

**Экспериментальные исследования алгоритмов выделения  
псевдооснов проекта *Snowball***

1. Изучить Интернет-проект «*Snowball*» URL: <https://snowballstem.org/>  
За основу использован взят Источники *Snowball: C-версии*  
библиотеки libstemmer:  
<https://snowballstem.org/download.html> (The C version of the libstemmer library).
2. Выбрать для исследования один из языков, реализованный в библиотеке, указанной в п.1., грамматику которого Вы представляете.
3. Загрузить и реализовать библиотеки (п.1.) для выбранного ЕЯ.
4. Составить два текстовых набора для выбранного ЕЯ, размерами допускающими делать выводы о статистической достоверности результатов:
  - А. Размеченный массив словоформ ЕЯ с выделенными псевдоосновами, для определения коэффициентов ***UI***, ***OI***, ***SW***.
  - В. Массив словоупотреблений из реальных текстов для вычисления коэффициентов ***MWC*** и ***ICF***.
5. Провести экспериментальные исследования с созданными текстовыми наборами с целью нахождения следующих характеристик:
  - а) коэффициент недостаточности отсечения ***UI***;
  - б) коэффициент избыточности отсечения ***OI***;
  - с) вес алгоритма выделения основ ***SW***;
  - д) Среднее количество слов, приведенных к одной основе ***MWC***;
  - е) Коэффициент сжатия индекса ***ICF***.
6. Оформить отчет (MSWord), включающий:
  - 1) Описание алгоритма выделения псевдооснов (по материалам сайта).
  - 2) Описание процедур вызова модулей и организации тестирования.
  - 3) Описание тестовых массивов.
  - 4) Описание итогов тестирования для ЕЯ с представлением и анализом результатов, включающих все характеристики, перечисленные в п.5.
  - 5) Результирующие выводы о качестве работы алгоритмов.