

## **Судебная лингвистика, определение авторства текста, анализ характера по написанному тексту.**

Как сообщает ряд специализированных изданий, специалисты по работе с данными смогли декодировать неструктурированные данные и точно спрогнозировать, например, в каких местах Афганистана будут происходить акты насилия. Теперь они также способны, проанализировав неструктурированные данные, определить личность автора документа. Похоже, что каждый из нас имеет "отпечаток письма" такой же уникальный, как и отпечатки пальцев.

По заявлениям судебных лингвистов, экспертов, которые расследуют происхождение текстов, если они получают образцы письма конкретного человека, они смогут определить с точностью до 95% авторство этого человека в других документах. Судебные эксперты вызывались в качестве свидетелей в громком процессе Пола Сеглиа, подавшего в суд на Марка Цукерберга, заявившего, что ему принадлежит половина Facebook. Они также использовались в качестве экспертов в расследовании убийств.

Хотя область судебной лингвистики развивалась до появления больших объемов данных, тот поток информации, который генерируется сейчас в Интернете, открывает новые возможности для автоматизации анализа. Компания, использующая эти возможности, утверждает, что она может обнаружить автора документа и узнать о нем очень многое - от пола, возраста и образования, до достоверности содержимого документа.

При этом некоторые виды анализа даже не требуют знакомства с конкретным человеком, чтобы определить авторство документа. Так, используя сотни тысяч доступных сообщений электронной почты от сотрудников компании Enron, группа компьютерных специалистов из Университета Согласия в Монреале проверила свой подход к кластеризации документов неизвестного происхождения. Их цель заключалась в определении документов, написанных одним человеком. Хотя они и отмечают необходимость в дальнейших исследованиях, они убеждены, что их метод кластеризации может быть использован при расследовании кибернетических преступлений, где все, что имеют в своем распоряжении следователи, - это огромные объемы подозрительной электронной почты, текстовых сообщений или другого письменного материала.

Хотя судебные лингвисты и говорят, что их анализ похож на сравнение ДНК человека для целей идентификации, они уверены, что могут находить те стилистические особенности в нашей прозе, которые делают нас уникальными, даже если писатель преднамеренно стремится сбить с толку того, кто пытается определить его или ее пол, или притворяется неграмотным, когда, фактически, имеет университетский диплом.

Большие объемы данных ускоряют развитие судебной лингвистики в качестве инструмента, который можно использовать в расследовании преступлений и в суде для выяснения истины. Судебная лингвистика становится еще одним методом анализа для улучшения современной жизни.

<http://life-24.com/post/6895#.Vf2aVpeyDX4>