

# TRX SŁOWIK

## INSTRUKCJA OBSŁUGI

System firmy TRX zamieniający słowo mówione na słowo pisane.

## 1. WPROWADZENIE

Aplikacja **Słowik** służy do transkrypcji, czyli automatycznej zamiany mowy zawartej w nagraniach na rejestratorach TRX na tekst. Możliwy jest także odsłuch nagrań i ich archiwizacja.

## PRYWATNOŚĆ I INFORMACJE WRAŻLIWE

Aplikacja działa całkowicie offline, nie korzysta z zewnętrznych zasobów sieciowych (np. obliczeń w chmurze). Informacje wrażliwe potencjalnie zawarte w nagraniach nie są przesyłane na zewnętrzne serwery. Pliki transkrypcji zapisywane są na lokalnym folderze w aplikacji i podlegają 30-dniowej retencji. Pliki tymczasowe używane podczas przetwarzania są kasowane na bieżąco.

W związku z powyższym, aplikacja ma duży rozmiar, gdyż zawiera w całości wbudowany lokalny model Automatic Speech Recognition do przetwarzania nagrań. Więcej informacji o nim w dalszej części instrukcji.

## 2. WYMAGANIA SYSTEMOWE

- **Rejestrator:**
  - oprogramowanie w wersji 2.2.4.12/05.12.2025 lub nowsze,
  - włączona usługa TRX API.
  - aktywna licencja na używanie aplikacji *Słowik*.
- **System operacyjny:** Windows 10 lub Windows 11.
- **Pamięć RAM:** minimum **16 GB** (aplikacja korzysta z ASR *parakeet-tdt-0.6b-v3* firmy *nVidia*).
- **Karta graficzna:** nie jest wymagana do poprawnej pracy aplikacji. Zastosowanie karty nVidia z minimum 4GB VRAM pozwala osiągnąć maksymalną wydajność przetwarzania. W przypadku braku zgodnej karty, aplikacja użyje procesora, co wydłuży czas przetwarzania.
- **Dysk:** ok. 6,5 GB wolnego miejsca na pliki aplikacji.

## 3. WYDAJNOŚĆ PRZETWARZANIA

Wydajność przetwarzania zależy bezpośrednio od zastosowanego sprzętu (procesor i ewentualnie karta graficzna). Testowy zbiór danych składał się ze 154 nagrań o łącznej długości 6:15:44 i średniej długości 2:26. Wydajność 100% oznacza, iż przetworzenie nagrań zajmuje dokładnie tyle czasu ile te nagrania trwają.

Procesor	Karta graficzna	czas przetwarzania [hh:mm:ss]	Wydajność [%]
i7-8700	-	49:21	760 (x7.6)
i7-8700	nVidia GTX1650 4GB	22:32	1670 (x16.7)
Ryzen 9 5950X	-	33:21	1130 (x11.3)
Ryzen 9 5950X	nVidia RTX3070Ti 8GB	9:49	3830 (x38.3)

Wyniki testów pokazują, że nawet tania karta graficzna daje lepsze rezultaty niż topowy procesor pracujący samodzielnie.

## 4. INSTALACJA PROGRAMU, STRUKTURA APLIKACJI

Instalacja programu odbywa się standardowo. Pliki programu umieszczane są typowo w folderze **C:\Program Files (x86)\Słowik**.

Podfolder **logs** zawiera pliki logów aplikacji. Pliki tymczasowe procesu transkrypcji są zapisywane w katalogu **files\_temporary**, natomiast wszystkie pliki wynikowe transkrypcji zapisane są w katalogu **files\_transcriptions** i mają format **recording\_{adres\_IP\_rejestratora}\_{indeks\_nagrania}.txt**. Każde uruchomienie aplikacji pozostawia tylko pliki transkrypcji utworzone w ciągu ostatnich 30 dni.

Deinstalacja aplikacji powoduje skasowanie wszystkich plików aplikacji oraz wszystkich zapisanych do tej pory plików transkrypcji.

## 5. LICENCJONOWANIE

Do poprawnej pracy systemu transkrypcji wymagana jest licencja na rejestratorze na używanie aplikacji *Słowik*. Brak takowej spowoduje pojawienie się błędu autoryzacji przy próbie uruchomienia aplikacji.

Pojedyncza licencja obejmuje jedną sesję połączenia pomiędzy aplikacją i urządzeniem. Licencja jest odejmowana z puli licencji na urządzeniu od momentu nawiązania połączenia z rejestratorem do momentu zamknięcia aplikacji albo zamknięcia rejestratora. Czyli na jednej licencji, w danym momencie z rejestratorem może pracować jedna instancja aplikacji.

## 6A. KORZYSTANIE Z APLIKACJI

Głównym plikiem aplikacji jest *Słowik.exe*. Po uruchomieniu pojawi się okno startowe. Użytkownik ma możliwość zdefiniowania adresu IP rejestratora, z którego będą analizowane nagrania oraz podania danych logowania do tego rejestratora.

### URUCHOMIENIE SYSTEMU

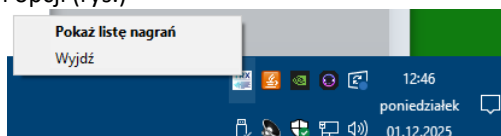
System transkrypcji uruchomiany jest przyciskiem "**Uruchom interfejs WWW**".

Uruchomienie może potrwać od kilkunastu do kilkudziesięciu sekund. W tym czasie aplikacja ładuje do pamięci RAM duże biblioteki oraz model ASR. Postęp uruchamiania systemu transkrypcji będą pokazywały animowane kropki w dzienniku zdarzeń.

Po pomyślnym załadowaniu, aplikacja automatycznie zminimalizuje się do *zasobnika systemowego* (obok zegara), a w domyślnej przeglądarce internetowej otworzy się interfejs sterowania dostępny pod adresem url: <http://127.0.0.1:31337/>. Tam odbywa się cała obsługa procesu transkrypcji.

### ZAMKNIĘCIE APLIKACJI

Zamknięcie aplikacji (i tym samym zwolnienie licencji na rejestratorze dla innej aplikacji *Słowik*) możliwe jest tylko przy pomocy ikony umieszczonej w zasobniku. Kliknięcie prawym klawiszem myszy na ikonie aplikacji pokaże listę dostępnych opcji (rys.)



Jednocześnie aplikacja po uruchomieniu serwera monitoruje stan połączenia z rejestratorem. Gdy rejestrator stanie się niedostępny przez dłuższy okres czasu, aplikacja jest automatycznie zamykana.

### PRZYWRACANIE OKNA Z LISTĄ NAGRAŃ

Kliknięcie lewym klawiszem myszy na ikonie aplikacji otwiera okno przeglądarki z listą nagrań odpowiadającą ostatniemu wyborowi zakresu czasu.

## 6B. OBSŁUGA INTERFEJSU WWW

### A. LISTA NAGRAŃ

W górnej części okna pokazana jest informacja o urządzeniu, z którego pobierane będą nagrania – adres IP oraz jakość łączności z rejestratorem. Gdy pojawiają się problemy z łącznością – prezentowana jakość spada. Gdy spadnie do 0 – rejestrator uznawany jest za niedostępny i aplikacja jest automatycznie zamykana.

Przyciski wyboru zakresu czasowego nagrań pozwalają wybrać okres czasu, z którego zostaną pobrane nagrania z rejestratora. Domyślnie pobierana jest ostatnia godzina nagrań. **Godzina** – oznacza pobranie nagrań z ostatniej godziny (wg czasu komputera na którym uruchomiona jest aplikacja). Analogicznie działają pozostałe opcje (ostatnie **8 godzin**, itd.).

Jeżeli istnieje potrzeba precyzyjnego wyboru zakresu dat i czasów nagrań – można to zrobić przy użyciu przycisku **Własny**. Data końcowa musi być późniejsza niż początkowa. Lista nagrań zostanie odświeżona po

naciśnięciu przycisku **Pokaż**. Można ustalić zakres czasowy o długości nieprzekraczającej 90 dni. Nie ma ograniczeń na daty początku i końca nagrania.

Główną część okna zajmuje tabela przedstawiająca informacje o nagraniach ściągniętych z rejestratora. W pierwszym wyróżnionym na zielono wierszu (nagłówek) listy znajdują się informacje opisujące parametry nagrań oraz dostępne funkcjonalności. Drugi wyróżniony wiersz zawiera elementy pozwalające kontrolować zawartość listy (filtracja / sortowanie).

Nagrania prezentowane są w sukcesywnych wierszach listy.

Pod listą dostępne są przyciski do kontroli procesu transkrypcji wszystkich nagrań widocznych na liście (filtrowanie jest uwzględnione).

Poniżej przedstawione są podstawowe statystyki dla nagrań: ich ilość oraz długości: sumaryczna i średnia.

Listę nagrań wieńczy informacja o użytym modelu językowym i jego licencji.

## B. OPERACJE NA LIŚCIE NAGRAŃ

### SORTOWANIE LISTY NAGRAŃ

Listę nagrań można sortować wg indeksów nagrań lub ich długości. Kierunek sortowania wybierany jest przy pomocy odpowiedniej strzałki (▼ / ▲) w drugim wierszu listy.

### FILTROWANIE LISTY NAGRAŃ

Możliwe jest filtrowanie listy rozmów wg kierunku rozmowy, numerów telefonów i nazwie kanału. W ten sposób łatwo dotrzeć do interesującej użytkownika rozmowy. Na liście przedstawione są nagrania spełniające wszystkie warunki filtrowania.

## C. OPERACJE NA NAGRANIACH

### ODSŁUCH NAGRAŃ

Odtworzenie nagrania możliwe jest po kliknięciu w **Indeks** nagrania (**niebieski numer**). Wówczas na dole strony pojawi się wbudowany odtwarzacz multimedialny. Odsłuch nagrania można zatrzymać oraz można poruszać się skokowo (co 5 s) w jego obrębie.

W trakcie odsłuchu wybrane nagranie posiada wyróżnione zielonym podświetleniem pole indeksu.

### TRANSKRYPCJA (ZAMIANA MOWY NA TEKST)

Proces transkrypcji odbywa się 100% offline (lokalnie) na komputerze użytkownika aplikacji.

Możliwe jest dodanie znaczników czasu do pliku tekstu (**Dodaj znaczniki czasowe**). Można także ustawić rzeczywisty czas znaczników w nagraniu (**Czas rzeczywisty**). Opcje te należy ustawić przed rozpoczęciem procesu transkrypcji.

### TRANSKRYPCJA POJEDYNCZEGO NAGRANIA

W kolumnie "Transkrypcja" należy przycisnąć przycisk     .

Wygląd przycisku będzie zmieniał się w zależności od statusu nagrania.






Wygląd przycisku	Status nagrania
    	gotowe do przetwarzania
    	w kolejce do przetwarzania
    	jest właśnie przetwarzane (animacja i podświetlenie).
    	jest już przetworzone. Plik transkrypcji jest gotowy do pobrania

## TRANSKRYPCJA WIELU NAGRAŃ



Nie ma konieczności czekania na zakończenie przetwarzania poprzedniego nagrania przed dodaniem kolejnego. Można po prostu zaznaczyć kliknięciem wybrane nagrania z listy i poczekać, aż system przetworzy je wszystkie.

W niektórych scenariuszach przydatne może być kliknięcie przycisku *Transkrybuj wszystkie* znajdującego się na spodzie listy nagrań. Wszystkie widoczne na liście nagrania (np. po odfiltrowaniu) zostaną dodane do kolejki przetwarzania.

## POBRANIE PLIKU Z TRANSKRYPCJĄ NAGRANIA

Po zakończonym procesie transkrypcji, należy kliknąć przycisk     . Pojawi się okno dialogowe przeglądarki pozwalające pobrać plik.



## KASOWANIE PLIKU TRANSKRYPCJI

Plik transkrypcji można skasować z pamięci aplikacji przy pomocy przycisku  . Proces transkrypcji można następnie powtórzyć zmieniając parametry operacji.

## ZATRZYMYWANIE PROCESU TRANSKRYPCJI WIELU NAGRAŃ

Przycisk *Zatrzymaj kolejkę* anuluje wszystkie oczekujące zadania przetwarzania (nie przerywa jednak aktualnego procesu przetwarzania nagrania).

## POBRANIE NAGRANIA Z REJESTRATORA

Aby pobrać nagranie w jego natywnym źródłowym formacie należy kliknąć przycisk   (w kolumnie **Zapisz do wav**). Pojawi się okno dialogowe przeglądarki pozwalające pobrać plik.

## 9. ROZWIĄZYWANIE PROBLEMÓW

"Błąd autoryzacji" - podane dane logowania nie są poprawne albo brakuje licencji na rejestratorze.

„Wystąpił nieoczekiwany błąd: timed out” - adres IP jest niepoprawny albo problemy z połączeniem sieciowym.

Ostrzeżenie o RAM: Jeśli Twój komputer ma mniej niż 16GB RAM, aplikacja wyświetli ostrzeżenie. Transkrypcja będzie działać, ale jej wydajność znacznie spadnie.

Brak ikony aplikacji w zasobniku po uruchomieniu serwera: Windows 11 domyślnie ukrywa ikony. Kliknij strzałkę '^' na pasku zadań, aby ją znaleźć.

## 10. INFORMACJE O UŻYTYM MODELU ASR

W projekcie zastosowano model rozpoznawania mowy Parakeet-TDT 0.6B v3, opracowany i udostępniony przez firmę NVIDIA na licencji CC BY-4.0.

- Wydajna architektura w porównaniu z innymi modelami ASR (zwłaszcza z wykorzystaniem GPU).
- Składa się z ok. 600 mln wytrenowanych parametrów sieci neuronowej.
- Zwraca znaczniki czasowe na poziomie pojedynczych jednostek mowy.
- Obsługuje 25 języków europejskich (m.in. polski, angielski, niemiecki, francuski, hiszpański, czeski, ukraiński). Język rozpoznaje automatycznie.
- Poprawny ortograficznie (interpunkcja i wielkie litery).
- Licencja CC-BY-4.0 (model otwarty i bezpłatny).

## SPIS TREŚCI

1. Wprowadzenie .....	1
Prywatność i informacje wrażliwe .....	1
2. Wymagania systemowe .....	1
<b>3. Wydajność przetwarzania .....</b>	<b>1</b>
4. Instalacja programu, struktura aplikacji .....	1
5. Licencjonowanie .....	2
6a. Korzystanie z aplikacji .....	2
Uruchomienie systemu .....	2
Zamknięcie aplikacji .....	2
Przywracanie okna z listą nagrań .....	2
6b. Obsługa Interfejsu WWW .....	2
A. Lista nagrań .....	2
B. Operacje na liście nagrań .....	3
C. Operacje na nagraniach .....	3
9. Rozwiązywanie problemów .....	4
10. Informacje o użytym modelu ASR .....	4