

Презентация, 52 слайда

Вопросы систематизации терминологии

проф. С.Ю.Соловьев

МГУ им.М.В.Ломоносова, ф-т ВМК

кафедра алгоритмических языков

Терминология

Друзы - кристаллические включения в виде сростков кристаллов, накапливаемых в вакуолях растительных клеток.

Ценовая политика – совокупность мероприятий и стратегий, которые использует предприятие при установлении цены.

Как распространяется терминология?



Эпицентр пожара находился на втором этаже.

/из СМИ/



Эпицентр – место на земной поверхности, расположенное непосредственно над гипоцентром очага землетрясения.

Гипоцентр – центральная точка очага землетрясения.

Терминология

- теория: терминоведение
 - практика:
 - онтологии: ИИ, бизнес-аналитика
 - тезаурусы: ИПС
 - интернет-глоссарии: www
- и др.

Терминоведение

разделы

- теоретическое
- прикладное

+ терминография

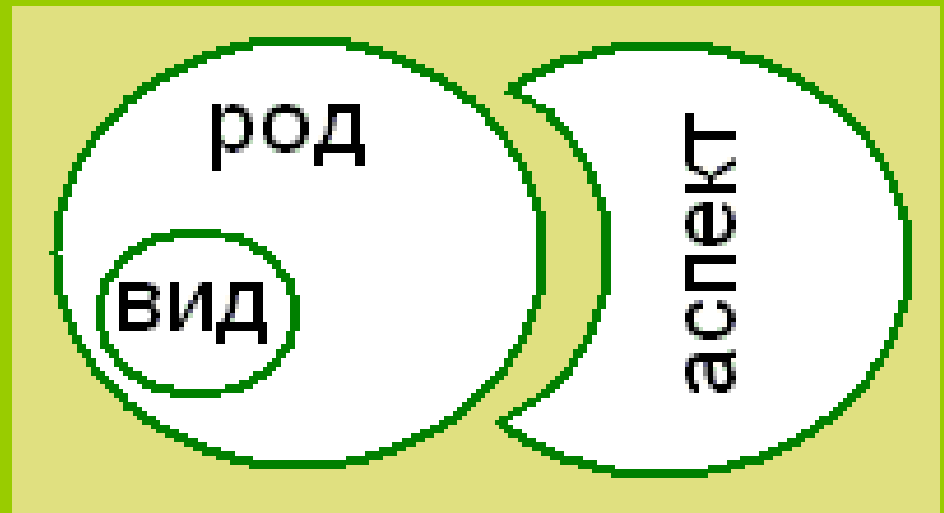
- общее | частное
- типологическое
- сопоставительное
- семасиологическое
- ономасиологическое
- историческое
- функциональное
- когнитивное

организации

- КНТ РАН
 - Комитет научной терминологии в области фундаментальных наук
 - бывш. КНТТ РАН
- ВНИИКИ Госстандарта
 - Всероссийский научно-исследовательский институт классификации, терминологии и информации по стандартизации и качеству

Типы терминов

- термины категорий: **материя, время, пространство**
- общенаучные и общетехнические термины: **метод**
- межотраслевые термины: **аналитический метод**
- специальные термины: **метод Зейделя**
 - родовые термины
 - видовые термины
 - аспектные термины
(аксессуарары)



Классификация терминов

по содержанию

- эмпирические термины: **зеленый**
- теоретические термины: **абсолютно твердое тело**

по областям

- научные термины
 - термины общественных наук
 - термины естественных наук
 - термины технических наук
- технические термины
- термины экономики
- термины политико-социальной сферы

Классификация терминов

по логической категории

- термины предметов: **млекопитающие**
- термины процессов: **делопроизводство**
- термины признаков, свойств: **теплоемкость**
- термины величин и их единиц: **сила тока, ампер**

по семантической структуре

- однозначные термины: **хромосома**
- многозначные термины: **суд**

по семантике

- свободные словосочетания: **справка с работы**
- устойчивые словосочетания: **всемирное тяготение**

Классификация терминов

по формальной структуре

- термины-слова [усечение]
- термины-словосочетания [сокращение]
 - существительное + прилагательное
 - существительное-существительное и др.
- термины специальной структуры: **i-балка**



по мотивированности

- полностью мотивированные термины: газопровод
- частично мотивированные: **болезнь Паркинсона**
- полностью немотивированные термины: ромб
- ложномотивированные термины: **громоотвод**

Классификация терминов

относительно языка-источника

- исконные термины: датчик
- заимствованные термины: квершлаг
- гибридные термины: антиобледенение

относительно периода

- термины-архаизмы
- термины-неологизмы



по частоте

- высокочастотные термины
- низкочастотные термины

Классификация терминов

по частям речи

- термины-существительные
- термины-прилагательные
- термины-глаголы
- термины-наречия: **пиано**

по авторству

- коллективные термины
- индивидуальные термины:
 - вертолет - Леонардо да Винчи
 - промышленность - Н.М.Карамзин
 - социология - О.Конт



Классификация терминов

по сфере использования

- универсальные термины для многих областей
- уникальные термины для одной области
- концепциально-авторские термины



по роли в познании

- термины для фиксации знания
- термины-инструменты познания: **x-лучи**
- термины обучения

Классификация терминов

по нормативности

- стандартизируемые термины
- стандартизированные термины → ГОСТ и др.
 - международные термины
- недопустимые термины → ГОСТ
- рекомендуемые термины
- рекомендованные термины
- параллельно допустимые термины
- отклоненные термины



Источники терминологии

- авторская терминология
- признанная терминология:

--- стандарты:

ГОСТ Р 51303-99 Торговля. Термины и определения. (148 терминов)

--- законы

--- публикации

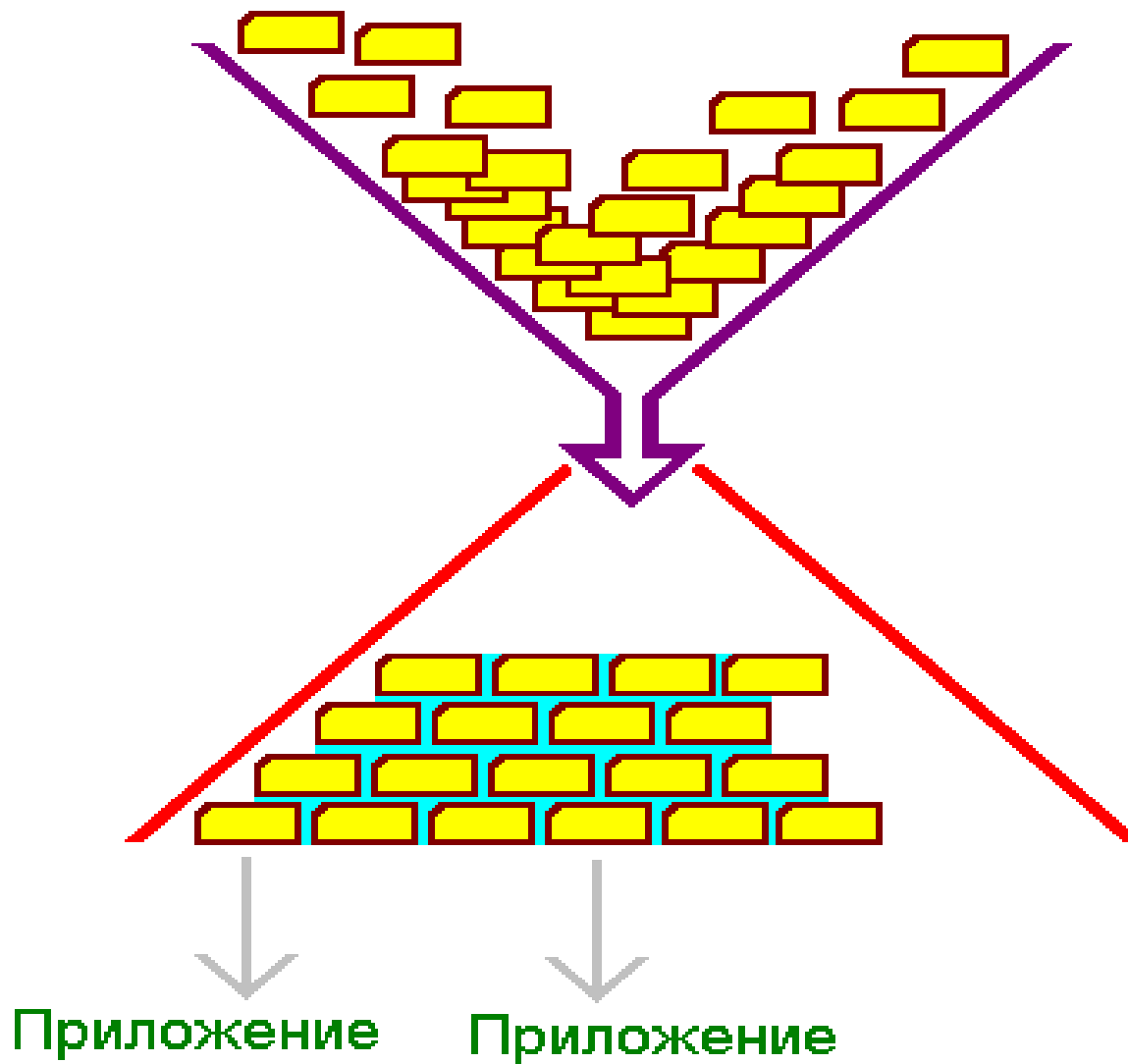


Формы существования терминологии

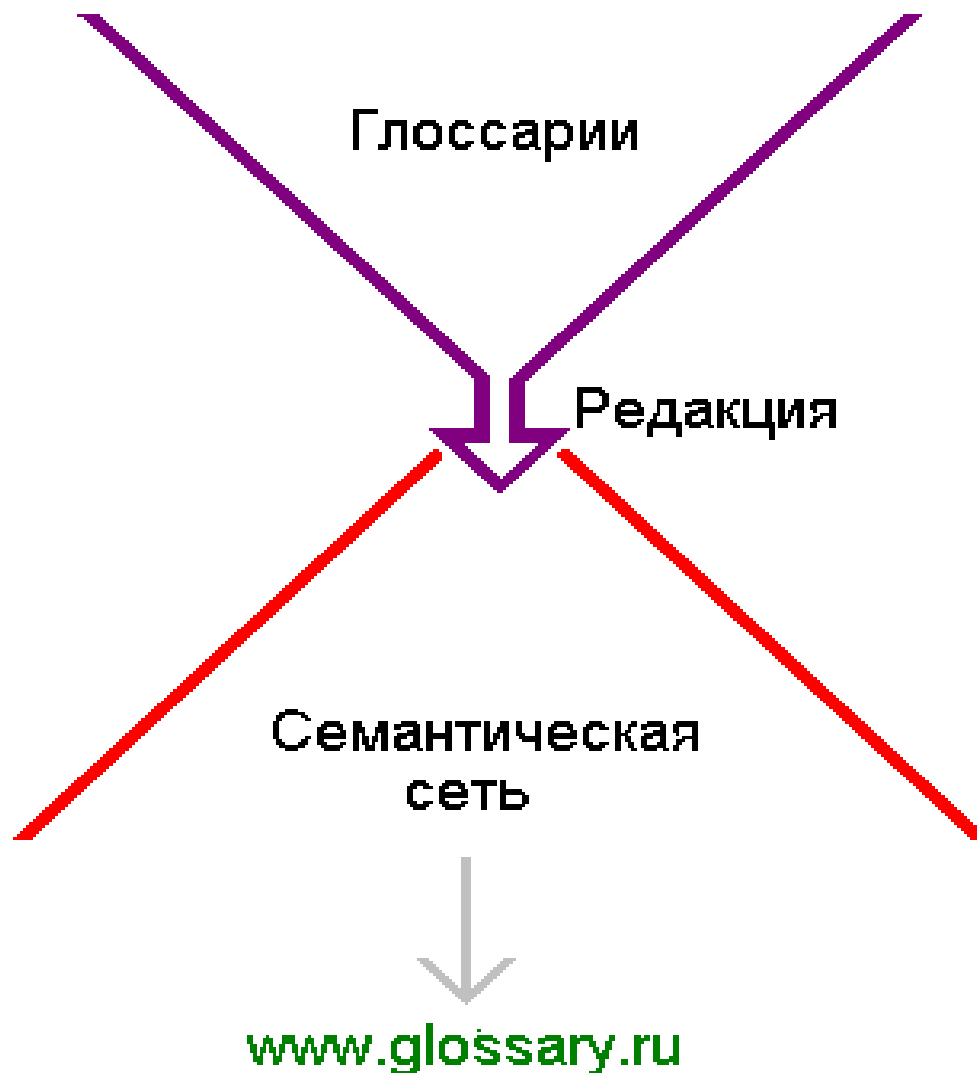
- словари: англо-русские ...
- ЭНЦИКЛОПЕДИИ (развернутые определения)
- Глоссарии

понятие, заданное последовательностью основных определений
--
- словари иностранных слов и пр.

Идея проекта



Идея проекта (уточнение)



Редакция

- **Ю**ридическое
лицо
- **Ш**тат
научных
редакторов
- **С**пециальное
программное
обеспечение



Типы терминологических статей

1. $\langle A \rangle$ есть $\langle B \rangle +$ особенности $\langle C \rangle$
- $\langle A \rangle$ есть $\langle B_1 \rangle | \langle B_2 \rangle | \dots | \langle B_n \rangle$
- $\langle A \rangle$ есть совокупность $\langle B_1 \rangle, \dots \langle B_n \rangle$

*. $\langle A \rangle$ есть

1 | 2 | 3

Различают (подклассы A)

Обычно (типичная ситуация)

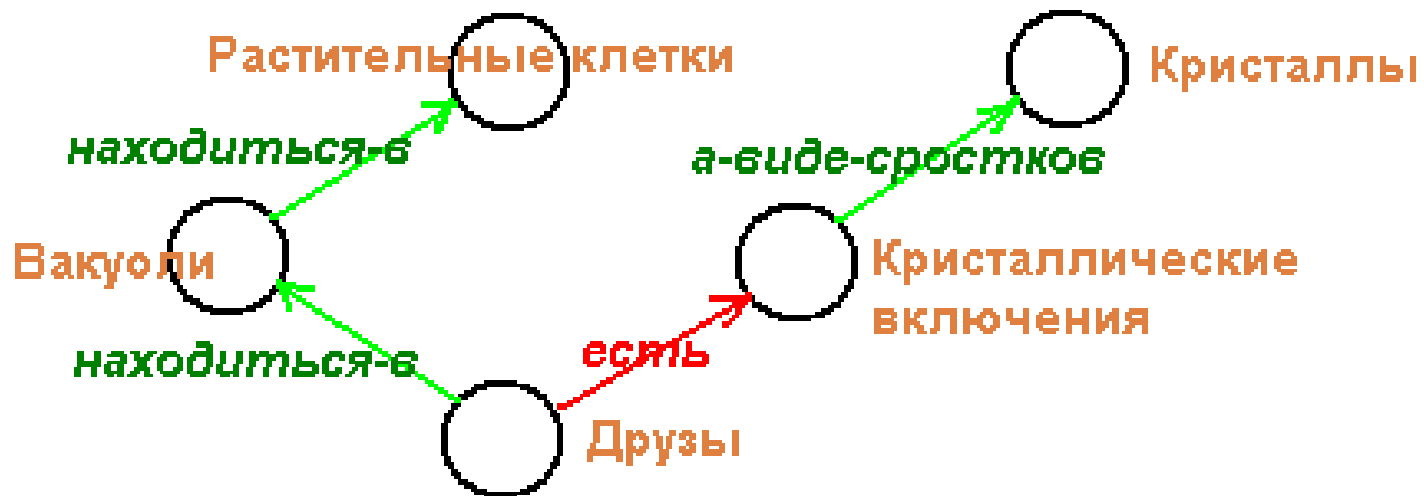
... для (прагматика)

$\langle A \rangle$ в им.падеже; ед. или мн. число 21

Статья (тип 1)

<А> есть <В> + особенности <С>

Друзы - кристаллические включения в виде сростков кристаллов, накапливаемых в вакуолях растительных клеток.

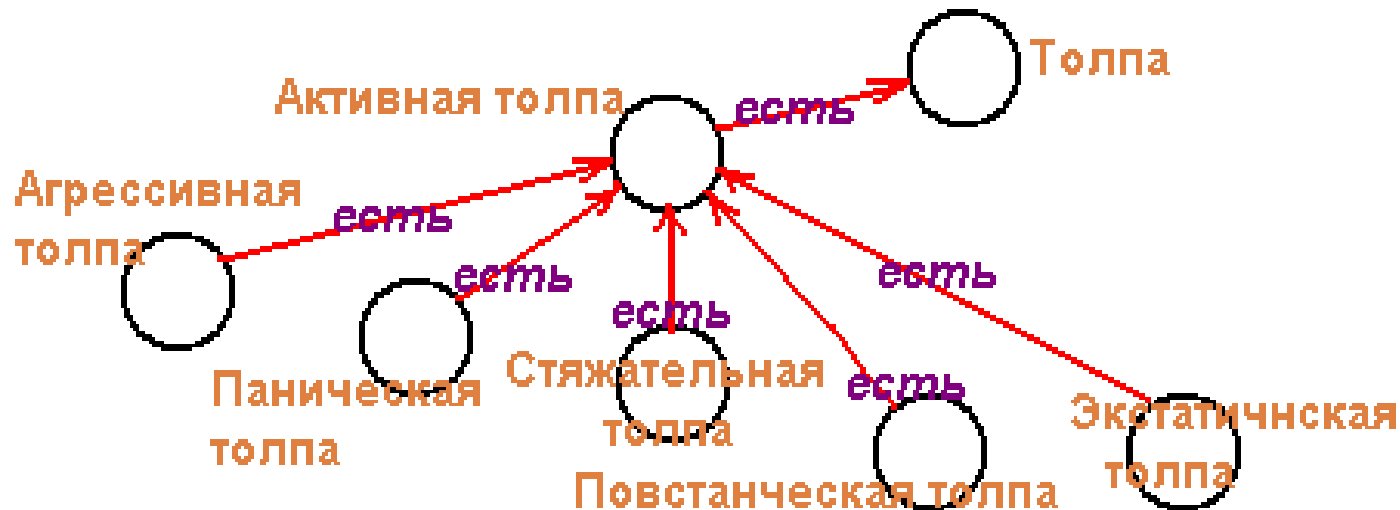


Статья (тип 2)

$\langle A \rangle$ есть $\langle B_1 \rangle$ | $\langle B_2 \rangle$ | ... | $\langle B_n \rangle$

Активная толпа – агрессивная, паническая, стяжательная, повстанческая или экстатическая толпа.

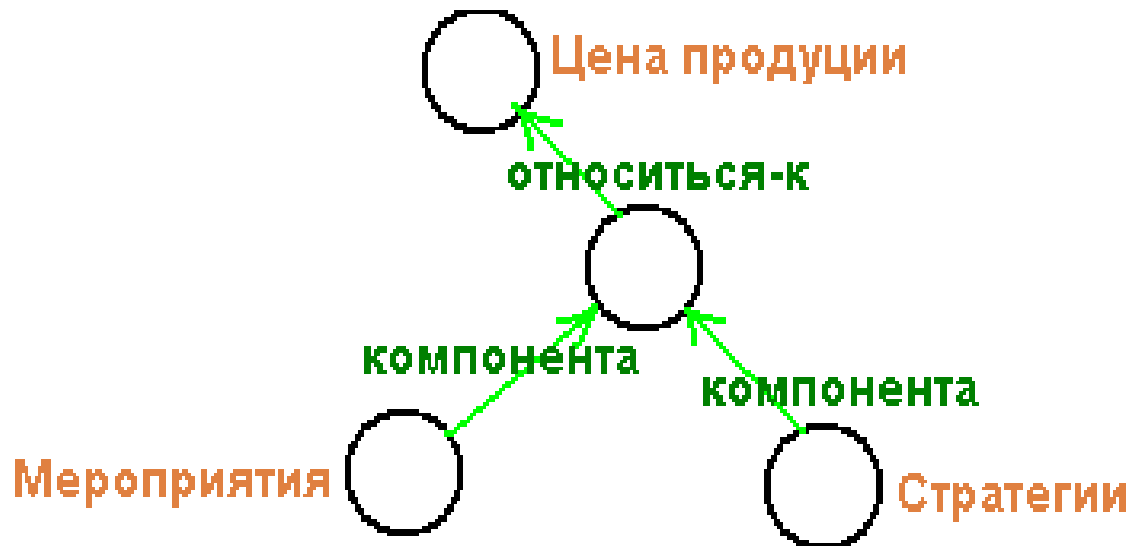
(n=5)



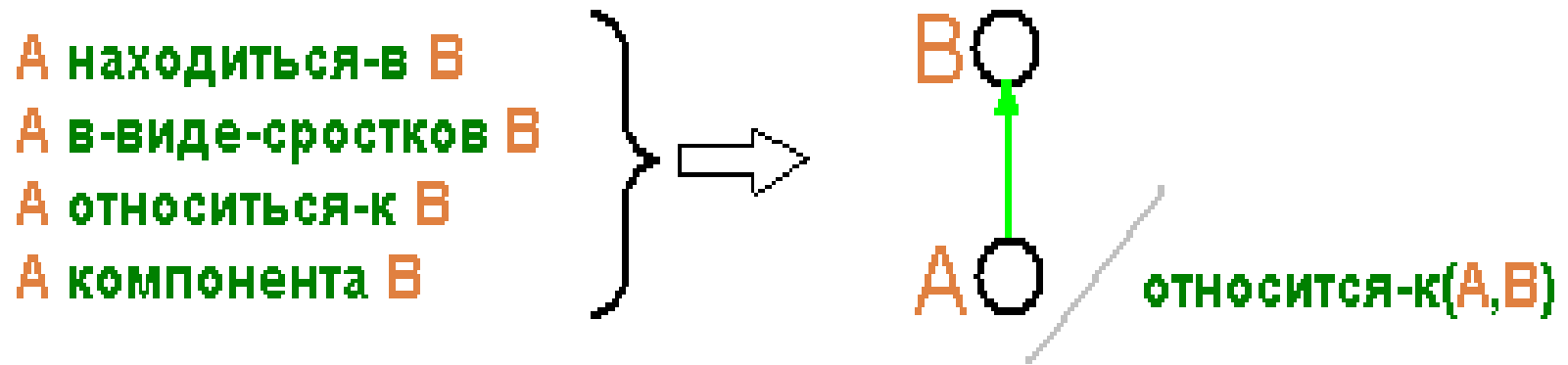
Статья (тип 3)

$\langle A \rangle$ есть совокупность $\langle B_1 \rangle, \langle B_2 \rangle \dots \langle B_n \rangle$

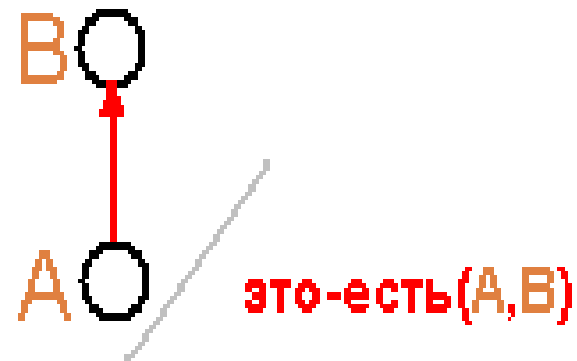
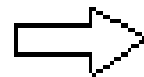
Ценовая политика – совокупность мероприятий и стратегий, которые использует предприятие при установлении цены. (n=2)



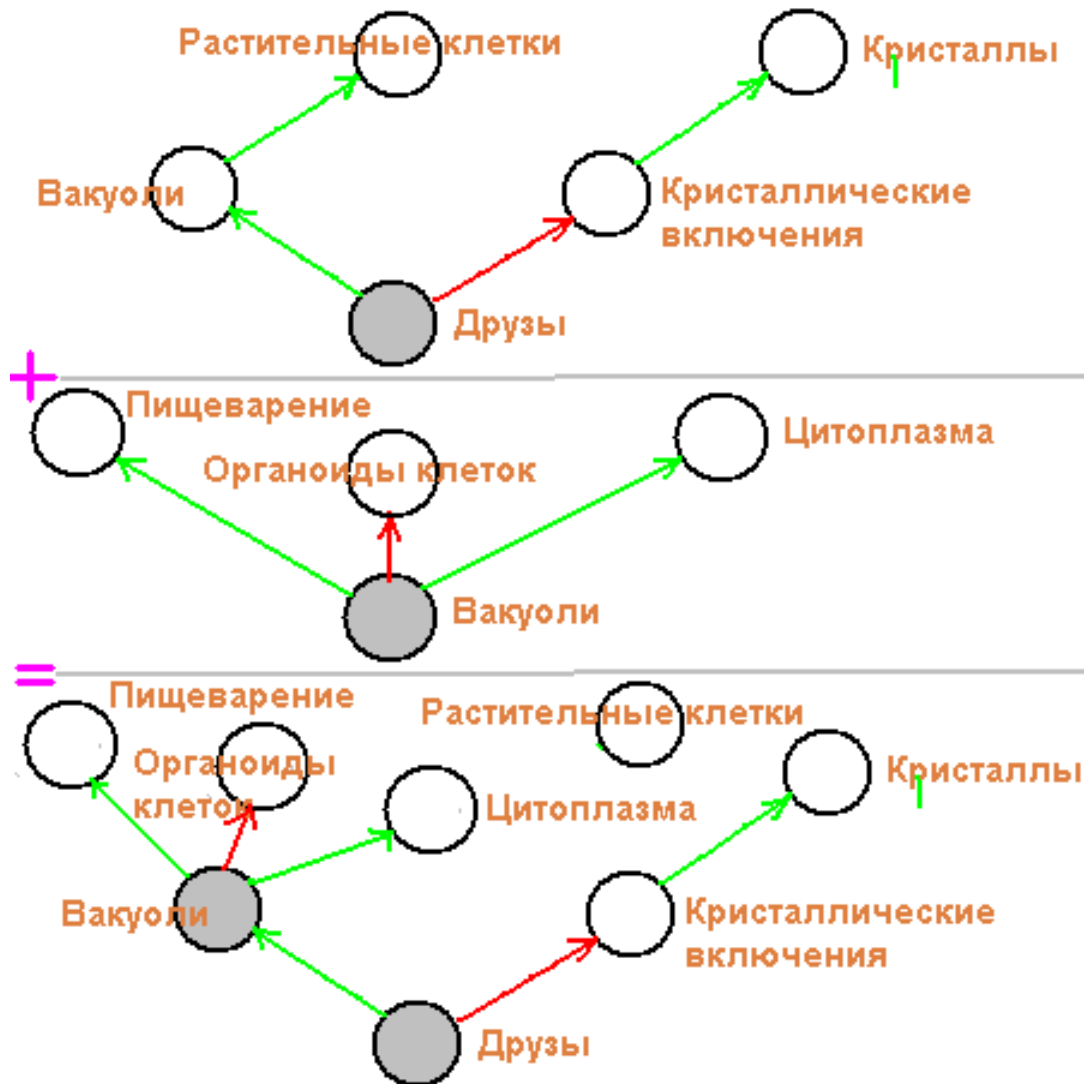
Соглашения



A *есть* **B**



Построение сети



<< Было

<< Новая статья
+ Редактор +

<< Стало

+ Редактор +

+ Редактор-программа:

Надстройка: поиск и выбор статей фрагмента

Текстовый редактор:

модификация фрагмента

Исполнитель фрагмента

+ Редактор-человек: 50 статей в сутки

эффект обучения

Структуры данных

:П: Вещное право

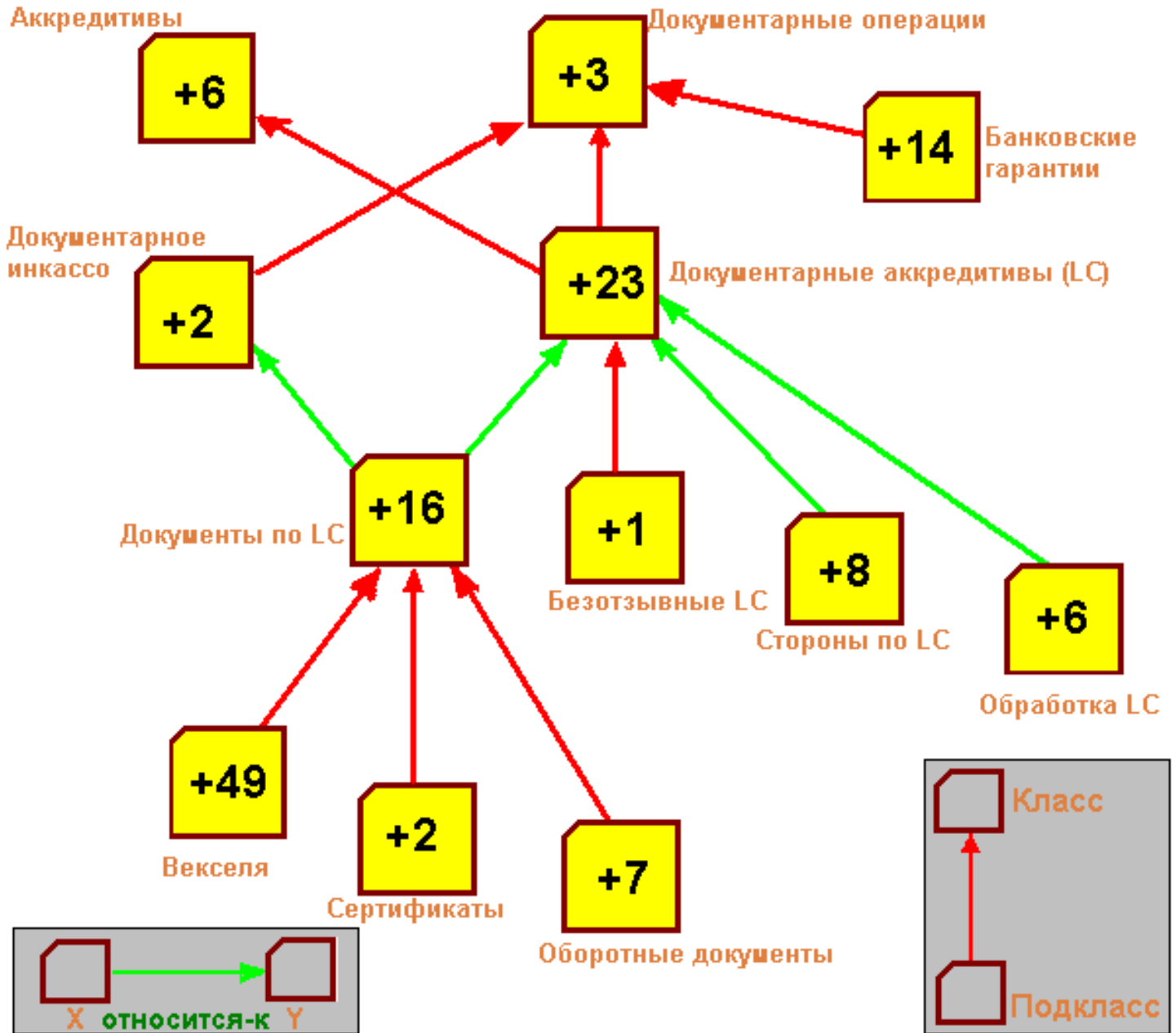
:Ч: Гражданское право

:О: Вещи

:Т: Вещное право

:В: Вещное право – правовые нормы, регулирующие имущественные отношения. Вещное право дает основание имеющему его лицу действовать в отношении имущества, не прибегая к согласию и поддержке других лиц. Объектом вещного права является конкретная вещь.

Ф
Р
А
Г
М
Е
Н
Т



Понятийные вершины

Понятийная вершина – формально - вершина, в которую входит одно или несколько ребер.

Неформально: Понятийной вершине соответствует класс объектов.

Статья *Кредит*



понятие *Кредиты*

январь 2009 г.

Количество статей = 53'477 шт.

в т.ч. «понятий» = 9'044 шт.

Количество связей = 75'569 шт.

в т.ч. «это-есть» = 34'566 шт.

«относится-к» = 41'003 шт.

Свойства сети

1. 1 статья - 1.4 связей

75569 / 53477

2. 6 статей - 1 понятие

53477 / 9044

3. Число 19

4. СВЯЗНОСТЬ

5. Циклы

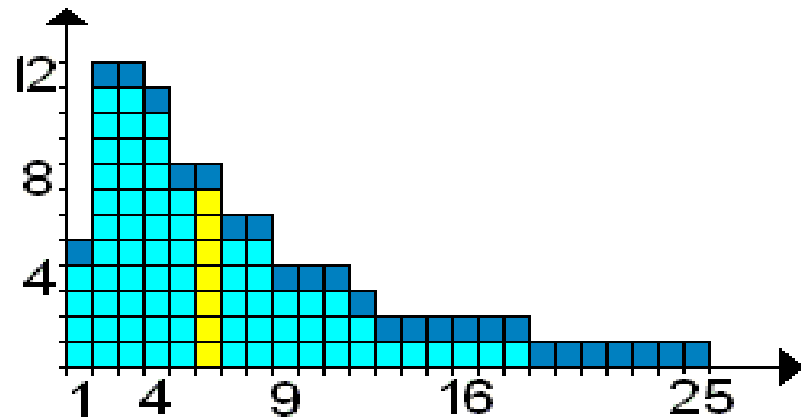
6. Ошибки

7. Влияние социального контекста

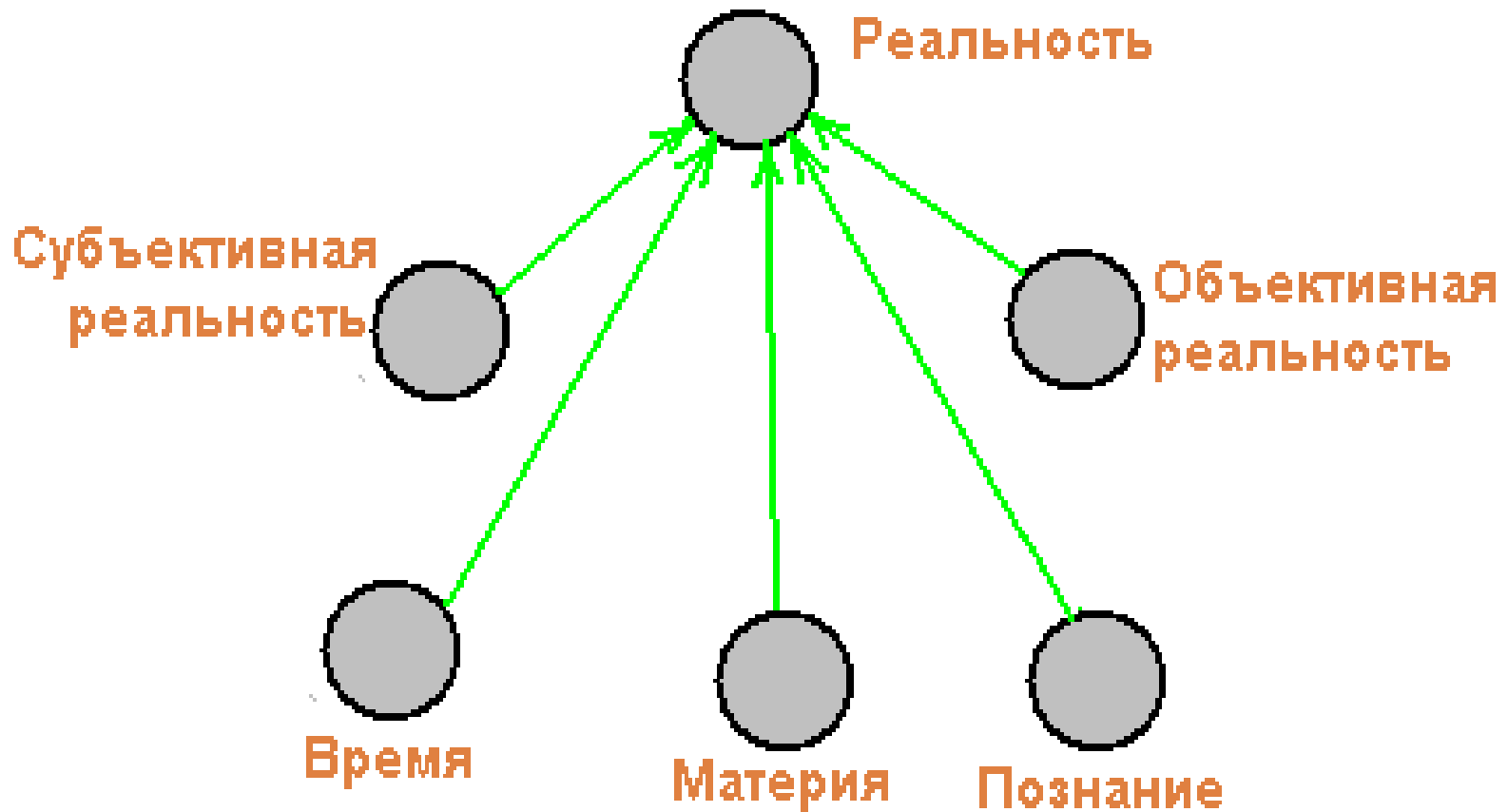
8. Субъективизм (5%)

9. Краевые эффекты: смысловые, структурные

10. 92% уникальной терминологии



Материализм



Прагматика

1. Сеть → Интернет: www.glossary.ru

- задача No.1 Единица показа?

- задача No.2 Навигация?

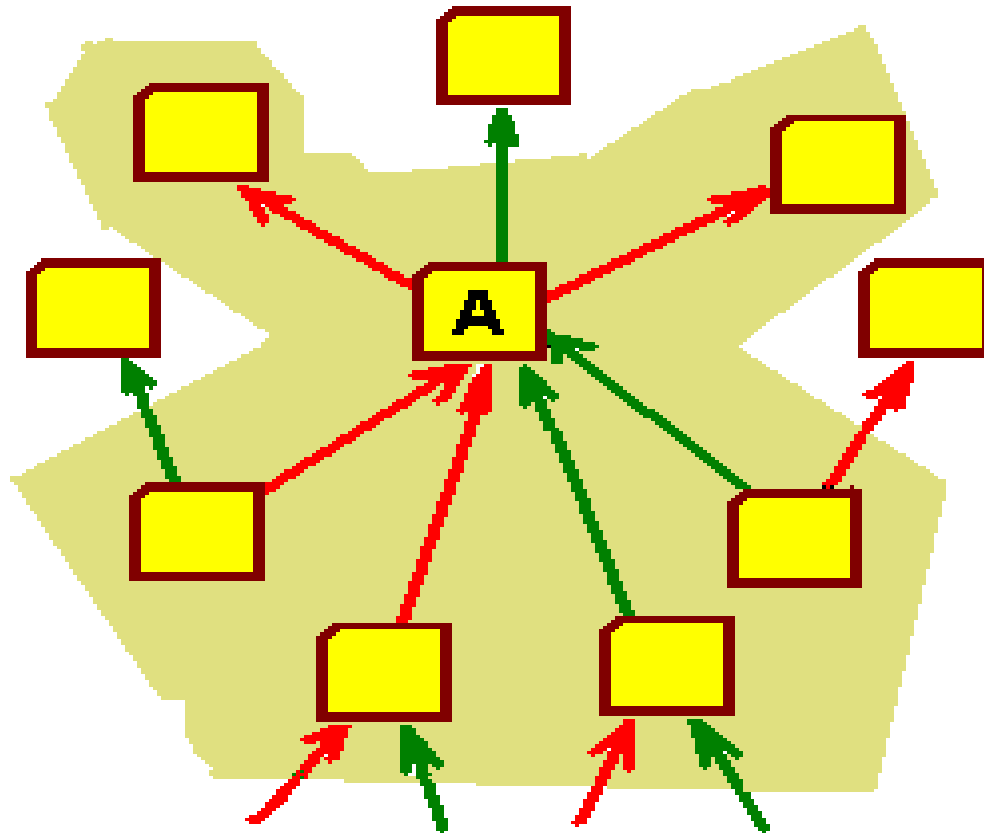
- задача No.3 Интеллектуальные функции?

2. Другие

1. Единица показа: Глоссарий

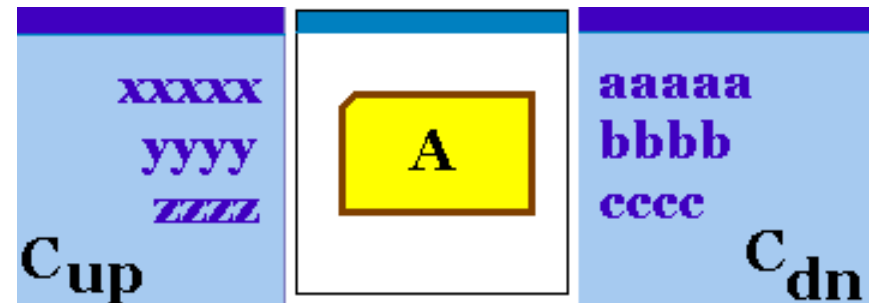
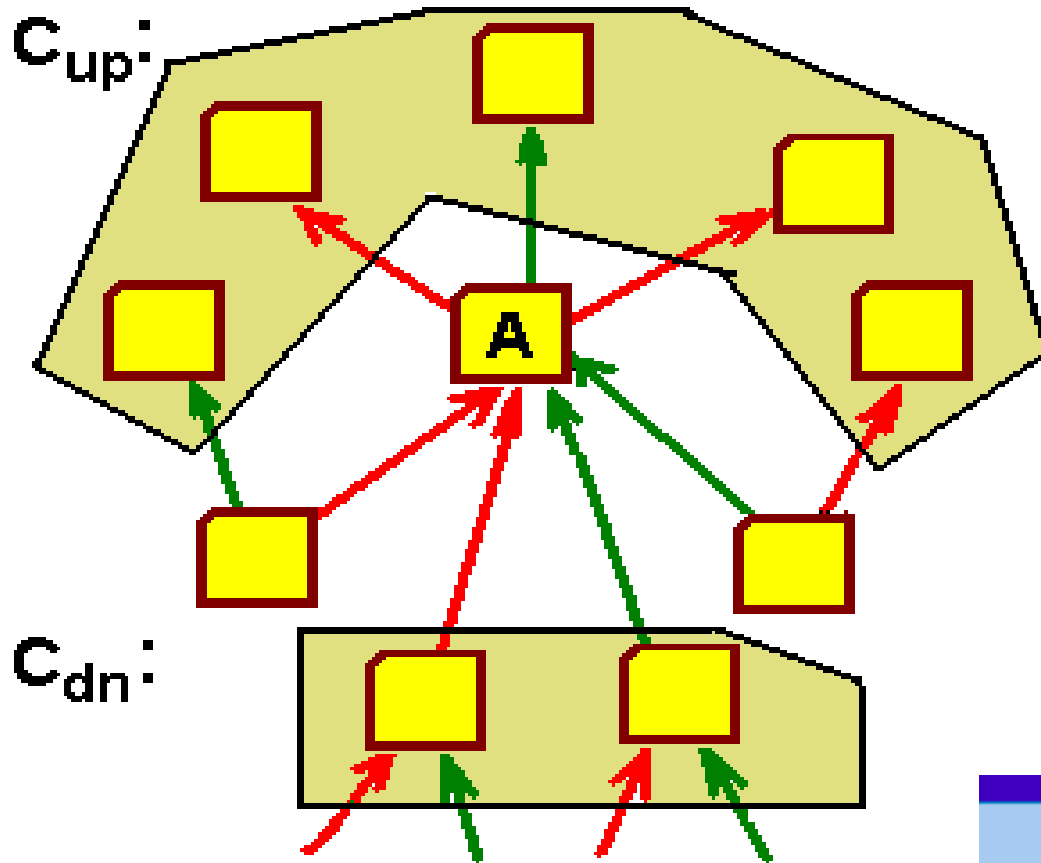
Пусть A – понятийная вершина.

$G(A)$ – множество статей глоссария



2. Навигация: Смежные глоссарии

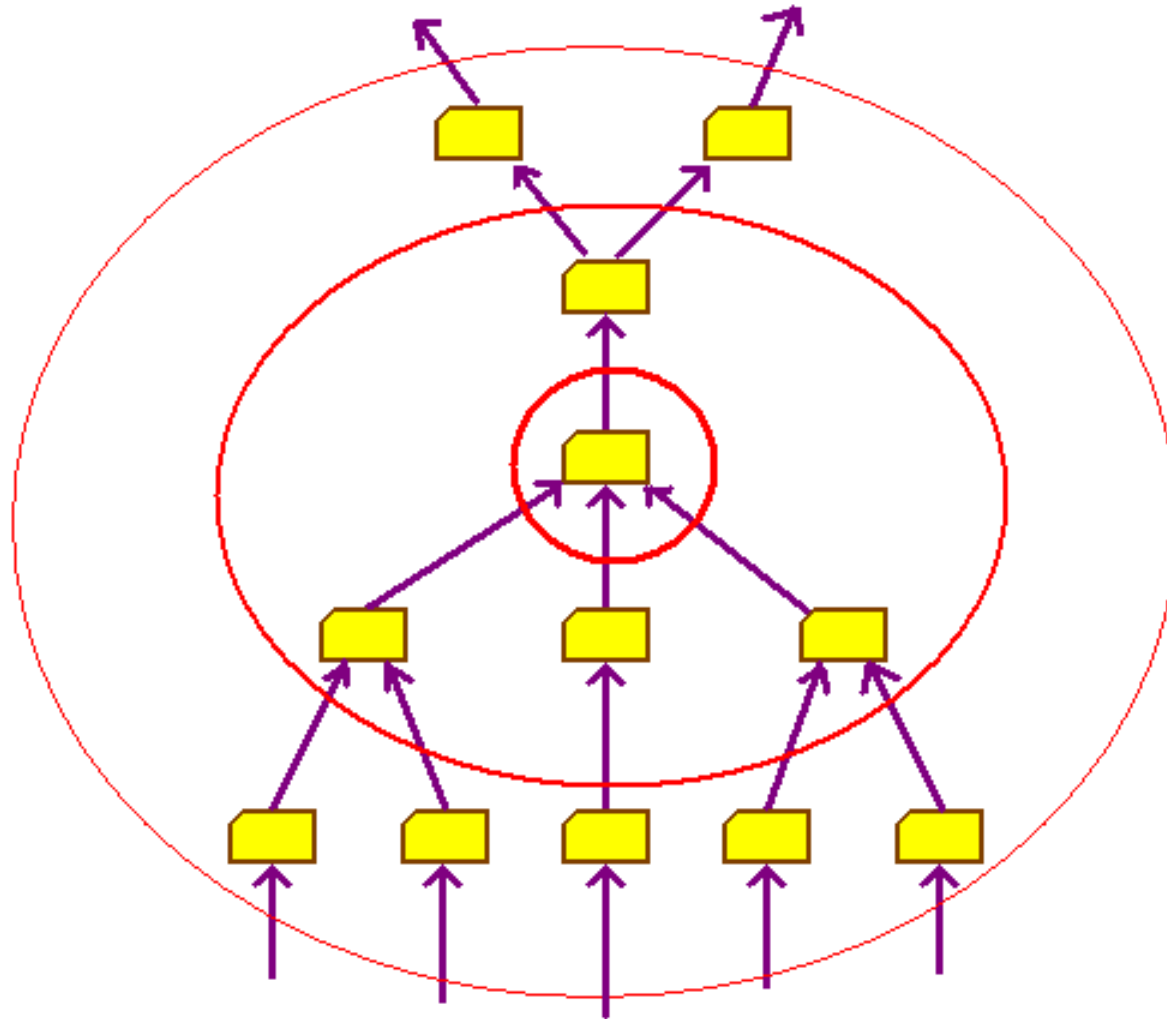
Множества понятий $C_{up}(A)$ и $C_{dn}(A)$



3. Интеллектуальные функции:

1. Тематическая группировка
2. Метод расширения (наследование)
3. Конструктор поисковых запросов к ИПС
4. Веб-карты
5. Визуализация фрагмента сети

3.1 Тематическая группировка

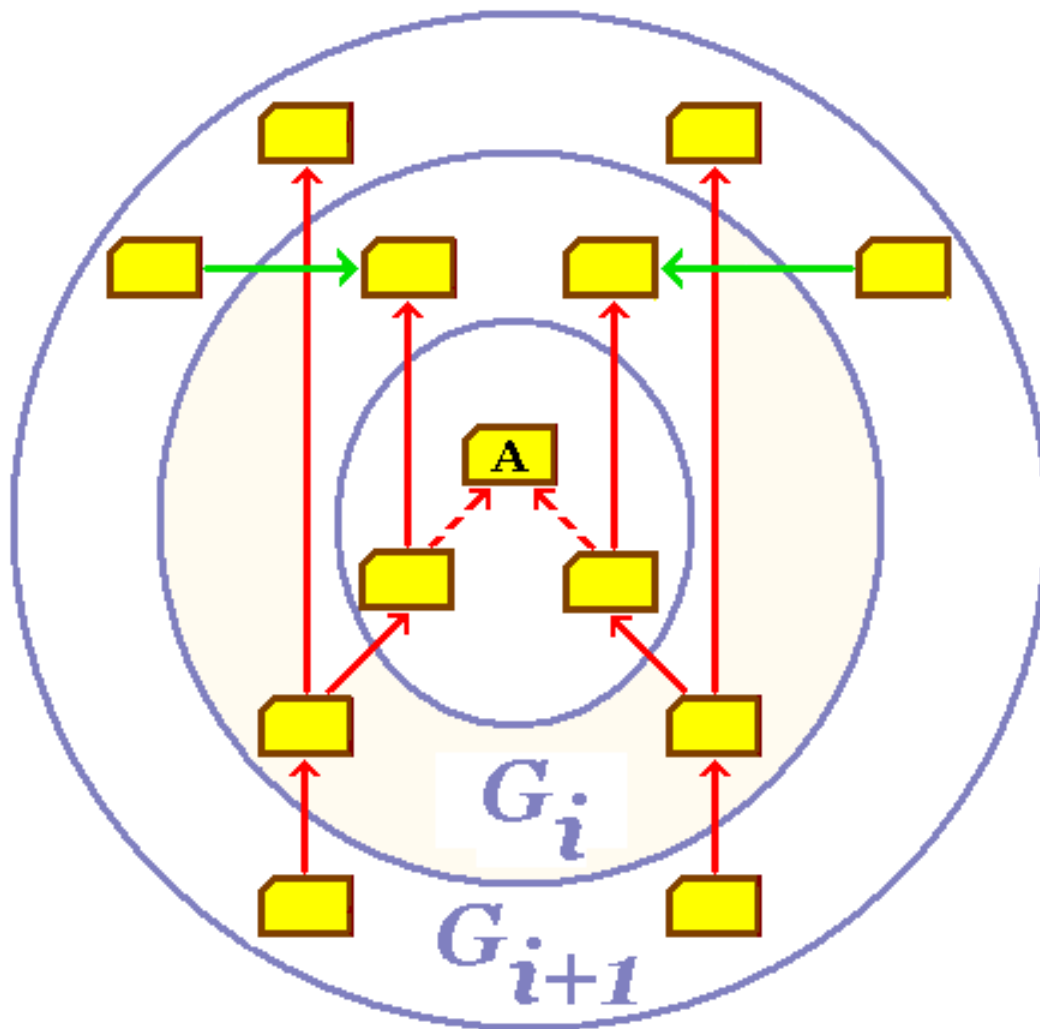


3.2. Метод расширения

$$G_0(A) \subseteq G_1(A) \subseteq \dots \subseteq G_n(A)$$

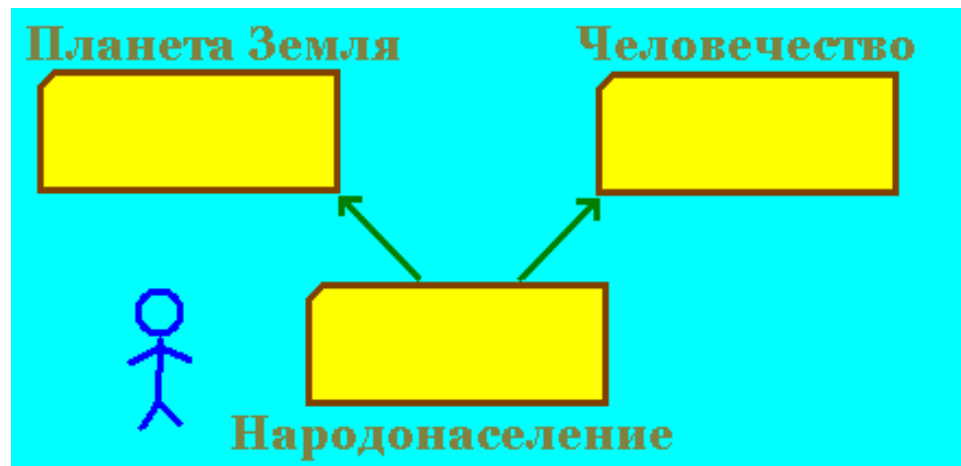
$$G_0(A) = \{A\}$$



$$G_1(A) = G(A)$$



Наследование

3.3 Запросы к ИПС - 1



Планета Земля 
Человечество 

Народонаселение

Население
Population

Народонаселение - совокупность людей, живущих на Земном шаре (человечество) или в пределах конкретной территории - континента, страны, области и т.д. Народонаселение непрерывно возобновляется в ходе воспроизводства.

Развернуть глоссарий

Конструктор поисковых запросов к ИПС Яндекс

Я ИЩУ ДОКУМЕНТ, КОТОРЫЙ


содержит не содержит содержит с сохранением порядка слов



В ОДНОМ ПРЕДЛОЖЕНИИ УПОМИНАНИЕ:

народонаселение

ИЛИ ЕГО СИНОНИМЫ:

- население
 population

содержат упоминание  планета земля  и его синонимы

игнорируют упоминание  человечество  и его синонимы

Найти в Интернет

3.3 Запросы к ИПС - 2

Планета Земля

Человечество

Народонаселение

Население
Population

Народонаселение - совокупность людей, живущих на Земном шаре (человечество) или в пределах конкретной территории - континента, страны, области и т.д. Народонаселение непрерывно возобновляется в ходе воспроизводства.

Развернуть глоссарий

Конструктор

поисковых запросов
к ИПС Яндекс

я ищу ДОКУМЕНТ, который

содержит не содержит содержит с сохранением порядка слов

В ОДНОМ ПРЕДЛОЖЕНИИ УПОМИНАНИЕ:

народонаселение

ИЛИ ЕГО СИНОНИМЫ:

- население
 population

ПОИСК ОГРАНИЧИТЬ
ДОКУМЕНТАМИ,
которые

содержат упоминание планета земля и его синонимы

игнорируют упоминание человечество и его синонимы

Найти в Интернет



<http://yandex.ru/yandsearch?text=%28%D0%F0%E4%EE%ED%F1%E5%EB%E5%ED%E8>

Поиск [Почта](#) [Новости](#) [Маркет](#) [Карты](#) [Словари](#) [Блоги](#) [Картинки](#) [ещё](#) ▼

Яндекс

Нашёлся
1 млн страниц

(народонаселение | население) && ((планета земля) | earth)

Найти

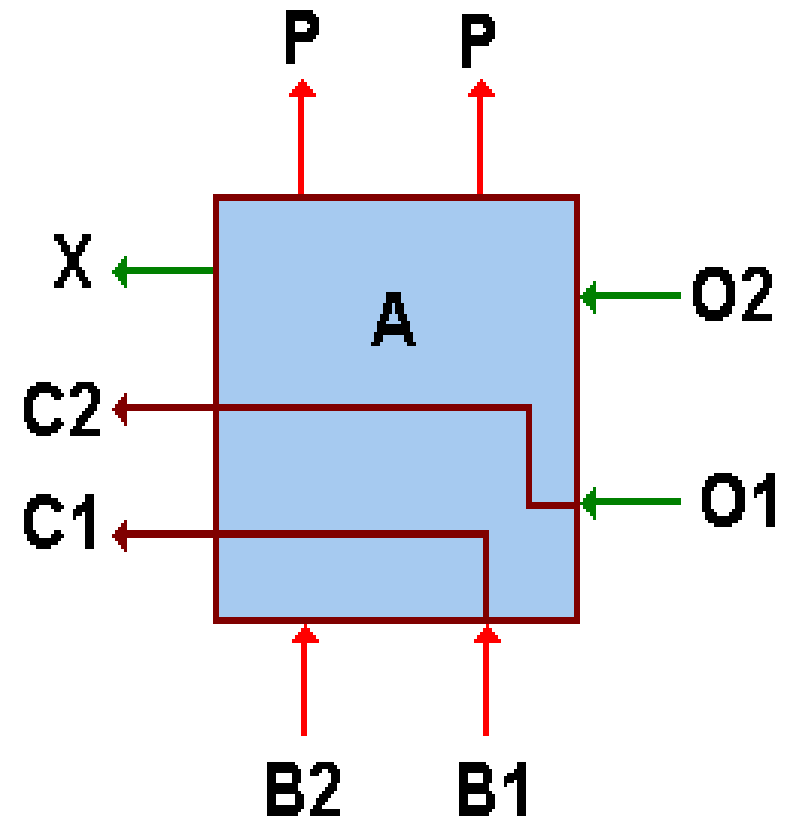
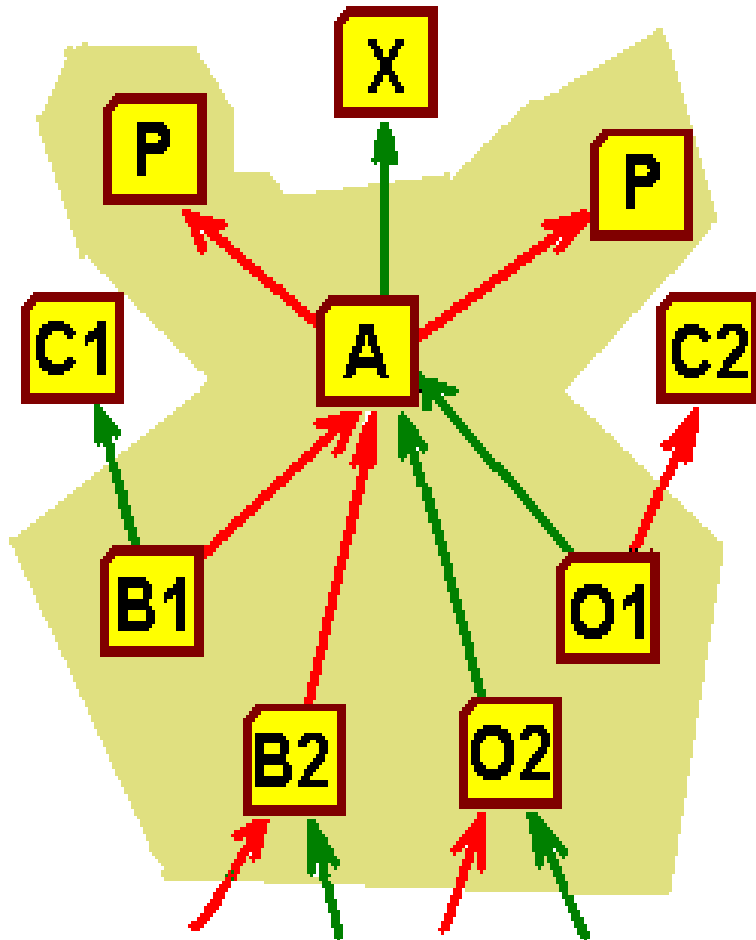
в найденном в Москве

[расширенный поиск](#)

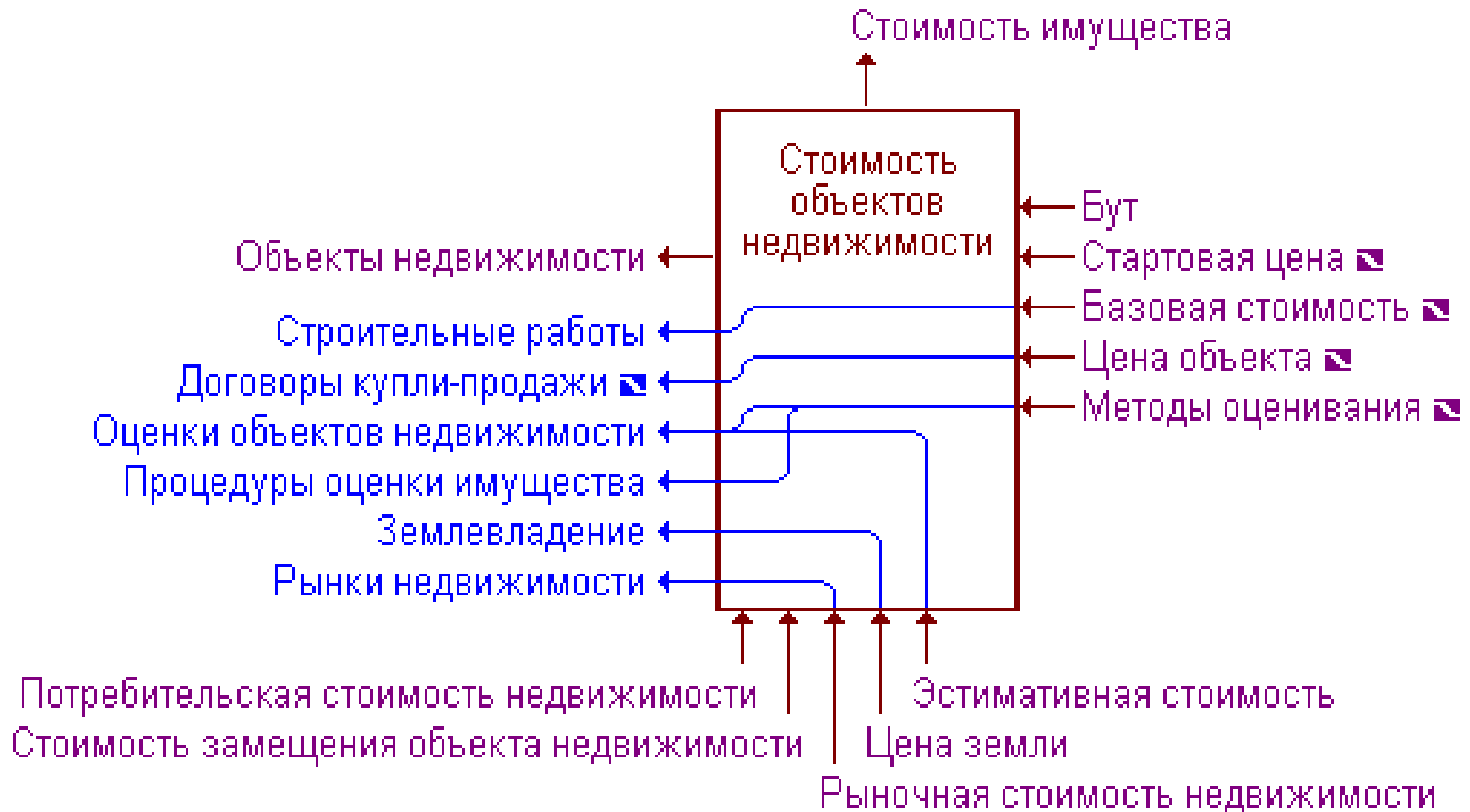
3.4 Веб-карта



3.5 Принцип визуализации



Результат визуализации



Формула глоссария $1 + 5/2 + 5/2 + 1 + 6$

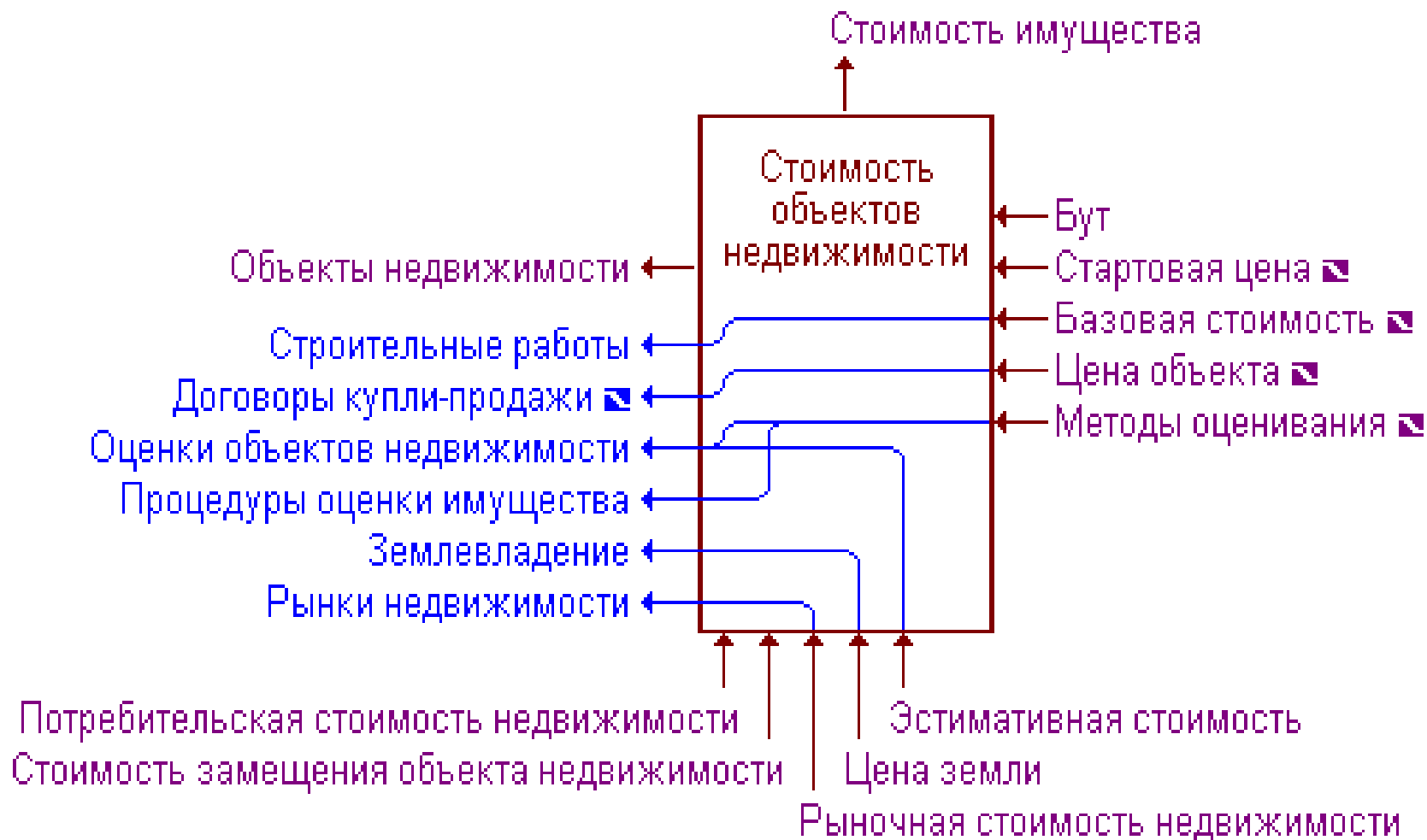


Таблица формул

1. $1+1/0+4/1+1+4$ CGI-приложения

...

5826. $1+5/2+5/2+1+6$ Стоимость об...

...

7351. $1+4/4+4/3+0+2$ Ячеистые бетоны

www.park.glossary.ru/serios/formulas.zip

93% понятий содержат некоторое количество собственных подвигов или собственных аксессуаров.

68% понятий одновременно содержит подвиды и аксессуары.

21% понятий имеют сильно востребованный состав.

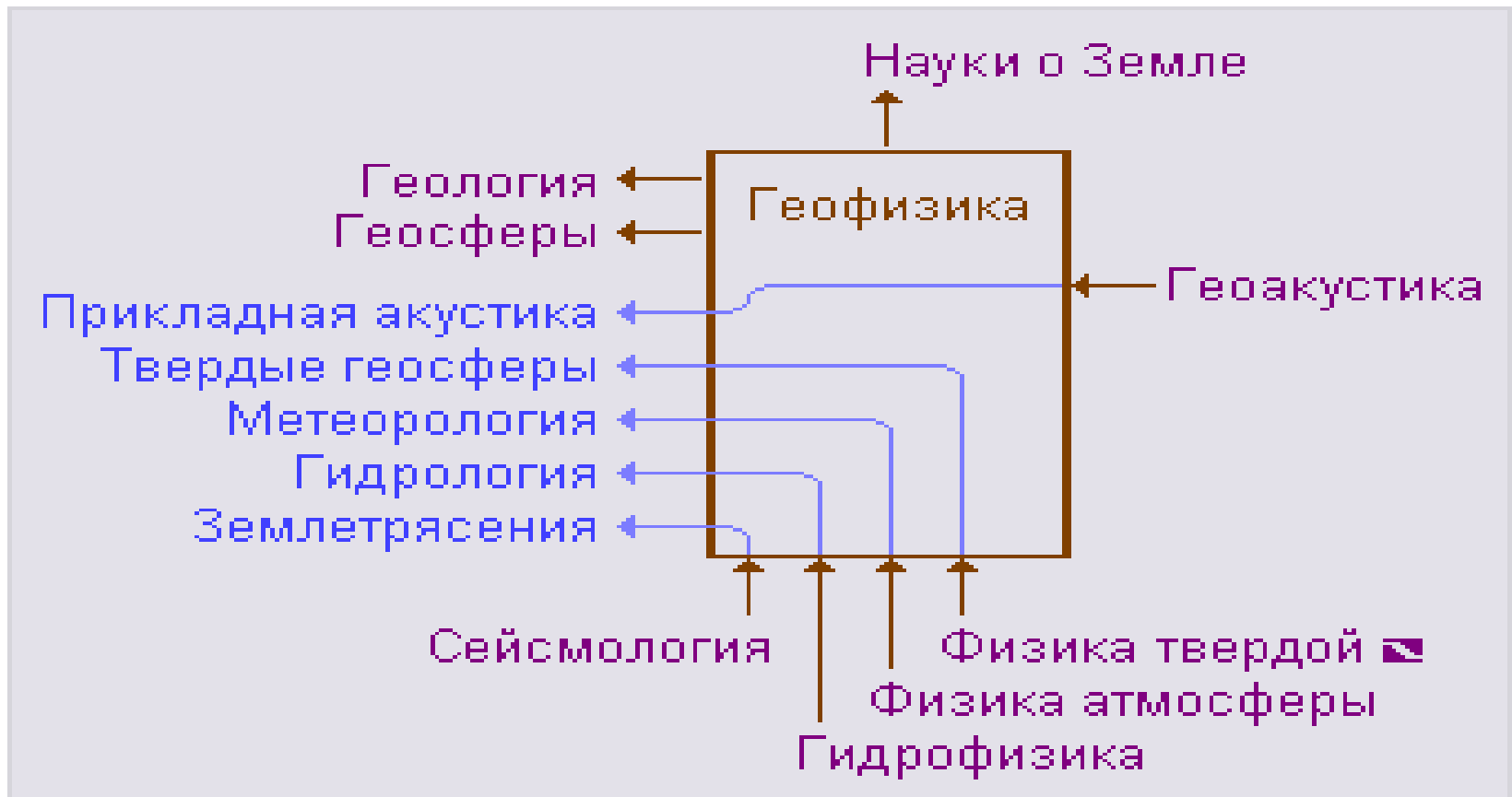
10% понятий имеет строение общего вида.

7%

68%

21%

10%

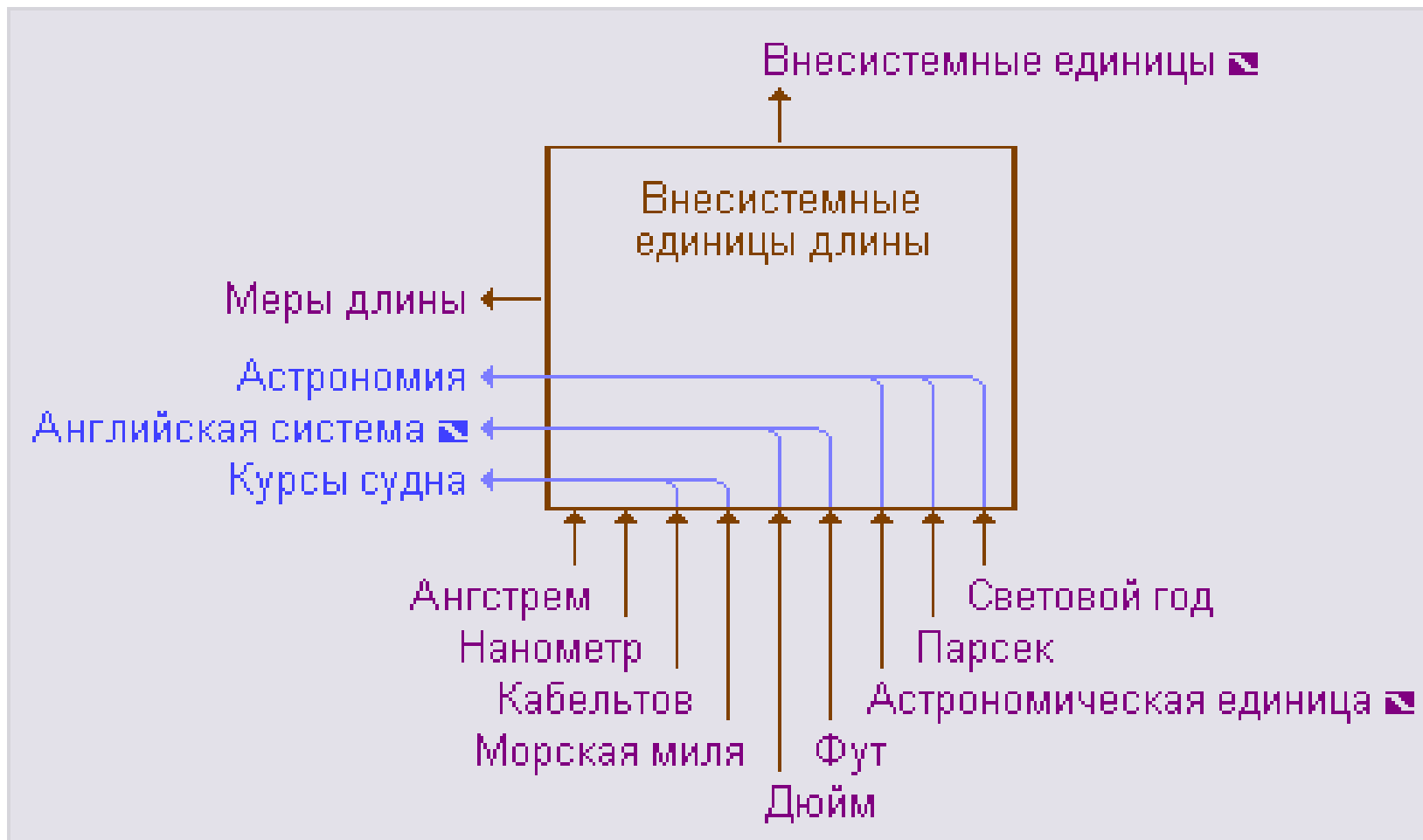


93%

32%

79%

90%



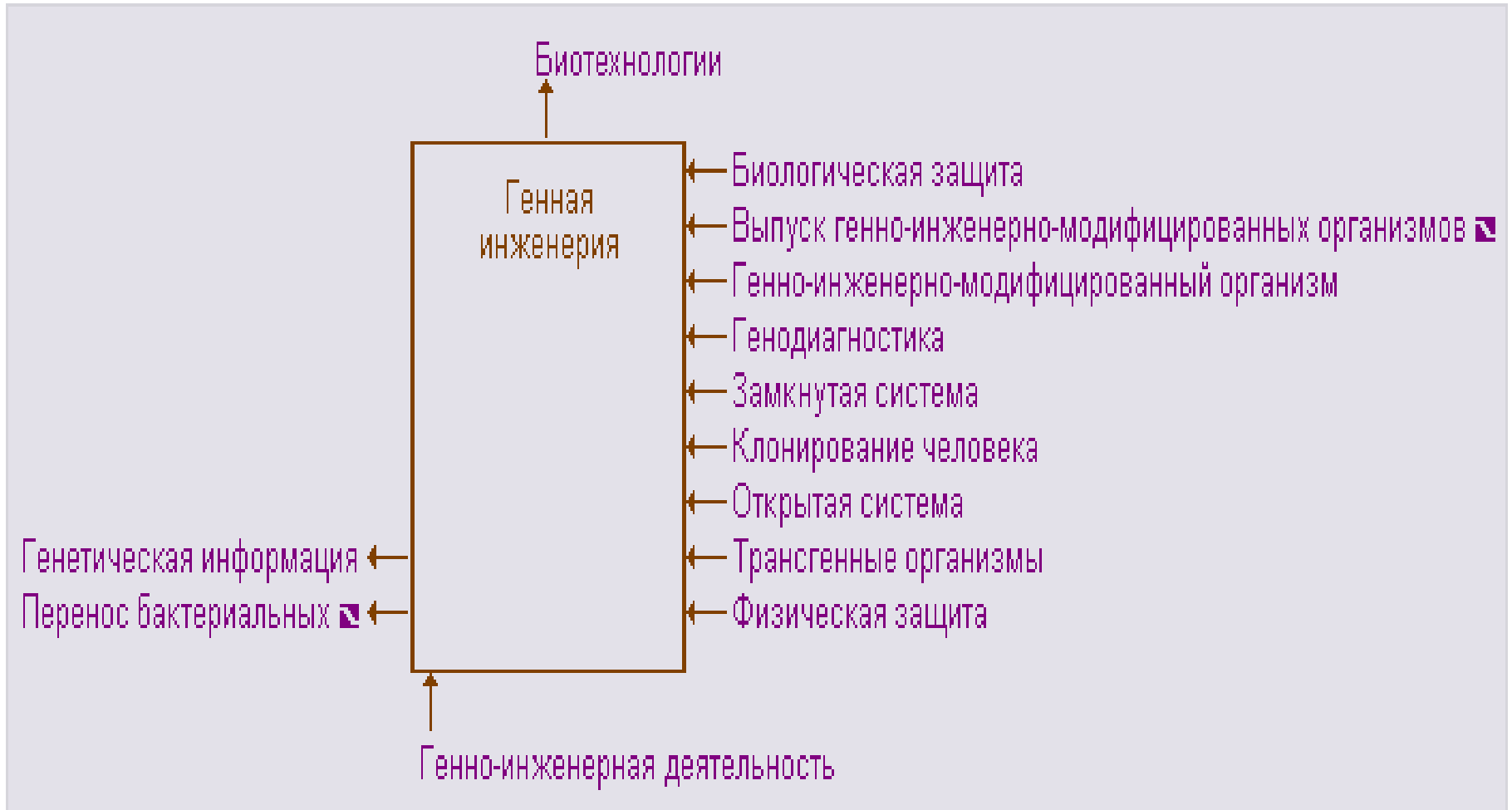
93%

68%

79%

90%

Кумулятивность?



Нерешенные задачи

1. Кластеризация понятий
2. Навигация
3. Словарное представление фрагмента сети
4. Порядок предъявления статей
5. Алгоритм формирования запросов к ИПС (часть 2)
6. Автоматизация формирования сети
7. Применение сети в ИПС

**Темы рефератов
на www.park.glossary.ru/reports.htm**

Вопросы?

проф.С.Ю.Соловьев soloviev@jscc.ru