



# **Erfahrungen und Erkenntnisse aus dem Einsatz automatischer Verfahren in der Mediendokumentation**

Günter Peters, 8. September 2017



# Zur Person

- Studium Politologie, Germanistik, Pädagogik, Philosophie (Höheres Lehramt) an der Universität Göttingen – Staatsexamen 1978
- Von 1978 bis 2015 in der G+J Dokumentation
  - 1978 – 1989 Fachdokumentar für TV, AV-Medien, Medienpolitik
  - 1989 – 1993 Leiter Referat Naturwissenschaft, Medien, Sport der G+J Textdokumentation
  - 1993 – 2000 Leiter der G+J Textdokumentation
  - 2000 – 2008 Stellvertretender Leiter der G+J Dokumentation
  - 2008 – 2015 Leiter der G+J Dokumentation



# Ein wenig Historie

- Automatische Verfahren werden seit 2002 professionell in deutschen Dokumentationen eingesetzt (Kooperation von Gruner + Jahr und IBM).
- Der Einsatz blieb für die folgenden ca. 10 Jahre auf Volltexte beschränkt.
- Neben Text Minern wurden semantische Verfahren entwickelt, die die Ergebnisse automatischer Verfahren verbesserten.
- Temis ([temis.org](http://temis.org)) war der Kooperationspartner und Dienstleister für G+J beim Einsatz des TextMiners.



# Der Anfang

- In dieser ersten Phase des Einsatzes von automatischen Verfahren mussten auf Seiten der Dokumentationen und der Technologie folgende Voraussetzungen vorhanden sein.
  - Volltexte mussten in großer Menge vorhanden sein und
  - leistungsfähige Rechner mussten große Mengen von Texten verarbeiten können.
  - Algorithmen zur Textanalyse, Spracherkenner etc. mussten entwickelt worden sein.
- Dies war ab Mitte der 1990er Jahre der Fall – allerdings in Anfängen (Windows 3.11, lokale Speicher).



# Grenzen der Technologie

- Die vergleichsweise geringe Speicherkapazität, kleine Arbeitsspeicher, langsame Prozessoren und Softwares in den Kinderschuhen begrenzte in dieser Phase die Einsatzmöglichkeiten der Technologie.
- Ideen und Ansätze zur Erweiterung der Einsatzfelder automatischer Verfahren waren vorhanden, stießen aber an Grenzen.
- So waren Inhalte und Konnotationen von Fotos und Videos damals automatisiert schwer zu erschließen. Das hat sich erst in den letzten 2 – 3 Jahren geändert.



# Grenzen der Technologie

- Neben technischen Grenzen gab (und gibt) es auch Probleme mit den Inhalten der Texte:
- Homonymie und Polysemie (z. B. Läufer, Bank, Ball) führen zu Fehlern bei automatisch erzeugten Verschlagwortungsvorschlägen.
- Die inhaltlichen Zuordnungen in den Systematiken sind in den einzelnen Medienhäusern spezifisch und unterschiedlich.
- In den Medien tauchen täglich neue Begriffe auf oder werden bekannte Begriffe in neuen Zusammenhängen gebraucht, was automatische Verfahren nur nach Training erkennen können.



# Grenzen der Technologie

- Die Unterscheidung zwischen relevanten und irrelevanten Fakten und Namen in Texten ist für Text Miner nicht möglich.
- Bei Audiofiles (Sprache) sind oft Dialekte, Nebengeräusche und andere technische Beeinträchtigungen Hindernisse für eine automatisierte Inhaltserschließung.
- Textinhalte waren und sind noch am besten automatisiert zu erschließen, weshalb die Herstellung einer textuellen Inhaltsbeschreibung zur Zeit eine Möglichkeit für eine inhaltliche Erschließung von Medieninhalten, also auch von Fotos, Videos und Sprache zu sein scheint.



# Die Entwicklung

- In den letzten Jahren hat sich die Entwicklung automatischer Verfahren beschleunigt.
- Die Algorithmen und Softwares wurde weiter entwickelt, es entstanden Ideen zur inhaltlichen Beschreibung von Webinhalten (Semantic Web - [https://de.wikipedia.org/wiki/Semantic\\_Web](https://de.wikipedia.org/wiki/Semantic_Web)).
- Durch Suchmaschinenfirmen wie Google kamen neue Player mit Geld, technischem Know-How und Dynamik hinzu.
- Das technisch Mögliche soll erforscht und umgesetzt werden.
- Dazu passt, dass die Hardware- und Speicherprobleme de facto nicht mehr vorhanden sind.





# Erfahrungen

- Der Einsatz von automatischen Verfahren geschah und geschieht in der Regel halbautomatisch, das heißt
  - Die eingesetzte Software erstellt Verschlagwortungsvorschläge, die die Dokumentare überprüfen und entweder korrigieren oder akzeptieren.
  - Die Gesamtheit der so intellektuell verschlagworteten bzw. überprüften Artikel bildet die Trainingsmenge für die eingesetzten Softwares.
- Die Effekte der automatischen Verschlagwortung bestehen in
  - der Verbesserung der Konsistenz der Verschlagwortung (ein qualitativer Aspekt – wirkt langfristig)
  - und der schnelleren Bearbeitung der zu bearbeitenden Artikelmenge (ein quantitativer Aspekt – wirkt kurzfristig).



# Erfahrungen

- Durch automatische Verfahren gewinnt man Arbeitszeit, sie werden als Rationalisierungsinstrumente eingesetzt.
- Allerdings erfordern sie neue und zusätzliche Tätigkeiten für Training, Projektkoordination und Systematikpflege.
- In der Summe bewirken automatische Verfahren Einsparungen von Aufwänden im Lektorat bzw. Input von Dokumentationen.



# Erfahrungen

- Der Einsatz automatischer Verfahren ist im Prinzip die Lösung von quantitativen Problemen bei der inhaltlichen Erschließung von Medieninhalten.
- Mit solchen Verfahren wird Dokumentation zu Data Management – es ist möglich, große Mengen an Dokumenten aktuell zu bearbeiten und mit Metainformationen angereichert zur Verfügung zu stellen.
- Dies ist angesichts der Datenmengen, die von Archiven und Dokumentationen jetzt und zukünftig verarbeitet werden müssen, auch nötig.



# Aktuelle Tendenzen

- Zur Zeit erleben wir einen Entwicklungsschritt in Richtung Automatisierung, der auch die bisherige halbautomatische Arbeitsweise modifizieren könnte.
- Denn sowohl Internetfirmen als andere große Medienfirmen haben ein großes Interesse an einer automatisierten Erschließung aktueller und historischer Bestände.
- Die technischen Voraussetzungen sind auch vorhanden, um einen größeren Grad an Automatisierung zu realisieren als wir ihn bislang kennen.



# Aktuelle Tendenzen

- Dazu kommt, dass die öffentlich-rechtlichen Rundfunkanstalten unter ökonomischen und politischen Druck geraten sind und in vorweggenommenen Sparbemühungen auf automatische Verfahren als Rationalisierungsinstrumente setzen.
- Private Fernsehanstalten sind abhängig von den Gewinnerwartungen ihrer Anteilseigner, was zu einem internen Kostendruck führt.



# Erkenntnisse

- Automatische Verfahren sind aus der inhaltlichen Erschließung von Medien (vor allem von Texten) nicht mehr wegzudenken.
- Bei der Erschließung von Audiodaten ist die Transskribierung und somit die Herstellung einer Textgrundlage die Voraussetzung für eine adäquate inhaltliche Erschließung.
- Bei der Erkennung von Bild- und Videoinhalten steht noch am Anfang eines professionellen Einsatzes.



# Erkenntnisse

- Der Einsatz automatischer Verfahren hat bei der Textdokumentation nicht zum Verschwinden der intellektuellen Verschlagwortung geführt.
- Eine ähnliche Entwicklung erwarte ich auch beim Einsatz automatischer Verfahren in anderen Medien.
- Auf Dokumentare kommen neue Herausforderungen zu: neben den bisherigen Tätigkeiten die inhaltliche Steuerung automatischer Verfahren durch halbautomatischen Einsatz oder andere Formen des Trainings.



# Berufsbild

- Der Einsatz automatischer Verfahren ist nur ein Teil der technischen Entwicklung der letzten Jahre, die Veränderungen in den Tätigkeiten und im Berufsbild von Dokumentaren bewirkt.
- Das Internet und vor allem Suchmaschinen wie Google beeinflussen die Tätigkeit von Dokumentaren positive wie negativ.
- Die Ökonomie (Pressekrise, Gewinnerwartungen Gesellschafter privater TV-Anstalten, Gebührendruck auf öffentlich-rechtliche Rundfunkanstalten) erzwingt Veränderungen in der Arbeit und Anzahl von Dokumentaren und Dokumentationseinrichtungen.





# Auswirkungen auf das Berufsbild

- Diese Entwicklungen haben Auswirkungen auf das Berufsbild von Dokumentaren und Archivaren:
  - Weg vom Fachmann für inhaltliche Erschließung und Aufbereitung, hin zum Datenmanager;
  - Weg vom qualifizierten Rechercheur, hin zum Berater von Nutzern (impliziert auch eigene Recherchen);
  - Aneignung von grundlegenden IT-Kenntnissen für die Steuerung diverser automatischer Verfahren;
  - Aneignung von grundlegenden urheberrechtlichen Kenntnissen für die Beurteilung rechtlicher Probleme.



# Ende

- Danke für Ihre Aufmerksamkeit.
- Haben Sie Fragen?