

163(7) заседание Межпредметного семинара
состоится в среду 27 октября 2010 г. в 18:35 в аудитории 202НК

Статистический анализ литературных текстов

д.ф.-м.н. Орлов Юрий Николаевич

(ИПМ РАН, кафедра высшей математики)

Рассматриваются задачи классификации и идентификации литературных текстов на основе анализа статистических закономерностей буквенных распределений, т.е. вероятностей встречаемости букв и буквосочетаний. Тексты классифицируются по жанрам и авторам.

Основным инструментом анализа являются функции распределения законченных произведений по буквам и буквосочетаниям, а также выборочные буквенные распределения, построенные по фрагментам текстов. Для набора текстов, объем наименьшего из которых достаточен для того, чтобы считать буквенное распределение каждого из них квазистационарным, вычисляются расстояния между распределениями букв в пространстве суммируемых функций.

В докладе будут введены основные понятия статистического анализа нестационарных временных рядов, каковыми являются последовательности букв в книге. Будет описан также метод установления автора текста, использующий спектральный анализ оператора трансляций между двумя выборочными функциями распределения букв. Этот метод, в котором строится хаусдорфово множество для оператора трансляций, является на сегодня наиболее точным методом идентификации автора, позволяющий правильно определить 95% текстов.

Литература

1. Орлов Ю.Н., Осминин К.П. Определение жанра и автора литературного произведения статистическими методами // Прикладная информатика, 2010. Т. 26. № 2. С. 95-108.
2. Орлов Ю.Н., Осминин К.П. Нестационарные временные ряды: методы прогнозирования с примерами анализа финансовых и сырьевых рынков. – М.: Книжный дом «ЛИБРОКОМ», 2010. – 384 с.

Следующее заседание:

- 03.11.2010. "Изотопы в современных физических экспериментах - GERDA и другие низкофононовые эксперименты." Л.В.Инжечик (кафедра общей физики)

Межпредметный семинар может быть зачтён как **технический курс по выбору**.

Страничка Межпредметного семинара: <http://www.theorphys.fizteh.ru/mezhpr/> . На страничке регулярно выкладываются аннотации предстоящих докладов, а в некоторых случаях тезисы/конспекты/презентации/рефераты прошедших докладов. Там же можно подписаться на рассылку и обсудить тематику семинара в форуме.

Межпредметный семинар проходит еженедельно (кроме каникул, сессий, первых и последних недель семестра) **по средам в 18:35 в аудитории 202НК**.