

Разработка сервиса коротких мультязычных переводов на основе распределенной системы обработки аннотированной информации

К. Ф. Край, А. З. Энес

Институт информатики и проблем регионального управления –
филиал Кабардино-Балкарского научного центра Российской академии наук
360000, Россия, г. Нальчик, ул. И. Арманд, 37-а

Аннотация. Разработка сервиса коротких двуязычных переводов представляет собой актуальную задачу создания систем автоматического перевода текстов для средних и малых языков, для которых не существуют аннотированные базы данных, необходимые для обучения. В рамках данной работы реализуется система для автоматического перевода коротких текстов с одного языка на другой, особенностью которой является использование распределенной архитектуры для сбора, обработки, хранения и использования информации, предоставляемой пользователями системы на основе оригинальной модели мотивации, что позволяет значительно увеличить скорость создания обучающей выборки и качество обработки переводов.

Ключевые слова: машинное обучение, распределенная система, обработка естественного языка, сбор данных, веб-сервис, краудсорсинг, машинный перевод, искусственный интеллект, аннотирование информации, анализ текстов, перевод, текст, обработка информации

REFERENCES

1. Schultz T., Kirchhoff K. Multilingual speech processing. 2006. ISBN: 9780080457628
2. Bisani M., Ney H. Joint-sequence models for grapheme-to-phoneme conversion. *Speech Communication*. 2008. Vol. 50. No. 5. Pp. 434–451. DOI: 10.1016/j.specom.2008.01.002.hal-00499203
3. Kartseva E.Yu., Margaryan T.D., Gurova G.G. *Razvitie mashinnogo perevoda i ego mesto v professional'noy mezhkul'turnoy kommunikacii* [Development of machine translation and its place in professional intercultural communication]. *RUDN journal of language studies, semiotics and semantics*. 2016. No. 3. Pp. 155–164. EDN: WHBQCV. (In Russian)
Карцева Е. Ю., Маргарян Т. Д., Гурова Г. Г. Развитие машинного перевода и его место в профессиональной межкультурной коммуникации // Вестник Российского университета дружбы народов. Серия: Теория языка. Семиотика. Семантика. 2016. № 3. С. 155–164. EDN: WHBQCV
4. Corney J., Lynn A., Torres-Sanchez C. et al. Towards crowdsourcing translation tasks in library cataloguing, a pilot study. *4th IEEE International conference on digital ecosystems and technologies*. Dubai. United Arab Emirates. 2010. Pp. 572–577. DOI: 10.1109/DEST.2010.5610593
5. Jiménez-Crespo M.A. Crowdsourcing and online collaborative translations. 2017. Amsterdam/Philadelphia: John Benjamins Publishing Company. 320 p. ISBN: 1932–8036/2017BKR0009
6. Certificate of state registration of a computer program No. 2024610266 Russian Federation. *Programma dlya sozdaniya i podderzhki infrastruktury mashinnogo obucheniya mul'tiagentnoy nejrokognitivnoy sistemy prinjatiya resheniy ponimaniyu vyskazyvaniy* [Program for creating and supporting machine learning infrastructure for a multi-agent neurocognitive decision-making system for understanding statements]. No. 2023688479: application. 12/19/2023: publ. 01/09/2024 / A.Z. Enes, Z.V. Nagoev, K.F. Krai. Kabardino-Balkarian Scientific Center of the Russian Academy of Sciences. EDN: IRFXLN. (In Russian)
Свидетельство о государственной регистрации программы для ЭВМ № 2024610266 РФ. Программа для создания и поддержки инфраструктуры машинного обучения мультиагентной

нейрокогнитивной системы принятия решений пониманию высказываний: № 2023688479, заявл. 19.12.2023; опублик. 09.01.2024 / А. З. Энес, З. В. Нагоев, К. Ф. Край. Кабардино-Балкарский научный центр РАН. EDN: IRFXLN

7. Certificate of state registration of a computer program No. 2024610267 Russian Federation. *Programma dlya raspredelennoy platformy audio perevodov* [Program for a distributed audio translation platform]: No. 2023688483: application. 12/19/2023; publ. 01/09/2024 / A.Z. Enes, Z.V. Nagoev, K.F. Krai. Kabardino-Balkarian Scientific Center of the Russian Academy of Sciences. EDN: GGOBJH. (In Russian)

Свидетельство о государственной регистрации программы для ЭВМ № 2024610267 РФ. Программа для распределенной платформы аудио переводов: № 2023688483 заявл. 19.12.2023, опублик. 09.01.2024 / А. З. Энес, З. В. Нагоев, К. Ф. Край; Кабардино-Балкарский научный центр РАН. EDN: GGOBJH

8. McDonough D.J. Analyzing the crowdsourcing model and its impact on public perceptions of translation. *The Translator*. 2012. No. 18(2). Pp. 167–191. DOI: [10.1080/13556509.2012.10799507](https://doi.org/10.1080/13556509.2012.10799507)

9. Yasir M., Akbar Kh., Bashir M. The Impact of artificial intelligence on language translation: A Review. *IEEE Access*. 2024. No. 12. Pp. 25553–25579. DOI: [10.1109/ACCESS.2024.3366802](https://doi.org/10.1109/ACCESS.2024.3366802)

Информация об авторах

Край Карина Фaeзовна, мл. науч. сотр. отдела «Компьютерная лингвистика», Институт информатики и проблем регионального управления – филиал Кабардино-Балкарского научного центра РАН;

360004, Россия, г. Нальчик, ул. И. Арманд, 37-а;

kraykarina@mail.ru, ORCID: <https://orcid.org/0000-0002-6927-7361>, SPIN-код: 5967-0267

Энес Ахмед Зюльфикар, мл. науч. сотр. отдела «Компьютерная лингвистика», Институт информатики и проблем регионального управления – филиал Кабардино-Балкарского научного центра РАН;

360004, Россия, г. Нальчик, ул. И. Арманд, 37-а;

ahmedenes@mail.ru, ORCID: <https://orcid.org/0000-0003-3633-4910>, SPIN-код: 3643-1808