



Определение ЛИЧНОСТНЫХ характеристик автора текста

Ефремова Наталья Эрнестовна
Грацианова Татьяна Юрьевна
Большакова Елена Игоревна

Содержание

- Виды анализа
- Способы определения пола, возраста, профессии
- Фоносемантический анализ
- Сравнительный анализ
- Коммерческие системы



Области применения



- ✓ Криминалистика
- ✓ Литературоведение
- ✓ Психиатрия
- ✓ Веб-аналитика
- ✓ Подбор кадров

Многие методы можно использовать также в «обратном направлении» - для оценки соответствия текста адресату

Виды анализа



- Анализ личности по семантике текста
 - явный анализ (что человек сообщает о себе)
«Я, мать Иванова Пети, говорила...»
 - неявный анализ (анализ содержания и стиля)
«Котенок был маленький и хороший»
- Дополнительный анализ
 - фоносемантический (оценка эмоциональной окраски)
 - сравнительный (похожесть на стиль других людей)

Явный анализ



Сопоставление с синтаксическими шаблонами

❖ Пол

Я был прописал по адресу ...

❖ Возраст

Шаблоны:

[наречие] «мне» [глагол] [наречие] <числительное> [« лет» / « года »]

□ («Когда мне было 16 лет»; «Мне уже 17!»; «Мне исполнилось двадцать два года»)

❖ Политические взгляды

❖ Профессия

Я работаю юристом ...

❖ Положение в обществе

Неявный анализ



- Вид анализа, зависящий только от структуры и содержания текста и не принимающий во внимание явные описания себя автором.
- Полезен в криминалистике и авторской экспертизе, где автор может умышленно не давать о себе информации, исказить ее или пытаться имитировать чужой стиль изложения
- Признаки предлагаются лингвистами на основе исследований.
Сложность реализации

Исследуемые признаки

- процентное соотношение частей речи в тексте
- слова-маркеры
- размер словарного запаса
- сложность и структура предложений
- наличие профессионализмов

Определение пола автора



- анализ процентного соотношения частей речи;
- наличие уменьшительно-ласкательных суффиксов у существительных и прилагательных;
- наличие наречий-маркеров, выражающих степень уверенности/неуверенности («*может быть*», «*по-видимому*», «*по-моему*»); ;
- наличие специфических названий цветов;
- расчлененные вопросы («*..., не так ли?*»);
- употребление вводных слов, особенно имеющих значение констатации: («*очевидно*», «*несомненно*», «*конечно*»);
- слова гиперболизированной экспрессивности («*жутко*», «*колоссально*»).

Анализ возраста



Формула читабельности Рудольфа Флеша, адаптированная для русского языка

$$K=206.836 - 65.14*W - 1.52*S$$

- K — оценка трудности текста
- S — средняя длина предложений в словах,
- W — среднее число слогов в слове

Средняя школа $K=60...70$

Высшее образование $K= 0...30$

Фоносемантический анализ



Теория А.П.Журавлева

позволяет узнать скрытый настрой и эмоции, которые автор внес в свой текст.

- Качественные характеристики для каждого звука русской речи по 25 шкалам
(большой-маленький, светлый-темный, радостный-печальный)
- Формулы расчета фоносемантической характеристики слова
 - характеристики входящих в слово звуков
 - место звука в слове
 - ударный ли звук
 - частотность звуков в слове

Пример «Родина-мать зовет!»:

мужественный, сильный, яркий, громкий

Сравнительный анализ



Сравнение характеристик исследуемого текста со значимыми характеристиками, описывающими авторский стиль известных писателей.

- Средняя длина предложения
- Процент частей речи
- Доля диалогов в тексте
- Активный словарный запас (количество уникальных словарных слов в тексте)
- Активный несловарный запас (количество уникальных выдуманных слов в тексте — сюда входят все имена, названия и неологизмы)
- Биграммы частей речи — частота употребления пар «существительное-глагол», «наречие-прилагательное» и т.д.
- Части речи на позициях в предложении.
- Частоты всех известных знаков препинания.

Коммерческие системы



- ✓ Проект ВААЛ (www.vaal.ru) Дымшиц М., Шалак В.
 - ✓ Набор DLL-библиотек, подключающихся к Word
 - ✓ Стоимость каждого модуля 300-400 Евро
 - ✓ фоносемантический анализ
 - ✓ Контент-анализ
 - ✓ Анализ «жанра»
- ✓ АнализПисем.RU (www.analizpisem.ru), Научно-исследовательского Центра «ЯМИР».
 - ✓ Настройка автора письма
 - ✓ Скрытые характеристики автора текста
 - ✓ «Дюжий бренд» - генерация и анализ бренда
- ПСЕВДОНИМ.RU (www.psevdonim.ru)
- Проект «Текстология» (www.textology.ru)

Задание



Подсчитайте индекс читабельности (можно использовать библиотеки)

- ✓ Своей аннотации к ВКР
- ✓ Первых абзацев введения
- ✓ Детской книжки или новостного сообщения

Считать можно вручную, можно использовать программные средства (напишите, как считали)

Проанализируйте получившиеся числа.