

# СИСТЕМА ЛИНГВИСТИЧЕСКОГО АНАЛИЗА ДАННЫХ

В.Б. Соколов

*Учреждение образования «Белорусский государственный университет информатики и радиоэлектроники», Минск, Беларусь*

Предлагается разработать систему, позволяющую в обрабатываемых данных находить определенные закономерности, дающие возможность маркировать людей по группам интересов, а также выявлять скрытые закономерности о их намерениях.

Целью данной системы является оказание противодействия современным методам ведения информационной войны, а также нейтрализации ее последствий в первую очередь на высокотехнологичный сектор.

Задачей представляемой системы является выявление уникальных словарей групп, подвергшихся влиянию, а также, определение по нечеткому анализу степени внедрения этих словарей в социум для дальнейшей ретрансляции.

Реализация системы лингвистического анализа данных позволит выделить группы без семантического анализа, который без углубленной проработки группы, а также ее целей просто не возможен.

В случае явного подозрения на криминал можно использовать системы допроса, основанные на применении интерфейса мозг – компьютер, с демонстрацией ряда изображений с высокой частотой следования – так называемый визуальный анализатор человека на узнавание, срабатывающий быстрее, чем происходят процессы мышления. Ряд изображений составляется на основе словарей и предполагаемой семантики. Это позволит работать значительно эффективней.

Реализация системы лингвистического анализа базируется на использовании определенного и необходимого функционала, а именно - анализа контента социальных сетей, отрисовка карты распределения словарей по типу карт Кохонена, K-Means, K-Means++ и K-Medoids, Partitioning Around Medoids (PAM), а также буферных зон (частичное влияние).

Система лингвистического анализа данных представляет собой самообучающуюся систему, способную маркировать пользователей соцсетей в соответствии с их предпочтениями на основе анализа данных активности в соцсетях: текст (в т.ч. графические элементы – эмодзи), реакции (лайки, репосты и т. д.), производить оценку активности в социальных группах.

Сбор исходных данных для обработки системой лингвистического анализа осуществляется применением парсинга данных в соц. сетях

Для реализации системы лингвистического анализа используются различные интерфейсные системы, позволяющая работать с данными сторонним пользователям, различные веб интерфейсы, приложения.

Обучение системы лингвистического анализа данных предусматривает использование известных алгоритмов и систем, позволяющих взаимодействовать человеку и системе лингвистического анализа для обучения последней.

Подсистема хранения и обработки данных системы лингвистического анализа данных использует хорошо известные инфраструктурные решения для вычислений, хранения, сбора и обработки данных. Облачные и стационарные системы.

Методы реализации системы лингвистического анализа данных базируются на двухэтапном принципе. 1. Создание базы данных (Big Data). 2. Обработка текста VSM, LSA. Алгоритмическая реализация системы лингвистического анализа данных представляет собой классическую последовательность действий, состоящих из следующих этапов. 1. Сбор экспериментальных данных. 2. Выполнение кластеризации обозначенными методами, определение подходящего значения  $k$ . 3. Анализ полученных результатов. 4. В случае если остаются неясности с группами – изменение параметров кластеризации, до момента, когда число несоответствий будет укладываться в допустимый диапазон. 5. Проработка целевых групп.