

TextQuest – computerunterstützte Inhaltsanalyse

Autor:
Dr. Harald Klein, M.A.
Osnabrück
Deutschland

Mai 2026

Entwicklung von TextQuest

- erste PL/1 Version 1983 unter dem Namen Intext – Inhaltsanalyse von Texten auf IBM Mainframe unter MVS
- erste MS-DOS Version 1988, geschrieben in C
- erste Windows-Version 1999, Umbenennung in TextQuest, geschrieben in C/C++ mit wxWidgets
- erste Version für Apple Mac-OS 2013

„Unter dem Begriff 'Computerunterstützte Inhaltsanalyse' versteht man Verfahren, die die Analyse und Vercodung der zu untersuchenden Inhalte mittels EDV-Programmen vornehmen. Voraussetzung für die Anwendung derartiger Verfahren ist, dass das Untersuchungsmaterial vertextet, d.h. in schriftlicher Form vorliegt.“ Hörschinger 1985, S. 18

Generelle Unterschiede der Arten von Inhaltsanalyse

konventionell	computerunterstützt
Text durch Lesen eines kleinen Teils kennenlernen	Text digitalisieren und dabei kennenlernen
entfällt	Wörterliste als Basis des Kategoriensystems erzeugen
Kategoriensystem durch Definitionen und Beispiele aufbauen	Kategoriensystem durch vollständige Aufzählung der Suchbegriffe aufbauen
Kodiererschulung durchführen	entfällt
Daten auf Codesheet eintragen	entfällt
Daten vom Codesheet in Computer eingeben	entfällt
Daten bereinigen	entfällt
Daten auswerten	Daten auswerten

Spezielle Unterschiede

konventionell	computerunterstützt
bearbeitbare Textmenge von Anzahl der Kodierer abhängig	bearbeitbare Textmenge nahezu unbegrenzt
bedingt komplexe Kategoriensysteme verwendbar	komplexe Kategoriensysteme verwendbar
Kodierung innerhalb von Wochen	Kodierung innerhalb von Minuten
meist nur eine Kodierung	beliebig viele Kodierungen
Kategoriensystem kann während der Kodierung nicht geändert werden, da eine erneute Analyse oft unmöglich ist. Erweiterungen sind möglich.	Kategoriensystem kann leicht geändert und erweitert werden, da beliebig viele Kodierungen möglich sind.

Probleme der Arten von Inhaltsanalyse

konventionell	computerunterstützt
alle Arten möglich	nur bestimmte Arten möglich
Kategoriensystem für eine Textsorte gültig	Kategoriensystem nur für einen Materialkorpus gültig
geringe Transparenz des Kategoriensystems durch Beispiele und Kodierregeln	totale Transparenz des Kategoriensystems durch alle Suchbegriffe
Reliabilitätsproblem durch menschliche Kodierende	100 Prozent Reliabilität
Ambiguität und Negation unproblematisch	Ambiguität und Negation müssen aufwändig behandelt werden
kaum Transfer des methodischen <i>Wie</i>	Publizierung von Kategoriensystemen und der Suchbegriffe, akkumulative Forschung

Arbeitsablauf computerunterstützten Inhaltsanalyse

- ➊ Definition der Texteinheit und der externen Variablen
- ➋ Digitalisierung des Untersuchungsmaterials
- ➌ Definition der (Ab)Schreibkonventionen
- ➍ Erstellung einer Wörterliste als Basis für die Entwicklung eines a-posteriori Kategoriensystem
- ➎ Pretest des Kategoriensystems
- ➏ Disambiguierung potentiell mehrdeutiger Suchbegriffe mittels Konkordanzen
- ➐ Kontrolle von Negation
- ➑ Codierung des Textes
- ➒ Weiterentwicklung des Kategoriensystems

Möglichkeiten der Digitalisierung

- Abschreiben
- Lesen mit Texterkennungssoftware
- Diktieren mit Diktiersoftware
- Scannen mit Scannern oder Scanrobotern
- Texte im Internet: Webseiten, PDF-Dateien
- Verarbeiten digitalisierter Texte
 - Texte sind flüchtig
 - Texte auf Datenträgern wie DVD, CD oder Diskette
 - Texte aus Archiven
 - Problem der Steuerzeichen, Konvertierungsprogramme
- Standardisierung von Textdaten
- Archive für Textdaten

- Wie ist die Texteinheit definiert?
- Welche externen Variablen werden gebraucht?
- Texteinheit: der statistische Fall, auf den sich die Hypothesen beziehen
- externe Variablen: werden für Hypothesenüberprüfung gebraucht, anderer Name: Identifikatoren

- Beschaffung der Texte
- Erstellen der Rohtexte
- Erstellen der Systemdatei aus dem Rohtext
- Erstellen einer Wörterliste bzw. deren Derivaten:
 - Kontrolle der Orthographie
 - Kontrolle der Transkriptionsregeln
 - Basis zur Konstruktion von Kategoriensystemen
- Disambiguierung von mehrdeutigen Wörtern/Suchbegriffen mittels Konkordanzen

- mit Steuerzeichen: bis zu 50 externe Variablen, maximale Länge 10 Zeichen – alle Versionen
 - Steuerzeichenformat
 - festes Format, z.B. als CSV oder Excel.Datei
- ohne Steuerzeichen: automatische Vergabe der externen Variablen – bis Version 4.2
 - Zeilenformat: Zeilenzähler
 - Absatzformat: Absatzzähler
 - Seitenformat: Seitenzähler
 - Satzformat: Satzzähler

- Textdatei importieren
- Wörterliste
- Wortsequenzlisten
- Wortpermutationslisten
- Vergleich von Wörterlisten, auch mehr als zwei gleichzeitig
- Inhaltsanalyse

- Reduzieren: Wörter auf Ausschlusslisten (STOP-Wörter)
- Reduzieren: Kriterien Länge und Häufigkeit

Formulierung von Suchbegriffen

- einzelne Wörter (' attraktiv ')
- Wortteile ('attraktiv')
- Wortanfänge (' attraktiv')
- Wortenden ('politik ')
- Wortfolgen ('gut aussehen')
- Wortstammfolgen ('gut> aussehen>')

Optionen bei der Formulierung von Suchbegriffen

- C = Codierkontrolle: potentiell mehrdeutiger Suchbegriff, diese Markierung ist wichtig für die interaktive Codierung und/oder die entsprechende Protokolldatei
- U = Upper-/lowercase: Unterschiede, die nur auf Groß-/Kleinschreibung beruhen, werden nicht berücksichtigt
- D = direkt: die Teile einer Wortfolge müssen direkt hintereinander vorkommen
- F = folgend: die Teile einer Wortfolge müssen hintereinander vorkommen, Abstand ist egal
- S = simultan: die Teile einer Wortfolge müssen vorkommen, Reihenfolge und Abstand sind egal

- Zähler der Codes in der Analyseeinheit (TAB-Datei)
- Reihenfolge der Codes in der Analyseeinheit (VEC-Datei)

Ablauf der Codierung

- Lesen und internes Speichern der Suchbegriffe und deren Etiketten
- Lesen der 1. Texteinheit
- Suchen des 1. Suchbegriffes
 - Wenn gefunden, Zähler hochzählen und nach der Fundstelle weitersuchen
 - Wenn nicht gefunden, nächsten Suchbegriff suchen
 - solange durchführen, bis alle Suchbegriffe abgearbeitet sind, dann Zählermatrix in die Ausgabedatei schreiben
- nächste Texteinheit einlesen und 1. Suchbegriff suchen
- solange Texteinheiten lesen, bis alle durchsucht wurden

Probleme bei der Codierung

- Mehrdeutigkeit von Suchbegriffen, z.B. Homonyme
Beispiele: Schloss, Bank, Strauß
- Pronomenproblem: Suchbegriff wird durch Pronomen ersetzt
(*sie* statt *Bundeskanzlerin*)
- Negation von Suchbegriffen
 1. Beispiel:
Textstelle: Sie waren *keine* **Arbeiter**, sondern Helfer.
Suchbegriff: arbeit
 2. Beispiel:
Textstelle: Sie waren *nicht* **arbeitslos** geworden.
Suchbegriff: arbeit
- Mehrfachcodierung: bei Suchbegriffen und Textstellen
- Kontrolle von Mehrfachcodierung
 - im Kategoriensystem vor der Codierung
 - während oder nach der Codierung

- Datei der codierten Texteinheiten
- Datei der uncodierten Texteinheiten
- Datei der kompletten Codierkontrolle
- Datei der Texteinheiten mit Negation
- Datei der mehrdeutigen Suchbegriffe
- interaktive Codierung

- kostenlose 30 Tage Testversion, die mit einem Schlüssel freigeschaltet werden kann
- einfache, mehrfache und Netzwerklizenzen
- Laufzeit: 3 Monate, ein Jahr oder unbegrenzt – MacOS nur Einzelversionen mit unbegrenzter Laufzeit
- Nutzer: kommerzielle, NGOs, (Aus)Bildung wie Schulen oder Universitäten

Bestellungen:

<https://www.social-science-consulting.com> oder

<https://www.textquest.de>