

[Wiki »](#)

## Рекомпиляция ASR модели

### Да кто такая эта ваша "рекомпиляция"?

**Рекомпиляция модели** - это процесс создания ASR модели на основе уже существующей, в которую будут добавлены пользовательские слова и словосочетания. Т.е. есть моделька, вроде работает, но вот имена криво распознает. Почему? Потому что в ней таких имен нет. Поэтому мы просто берем и добавляем отсутствующие слова (хотя не факт, что поможет).

### Откуда берутся "новые слова"?

Новые слова берутся из двух источников - телефонной книги и таблицы **extended\_words**, которой рулит ecss-clerk(автосекретарь). На основе телефонной книги строится корпус слов с перестановками, добавлением вариантов имен с сокращениями (которые берутся из таблицы aliases). Т.е. по итогу после рекомпиляции получается модель, в которой есть все имена из телефонной книги.

### Запуск рекомпиляции

#### - Ок, это всё очень интересно, но мне надо слова добавить в модель, чего делать?

Рассказываю. Во-первых, надо убедиться, что ecss-clerk живой (systemctl status ecss-clerk.service) и не жалуется в логах(/var/log/ecss/ecss-clerk/) на что-нибудь связанное с AsrManager и Phonebook. Во-вторых, проверить что HTTP сервер с Kaldi тоже активен (systemctl status kaldi-ru.service). В-третьих, если есть необходимость, добавить своих алиасов и доп. слов, которые будут включены в корпус новой скомпилированной модели. Почитать про последнее можно [тут](#)

Главное, чтобы был верно указан url для импорта книги. Он указывается при установке ecss-clerk deb пакета.

Теперь можно запустить сам процесс:

```
curl localhost:9010/model/recompile
```

```
{
  "msg": "Recompile process start sucessfully"
}
```

#### - Круто. А мониторить это как?

С этим все не очень радужно. После успешной/неуспешной компиляции уведомлении, конечно, доходит до ecss-clerk, но искать результат в его логах тяжело и долго.

Поэтому проще следить за самим HTTP сервером Kaldi. Во-первых все запросы на его имя можно глянуть в journalctl:

```
journalctl -u kaldi-ru.service
```

Лог компиляции можно глянуть в /var/log/kaldi/model/build\_model.log. Если там пусто, смотрим в journalctl и ищем ошибку.

После успешной компиляции kaldi-ru поднимет новый инстанс с tcp сервером Kaldi, отправит новый порт ecss-clerk и остановит старый.

### Ручная настройка

#### - Ну рекомпилировал я, а распознается все равно какая-то хрень.

Бывает. Добавление слов в модель не всегда может пройти гладко, т.к. сам процесс это не просто запись в какой-то файл новых слов (хотя в том числе это тоже :)).

Чтобы бороться такие проблемы была добавлена ручная настройка через таблицу **hard\_aliases**

Суть: таблица используется по сути как структура "ключ-значение": Ключ - "хрень", которая распознается при произношении имени, значение - полное имя из телефонной книги. После добавления такого жесткого алиаса и нахождении его в результате распознавания (причем именно его, а не как подстроки), будет производится поиск по полному имени, указанному для жесткого алиаса.