

HTK – czego NIE robić

Bartłomiej Chojnacki

Wstęp

Niniejszy poradnik proponuję jako uzupełnienie do istniejącego już świetnego dokumentu Adriana Sekuły „Poradnik HTK”, gdyby nie on zapewne do tej pory grzebałbym się z tym systemem, tymczasem w trakcie moich osobistych wojen z tym toolkitem zwiedziłem sporo zakamarków Internetu oraz linijek kodu źródłowego HTK w poszukiwaniu wyjaśnień do otrzymanych błędów, proponuję więc garść porad i uwag odnośnie istniejącego poradnika, które zapewne pozwolą wielu z Was uniknąć dużych błędów przy pracy z tym systemem i zminimalizować ryzyko konieczności (czasem wielokrotnego) zaczynania od nowa.

Porady ogólne

- każdy plik, który tworzycie, trzeba pilnować żeby zaczynał się od razu tekstem z komendami, a kończył zawsze enterem, który oznacza dla HTK koniec pliku (znalezione w kodzie źródłowym)
- foldery, które będą po drodze do Waszego pliku z projektem, nie mogą mieć spacji w nazwie, inaczej prędzej czy później coś się zepsuje, czyli jeśli będziecie to trzymać w folderze „technologia mowy” niech to będzie „technologiamowy”.
- unikajcie w nazwach plików, słów, czegokolwiek, polskich liter (co oczywiste) i cyferek (co już mniej oczywiste ale prawie tak samo komplikujące). Jest jakaś komenda która pozwala htk w danym miejscu rozpoznać cyferkę jako tekst nie liczbę, ale różnie z tym bywa, po prostu – nie. Na którymś etapie, jeśli macie np. w słowniku coś z cyferkami albo gdzieś w gramie, htk będzie próbować rozpoznać Wasze cyferki jako numery próbek albo czas, co nie skończy się dobrze.
- uważać przy przepisywaniu komend, czasem małe L wygląda identycznie jak duże I
- **nie kopiować nic ctrl+c / ctrl+v bezpośrednio z pdfa!** Bo po prostu często nie działa... przepisujcie komendy ręcznie. Najlepiej niczego nie kopiować bo może to wywołać błędy, kusi skopiowanie kodu do proto ale lepiej dogadać się z kolegą obok i we dwie osoby to przepisać w 5 minut niż potem się zastanawiać co to za błąd i jak to naprawić.
- przypomnieć sobie podstawy obsługi konsoli cmd, wygooglować podstawowe komendy, sprawdzać czy jesteśmy aktualnie w tej lokalizacji którą chcemy zanim zacznie się panikować że „czegoś mi nie znajduje”
- średnie rozpoznanie u studentów to ~40% dla słów, nie martwcie się więc że to zły wynik

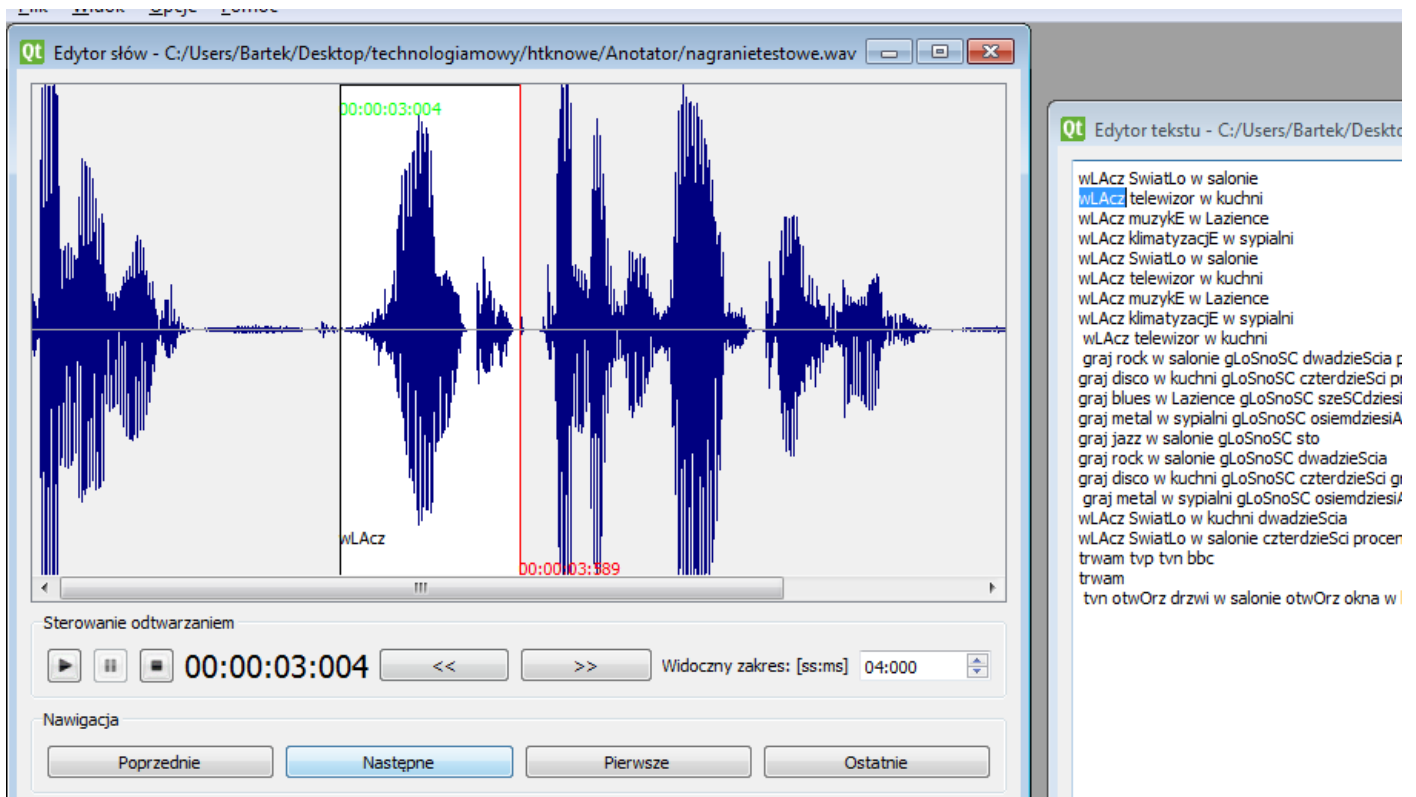
Gramatyka i słownik

Tutaj mogę mieć parę porad.

- konstruuując gramatykę, szczególnie zdania, które HTK ma rozpoznawać, unikajcie długich konstrukcji jeśli nie chcecie na koniec mieć wszędzie 0% rozpoznania zdań, przy czym skomplikowane zdanie to takie powyżej trzech słów. Słowa opcjonalne również mocno mieszają
- starajcie się mieć część gramatyki stałej, np. zaczynać tym samym słowem, to znacząco zwiększy Wasze końcowe rozpoznanie
- już od tego etapu trzeba się pilnować żeby **wszędzie stosować ten sam, htk-friendly zapis słów**, z tymi dużymi literami jako polskie itd. Szczególnie pilnować się np. odmiany przez przypadki
- transkrypcja fonetyczna, robi się to wcześniej i zazwyczaj większość osób nie przykłada się do tego, a tymczasem warto to zrobić dobrze na początku, a nie za pomocą losowych zestawień znaków bo wbrew pozorom to ma dość duże znaczenie. Polecam alfabet Corpora Grocholewski, chyba najprościej
- uważać z tą ciszą, bo różnie bywa, można spróbować to zamodelować, ale potem koniecznie trzeba pamiętać o traktowaniu tego jako normalny fonem („sp” i „sil”), czyli musi występować w anotacji na poziomie fonemów, w reestymacjach i w liście fonemów. Poza tym program do anotacji na poziomie fonemów różnie do tego podchodzi, w moim wypadku nie wykrył żadnego sp i tylko dwa sil co oczywiście powodowało błąd (brak fonemu sp) i warningi przy reestymacji = nikłe rozpoznanie. Musiałem ręcznie wyszukiwać ciszę i dodawać przy anotacji na poziomie fonemów.

Anotacja

Dość trudny etap, trzeba być bardzo ostrożnym. Klucz do sukcesu to dobrze przygotowany plik z zapisem nagrania – żadnych kropek, znaków specjalnych, cyferek, same słowa w zapisie w stylu dict i gram.



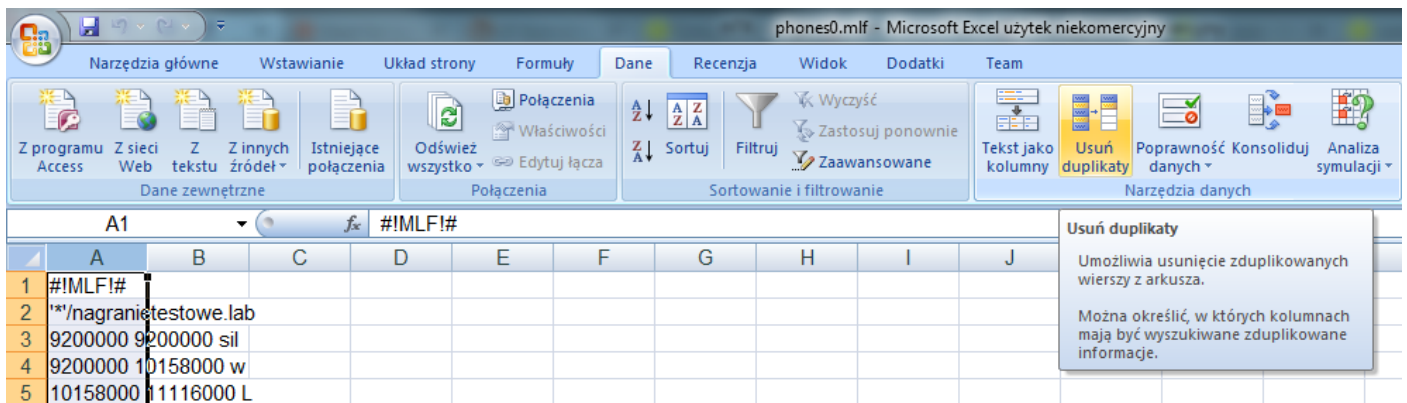
W takim układzie wszystko powinno pójść w miarę sprawnie. Niemniej trzeba pamiętać:

- forma zapisu słów koniecznie jak ta z dicta i grama
- częste zapisywanie, nie tylko samego projektu, ale też eksport tego pliku .mlf który ma być wynikiem anotacji, program bardzo lubi się wysypać, a jak ktoś ma dłuższe nagranie to zawieszenie programu może boleć
- bardzo ostrożnie zmieniać słowa, koniecznie za pomocą tych guzików! Tj. zawsze po zaznaczeniu fragmentu dla danego słowa klikamy „następne” i sprawdzamy czy przeskoczyło na kolejne słowo i to właściwe, tutaj program też lubi się wysypywać i pomijać słowa, trzeba wtedy cofnąć się do ostatniego, dobrze zaznaczonego i robić dalej, zaznaczając te słowa które są podświetlone na liście

Anotacja na poziomie fonemów i lista fonemów

Tutaj w poradniku jest pewne niedopowiedzenie, mianowicie lista fonemów w kolejności takiej jak w nagraniu nie jest „może potrzebna”, tylko potrzebna i trzeba ją zrobić. Sam poradnik tego nie wyjaśnia, informacje o używanej do tego komendzie jest na **stronie 30 htkbook** (39 strona pdfa, HLEd).

Na podstawie tego można też utworzyć wymaganą potem przy reestymacji listę wszystkich fonemów jakich używamy (w poradniku jako monophones0) – na jej podstawie można też robić sam prototypowy model hmm, żeby mieć pewność że niczego nie pominęliśmy. Proponuję użyć do tego Excela – otwieramy plik mlf z utworzoną za pomocą HLEd anotacją na poziomie fonemów, kasujemy zbędne kolumny z cyferkami zostawiamy same nazwy fonemów i używamy opcji usuń duplikaty.



The screenshot shows the Microsoft Excel interface with the 'Dane' (Data) ribbon selected. The spreadsheet contains the following data:

	A	B	C	D	E	F	G	H	I	J
1	#!MLF!#									
2	*!/nagranietestowe.lab									
3	9200000	9200000	sil							
4	9200000	10158000	w							
5	10158000	1116000	L							

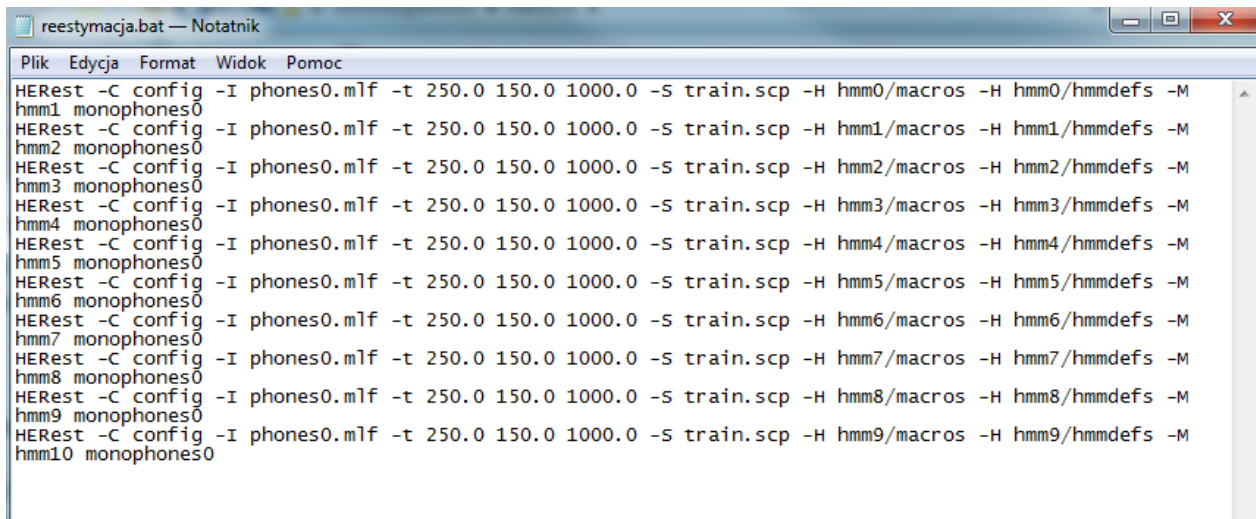
The 'Usuń duplikaty' (Remove Duplicates) dialog box is open, displaying the following text:

Usuń duplikaty
Umożliwia usunięcie zduplikowanych wierszy z arkusza.
Można określić, w których kolumnach mają być wyszukiwane zduplikowane informacje.

UWAGA! Ta opcja nie rozróżnia wielkości liter! W wypadku, gdy macie gdzieś fonem o i O (ó dla htk) to jedno z nich spadnie z rowerka i trzeba będzie ręcznie dopisać, ale i tak łatwiej to kontrolować i zajmie to mniej roboty niż robienie tego ręcznie.

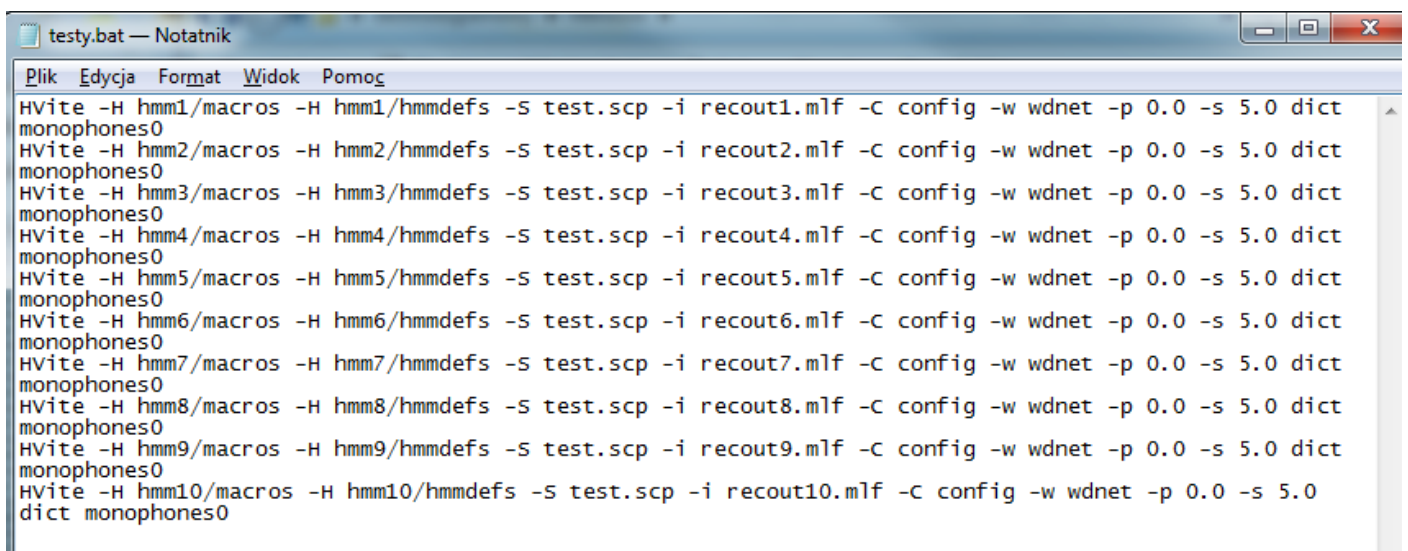
Reestymacja, testowanie i opracowanie

To już niemal końcowy etap pracy z htk i już całkiem szybki jak się zrobi to poprawnie. Jeśli wszystko robiło się z „poradnikiem htk” i z zastosowaniem moich wskazówek wszystko powinno działać, nawet sama reestymacja z którą często są problemy. Jedyne co mogę tu zaproponować to nie wklepywać komend z użyciem konsoli cmd, tylko w tym miejscu przesiąść się na pliki wykonywalne .bat (wcześniej nie było to tak istotne, bo dużo pracy było wykonywanej ręcznie i często mogły pojawić się błędy które można było prościej poprawić). Wygląda to mniej więcej tak:



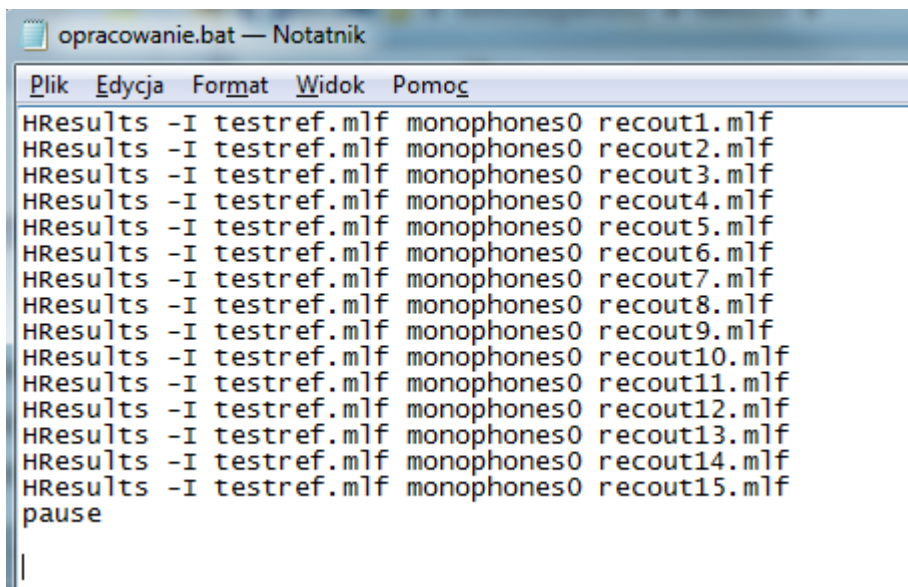
```
reestymacja.bat — Notatnik
Plik Edycja Format Widok Pomoc
HERest -C config -I phones0.mlf -t 250.0 150.0 1000.0 -s train.scp -H hmm0/macros -H hmm0/hmmdefs -M
hmm1 monophones0
HERest -C config -I phones0.mlf -t 250.0 150.0 1000.0 -s train.scp -H hmm1/macros -H hmm1/hmmdefs -M
hmm2 monophones0
HERest -C config -I phones0.mlf -t 250.0 150.0 1000.0 -s train.scp -H hmm2/macros -H hmm2/hmmdefs -M
hmm3 monophones0
HERest -C config -I phones0.mlf -t 250.0 150.0 1000.0 -s train.scp -H hmm3/macros -H hmm3/hmmdefs -M
hmm4 monophones0
HERest -C config -I phones0.mlf -t 250.0 150.0 1000.0 -s train.scp -H hmm4/macros -H hmm4/hmmdefs -M
hmm5 monophones0
HERest -C config -I phones0.mlf -t 250.0 150.0 1000.0 -s train.scp -H hmm5/macros -H hmm5/hmmdefs -M
hmm6 monophones0
HERest -C config -I phones0.mlf -t 250.0 150.0 1000.0 -s train.scp -H hmm6/macros -H hmm6/hmmdefs -M
hmm7 monophones0
HERest -C config -I phones0.mlf -t 250.0 150.0 1000.0 -s train.scp -H hmm7/macros -H hmm7/hmmdefs -M
hmm8 monophones0
HERest -C config -I phones0.mlf -t 250.0 150.0 1000.0 -s train.scp -H hmm8/macros -H hmm8/hmmdefs -M
hmm9 monophones0
HERest -C config -I phones0.mlf -t 250.0 150.0 1000.0 -s train.scp -H hmm9/macros -H hmm9/hmmdefs -M
hmm10 monophones0
```

Utworzyć taki plik, zapisać w notatniku jako .bat i odpalić. Przypominam tylko, że używany tutaj mlf to segmentacja na poziomie fonemów (ta z HLEd). To samo polecam przy robieniu testów...



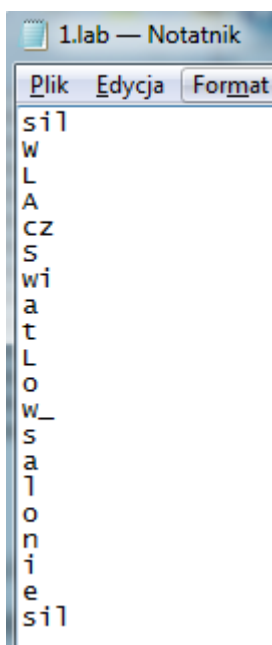
```
testy.bat — Notatnik
Plik Edycja Format Widok Pomoc
HVite -H hmm1/macros -H hmm1/hmmdefs -s test.scp -i recout1.mlf -C config -w wdnet -p 0.0 -s 5.0 dict
monophones0
HVite -H hmm2/macros -H hmm2/hmmdefs -s test.scp -i recout2.mlf -C config -w wdnet -p 0.0 -s 5.0 dict
monophones0
HVite -H hmm3/macros -H hmm3/hmmdefs -s test.scp -i recout3.mlf -C config -w wdnet -p 0.0 -s 5.0 dict
monophones0
HVite -H hmm4/macros -H hmm4/hmmdefs -s test.scp -i recout4.mlf -C config -w wdnet -p 0.0 -s 5.0 dict
monophones0
HVite -H hmm5/macros -H hmm5/hmmdefs -s test.scp -i recout5.mlf -C config -w wdnet -p 0.0 -s 5.0 dict
monophones0
HVite -H hmm6/macros -H hmm6/hmmdefs -s test.scp -i recout6.mlf -C config -w wdnet -p 0.0 -s 5.0 dict
monophones0
HVite -H hmm7/macros -H hmm7/hmmdefs -s test.scp -i recout7.mlf -C config -w wdnet -p 0.0 -s 5.0 dict
monophones0
HVite -H hmm8/macros -H hmm8/hmmdefs -s test.scp -i recout8.mlf -C config -w wdnet -p 0.0 -s 5.0 dict
monophones0
HVite -H hmm9/macros -H hmm9/hmmdefs -s test.scp -i recout9.mlf -C config -w wdnet -p 0.0 -s 5.0 dict
monophones0
HVite -H hmm10/macros -H hmm10/hmmdefs -s test.scp -i recout10.mlf -C config -w wdnet -p 0.0 -s 5.0
dict monophones0
```

... i opracowaniu statystycznym. Tutaj pause na końcu, żeby okienko cmd nie zniknęło i można było przeanalizować wyniki rozpoznania.



```
opracowanie.bat — Notatnik
Plik Edycja Format Widok Pomoc
HResults -I testref.mlf monophones0 recout1.mlf
HResults -I testref.mlf monophones0 recout2.mlf
HResults -I testref.mlf monophones0 recout3.mlf
HResults -I testref.mlf monophones0 recout4.mlf
HResults -I testref.mlf monophones0 recout5.mlf
HResults -I testref.mlf monophones0 recout6.mlf
HResults -I testref.mlf monophones0 recout7.mlf
HResults -I testref.mlf monophones0 recout8.mlf
HResults -I testref.mlf monophones0 recout9.mlf
HResults -I testref.mlf monophones0 recout10.mlf
HResults -I testref.mlf monophones0 recout11.mlf
HResults -I testref.mlf monophones0 recout12.mlf
HResults -I testref.mlf monophones0 recout13.mlf
HResults -I testref.mlf monophones0 recout14.mlf
HResults -I testref.mlf monophones0 recout15.mlf
pause
|
```

Na koniec jeszcze, wymagane do opracowanie statystycznego, zapis fonemami słów wypowiedzanych w nagraniu testowym, nie było dla mnie do końca jasne jak to ma wyglądać.



```
1.lab — Notatnik
Plik Edycja Format
sil
w
L
A
cz
S
wi
a
t
L
o
w_
s
a
l
o
n
i
e
sil
```

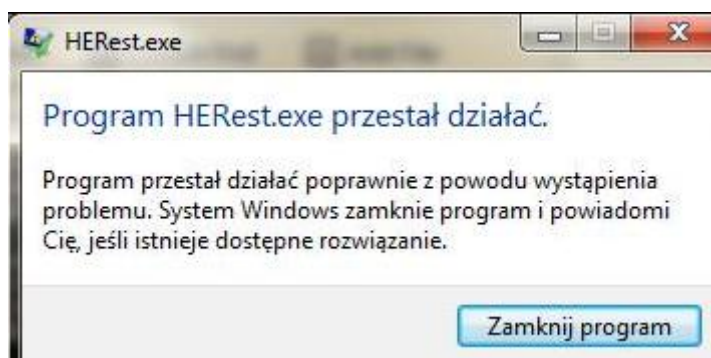
Sposoby rozwiązywania błędów oraz wybrane przykłady

Mimo nawet bardzo ścisłego stosowania się do instrukcji zawsze coś może pójść nie tak, ten poradnik pozwoli Wam uniknąć większości, powiedzmy, standardowych błędów, reszta tych standardowych jest opisana w linku, który podaję na końcu. Gdy mój rocznik zmagał się z projektem w HTK udało mi się rozwiązać szereg także tych mniej standardowych, opis poniżej, a także kilka porad ogólnych.

Etapy postępowania w wypadku wystąpienia błędu:

- sprawdzasz czy nie masz błędu w samej komendzie
- jeżeli wpisujesz komendę i ona nie działa, to problem leży w plikach do których się ta komenda odwołuje, w pierwszej kolejności sprawdzasz właśnie je, problemem może być np. dopisany gdzieś przypadkowo znak lub spacja
- sprawdzasz czy na końcu każdego pliku dopisany jest enter
- googlujesz error, który wystąpił lub sprawdzasz na stronie (link na końcu)
- pamiętaj, że w momencie znalezienia błędu w jednym z plików trzeba powtórzyć wszystkie operacje wykonywane za jego pośrednictwem! Np. błąd w słowniku oznacza konieczność powtórzenia segmentacji na poziomie fonemów, a to z kolei wymaga powtórzenia reestymacji

Przykładowe błędy z którymi się spotkałem:



Przy próbie wywołania komendy reestymacji wyskakiwał taki błąd. Rozwiązaniem okazał się plik .lab zawierający fonemy po kolei z nagrania (ten tworzony z pliku mlf zawierającego segmentację na poziomie fonemów). Wszystkie fonemy w nim występujące muszą być oddzielone enterami, osoby, u których ten błąd występował, miały je oddzielane spacjami, htk wtedy traktuje to jak jeden i macie fonem o rozmiarze 1000... nic dziwnego że program się wiesza. Wystarczy skopiować właściwie te nazwy fonemów z Excela, albo otworzyć ten wadliwy plik .lab w Excelu, zaznaczyć wiersz w którym są nazwy fonemów, skopiować, wkleić z opcją „wklej specjalnie” i zaznaczoną transpozycją. I wtedy skopiować czy tam zapisać w formacie .lab.

```
F:\STUDIA\U_semestr\Technologia Mowy\Projekt>HUnit -H hmm1/macros.txt -H hmm1/
ndefs.txt -C config.txt -w wnet -p 0.0 -s 5.0 dict.txt monophones0.txt
ERROR [+8231] GetHCIModel: Cannot find hmm [???-leV_[+???]
FATAL ERROR - Terminating program HUnit
```

Błąd, który zabił ćwieka nie tylko mi na dłuugi czas, ale w końcu udało mi się go rozwiązać. Taka formuła błędu wskazuje na błąd w słowniku, natomiast wielokrotne przepatrzenie i przeanalizowanie WSZYSTKIEGO nic nie wykazało, nawet sprawdzanie osobno każdego fonemu e_. Dopiero po kilku dniach wpadłem co może być nie tak, przekleiłem słownik do Worda i włączyłem wyświetlanie znaków specjalnych i...

czwartA → cz·f·a·r·t·o·e·_·sp¶

Koleżanka przy tworzeniu słownika przez przypadek wcisnęła coś z klawiatury, przez co przed _ dopisał się jakiś znak, co ciekawe, niewidzialny... Udało mi się ustalić że coś tam jest bo skasowanie e_ wymagało trzykrotnego wciśnięcia backspace, nie dwukrotnego, potem już poszło łatwo, znaleźć i skasować dodatkowe znaki, błąd ustąpił. **Pamiętajcie, że po zmianie w słowniku trzeba powtórzyć wszystkie czynności które go używają, czyli segmentację na poziomie fonemów, a co za tym idzie też reestymację.**

Koniec

I to by było na tyle, życząc wszystkim owocnej pracy z htk pragnę jeszcze tylko dorzucić bardzo użyteczną stronę, która czasem odpowie na Wasze pytania, można ją znaleźć też w „Poradniku HTK”.

http://www.ling.ohio-state.edu/~bromberg/htk_problems.html