

Hinweise zum Praktikum

Spracherkennung

Allgemeines

- Jede Gruppe besitzt ihr eigenes Gruppenverzeichnis unter dem Verzeichnis `HTK-KURS/` (`GROUP1` z.B. für die erste Gruppe). Mit dieser Nummer ist auch der Rechner zugeordnet, den man verwenden sollte (`cip1` - `cip7`).
- Hilfreich (für Euch bzw. mich) ist es, daß in Eurem Gruppenverzeichnis in der Datei `Besitzer.txt` Eure Namen aufgeführt sind. Falls diese noch nicht existiert, bitte anlegen. Wer noch keinen Texteditor unter Linux kennt, ist beim `kwrite` (Eingabe dieses Befehls in der Kommandozeile) gut aufgehoben.
- Bitte nichts in anderen Gruppenverzeichnissen oder im Hauptverzeichnis (`HTK-KURS/`) verändern. Fairerweise auch nicht ungefragt bei anderen Gruppen “anschauen”.
- Bitte nicht das Passwort ändern, auch keine Punkt-Dateien im Home-Verzeichnis von `htk` oder Einstellungen des Desktops (Farben und ähnliches) verändern!
- Private Nutzung des Accounts bitte in Maßen halten und private Dateien nicht hier speichern. Bitte keine Emails vom Account `htk` versenden; möglichst auch bei der Verwendung von Webmail wie z.B. Yahoo darauf achten, daß Eure Login-Informationen nicht dauerhaft im Browser gespeichert werden.

Info zu den Trainingsdaten

Die Partitur-Files in `TRAIN.Partitur`, `DEV.Partitur` und `TEST.Partitur` sind nicht die aktuellen VM-Partituren. Daher sind dort Schreibweisen enthalten (z.B. `Hansjörg`), die im offiziellen VM Lexikon nicht mehr vorkommen. Daher muss unbedingt das hier gespeicherte Lexikon `vm_ger.lex` verwendet werden.

Die 'Skripten' (File Listen) sind bereits von 'leeren' Turns bereinigt.

In den Trainings- und Test-Korpora befinden sich die `htk` files sowohl mit alter als auch mit neuer Schreibweise des Dateinamens gelinkt. Der Kurs verwendet die alte Schreibweise.

Dokumentation

Das HTK-Buch ist im Unterverzeichnis `Kurs-Unterlagen` in der PDF-Datei `HTK_Book.3.2.pdf`; Hilfe zu Unix gibt es unter anderem in den PDF-Dateien `unixkurs.pdf` und `unixkurs_lang.pdf`.

Bitte diese großen PDF-Dateien nicht an den Druckern am Institut ausdrucken; einzelne Seiten ist OK.