

ПОСТРОЕНИЕ СЕМАНТИЧЕСКИХ СВЯЗЕЙ МЕЖДУ ЗНАЧЕНИЯМИ СЛОВ

Усталов Д.А.^{1,2}

¹⁾ Уральский федеральный университет имени первого Президента России
Б.Н. Ельцина, г. Екатеринбург, Россия

²⁾ Институт математики и механики им. Н.Н.Красовского УрО РАН,
г. Екатеринбург, Россия

E-mail: dmitry.ustalov@urfu.ru

TOWARDS WORD SENSES AND LINKS BETWEEN THEM

Ustalov D.A.^{1,2*}

¹⁾ Ural Federal University, Yekaterinburg, Russia

²⁾ Krasovskii Institute of Mathematics and Mechanics, Yekaterinburg, Russia

Annotation. In this study, we demonstrate an unsupervised approach for constructing a semantic network uniting word senses (or word concepts) rather than the coarse-grained concepts. The reported study was funded by RFBR (project no. 16-37-00354 мол_a) and by RFH (project no. 16-04-12019).

Метод «Извлечение, кластеризация, онтологизация», предложенный для построения лексической онтологии португальского языка [1], автоматически формирует понятия и отношения между ними при помощи методов обучения без учителя по материалам семантических словарей. На этапе онтологизации требуется тезаурус высокого качества для вычисления пути между понятиями. Доступность такого ресурса ограничена. На рис. 1 представлено развитие данного метода с целью выделения отдельных значений слов [2], объединения их в понятия (множества синонимов), и построения семантических отношений между значениями слов, а не между отдельными словами или понятиями.

При построении отношений происходит формирование иерархических контекстов понятий и их расширение путём линейного преобразования векторов слов с поиском ближайших соседей [3]. Затем осуществляется связывание значений слов в таких контекстах с использованием косинусной меры близости [4]. Это позволяет построить семантическую сеть слов, то есть ориентированный граф, вершины которого — значения слов («концепты слов»), а рёбра — отношения между ними.

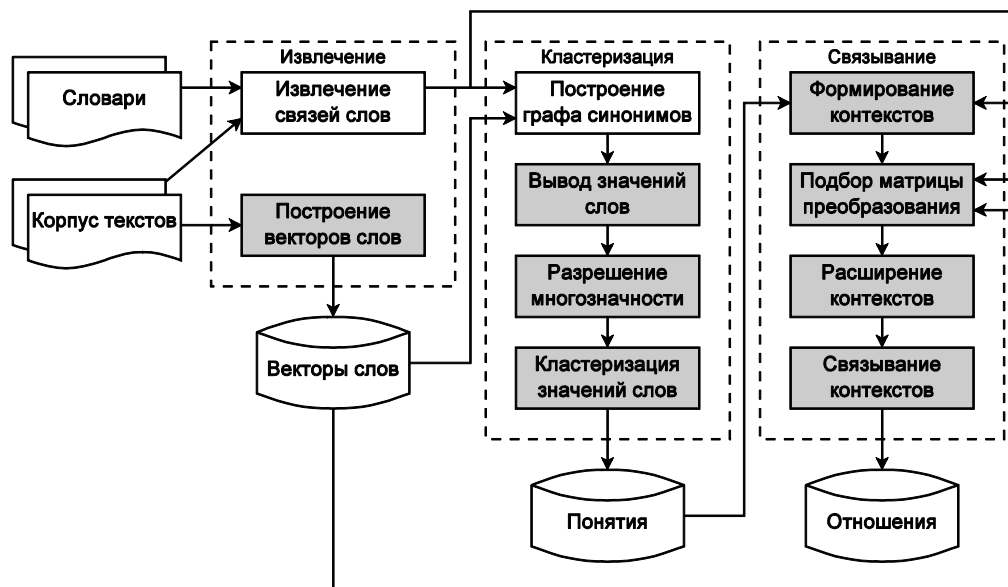


Рис. 1. Схема подхода «Извлечение, кластеризация, связывание»

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 16-37-00354 мол_a и при финансовой поддержке РГНФ в рамках научного проекта № 16-04-12019 «Интеграция тезаурусов RussNet и YARN».

1. Gonçalo Oliveira H., Gomes P., Language Resources and Evaluation, 48(2), 373 (2014).
2. Quillian M.R., Behavioral Science, 12(5), 410 (1967).
3. Ustalov D. et al., Proc. EACL 2017 – Short Papers (2017).
4. Faralli S. et al., Proc. The Semantic Web – ISWC 2016, Part II, 56 (2016).