# Дифференцированное измерение эмоциональности текстов с помощью алгоритма PMI-IR

Андрей Четвериков

# Зачем оценивать эмоции в тексте?

- Необходимы практичные и эффективные методы оценки эмоционального состояния
- Физиологические методы и самоотчеты не подходят
- Развитие affective computing требует методов оценки эмоций в тексте
- Компьютеры + интернет => много «живого» материала для анализа

# PMI-IR

Pointwise Mutual Information - Information Retrieval

- Автор метода – Peter D. Turney (2001)
- Первоначальная область применения:

поиск синонимов

$$PMI(слово, шкала) = \log_2\left(\frac{f(слово\ шкала)}{f(слово)\ f(шкала)}\right)$$

# Pro et contra

За:

1) Гигантский корпус текстов => высокая точность
2) Не требует «словаря эмоций»
3) Высокая скорость работы
4) Не требует предварительной экспертной оценки
5) Возможен выбор любых шкал оценки

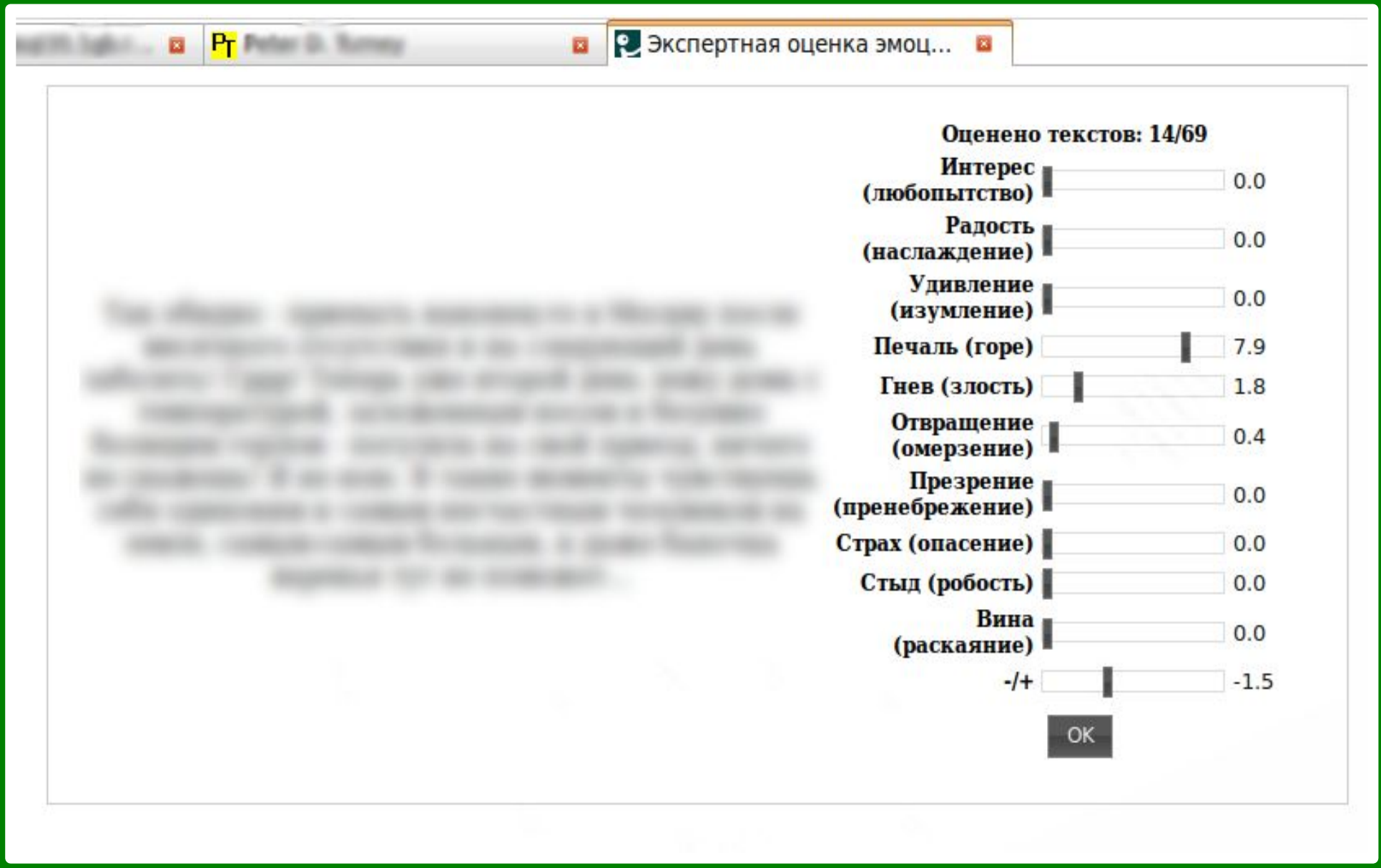
Против:

1) Анализ в основном «языка описания эмоций»
2) Не учитывает контекст (на данном этапе)
3) Поисковый спам

# Программа «ЭСКА»

1. Шкалы – 10 эмоций К. Изарда, по 2 слова на эмоцию.
   Пример: *радость*|*наслаждение* для эмоции «радость».
2. Обучение: 400 записей из LiveJournal (ЖЖ), около 11000 уникальных слов.
3. Проверка: 70 записей из LiveJournal, оцененных 4 экспертами по 11 шкалам (10 эмоций + валентность)

# Экспертная оценка




(тексты брались из общедоступных записей)

# Эксперты vs ЭСКА

## r Пирсона * 100

| Эмоции | Эксперты~ЭСКА | Эксперты | SE-07 PMI-IR | SE-07 Top |
|---|---|---|---|---|
| Интерес | 01 | 40 | | |
| Радость | 27 | 77 | 02 | 26 |
| Удивление | 26 | 54 | 08 | 16 |
| Горе | 26 | 60 | 12 | 41 |
| Гнев | 33 | 68 | 23 | 32 |
| Страх | 24 | 49 | 23 | 45 |
| Отвращение | 19 | 54 | 16 | 18 |
| Презрение | 20 | 38 | | |
| Стыд | 08 | 41 | | |
| Вина | 29 | 77 | | |
| Валентность | 39 | 84 | | 48 |

# Выводы

1. PMI-IR работает на русскоязычных текстах
2. Точность оценки увеличивается за счет использования композитных шкал
3. Есть большой простор для дальнейших разработок

# Спасибо за внимание

andrey@chetvericov.ru