

# Формирование и кластеризация контекстов для существительных русского языка в рамках Расщепленных Значений.

Д. В. Михайлов, Г. М. Емельянов, Н. А. Степанова

Новгородский Государственный Университет имени Ярослава Мудрого

## Цель.

*Разработка и исследование методов выявления Расщепленных Значений (РЗ) в составе синтаксических контекстов существительных для использования указанных контекстов как исходных данных семантической кластеризации текстов Естественного Языка (ЕЯ).*

## Задачи исследования.

- 1) Формулирование требований к синтаксическому контексту слова как базовой структуре семантической кластеризации ЕЯ-текстов.
- 2) Анализ особенностей и построение формальной концептуальной модели процесса семантической кластеризации ЕЯ-текстов.
- 3) Разработка и программная реализация механизма выявления и кластеризации Расщепленных Предикатных Значений (РПЗ) на основе результатов синтаксического разбора предложений.
- 4) Проведение машинных экспериментов и выработка практических рекомендаций по качественному составу синтаксических контекстов существительных для известных в лингвистике расщеплений Лексического Значения слова.

# Установление Семантической Эквивалентности как задача кластеризации ЕЯ-текстов.

Дано :

$G = \{T_1, \dots, T_{n(G)}\}$  — множество анализируемых текстов,  $n(G) = |G|$ .

Требуется :

- 1) По результатам синтаксического разбора  $T_1, \dots, T_{n(G)}$  выявить для каждого  $T_i \in G, i = \overline{1, n(G)}$  :
  - множество  $V(T_i)$  ситуаций, описываемых  $T_i$ ;
  - множество  $M(T_i)$  объектов (понятий), значимых в ситуациях из множества  $V(T_i)$ ;
  - тернарное отношение  $I \subseteq G \times M \times V$ , которое ставит в соответствие каждому  $m \in M: M = \bigcup_i M(T_i)$  ситуацию  $v \in V: V = \bigcup_i V(T_i)$ , в которой он фигурирует относительно заданного  $T_i$ .
- 2) На основе выявленного отношения  $I$  выделить в  $G$  группы текстов, сходных по встречаемости объектов в одних и тех же ситуациях.

# Ситуационный контекст как основа кластеризации ЕЯ-текстов.

Пусть

$$S_{ki} = \{v_1, \dots, v_{n(ki)}, m_{ki}\} \quad (1)$$

есть последовательность соподчиненных слов, получаемая как результат синтаксического анализа некоторого  $T_i \in G$ .

При этом :

$v_1 \in V(T_i)$  обозначает некоторую ситуацию и является глаголом (отглагольным существительным);

$m_{ki}$  — существительное и обозначает некоторое понятие, значимое в  $v_1$ ;

$\forall v_l \in \{v_2, \dots, v_{n(ki)}\}, l = \overline{2, n(ki)}$  — некоторое существительное.

**Утверждение 1.** Пусть  $R_q$  — отношение синтаксического подчинения,  $q$  — тип этого отношения, характеризуется падежом зависимого слова и предлогом для связи синтаксически главного и зависимого слова. При  $v_1 R_q v_2$  возможно существование отношения  $R_q$  между  $v_1$  и любым словом последовательности (1) вне зависимости от уже существующих отношений подчинения между словами этой последовательности.

**Доказательство.** Следует из соотношения смыслов соподчиненных слов.

**Замечание 1.** На основе Утверждения 1 справедливым будет утверждать, что  $\forall v_l \in \{v_2, \dots, v_{n(ki)}\}$  так же, как и  $m_{ki}$ , обозначает некоторое понятие, значимое в  $v_1$ . Таким образом, для  $\forall S_{ki}$  (1)  $\{v_2, \dots, v_{n(ki)}, m_{ki}\} \subset M(T_i)$ . Причем  $V(T_i) = \bigcup_k (S_{ki} \setminus \{m_{ki}\})$ .

**Замечание 2.** Роль, в которой  $m \in M(T_i)$  выступает относительно некоторой ситуации  $v \in V(T_i)$ , определяется типом  $q$  отношения  $R_q$  между  $v$  и словом справа от него в последовательности (1), где  $v_1 = v$ .

---

**Алгоритм 1** Формирование троек-кандидатов на включение в отношение  $I$ .

---

**Вход:**  $P_i^S = \{S_{ki} : S_{ki} \text{ — последовательность вида } (1)\};$

**Выход:**  $P_i^K = \{P_{ki}^K : P_{ki}^K = \{(g_i, m, v) : (g_i, m, v) \in I\}\};$

//  $g_i$  есть некоторая пометка для  $T_i \in G$

$P_i^K := \emptyset;$  // Инициализация

**пока**  $P_i^S \neq \emptyset$

Выбрать  $S_{ki}$  из  $P_i^S$ ;

$P_{ki}^K := \emptyset;$

**для**  $l = 1, \dots, n(ki)$

$P_{ki}^K := P_{ki}^K \cup \{(g_i, m_{ki}, v_l)\};$  //  $S_{ki} = \{v_1, \dots, v_{n(ki)}, m_{ki}\}$  в соответствии с (1)

$j := n(ki);$

**пока**  $j > l$

$P_{ki}^K := P_{ki}^K \cup \{(g_i, v_j, v_l)\};$

$j := j - 1;$

$P_i^K := P_i^K \cup \{P_{ki}^K\};$

$P_i^S := P_i^S \setminus \{S_{ki}\};$

---

# Анализ Формальных Понятий и концептуальная кластеризация.

Рассмотрим  $G$  как множество формальных объектов,  $M$  — как множество формальных признаков,  $V$  — как множество значений формальных признаков.

Отношению  $I \subseteq G \times M \times V$  ставится в соответствие формальный контекст :

$$K = (G, M, V, I) \quad (2)$$

**Определение 1.** Под Формальным Понятием (ФП) для (2) понимается пара  $(A, B)$ :  $A \subseteq G$ ,  $B \subseteq M \times V$ ,  $A = B'$ ,  $B = A'$ , причем

$$A' = \{(m, v) : m \in M, v \in V \mid \forall g \in A : m(g) = v\},$$

$$B' = \{g \in G \mid \forall (m, v) \in B : m(g) = v\}$$

**Определение 2.** ФП  $(A_1, B_1)$  является подпонятием для ФП  $(A_2, B_2)$ , если  $A_1 \subseteq A_2$ , а  $B_2 \subseteq B_1$  :  $(A_1, B_1) \leq (A_2, B_2)$ . При этом  $(A_2, B_2)$  называют суперпонятием для ФП  $(A_1, B_1)$ , а отношение  $\leq$  — отношением порядка для ФП.

**Определение 3.** Множество  $\mathfrak{K}(G, M, V, I)$  всех ФП контекста (2) вместе с отношением  $\leq$  называется решеткой Формальных Понятий.

**Замечание.** Каждое слово в множестве  $M = \bigcup_i M(T_i)$  представлено вместе с предлогом, посредством которого оно связывается с другим словом, синтаксически главным по отношению к нему в (1).

**Определение 4.** ФП  $C_1$  и  $C_2$  считаются сравнимыми, если либо  $C_1 \leq C_2$ , либо  $C_2 \leq C_1$ . Подмножество множества ФП, в котором каждые два элемента сравнимы, называют цепочкой.

---

**Алгоритм 2** Построение формального контекста для исходного множества текстов.

---

**Вход:**  $G$ ;

**Выход:**  $K(G, M, V, I)$ ;

- 1: Синтаксический анализ текстов из  $G$  с формированием  $P_i^S = \{S_{ki} : S_{ki} \text{ — последовательность вида (1)}\}$  для каждого  $T_i \in G$ ;
  - 2: Для каждого  $T_i \in G$  на основе  $P_i^S$  выделить  $M(T_i)$  и  $V_1(T_i) \subset V(T_i) : V_1(T_i) = \{v_1 : S_{ki} = \{v_1, \dots, v_{n(ki)}, m_{ki}\}, (1)\}$ ;
  - 3: На основе выделенных  $\{M(T_i) | i = \overline{1, n(G)}\}$  и  $\{V_1(T_i) | i = \overline{1, n(G)}\}$  найти одноименные ситуации  $v$ , принадлежащие различным  $V_1(T_i)$  и сходные по фигурирующим в них объектам  $m \in M : M = \bigcup_i M(T_i)$  в сходных ролях;
  - 4: Приписать названиям выявленных на Шаге 3 ситуаций одинаковые целочисленные индексы в соответствующих  $V_1(T_i)$  и  $P_i^S$ ;
  - 5: По аналогии с Шагом 3 на основе выделенных  $P_i^S$  найти разноименные ситуации  $v$ , принадлежащие различным  $V_1(T_i)$  и сходные по фигурирующим в них объектам  $m \in M : M = \bigcup_i M(T_i)$  в сходных ролях;
  - 6: По каждой группе синонимов  $\{v_1 : S_{ki} = \{v_1, \dots, v_{n(ki)}, m_{ki}\}, (1) | i = \overline{1, n(G)}\} := Syn$ , выявленной на Шаге 5, выделить канонический представитель  $v'_1$  с наибольшей частотой употребления и заменить все  $v_1 \in S_{ki} | S_{ki} \in Syn$  на  $v'_1$ ;
  - 7: Выполнить Шаги 3–6 для разноименных ситуаций, принадлежащих различным  $V_1(T_i)$  и сходных по фигурирующим в них  $m \in M : M = \bigcup_i M(T_i)$ , но с меной ролей (конверсивы);
  - 8: Для каждого  $T_i \in G$  сформировать  $V(T_i) = V_1(T_i) \cup (\bigcup_k (S_{ki} \setminus \{m_{ki}\} \setminus \{v_1\}))$  и установить отношение  $I$  в соответствии с Алгоритмом 1 с учетом результатов выполнения шагов 3–7;
-

# Расщепленные Предикатные Значения.

**Утверждение 2.** Пусть  $S_1 \subset T_1$ ,  $S_2 \subset T_2$ ,  $S_1 = \{S_{11}, \dots, S_{n(S_1)1}\}$ ,  $S_2 = \{S_{12}, \dots, S_{n(S_2)2}\}$ ,  
где  $n(S_1) = |S_1|$ ,  $n(S_2) = |S_2|$ ,  $n(S_2) = n(S_1) - 1$ .

Применительно к паре  $\{T_1, T_2\} \subset G$  анализируемых текстов Расщепленное Предикатное Значение (РПЗ) будет иметь место в следующих двух случаях.

**Случай (1).**

$$\left. \begin{array}{l} S_{11} = \{v_{11}, v_{12}, v_{13}, \dots, v_{1n_1}, m_{11}\} \\ S_{21} = \{v_{11}, v_{12}, v_{23}, \dots, v_{2n_2}, m_{21}\} \\ \dots \\ S_{k1} = \{v_{11}, v_{12}, v_{k2}, \dots, v_{kn_k}, m_{k1}\} \\ S_{(k+1)1} = \{v_{11}, v_{(k+1)2}, \dots, v_{(k+1)n_{(k+1)}}, m_{(k+1)1}\} \\ \dots \\ S_{n(S_1)1} = \{v_{11}, v_{n(S_1)2}, \dots, v_{n(S_1)n_{n(S_1)}}, m_{n(S_1)1}\} \end{array} \right\}, \quad \left. \begin{array}{l} S_{12} = \{v_{21}, v_{13}, \dots, v_{1n_1}, m_{11}\} \\ S_{22} = \{v_{21}, v_{23}, \dots, v_{2n_2}, m_{21}\} \\ \dots \\ S_{k2} = \{v_{21}, v_{k2}, \dots, v_{kn_k}, m_{k1}\} \\ S_{(k+1)1} = \{v_{21}, v_{(k+1)2}, \dots, v_{(k+1)n_{(k+1)}}, m_{(k+1)1}\} \\ \dots \\ S_{n(S_1)1} = \{v_{21}, v_{n(S_1)2}, \dots, v_{n(S_1)n_{n(S_1)}}, m_{n(S_1)1}\} \end{array} \right\}$$

**Случай (2).**

$$\left. \begin{array}{l} S_{11} = \{v_{11}, v_{13}, \dots, v_{1n_1}, m_{11}\} \\ S_{21} = \{v_{11}, v_{23}, \dots, v_{2n_2}, m_{21}\} \\ \dots \\ S_{(k-1)1} = \{v_{11}, v_{(k-1)2}, \dots, v_{(k-1)n_{(k-1)}}, m_{(k-1)1}\} \\ S_{k1} = \{v_{11}, v_{12}\} \\ S_{(k+1)1} = \{v_{11}, v_{(k+1)2}, \dots, v_{(k+1)n_{(k+1)}}, m_{(k+1)1}\} \\ \dots \\ S_{n(S_1)1} = \{v_{11}, v_{n(S_1)2}, \dots, v_{n(S_1)n_{n(S_1)}}, m_{n(S_1)1}\} \end{array} \right\}, \quad \left. \begin{array}{l} S_{12} = \{v_{21}, v_{13}, \dots, v_{1n_1}, m_{11}\} \\ S_{21} = \{v_{21}, v_{23}, \dots, v_{2n_2}, m_{21}\} \\ \dots \\ S_{(k-1)1} = \{v_{21}, v_{(k-1)2}, \dots, v_{(k-1)n_{(k-1)}}, m_{(k-1)1}\} \\ S_{(k+1)1} = \{v_{21}, v_{(k+1)2}, \dots, v_{(k+1)n_{(k+1)}}, m_{(k+1)1}\} \\ \dots \\ S_{n(S_2)1} = \{v_{21}, v_{n(S_2)2}, \dots, v_{n(S_2)n_{n(S_2)}}, m_{n(S_2)1}\} \end{array} \right\}.$$

## Концептуальная кластеризация РПЗ.

Пусть

$$\{\{v_{11}, v_{12}\} := SV_i \mid i \in \{1, n(G)\}\} := SV^G$$

есть множество РПЗ, выявляемых в соответствии с Утверждением 2 (здесь  $v_{11}$  — синтаксически главное,  $v_{12}$  — зависимое слово в РПЗ  $SV_i$ ).

Введем в рассмотрение формальный контекст :

$$K^{SV} = (G^{SV}, M^{SV}, I^{SV}), \quad (3)$$

в котором

$$\begin{aligned} G^{SV} &= \{v_{12} : v_{12} \in SV_i \mid i = \overline{1, n(G)}\}, \\ M^{SV} &= \{v_{11} : v_{11} \in SV_i \mid i = \overline{1, n(G)}\}. \end{aligned}$$

Отношение  $I^{SV} \subseteq G^{SV} \times M^{SV}$  ставит в соответствие каждому синтаксически зависимому слову  $v_{12} \in G^{SV}$  некоторого  $SV_i \in SV^G$  множество слов из  $M^{SV}$ , синтаксически главных по отношению к нему в тех или иных РПЗ из  $SV^G$ .

**Замечание.** С учетом возможного наличия конверсивов для  $v_{21}$  применительно как к Случаю (1), так и Случаю (2) РПЗ из рассматриваемых Утверждением 2, будем предполагать, что соответствующая замена уже выполнена, а  $S_1 \subset T_1$  и  $S_2 \subset T_2$  описывают одно и то же множество объектов относительно одной и той же ситуации, обозначаемой  $v_{21}$ , то есть без мены ролей.



---

**Алгоритм 3** Формирование множества объектов с наборами признаков.

---

**Вход:**  $P^S = \{P_i^S = \{S_{ki} : S_{ki} \text{ — последовательность вида (1), } i = \overline{1, n(G)}\}\};$

**Выход:**  $P^C$ ; // Множество объектов с наборами признаков

$P^K = \{((v_{12}, v_{11}), v_{21}) : \{v_{11}, v_{12}\} \text{ есть РПЗ для } v_{21} \text{ в соответствии с Утверждением 2}\};$

$P^C := \emptyset$ ;  $P^K := \emptyset$  // Инициализация

для  $i = 1, \dots, n(G)$

Выделить в  $P_i^S$  группы  $P_{ki}^S \subseteq P_i^S$  подряд идущих  $S_{ki}$  (1) с одним и тем же  $v_1$ ;

$P_i^{S'} := \{P_{ki}^S : P_{ki}^S = \{S_{ki} \in P_i^S, v_1 \text{ — одинаково}\}\};$

$P^{S'} := \{P_i^{S'} \mid i = \overline{1, n(G)}\};$

для всех  $P_i^{S'}$  таких, что  $i = \overline{1, n(G)}$

Выбрать  $P_j^{S'} \in P^{S'} : j \neq i$ ;

для всех  $P_{k_1 i}^S \in P_i^{S'}$

Найти  $P_{k_2 j}^S \in P_j^{S'} : \{P_{k_1 i}^S, P_{k_2 j}^S\}$  удовлетворяет условию Утверждения 2;

$P^K := P^K \cup \{((v_{12}, v_{11}), v_{21})\}$  в соответствии с Утверждением 2;

если  $P^C = \emptyset$  то

$P_k^C := \{v_{11}\}$ ;  $P^C := P^C \cup \{(v_{12}, P_k^C)\}$ ;

иначе

Найти  $(v_{12}, P_k^C) \in P^C$ ;

$P^C := P^C \setminus \{(v_{12}, P_k^C)\}$ ;  $P_k^C := P_k^C \cup \{v_{11}\}$ ;  $P^C := P^C \cup \{(v_{12}, P_k^C)\}$ ;

для всех  $P_i^S$  таких, что  $i = \overline{1, n(G)}$

Заменить  $\{v_{11}, v_{12}\} : \exists ((v_{12}, v_{11}), v_{21}) \in P^K$  в  $\forall S_{ki} = \{v_{11}, v_{12}, v_{k2}, \dots, v_{kn_k}, m_{ki}\}$  (1) на  $v_{21}$ ;

---

# Критерий полезности решетки ФП для РПЗ.

Пусть  $C = (A, B)$  — ФП для (3). При этом для  $A \subseteq G^{SV}$  и  $B \subseteq M^{SV}$  вводится пара отображений:  $A' = \{v_{11} \in M^{SV} \mid \forall v_{12} \in A : v_{12} I^{SV} v_{11}\}$  и  $B' = \{v_{12} \in G^{SV} \mid \forall v_{11} \in B : v_{12} I^{SV} v_{11}\}$ ,  $A' = B$ ,  $B' = A$ .

Коэффициент  $F$  показывает соответствие формируемого ресурса требованию иерархичности:

$$F = \max_{j=1}^J \left( \sum_{i=1}^{n_j} |A_i| \right), \quad (4)$$

где  $J$  — индексное множество цепочек,  $j \in J$  — номер цепочки,  $n_j$  — количество ФП в цепочке с номером  $j$ ,  $i$  — порядковый номер ФП в цепочке.

Смысл :  $\forall C = (A, B) : C \in \mathfrak{R}^{SV}(G^{SV}, M^{SV}, I^{SV})$ , должно входить в цепочку максимальной длины при  $|A| \rightarrow \max$ .

**Определение 5.** Под областью в решетке ФП понимается набор ФП, связанных отношением порядка с одним Наибольшим Общим Подпонятием (НОПП) и/или одним Наименьшим Общим Суперпонятием (НОСП). В роли НОПП может выступать наименьшее ФП в решетке, а в роли НОСП — вершинное ФП.

С учетом критерия (4) формирование  $\mathfrak{R}^{SV}(G^{SV}, M^{SV}, I^{SV})$  ведется по областям. С целью минимизации числа спорных ФП каждое следующее ФП в цепочке выбирается по принципу постепенного уменьшения  $|B|$  и максимизации количества общих признаков с потенциальным подпонятием при минимуме общих признаков с любым ФП, не входящим в цепочку.

В целях максимизации критерия (4) вводится частота  $Cnt(v_{11})$  встречаемости каждого  $v_{11} \in M^{SV}$  с различными  $v_{12} \in G^{SV}$ .

$Cnt(v_{11})$  используется для оценки информативности каждого признака из первоначально выявленных для  $\{v_{12}\}$  и равна числу соответствующих употреблений  $v_{11} \in M^{SV}$  в тексте.

---

**Алгоритм 4** Формирование цепочки в  $\mathfrak{R}^{SV}$  по масимуму критерия полезности.

---

**Вход:**  $P^C$  на выходе Алгоритма 3;

**Выход:**  $P_{Ch(j)}^C = \{(v_{12}, P_k^C) : (v_{12}, P_k^C) \in P^C | P_k^C \text{ — набор признаков для } v_{12}\};$

$P_{Rest}^C;$  // Подмножество исходного  $P^C$ , не вошедшее в  $P_{Ch(j)}^C$

$P_{Neigh(j)}^C \subset P_{Ch(j)}^C;$  // Соседние ФП для тех ФП, относительно которых  
// рассматривается отношение порядка

$P_{Ch(j)}^C := \emptyset;$  // Инициализация

$P_{Neigh(j)}^C := \emptyset;$

Выбрать  $(v_{12_{max}}, P_{max}^C)$  из  $P^C : |P_{max}^C| \rightarrow \max;$

$P^C := P^C \setminus \{(v_{12_{max}}, P_{max}^C)\};$

$P_{Ch(j)}^C := P_{Ch(j)}^C \cup \{(v_{12_{max}}, P_{max}^C)\};$

$P_{tmp}^C := P_{max}^C ;$

**ЦИКЛ**

Выбрать  $(v_{12}, P_k^C)$  из  $P^C : P_k^C \subset P_{tmp}^C$  и  $|P_{tmp}^C \cap P_k^C| =: Cr \rightarrow \max;$

**при**  $Cr = \emptyset$  **выход;**

$P_{tmp}^C := P_k^C;$

$P_{Ch(j)}^C := P_{Ch(j)}^C \cup \{(v_{12}, P_k^C)\};$

$P^C := P^C \setminus \{(v_{12}, P_k^C)\};$

Выбрать  $\{(v_{12_{Cr}}, P_{Cr}^C) | P_{Cr}^C \supseteq Cr\} =: P^{Cr} \subseteq P^C;$

$P_{Ch(j)}^C := P_{Ch(j)}^C \cup P^{Cr};$

$P_{Neigh(j)}^C := P_{Neigh(j)}^C \cup P^{Cr};$

$P^C := P^C \setminus P^{Cr};$

$P_{Rest}^C := P^C;$

---

**Алгоритм 5** Генерация множества цепочек для "соседних" ФП в  $\mathfrak{R}^{SV}$ .

---

**Вход:**  $P^C$  на выходе Алгоритма 3;

**Выход:**  $P_{Ch}^C = \{P_{Ch(j)}^C : P_{Ch(j)}^C \text{ — цепочка ФП,}$   
формируемая Алгоритмом 4};

$P_{Ch}^C := \emptyset;$  // Инициализация

**ЦИКЛ**

Сформировать  $P_{Ch(j)}^C, P_{Neigh(j)}^C \subset P_{Ch(j)}^C$  и  $P_{Rest}^C$   
Алгоритмом 4 на основе  $P^C$ ;

**при**  $|P_{Ch(j)}^C| \leq 1$  **выход**;

$P_{Ch}^C := P_{Ch}^C \cup \{P_{Ch(j)}^C\};$

$P^C := P_{Neigh(j)}^C \cup P_{Rest}^C;$

---

**Определение 6.** ФП  $C_2$  называется соседним по отношению к ФП  $C_1$  в решетке  $\mathfrak{R}$ , если они имеют НОСП, отличное от вершинного ФП в этой решетке.

---

**Алгоритм 6** Формирование частотного словаря для признаков из множества  $M^{SV}$ .

---

**Вход:**  $P^K = \{((v_{12}, v_{11}), v_{21}) \text{ на выходе Алгоритма 3}\}$ ;

**Выход:**  $P^{Cnt} = \{(v_{11}, Cnt(v_{11})) | v_{11} \in M^{SV}\}$ ;

$P^{Cnt} := \emptyset$ ;  $P^U := \emptyset$ ;    // Инициализация

    // Вспомогательный цикл — объединение списков пар  $(v_{12}, v_{11})$

**пока**  $P^K \neq \emptyset$

    Выбрать  $((v_{12}, v_{11}), v_{21})$  из  $P^K$ ;

$P^U := P^U \cup \{(v_{12}, v_{11})\}$ ;  $P^K := P^K \setminus \{((v_{12}, v_{11}), v_{21})\}$ ;

**пока**  $P^U \neq \emptyset$

$P^{Oth} := \emptyset$ ;

    Выбрать  $(v_{12}, v_{11})$  из  $P^U$ ;

$P^U := P^U \setminus \{(v_{12}, v_{11})\}$ ;

$Cnt(v_{11}) := 1$ ;

**пока**  $P^U \neq \emptyset$

        Выбрать  $(v'_{12}, v'_{11})$  из  $P^U$ ;

**если**  $v_{11} = x \odot \langle : \rangle \odot p_y$  **и**  $v'_{11} = x \odot \langle : \rangle \odot p_{y_1}$  **то**

$Cnt(v_{11}) := Cnt(v_{11}) + 1$ ;  $P^U := P^U \setminus \{(v'_{12}, v'_{11})\}$ ;

            //  $x$  — синтаксически главное слово,  $y$  — зависимое,  $\odot$  — операция конкатенации,

            //  $p_y$  — предлог между  $x$  и  $y$

**иначе**

$P^{Oth} := P^{Oth} \cup \{(v'_{12}, v'_{11})\}$ ;  $P^U := P^U \setminus \{(v'_{12}, v'_{11})\}$ ;

$P^{Cnt} := P^{Cnt} \cup \{(v_{11}, Cnt(v_{11}))\}$  ;

$P^U := P^{Oth}$ ;

---

**Алгоритм 7** Генерация формального контекста.

---

**Вход:**  $P^S = \{P_i^S = \{S_{ki} : S_{ki} \text{ — последовательность вида (1), } i = \overline{1, n(G)}\}\};$

**Выход:**  $K^{SV} = (G^{SV}, M^{SV}, I^{SV});$

Сформировать  $P^C$  и  $P^K = \{((v_{12}, v_{11}), v_{21}) : \{v_{11}, v_{12}\} \text{ есть РПЗ для } v_{21} \text{ в соответствии с Утверждением 2}\}$

Алгоритмом 3 на основе  $P^S$ ;

Сформировать  $P^{Cnt} = \{(v_{11}, Cnt(v_{11})) \mid v_{11} \in M^{SV}\}$

Алгоритмом 6 на основе  $P^K$ ;

$\Delta_F := 0$ ;

**пока**  $\Delta_F \leq 0$

$\Delta_F := |\Delta_F|$ ;

Сформировать  $P_{Ch}^C$  Алгоритмом 5 на основе  $P^C$ ;

Найти  $\max_{j=1}^J (|P_{Ch(j)}^C : P_{Ch(j)}^C \in P_{Ch}^C|) =: F_{tmp}$ ,

где  $J$  — индексное множество цепочек, (4);

$\Delta_F := \Delta_F - F_{tmp}$ ;

Найти  $v_{11_c} \in M^{SV} : (v_{11_c}, Cnt(v_{11_c})) \in P^{Cnt}$  и  $Cnt(v_{11_c})$  — максимально;

**для всех**  $(v_{12}, P_k^C) \in P^C$

$P_k^C := P_k^C \setminus \{v_{11_c}\}$ ;

$P^{Cnt} := P^{Cnt} \setminus \{(v_{11_c}, Cnt(v_{11_c}))\}$ ;

$K^{SV} := \bigcup_{j=1}^J P_{Ch(j)}^C$ ;

# Структура программного комплекса.

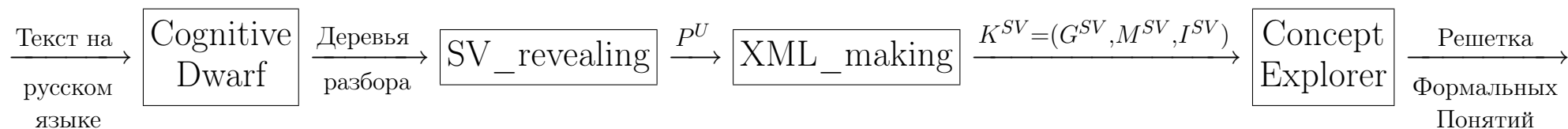


Рис. 1. Схема обмена данными между модулями программного комплекса

Здесь :

Cognitive Dwarf — программный пакет синтаксического анализа (ООО ”Когнитивные технологии“, <http://cs.isa.ru:10000/dwarf>);

SV\_revealing — модуль извлечения РПЗ из синтаксического дерева с формированием множества  $\{(v_{12}, v_{11}) : \{v_{11}, v_{12}\} \in SV^G, (3)\} := P^U$ ;

XML\_making — модуль генерации контекста  $K^{SV} = (G^{SV}, M^{SV}, I^{SV})$  по Алгоритму 7;

Concept Explorer — система анализа данных, реализующая методы Анализа Формальных Понятий, <http://conexp.sourceforge.net>.

# Пример выделения Лексической Функции на основе решетки Формальных Понятий для Расщепленных Предикатных Значений.

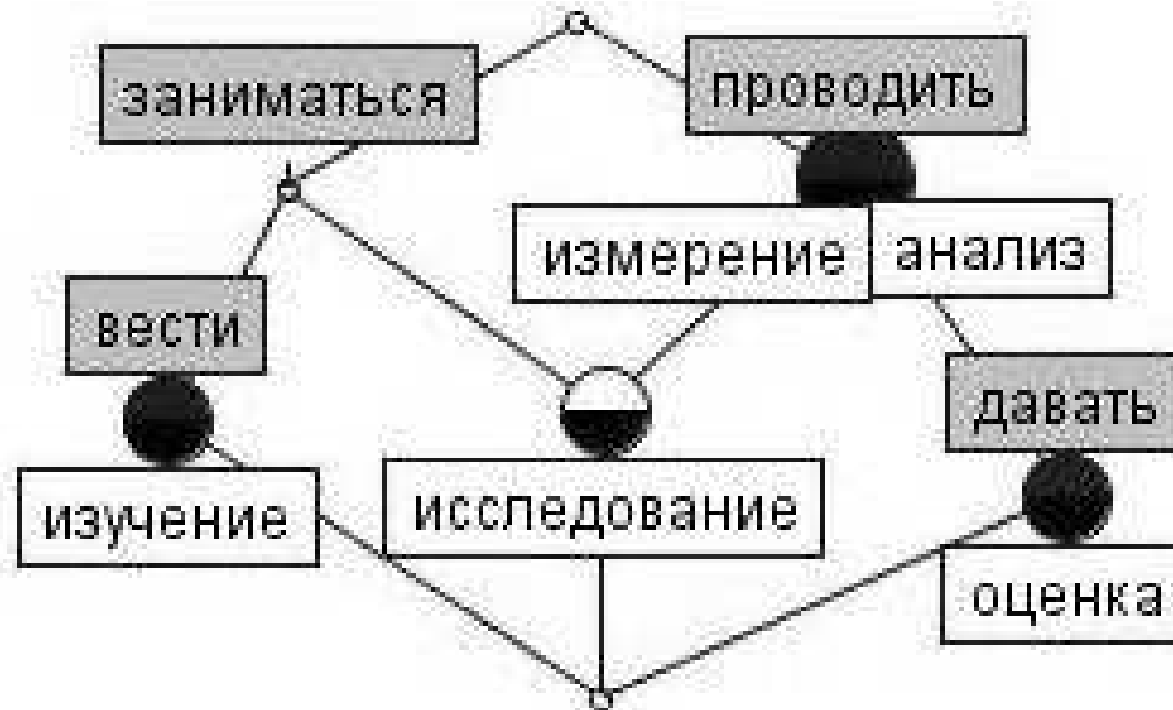


Рис. 2. РПЗ для ЛФ  $Oper_1$



# Выводы.

- Решетка Формальных Понятий для совокупности Расщепленных Предикатных Значений позволяет выделять группы Смысловых Отношений, из задаваемых Лексическими Функциями-параметрами. Значение критерия полезности решетки дает возможность делать выводы о сходстве ролевого состава ситуаций, обозначаемых словами-аргументами той или иной Лексической Функции в составе Расщепленных Предикатных Значений.
- При использовании последовательностей соподчиненных слов как основы выявления Расщепленных Значений отдельного исследования заслуживает включение наречий и прилагательных в состав указанных последовательностей. При этом за счет введения характеристик действий и дополнительных характеристик участников ситуаций могут быть выделены расщепления с оценочными адъюнктами, а также расщепления на основе синтаксической деривации.