

Checkliste

Datenbank-Info

- Projekt
- Kurzname
- Vollständiger Name
- Sprachdatenbank-ID
- Ersteller
- Datum

Spezifikation

Sprachlicher Inhalt

- Sprechsituation
- Sprachstil
- Szenario
- Vokabular
- Aufgabe

Sprecherpopulation

- Anzahl

Demographische Merkmale

- Muttersprache
- Fremdsprachen
- Dialekt, regionale Färbung
- Ausbildung, sozialer Status

Biometrische Merkmale

- Geschlecht
- Alter
- Größe, Gewicht
- Raucher
- Zahnspange
- Piercing im Vokaltrakt
- Sprach- oder Sprechfehler

Aufnahmetechnik und -umgebung

- Akustische Umgebung
- Skript
- Umgebungsgeräusch
- Mikrofone
- Hardware
- Unterlagen

Ablaufkontrolle

- automatisch, semi-automatisch, manuell
- überwachte bzw. nicht-überwachte Aufnahmen

Telefonaufnahmen

- Serverhard- und -software
- Telefonnetz
- Aufnahmeumgebung
- Endgerät, Freisprecheinrichtung

Feldaufnahmen

- Testaufnahme

Wizard of Oz

- Wizard-Simulation

Weitere Angaben

- Abtastrate
- Quantisierung und Kanäle
- Signaldatenformat
- Annotationsebenen und -verfahren
- Annotationsformat
- Metadateninhalt
- Metadatenformat
- Lexikonformat
- Korpusstruktur
- Terminologie
- Dokumentation
- Validierung
- Distributionsform und -medien

Vorbereitung der Datensammlung**Arbeitsteams zusammenstellen**

- Administration
- Aufnahme
- Datensicherung
- Annotation
- Dokumentation
- Validierung

Räume reservieren und einrichten

- Arbeitsplätze
- Aufnahmeräume, Studios, Fahrzeuge
- Warte- und Lagerräume

Rechner und Aufnahmetechnik installieren**Telefonaufnahmen**

- ISDN Anschluss (gebührenfrei?)
- ISDN Hardware
- Telefonserver
- Ansageprompts und Piepton
- Aufnahmeskript
- Pausenerkennung, Sprach-Erkennung
- Aufnahmeparameter justieren
- Echos vermeiden

Studio- und WOZ-Aufnahmen

- akustische Umgebungsbedingungen
- Mikrofone
- Vorverstärker, A/D-Wandler
- Aufnahmegeräte
- Aufnahmesoftware

Feldaufnahmen

- Batterien, Akkus
- Stromversorgung, Störungen
- Mobiltelefone ausschalten
- Aufnahmegeräte testen
- Ausweichplan für schlechtes Wetter
- tägliche Datensicherung

Wizard of Oz

- Beobachtungstechnik
- Schallisolierung Aufnahme- und Kontrollraum
- Sprachausgabe synthetisieren
- Dialogfluss und Ablaufplan für Simulation

Kommunikationssoftware und Datenfluss

- Aufnahme- und Annotationssoftware
- Wiki- und Forum-Software
- Datenfluss
- Datensicherung
- Probedurchlauf Aufnahme

Sprecher

- Sprecherrekrutierung
- Honorar
- Unterlagen, Formulare

Organisation

- Promptmaterial
- rechtliche Aspekte prüfen
- Kommunikation
- Projekt- und Terminkalender
- Dokumentvorlagen

Datensammlung

Aufnahmen und Protokolle

- Sprecherdatenblatt
- Aufnahmeprotokoll
- Speicherplatz
- Signalqualität
- Datensicherung
- Zusatzdaten erheben

Prävalidierung

- Daten auswählen und aufbereiten

Datenaufbereitung

- Datenübertragung auf Arbeitsrechner bzw. Server
- Dateisystem-Struktur und Dateibenennung
- Abtastrate, Quantisierung anpassen
- Schnitt, Filterung
- Formatkonversion Signaldaten
- Formatkonversion Text- und Annotationsdaten
- automatische Fehlerprüfungen
- Datenfluss initialisieren

Annotation

Annotierer

- Einweisung
- Annotiererkenntnisse
- Inter-Annotierer-Konsistenz

Annotationen

- Transkriptionsrichtlinien festschreiben
- Dateiformat für Annotationen
- Annotationseditoren
- Prozeduren für Annotation
- Annotationen in Datenfluss
- Implementation formaler Konsistenzkontrolle

Lexikon und Dokumentation

Aussprachewörterbuch

- Kategorien
- Liste der Einträge
- statistische Angaben
- Aussprache automatisch generieren, manuell überarbeiten
- Zweitkorrektur
- Abgabeformat und -kodierung

Dokumente

- Versionsnummer und Editionsdatum
- Spezifikation
- Aufnahme-Handbücher
- Annotations-Richtlinien
- Forums- und Wikibeiträge
- Softwareprotokolle
- Validierungsbericht(e)
- Konferenz- und Zeitschriftenartikel
- Pressebeiträge

Validierung

Prävalidierung

- Validierer bestimmen (intern/extern)
- Zeitplan und Termin Prävalidieren
- Empfehlungen aus Prävalidierungsbericht umsetzen

Abschlussvalidierung

- Validierer bestimmen
- Zeitplan und Termin Validierung
- Validierungsbericht in Dokumentation aufnehmen

Distribution

- Speichermedien und Datenträger für Signaldaten
- Kompression und Kodierung
- separate Distribution von symbolischen Daten
- Datensicherheit und Prüfdaten
- Katalogankündigung
- Metadaten bereitstellen
- Veröffentlichung