Худломер
Проект "Худломер" связан с задачей автоматической классификации стиля русскоязычных текстов. Автором были собраны и проанализированы 4 корпуса текстов, взятых из русской сети. Сюда вошли художественные произведения, публицистика, научные статьи и протоколы диалогов через ICQ и IRC. В результате были получены эмпирические кривые распределения длин слов в текстах, в зависимости от стиля. Эти кривые используются в качестве эталонов при классификации. On-line версия Худломера (на основе Perl-скрипта) может быть опробована здесь. Программа классифицирует стиль входного текста как: РАЗГОВОРНАЯ РЕЧЬ, ХУДЛО (худ.литература), ГАЗЕТНАЯ СТАТЬЯ или НАУЧНАЯ СТАТЬЯ.
На сайте есть статья с теоретическим описанием алгоритма классификации. Автор проекта широко известен как издатель литературного журнала DE-LIT-ZYNE, основатель сетевого конкурса ТЕНЕТА и энтузиаст русской сетературы.
Еще один проект автора, "Штампомер", связан с нахождением наиболее часто встречающихся фраз (штампов) в авторских текстах. К сожалению, этот проект в настоящее время заморожен.