

РАЗРАБОТКА МЕТОДА ВЫЯВЛЕНИЯ ИЗМЕНЕНИЙ ХАРАКТЕРИСТИК РЕЧИ ПОД ВЛИЯНИЕМ ПРОФЕССИИ

Матвеева А.А.

(Университет ИТМО)

Научный руководитель — к.т.н. Махныткина О.В.

(Университет ИТМО)

Введение. В данной работе рассмотрены методы извлечения терминологии (Learning Domain-Specific Information Extraction Patterns from the Web, Syntax-Based Skill Extractor for Job Advertisements), и разработан метод выявления изменений характеристик речи специалистов различных профессий, включающий анализ и обработку терминологии.

Исследования выполнены за счет финансирования университета ИТМО в рамках НИР № 622282 «Разработка русскоязычного персонифицированного диалогового агента с динамической долгосрочной памятью».

Основная часть. Профессия играет в жизни человека огромную роль в связи с тем, что большую часть времени люди проводят на работе. Продолжительная работа в одной профессиональной сфере оказывает заметное влияние на такие аспекты личности человека, как установки, ценностные ориентации, мотивы деятельности, отношение к другим людям [1]. В том числе профессия оказывает сильное влияние на речь человека. В большинстве случаев речь спортсмена, актера, политика и программиста будет отличаться даже в неформальном общении. Это может выражаться как в разнице обширности и специфики словарного запаса, использовании разных синтаксических структур, формальности речи, наличием слов паразитов и профессионализмов.

Для проведения исследований были выбраны несколько датасетов. Датасет, содержащий письменную речь персоны с профессией с «повышенной речевой ответственностью», и один датасет с письменной речью персоны с профессией без «повышенной речевой ответственности». Датасет National Hockey League Interviews на английском языке, состоящий из диалогов профессиональных хоккеистов. И датасет US 2020 Presidential Election Speeches, который содержит стенограммы выступлений различных политиков США в контексте президентских выборов в США 2020 года.

Для каждого из датасетов были выявлены изменения характеристик письменной речи и с помощью наиболее эффективного из методов извлечения терминологии (Learning Domain-Specific Information Extraction Patterns from the Web [2], Syntax-Based Skill Extractor for Job Advertisements [3]) выявлена зависимость соотношения использования профессиональной лексики в речи от профессии.

Выводы. В данной работе был реализован метод выявления изменений характеристик речи под влиянием профессии. Для выявления изменений были использованы такие методы как: анализ синтаксического, лексического состава выражений. Помимо этого, с помощью метода извлечения терминологий, выявляется соотношение использования профессиональной лексики в речи человека. На основе анализа выявляются статистические оценки для речи представителей различных профессий.

Список использованных источников:

1. Савинова Марина Сергеевна Профессия как часть социального статуса и ее отражение в характеристиках речи // Вестник КГУ. 2009. №4. URL:

<https://cyberleninka.ru/article/n/professiya-kak-chast-sotsialnogo-statusa-i-ee-otrazhenie-vprosodicheskikh-harakteristikah-rechi> (дата обращения: 26.01.2023).

2. Siddharth Patwardhan and Ellen Riloff. 2006. Learning Domain-Specific Information Extraction Patterns from the Web. In Proceedings of the Workshop on Information Extraction Beyond the Document, pages 66–73, Sydney, Australia. Association for Computational Linguistics.

3. E. Smith, M. Braschler, A. Weiler and T. Haberthuer, "Syntax-Based Skill Extractor for Job Advertisements," 2019 6th Swiss Conference on Data Science (SDS), Bern, Switzerland, 2019, pp. 80-81, doi: 10.1109/SDS.2019.000-3.