



НОВЫЕ ЛИНГВИСТИЧЕСКИЕ
ТЕХНОЛОГИИ

Опыт внедрения системы машинного перевода PROMT в бизнес-процессы АО «ОДК»



Александр Пивнык,
начальник отдела инженерного центра АО «ОДК»

Область сотрудничества

Работники АО «ОДК» регулярно сталкиваются в своей деятельности с информацией и источниками на иностранных языках. Чтобы вести активную научно-техническую деятельность и изучать опыт предприятий и научных центров по всему миру, наши конструкторы и инженеры должны владеть актуальной информацией, опубликованной в мировых научных изданиях, технической документации и отраслевой литературе.

Перевод специализированной лексики, применяемой в авиадвигателестроении, всегда вызывает определенные трудности.

Задачи

- Создание и внедрение решения по автоматическому переводу текстов и документов **любых форматов** с европейских и азиатских языков **непосредственно в корпоративной среде.**
- Обязательное обеспечение **полной конфиденциальности и безопасности** при работе с корпоративной информацией.

Решения

- с учетом внутренних требований АО «ОДК», настроена и внедрена нейронная система машинного перевода **PROMT Neural Translation Server (PNTS*)**

- специально для АО «ОДК» разработан **профиль «Авиадвигателестроение»** в языковых парах:
 - французский-русский / русский-французский
 - английский-русский / русский-английский

**PNTS – клиент-серверное решение для конфиденциального машинного перевода на основе нейронных сетей, натренированное на более чем 100 млн параллельных сегментов и обеспечивающее связный и качественный перевод для текстов и документов любого размера.*

Разработка профиля Авиадвигателестроение. Настройка

- Тексты АО «ОДК» не могли быть переданы в PROMT для настройки по причине их конфиденциальности.
- Специалисты PROMT самостоятельно нашли и подготовили релевантные данные для обучения в несколько этапов:
 1. Поиск и скачивание параллельных текстов в открытых источниках (сайты компаний-партнеров ОДК и других компаний в отрасли авиастроения).
 2. Сегментация по предложениям, токенизация.
 3. Прочистка (удаление дубликатов, пустых строк, неалфавитного мусора).
 4. Выравнивание по предложениям
- Всего были просмотрены сотни сайтов, выделены 34 условно-параллельных сайта, из которых позднее 6 сайтов были исключены после замечаний ОДК. Всего получено 144 898 параллельных строк на русском и английском языках, после фильтрации и прочистки объем данных, пригодных для обучения, значительно сократился, но и на нем удалось добиться значительного улучшения качества перевода.

Разработка профиля Авиадвигателестроение. Оценка качества

- После обучения моделей качество перевода было оценено на тематических предложениях, не участвовавших в обучении

Результаты автоматической оценки качества перевода по методике BLEU:

Направление	Размер тестовой части в предложениях	BLEU до настройки, %	BLEU после настройки, %
англо-русский	2000	29,27	58,44
русский-английский	2000	26,84	35,35
французский-русский	1000	16,86	27,83
русский-французский	1000	20,72	30,50

Разработка профиля Авиадвигателестроение. Оценка качества

→ После обучения моделей качество перевода было оценено на тематических предложениях, не участвовавших в обучении

Результаты экспертной (человеческой) оценки качества перевода:

Направление	Размер тестовой части в предложениях	Предложений с улучшением перевода после настройки	Предложений с ухудшением перевода после настройки	Предложений с переводом одинакового качества до и после настройки
англо-русский	100	52	15	33
русский-английский	100	36	14	50
французский-русский	200	95	6	99
русский-французский	200	94	19	87

Достигнутые результаты

- ✓ существенно облегчился процесс ознакомления с поступающей научно-технической информацией на иностранных языках;
- ✓ значительно увеличились объемы обработанной информации;
- ✓ решение задачи по использованию отечественного поставщика программного обеспечения в текущих реалиях времени, связанных с вопросами импортозамещения.

Планы по дальнейшему сотрудничеству

Проведение работ по усовершенствованию используемого решения PROMT:

- обогащение специализированной лексикой профиля «Авиадвигателестроение» для перевода с немецкого и китайского языков;
- достижение максимального сохранения исходного формата документа при переводе, минимальное постредактирование содержания;
- разработка решения Desktop с профилем «Авиадвигателестроение» для развертывания на «закрытых» ПК.



НОВЫЕ ЛИНГВИСТИЧЕСКИЕ
ТЕХНОЛОГИИ

**С PROMT
ВОЗМОЖНО МНОГОЕ!**

