

ПРОГРАММНЫЙ МОДУЛЬ ВЫЯВЛЕНИЯ ЛЖИ В ТЕКСТЕ

ВВЕДЕНИЕ

О важности изучения критериев лжи в письменном высказывании свидетельствует то, что маркеры лжи используются экспертами по безопасности, при приеме сотрудников на работу. В полиции, например, перед допросом подозреваемого просят сначала изложить на бумаге свою версию событий, далее ее читают, анализируют, а затем уже строят допрос, исходя из проведенного ими анализа.

I. Цель и задачи проекта

Цель проекта состоит в создании программного модуля анализа текста на выявление вероятной лжи. Для достижения указанной цели необходимо решить следующие задачи:

- изучить эффективные языковые маркеры лжи;
- ознакомиться с теоретическими положениями метода Саати;
- разработать базу данных, где будет храниться вся информация;
- разработать программный модуль анализа текста на основе объединения полученных данных о маркерах лжи и метода Саати.

II. Аналогии

Первый метод зародился в 1950-х годах, когда Удо Андеутш (немецкий психолог, который и ввел в оборот термин лингвистика лжи) разработал технику для анализа рассказов детей, содержащих обвинения взрослых в жестоком обращении. В широком смысле между ложными и правдивыми высказываниями есть существенные различия. Следовательно, лжецов можно распознать не по тому, что они говорят, а по тому, как они это говорят.

В 2012 году американские учёные из SAS Institute разработали своеобразный детектор лжи – программное обеспечение для проверки текстов на правдивость. Данная компания, известная своими разработками систем безопасности для банков и телефонных компаний, пред-

ставила программный продукт, который анализирует любой текст в электронном виде (например, содержание электронной почты) и находит в нём признаки лжи. Однако проверить валидность данной программы, к сожалению, невозможно, т.к. ее нет в открытом доступе.

В качестве аналога также можно считать детектор лжи (полиграф). Детектор лжи – техническое средство, используемое при проведении психофизиологических исследований для регистрации параметров дыхания, сердечно-сосудистой активности, электрического сопротивления кожи. Полиграф не читает мысли и не обнаруживает ложь, а всего лишь регистрирует физиологическую активность и изменения ее параметров. Он выявляет не ложь, а только возбуждение, которое с определенной долей вероятности может свидетельствовать о лжи.

III. Выводы

Проанализировав аналоги и различные научные статьи, были выделены четыре наиболее важных критерия лжи, которые следует использовать при разработке программного модуля анализа текста, например:

- редкое использование личных местоимений, чаще используют местоимения для наименования именно других людей;
- продуцирование максимально коротких предложений;
- употребление большого количества отрицаний;
- наличие общих слов или фраз, не несущих смысловой информации.

Следовательно, программный модуль будет сравнивать основные слова и грамматические конструкции с текстами, выбранными в качестве образцов или, так сказать, эталонов, в результате чего будет обнаруживаться вероятная неправда. Это может и не быть ложью – просто неуверенность или сомнение, но в результате будет показано, что текст подозрительный и требует повышенного внимания, потому что содержит вероятную ложь.

Бурая Валерия Владимировна, студентка кафедры информационных технологий автоматизированных систем БГУИР, valeriav@tut.by.

Научный руководитель: Герман Юлия Олеговна, ассистент кафедры информационных технологий автоматизированных систем БГУИР, jgerman@bsuir.by.