

Использование mystem

Вызов

Опции

Mystem — консольная программа. Ее следует запускать в терминале.

Программу можно использовать так:

Вызов

\$ mystem [опции] [входной файл] [выходной файл]

В результате mystem будет читать ввод и записывать результат своей работы в указанные файлы. Если имена файлов опущены, или если вместо имени указано "-", то будет использоваться стандартный ввод/вывод, соответственно.

Примеры

\$ mystem

на входе - стандартный ввод, на выходе -
стандартный вывод

\$ mystem input

на входе - файл input, на выходе -
стандартный вывод

\$ mystem - output

на входе - стандартный ввод, на выходе -
файл output

\$ mystem input output

на входе - файл input, на выходе -
файл output

Стандартный ввод и вывод, естественно, можно тоже перенаправлять:

\$ mystem <input>output

Опции

Опции надо указывать по правилам UNIX - до имен файлов, при этом можно склеивать, комбинировать, и т.д.

-c	<p>Копировать весь ввод на вывод. То есть, не только слова, но и межсловные промежутки.</p> <p>Опция необходима для возврата к полному представлению текста.</p> <p>В случае построчного вывода (когда задана опция <code>n</code>) межсловные промежутки вытягиваются в одну строку, символы перевода строки заменяются на <code>\r</code> и/или <code>\n</code>.</p> <p>Пробел для большей видимости заменяется на подчеркивание.</p> <p>Символ <code>\</code> заменяется на <code>\\</code>, подчеркивание на <code>_</code>.</p> <p>Таким образом можно однозначно восстановить исходный текст.</p>
-w	Печатать только словарные слова.
-l	Не печатать исходные словоформы, только леммы и граммы.
-i	Печатать грамматическую информацию, расшифровка ниже.
-g	Склеивать информацию словоформ при одной лемме (только при включенной опции <code>-i</code>).
-s	Печатать маркер конца предложения (только при включенной опции <code>-c</code>).
-e	Кодировка ввода/вывода. Возможные варианты: <code>cp866</code> , <code>cp1251</code> , <code>koi8-r</code> , <code>utf-8</code> (по умолчанию).
-d	Применить контекстное снятие омонимии.
--eng-gr	Печатать английские обозначения граммов.
--filter-gram	Строить разборы только с указанными граммами.
--fixlist	Использовать файл с пользовательским словарём.
--format	<p>Формат вывода.</p> <p>Возможные варианты: <code>text</code>, <code>xml</code>, <code>json</code>. Значение по умолчанию — <code>text</code>.</p>
--generate-all	Генерировать все возможные гипотезы для несловарных слов.
--weight	Печатать бесконтекстную вероятность леммы.