

Синтезаторы речи
в производстве аудиокниг
на языках малых народов России

Михаил Пожидаев, Елена Теплых

26 января 2023 г.

Новая модель литературы

Это не замена привычных аудиокниг

Новая модель распространения

Принципиальное отличие в том, что читатель может выбрать голос аудиокниги по своему предпочтению. Это сильное конкурентное преимущество, создающее рынок аудиокниг «по требованию».

Вспомните похожие примеры:

- ▶ цифровая фотография вытеснила «мокрую», хотя не стала её полной заменой;
- ▶ электронные пианино существенно потеснили акустические, хотя также не могут быть их заменой.

Порядок работы

Фазы создания синтезированной аудиокниги

1. Очистка исходного материала.
2. Формирование рубрики.
3. *Аннотирование текста с целью преобразования письменной речи в устную.*
4. *Синтезирование аудиоданных.*
5. Упаковка книги.

Выделены элементы, требующие новой реализации в зависимости от языка.

Аннотация текста

Построение устной речи на основе письменной

Пример аннотированного текста:

Пройдя по дороге TTS{2 км}{ДВА КИЛОМЕТРА}, он подумал, что TTS{дороги}{ДОР+ОГИ} здесь прекрасные.

Порядок обработки:

1. Сегментация и токенизация текста.
2. Выделение фрагментов, требующих аннотирования.
3. Аннотирование выделенных фрагментов с учётом контекста.

Категории аннотирования

Трудоёмкая часть с ручной работой

Типы фрагментов для обработки:

- ▶ обработка и снятие омонимии;
- ▶ обработка сокращений, включая снятие их неопределённости и пр.

Чрезвычайно трудоёмкая работа!

Категории для обработки могут быть в каждом языке свои. Не существует точной нормы для тестирования. Подобранные алгоритмы для русского языка в большинстве случаев не могут быть применены для других языков.

Синтезирование аудиоданных

Машинное обучение для анализа записей

Как бы ни звучало странно, процедура непосредственного получения аудиосигнала требует умеренное количество знаний о языке. В основе современных алгоритмов лежат рекуррентные нейросети и модели на базе Трансформера, выполняющие работу следующим образом:

1. Производится чтение носителем языка некоторого эталонного текста.
2. Производится анализ и сопоставление полученной записи с исходным текстом. Алфавит и другие лингвистические особенности большой роли не играют.
3. Обученная модель способна генерировать новый сигнал для нового текста.

Но процедура вычислительно трудоёмкая!

Спасибо за внимание!

E-mail: msp@luwrain.org

Домашняя страница: <https://marigostra.ru>

Канал в Телеграм: @MarigostraRu