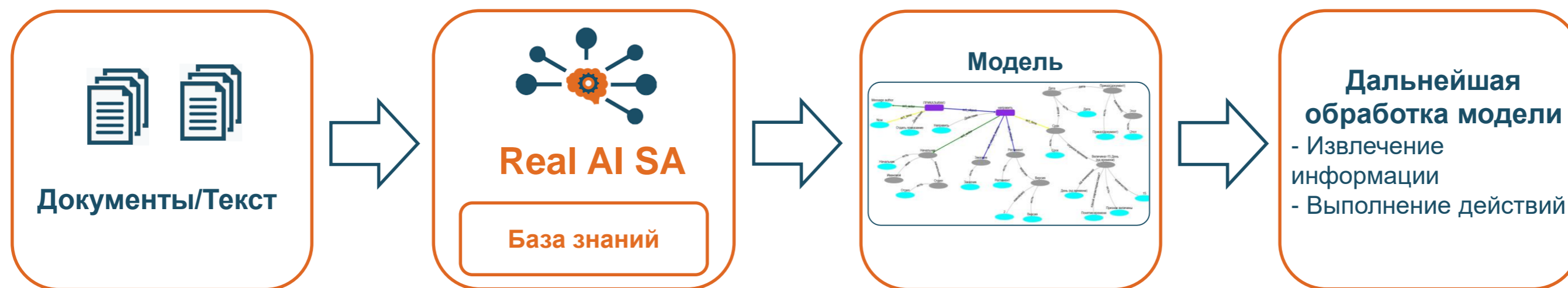
**Гурьев Александр** – автор концепции свойство-ориентированного подхода. Сооснователь и научный руководитель ООО «Риал AI Технологии», где мы развиваем технологии семантического анализа текста.

Программная платформа для извлечения поручений из распорядительных документов – создается при поддержке Фонда содействия инновациям.

Ядро системы - Real AI SA, Семантический анализатор.



# Real AI SA – Семантический анализатор



**Основная задача** – анализ текста на русском языке и построение семантической сети, отражающей смысл текста.

**Основные сложности:**

- Определение набора понятий для узлов и ребер
- Создание универсального алгоритма отображения текстовых элементов в эти понятия





## Парадигма активности

Все рассматривается как результат каких-то действий и, в свою очередь, может быть параметром в других действиях.

## Центральное понятие - действие

Техническое следствие



Методологическое следствие



Единый структурный элемент

Именованная объективированная  
активность

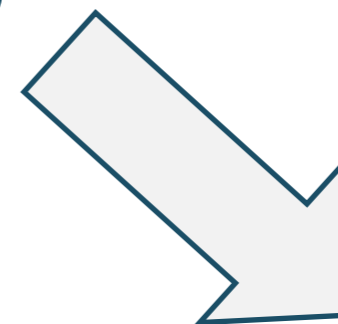
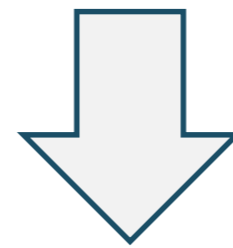
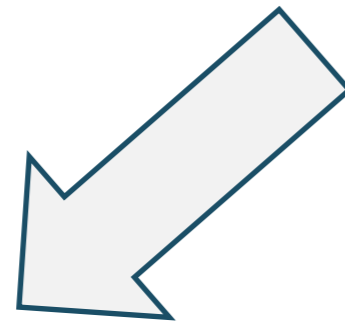
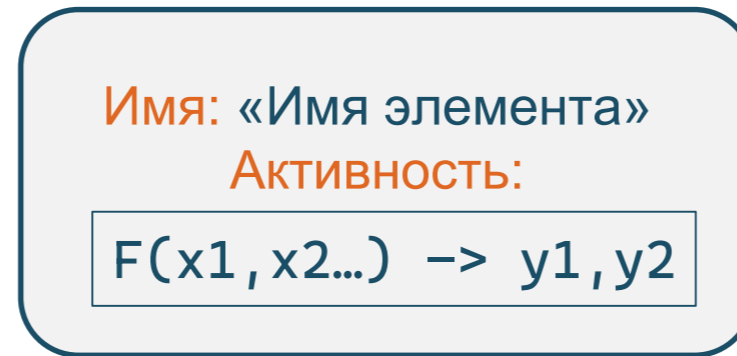
Рассмотрение всех понятий с точки зрения вида активности которую они представляют

Отсутствие активности (объект)  
Влияние на внешнюю среду (действие)  
Только внутреннее влияние (характеристика)

*Подробнее об этом - семинаре AGI Russia 28.04.2022.*



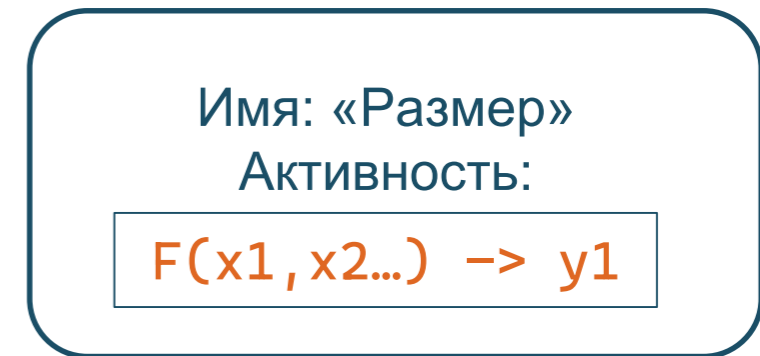
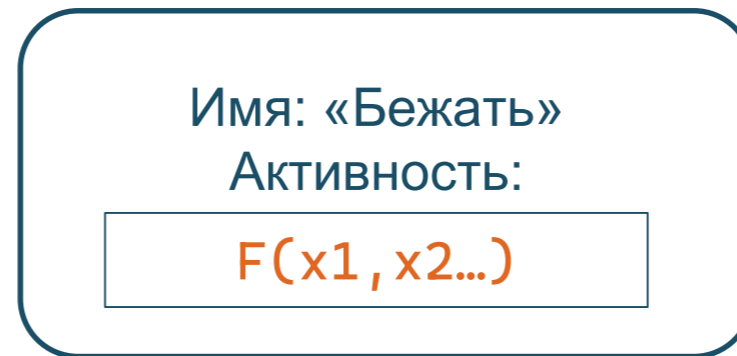
# Структура базового элемента



Объекты

Действия  
(в общепринятом смысле)

Характеристики



# Хранение знаний



Имя: «**Стол1**»  
Активность:  
*нет*

Имя: «Есть(являться)»  
Активность:  
 $F(x1) \rightarrow y1$   
*Стол1 -> Стол*

Имя: «Размер»  
Активность:  
 $F(x1) \rightarrow y1$   
*Стол1 -> Большой*

Имя: «Стол»  
Активность:  
*нет*

Имя: «Большой»  
Активность:  
*нет*





# Передача знаний. Свойство



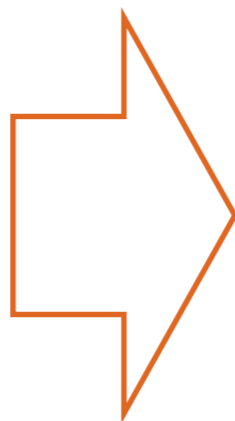
Вопрос:

«Что такое Стол1?»

Имя: «Стол1»

Активность:

*нет*



Ответ:

Для Стол1:

1. элемент(характеристика) «Есть(является)»  
возвращает элемент «Стол», и

2. элемент(характеристика) «Размер»  
возвращает элемент «Большой»

**Свойство** – описание значения характеристики  
для некоторого объекта(аргумента)

Имя: «Свойство X»

Активность:

*нет*

Аргумент свойства

Характеристика свойства

Значение свойства

**Текст** – передача знаний в виде свойств



# Свойство-ориентированный подход 1/2



- При построении семантической сети **НЕ** используется **объектно-ориентированный** подход (классы, наследование и т.д.).
- Узлы и ребра графа, образующего семантическую сеть, представляются **единым базовым элементом**.
- Базовый элемент – **это абсолютно абстрактное понятие**, содержащее функцию, которая может иметь в качестве аргументов и результата ссылки на другие элементы.





## Свойство-ориентированный подход 2/2



- Индивидуализация элемента происходит за счет участия как аргумент в функциях других элементов.
- Это участие описывается отдельным элементом отражающим некоторое свойство.
- Свойство подразумевает комбинацию трех обязательных элементов – действие, его аргумент и результат.
- Идея напоминает концепцию ECS (Entity, Component, System), но является, на наш взгляд, более универсальной и лаконичной.



# Семантический анализатор – реализация

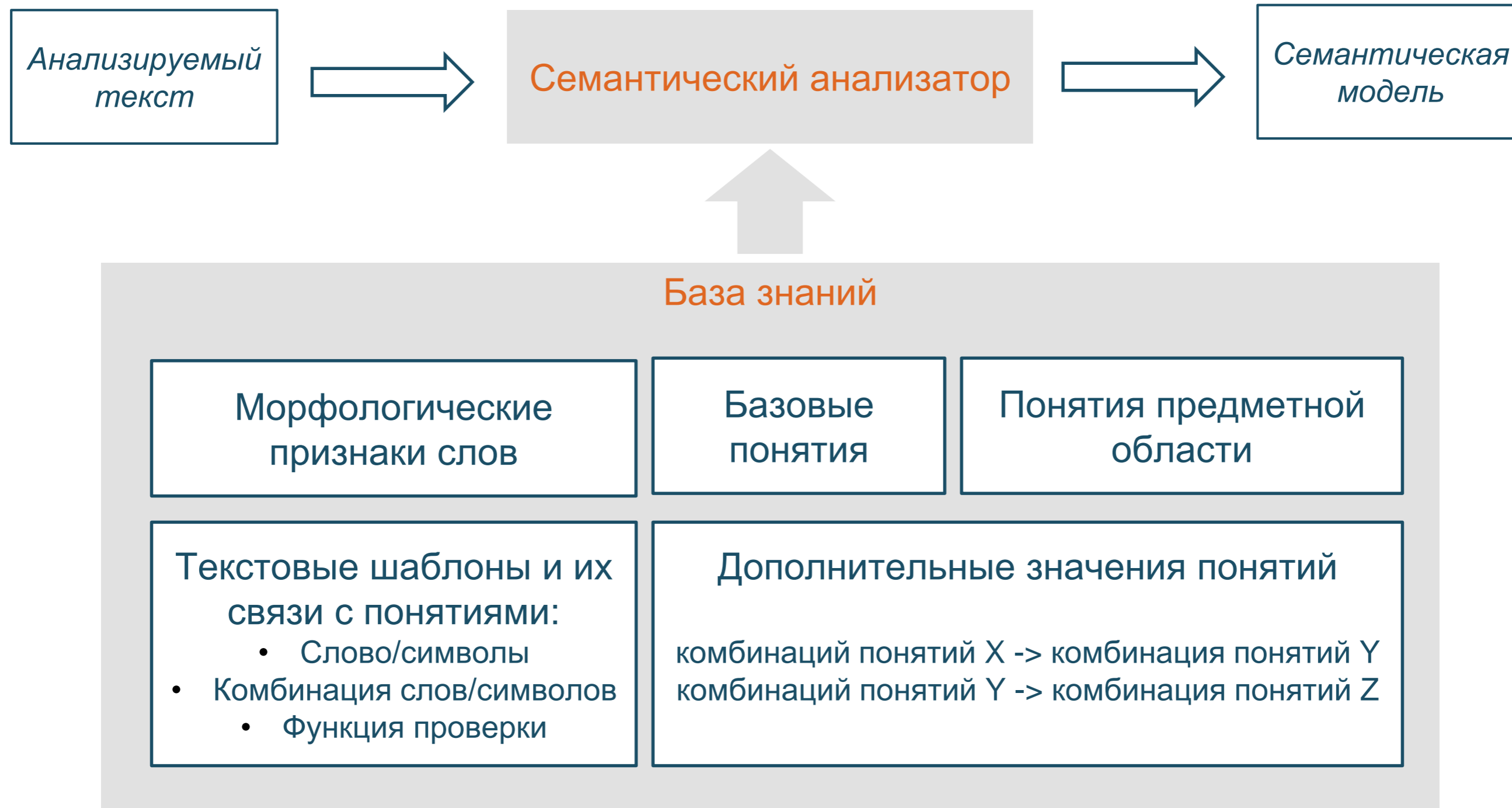


На данный момент реализованы или реализуются следующие проекты:

- **Выявление поручений** – продукт, в данный момент реализуется при поддержке Фонда содействия инновациям.
- Онлайн-демо **Овощи-Фрукты** – построение модели текста в ограниченной предметной области
- Онлайн-демо – **Капча** основанная на распознавании смысла текста.



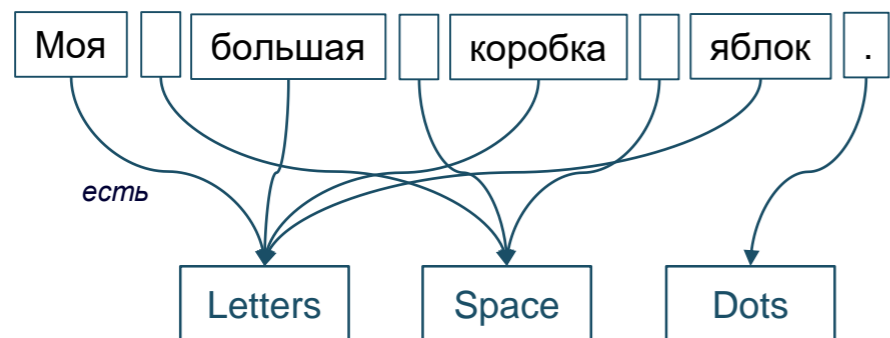
# Общая схема



# Этапы обработки

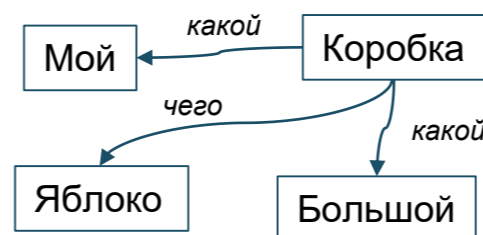


## Последовательность однотипных текстовых элементов (буквы, цифры, пробелов и т.д.)

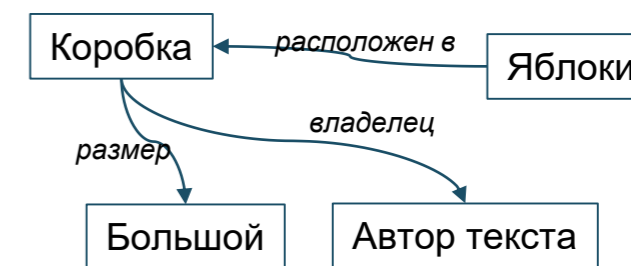


## Семантическая модель 1-го уровня

Понятия связанные грамматическими отношениями (какой, чего, на чем, в чем и т.д.)



## Выходная семантическая модель



Первичная классификация символов текста

Построение семантической модели 1-го уровня, разрешение омонимии

- Переход от грамматических отношений к смысловым
- Конвертация одних фрагментов графа в другие
  - Разрешение ссылок и гэпов

## База знаний

Морфологические признаки

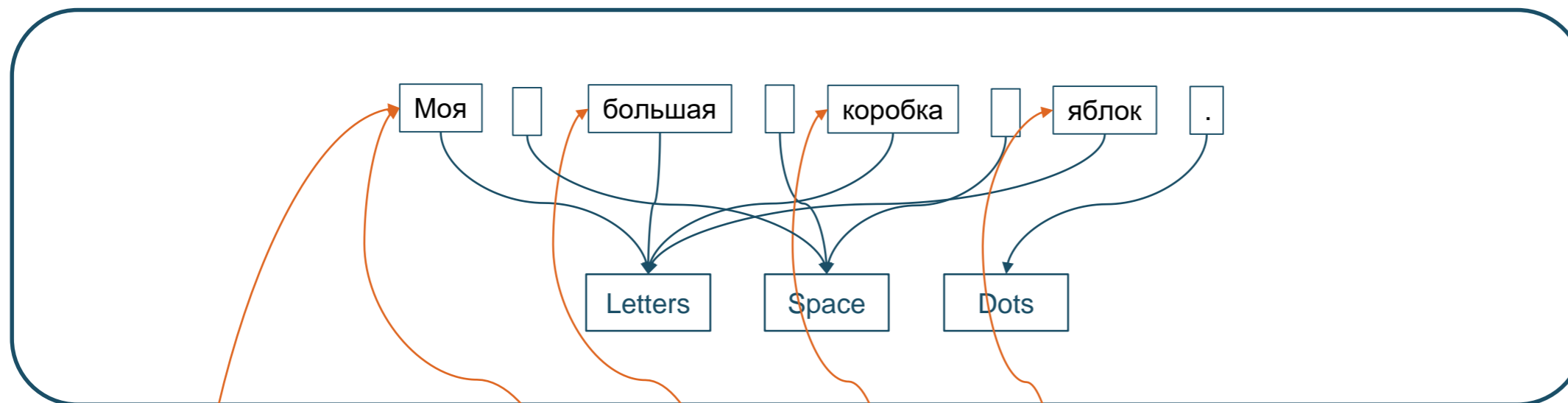
Текстовые шаблоны понятий

Базовые понятия и понятия предметной области

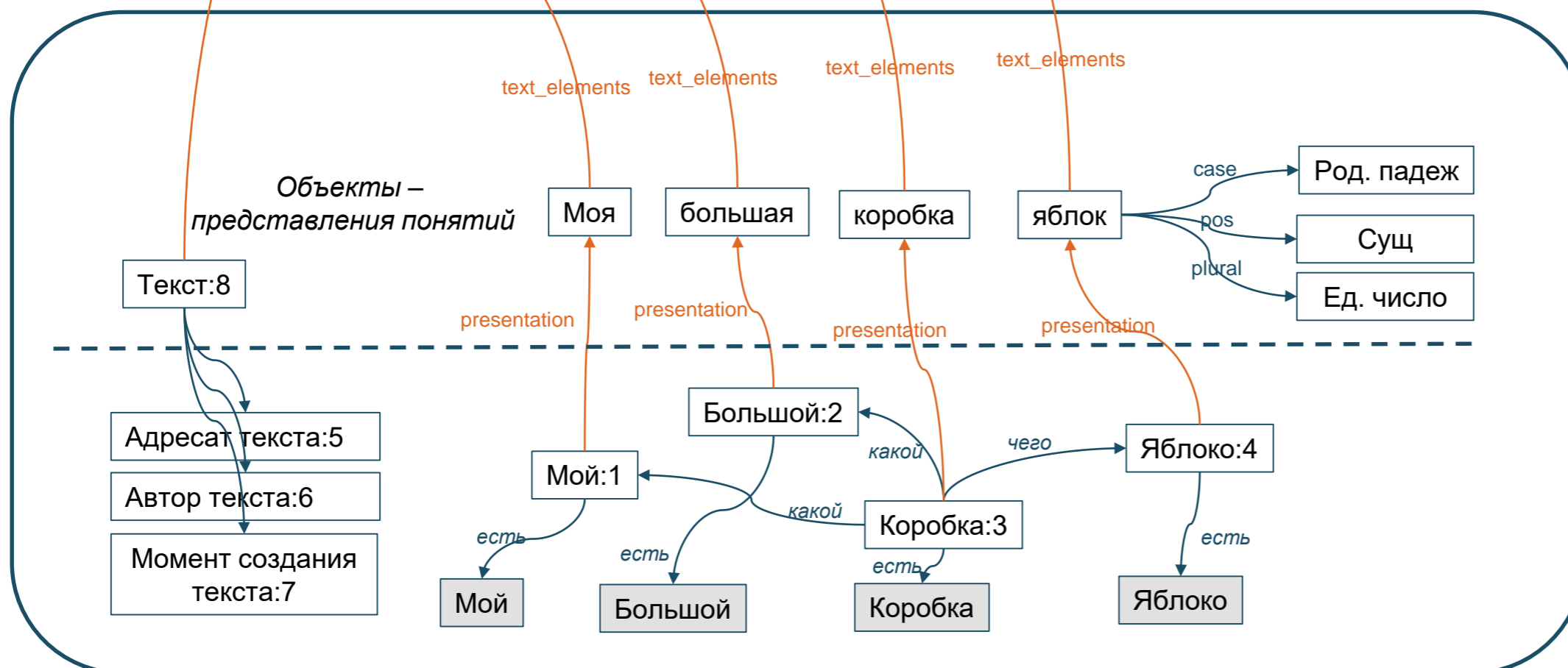
Соответствие одних понятий другим



# Связи между понятиями разного уровня



Последовательность текстовых элементов



Семантическая модель 1-го уровня.





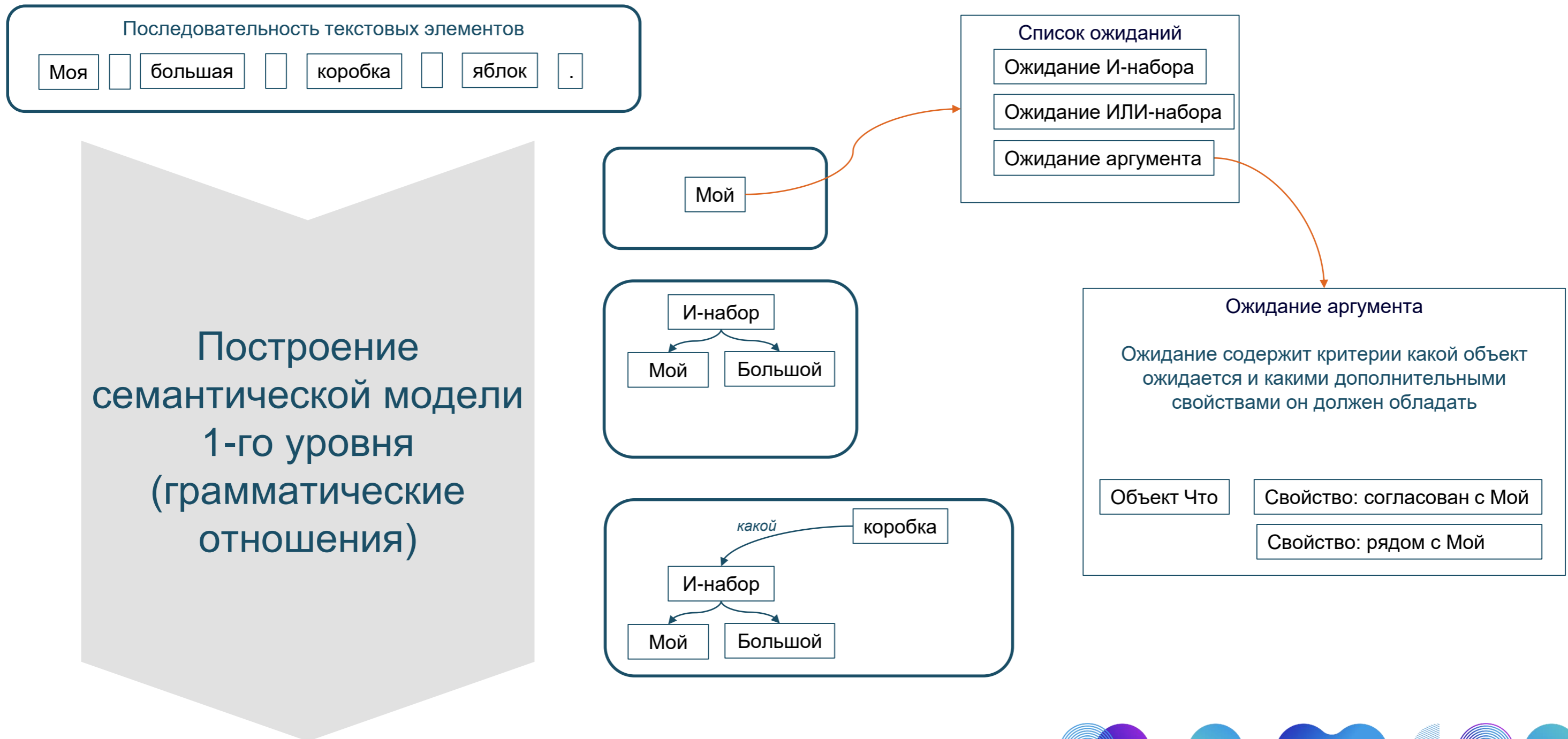
# Алгоритм работы. Концепция ожиданий



Построение семантической сети происходит с использованием концепции ожиданий («поиск в будущем»).

Прочитанные и распознанные текстовые элементы создают список «ожиданий», которые содержат некоторые критерии, на которые проверяются вновь появившиеся понятия.

Сбывшиеся ожидания приводят к созданию соответствующего объекта и встраивание его в модель с соответствующими отношениями





# Алгоритм работы. Перечень ожиданий



- И - Набор
- ИЛИ - Набор
- НЕ – Набор
- Какой(аргумент)
- Какой(результат)
- Чего
- Действие: актер
- Действие: объект
- Действие: Место(где)
- Действие: Место (куда)
- Действие: Время
- Действие: Результат



# Решение задач семантического анализа



Концепция ожиданий и имеющаяся взаимосвязанная модель понятий всех уровней – от текстовых элементов до смысловых понятий, позволяет решать сложные проблемы анализа текста, такие как:

- **Разрешение ссылок** (местоименные и через объект со свойствами).
- **Гэпы** (отсутствие некоторых смысловых частей).
- **Выбор значения для Омонимов.**



# Разрешение ссылок

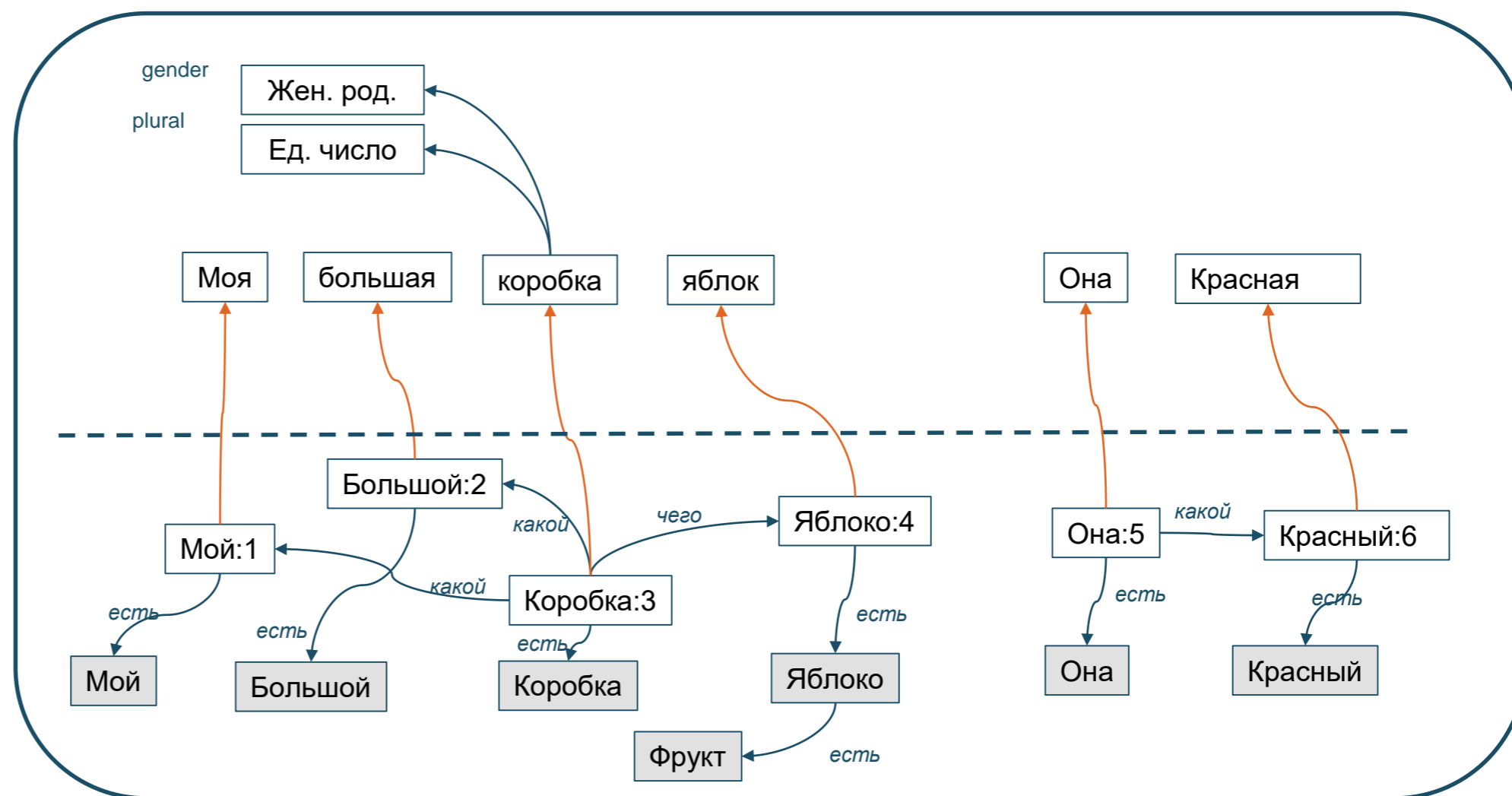


Поиск среди объектов модели объекта ссылки с учетом:

1. Расположения в тексте
2. Морфологических признаков (жен. род., ед. число)
3. Смысловых признаков (Яблоки есть фрукты)

- Моя большая коробка яблок.  
Она красная

- Моя большая коробка яблок.  
Фрукты красные

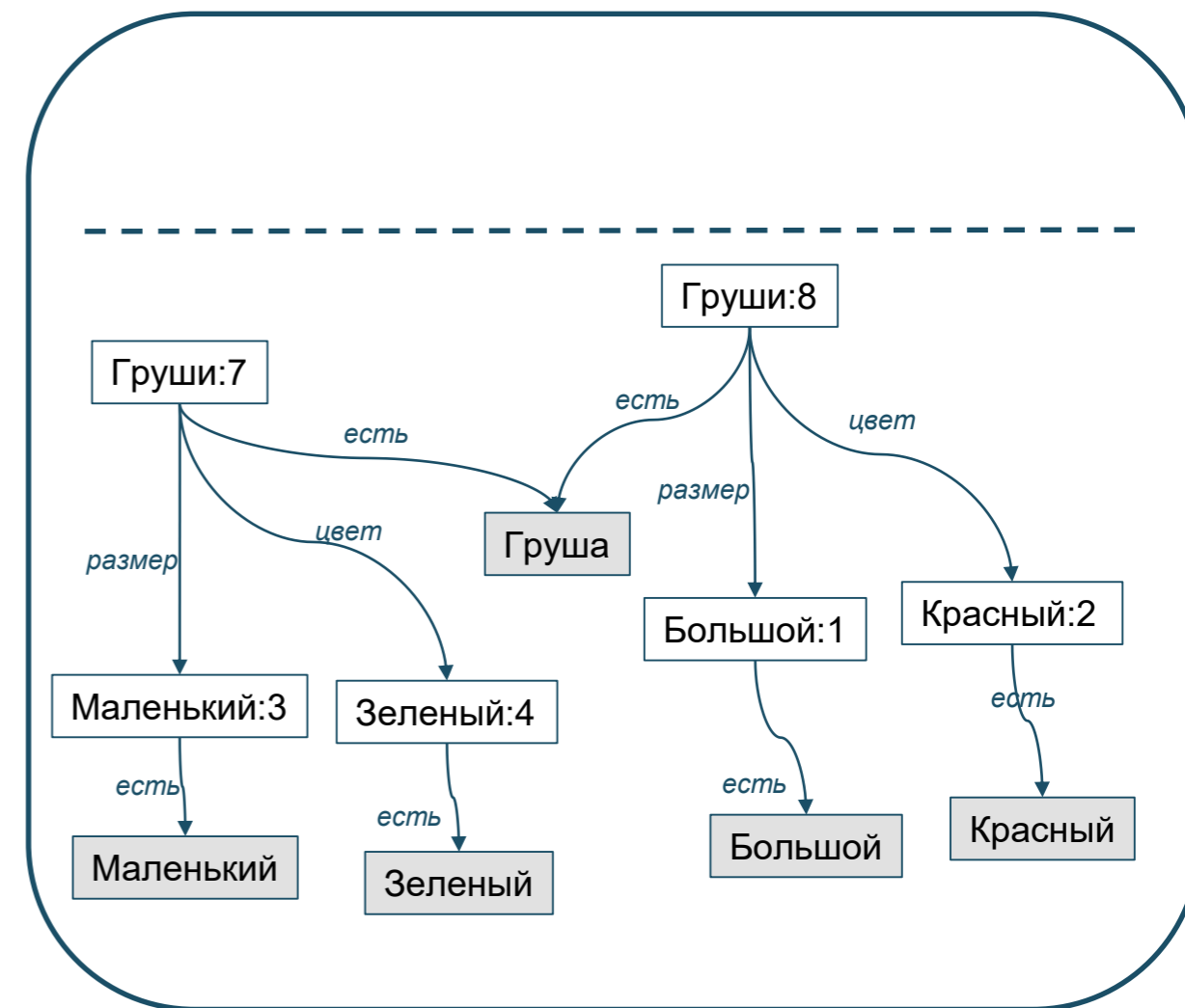
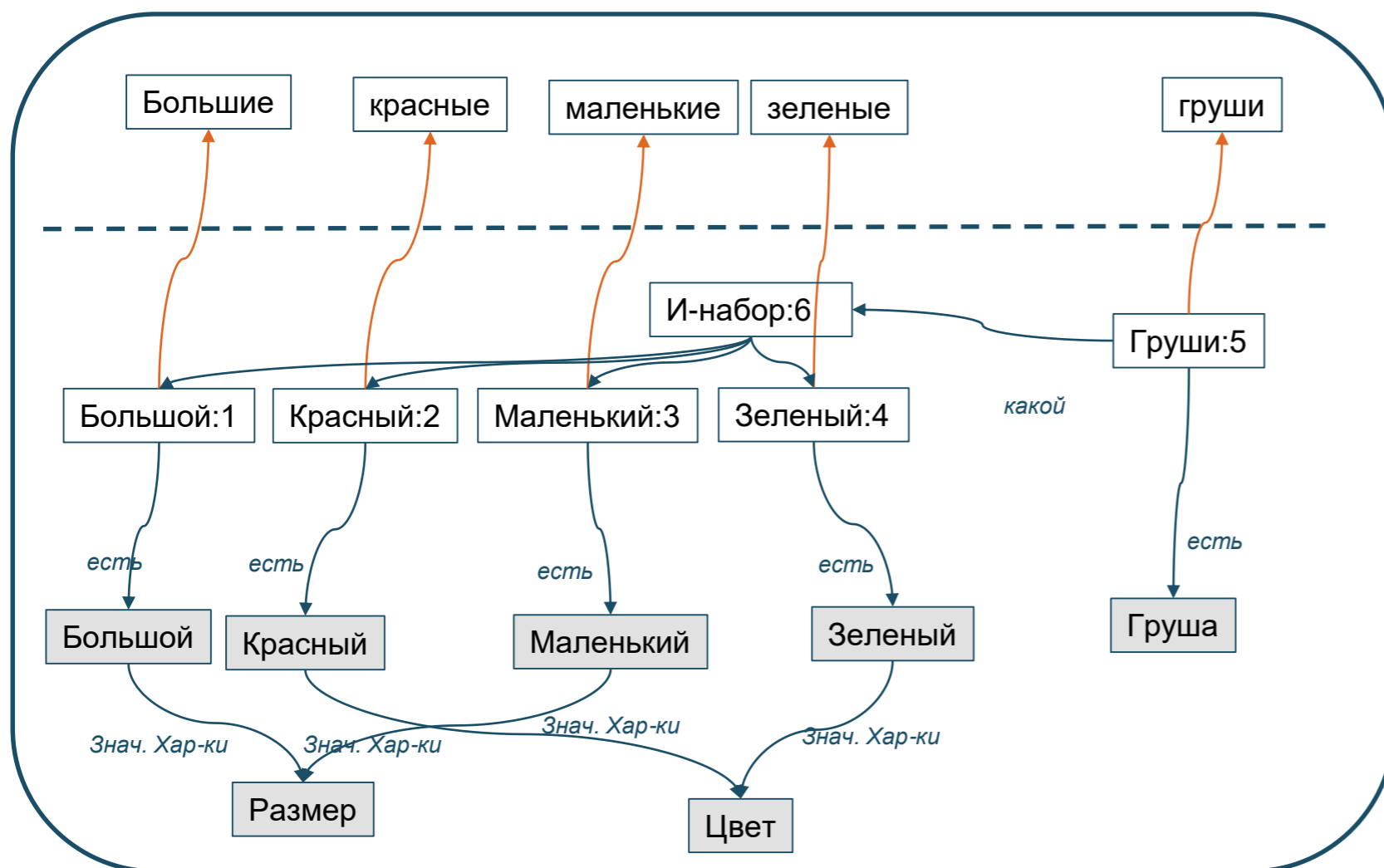


# Гэпы (пропуски понятий)



В случае присвоения перечня значений характеристик происходит проверка есть ли уже у объекта значение данной характеристики (в данном случае, размер или цвет) и если есть, то объект раздваивается и дальнейшее присвоение идет новому объекту.

- большие красные и маленькие зеленые груши



# Определение значения омонимов



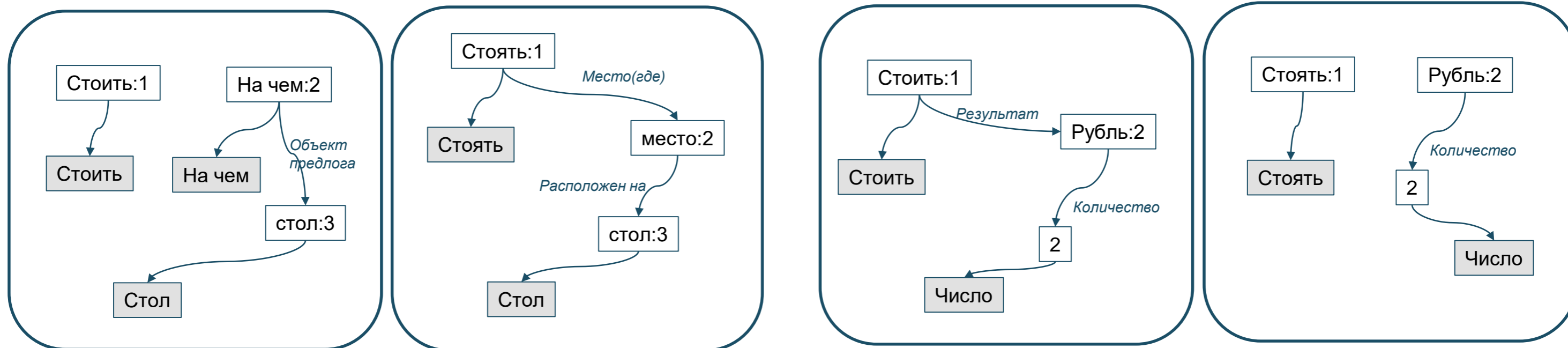
В случае если текстовый элемент имеет несколько разных значений, происходит выбор по следующим критериям:

1. Если одно из значений присутствует среди объектов Ожиданий, то выбирается оно. Например «**утвердить рекомендации**», «Рекомендации» – два разных понятия - документ и действие. Действие Утвердить порождает ожидание объекта действия – документа. Соответственно выбирается понятие, являющееся документом.

2. Происходит предварительный семантический анализ идущего следом фрагмента текста для каждого из значений. И выбирается значение, которое приводит к модели с меньшим числом вершин в графе.

Коробка яблок **стоит на столе**.

Она **стоит два рубля**





# Моделирование понятий предметной области



Необходимо ответить на следующие вопросы относительно понятия:

- Является ли это понятие **Объектом**, **Характеристикой** или **Действием**?



- Для характеристик – есть ли и какие **ограничения на область определения и область значения**?
- Для объекта – является ли он **значением какой-либо характеристики**?
- Для действий – какие **семантические и морфологические требования к параметрам** (Актор, Объект, Место, Время и т.д.)?
- Имеется ли, и нужно ли моделировать для данной предметной области, какое-то **дополнительное значение** связанное с этим понятием (например «серая зарплата», «мой», «сегодня» и т.д.)? Если да, то описать это.







## Служебные и базовые понятия

- Типы текстовых элементов (Буквы, Цифры и т.д)
- Виды ожиданий (И-набор, Или-набор, Не-набор, Аргумента, Значения, Актора действия, Объекта действия и.т.д)
- Списки разного рода
- Признаки характеристики
- Объект, Действие
- Есть
- Какой
- Чего
- Список свойств
- Свойство: аргумент
- Свойство: действие
- Свойство: результат
- и др.

## Понятия предметной области и их текстовые шаблоны

Объект

```
OBJ_POTATO = object_word("картофель")  
tmpl = template("картошка")  
add_template(OBJ_POTATO, tmpl)
```

Характеристика в двух проявлениях – как функция и как объект

```
COLOR = feature("цвет")  
OBJ_COLOR = object_word("Цвет")  
FEATURES_FOR_OBJ(OBJ_COLOR) = COLOR
```

Объект - значение характеристики

```
OBJ_RED = object_word("красный")  
FEATURES_FOR_VALUE(OBJ_RED) = COLOR
```

Характеристика с заданными областями для аргументов и значений

```
BUYER = feature("покупатель")  
FEATURE_ARG_OBJS(BUYER) = OBJ_PRODUCT  
FEATURE_RES_OBJS(BUYER) = OBJ_PERSON  
OBJ_BUYER = object_word("Покупатель")  
FEATURES_FOR_OBJ(OBJ_BUYER) = BUYER
```





## Действия

Действие LAY  
(шаблоны «лежать», «лежание»)

Семантические требования к Актору

Морф. требования к Актору, когда  
выражено глаголом

Морф. требования к Актору, когда  
выражено существительным

Семантические требования к Месту действия (любой объект)

Морф. требования к Месту действия (объекты с  
предлогами в соотв. падеже)

```
LAY = action_word("лежать")
tpl = template('лежание')
add_template(LAY , tpl)

// Лежать могут продукты и тара
ACT_ACTOR_CNCPTS(LAY) = [OBJ_PRODUCT,OBJ_PACKAGING]

// яблоки ("кто") лежат
ACT_ACTOR_V_OBJS(LAY) = OBJ_CHTO_NM_ACTOR

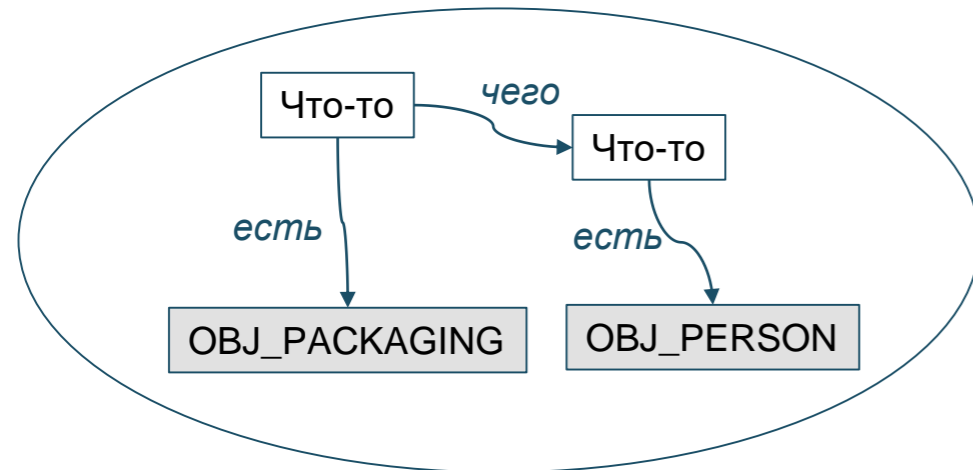
// лежание "чего"
ACT_ACTOR_N_OBJS(LAY) = OBJ_CHEGO

// Место действия (где)
ACT_LOCATION_CNCPTS(LAY) = [ОБЪЕКТ]
ACT_LOCATION_OBJS(LAY) = [OBJ_V_CHEM, OBJ_NA_CHEM,
OBJ_NAD_CHEM, OBJ_POD_CHEM]
```

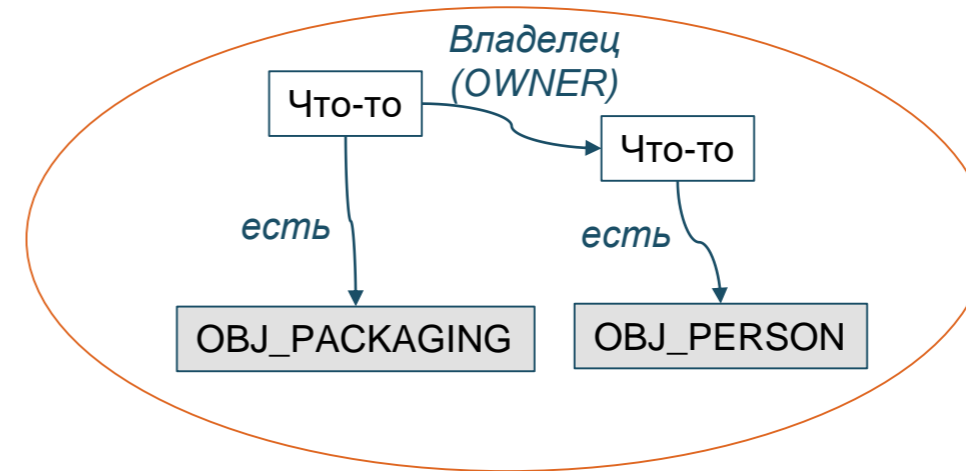




Тара – Чего - Человек



Тара – Владелец - Человек



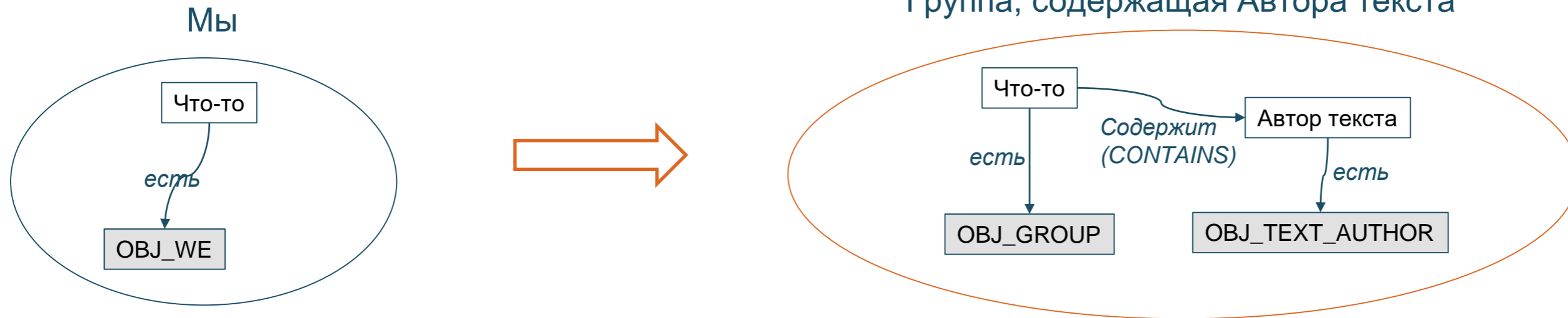
## Дополнительные значения отдельных понятий или фрагментов сети

// Описание значения понятия "Чего-Человек" для тары

```
MEANING_CHEGO_PERSON_PACK = object("Значение понятия 'чего-Человек' для тары")
meaning_chego_person_pack_prop1 = property("", CHEGO, OBJ_PERSON)
MEANING_ARG_PROPS(MEANING_CHEGO_PERSON_PACK) = meaning_chego_person_pack_prop1
meaning_chego_person_pack_prop2 = property("", IS, OBJ_PACKAGING)
MEANING_ARG_PROPS(MEANING_CHEGO_PERSON_PACK) = meaning_chego_person_pack_prop2
```

```
meaning_chego_person_pack_res = object("")
IS(meaning_chego_person_pack_res) = OBJ_MEANING_ARG
meaning_chego_person_pack_res_prop = property("", OWNER, OBJ_MEANING_ARG_PROP1_RES)
OBJ_PROPERTIES(meaning_chego_person_pack_res) = meaning_chego_person_pack_res_prop
MEANING_RES(MEANING_CHEGO_PERSON_PACK) = meaning_chego_person_pack_res
```





## Дополнительные значения отдельных понятий или фрагментов сети

// Описание значения понятия "мы"

```
MEANING_WE = object("Значение понятия 'мы'")
meaning_we_prop1 = property("", IS, OBJ_WE)
MEANING_ARG_PROPS(MEANING_WE) = meaning_we_prop1
```

```
meaning_we_res = object("")
IS(meaning_we_res) = OBJ_GROUP
meaning_we_res_prop = property("", CONTAINS, OBJ_TEXT_AUTHOR, OBJ_MEANING_RES)
OBJ_PROPERTIES(meaning_we_res) = meaning_we_res_prop
MEANING_RES(MEANING_WE) = meaning_we_res
```



# Примеры разбираемых фраз



*«Одна большая упаковка бананов стоит 2000 рублей.  
Она стоит под желтой полкой в овощном отделе»*

*«На моем столе стоит зеленая коробка. В ней лежит  
спелая красная клубника»*

*«Владелец манго Иванов, владельцев бананов много»*





# Ограничения на текущий момент



1. Все понятия должны быть описаны.
2. Не все варианты фраз корректно разбираются.
3. Есть неформализуемые без дополнительных знаний о предметных областях моменты в выборе вариантов построения модели.

*"Сладкие и свежие яблоки на **рынке в зеленых упаковках**"*

*"порядка и списка экспертов"*





# Перспективные направления развития



1. Реализация формирования модели 1-го уровня нейросетевыми средствами.
2. Автоматизированное формирование базы знаний на основе текстов требуемой предметной области с помощью технологий машинного обучения.
3. Синтез текстового представления на основе моделей.





# Семантический анализатор Real AI

Он-лайн демо: <https://real-ai.ru/demo/>

## Статьи на Хабр:

- *Смысл текста или представление знаний в системе, основанной на действиях* (<https://habr.com/ru/articles/554738/>)
- *Примеры представления смысла документа через действия* (<https://habr.com/ru/articles/575838/>)
- *Real AI семантический анализатор. Онлайн-демо – овощи, фрукты и их свойства* (<https://habr.com/ru/articles/712140/>)
- *Смысловая капча. Применение семантического анализатора Real AI для реализации капчи* (<https://habr.com/ru/companies/realai/articles/741486/>)

ООО «Риал АИ Технологии»  
<https://real-ai.ru/>