

# Vom Quellentext zur Datenbank : ein Konzept zur integrierten Verarbeitung quantitativer und qualitativer Daten

Autor(en): **Hagnauer, Stephan / Bartlome, Niklaus**

Objektyp: **Article**

Zeitschrift: **Geschichte und Informatik = Histoire et informatique**

Band (Jahr): **9 (1998)**

PDF erstellt am: **25.06.2024**

Persistenter Link: <https://doi.org/10.5169/seals-7236>

## **Nutzungsbedingungen**

Die ETH-Bibliothek ist Anbieterin der digitalisierten Zeitschriften. Sie besitzt keine Urheberrechte an den Inhalten der Zeitschriften. Die Rechte liegen in der Regel bei den Herausgebern.

Die auf der Plattform e-periodica veröffentlichten Dokumente stehen für nicht-kommerzielle Zwecke in Lehre und Forschung sowie für die private Nutzung frei zur Verfügung. Einzelne Dateien oder Ausdrucke aus diesem Angebot können zusammen mit diesen Nutzungsbedingungen und den korrekten Herkunftsbezeichnungen weitergegeben werden.

Das Veröffentlichen von Bildern in Print- und Online-Publikationen ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Die systematische Speicherung von Teilen des elektronischen Angebots auf anderen Servern bedarf ebenfalls des schriftlichen Einverständnisses der Rechteinhaber.

## **Haftungsausschluss**

Alle Angaben erfolgen ohne Gewähr für Vollständigkeit oder Richtigkeit. Es wird keine Haftung übernommen für Schäden durch die Verwendung von Informationen aus diesem Online-Angebot oder durch das Fehlen von Informationen. Dies gilt auch für Inhalte Dritter, die über dieses Angebot zugänglich sind.

# Vom Quellentext zur Datenbank – ein Konzept zur integrierten Verarbeitung quantitativer und qualitativer Daten<sup>1</sup>

---

Stephan Hagnauer und Niklaus Bartlome

Der Computer ist in der Geschichtswissenschaft als Arbeitsinstrument längst üblich geworden. Viele Historikerinnen und Historiker nutzen ihn vorwiegend oder ausschliesslich für die Textverarbeitung. Im Bereich quantitativer Studien werden zudem Datenverarbeitungssysteme eingesetzt, deren Stärke in der Regel in der Bearbeitung umfangreicher, als Zahlen vorliegender Informationen liegt. Solche Systeme sind damit allerdings eher auf die Bedürfnisse von Ökonomen oder Naturwissenschaftlern zugeschnitten und für Historiker/-innen nur teilweise geeignet, da in der Geschichtswissenschaft eine quantitative Analyse meist mit qualitativen Aspekten verknüpft wird. Hier soll nun eine Möglichkeit dargestellt werden, wie mit Hilfe herkömmlicher Software eine Kombination beider Methoden erreicht werden kann.

Das nachfolgend skizzierte Datenverarbeitungs-Konzept eignet sich also für Untersuchungen, bei denen die *systematische, quantitative* Auswertung einer Quelle mit einer *qualitativen* Textanalyse verbunden wird. Dies gilt beispielsweise auch für das Projekt zur Erforschung der Geschichte der Berner Staatsfinanzen,<sup>2</sup> für welches das hier beschriebene Verfahren entwickelt wurde. Zunächst waren Hunderte von Jahresrechnungen verschiedener Ämter mit insgesamt Zehntausenden von Buchungssätzen zu erfassen und auszuwerten. Diesen Quellen eignet jener Doppelcharakter, der oben beschrieben wurde. Sie erlauben einerseits eine Analyse der finanziellen Aspekte der staatlichen Tätigkeit,<sup>3</sup> geben also beispielsweise Aufschluss über die Besoldung der einzelnen Beamten, über die Auslagen der öffentlichen Hand für den Unterhalt der Gebäude, Strassen und Brücken oder über die Höhe der Einnahmen aus Zöllen. Andererseits sind in diesen Quellen auch zahlreiche qualitative Informatio-

---

1 Überarbeitete Fassung des Artikels: Hagnauer, Stephan: «Die Auswertung von Textquellen und quantifizierbaren Daten in einem textorientierten Datenverarbeitungskonzept». In: Guex, Sébastien; Körner, Martin; Tanner, Jakob (Hgg.): *Staatsfinanzierung und Sozialkonflikte (14.-20. Jh.)*. Zürich 1994, S. 87-104.

2 Leitung: Prof. Dr. Martin Körner (Bern).

3 Solche Analysen finden sich beispielsweise in Körner, Martin: *Luzerner Staatsfinanzen 1415-1798. Strukturen, Wachstum, Konjunkturen*. Luzern/Stuttgart 1981; und in Hagnauer, Stephan: *Die Finanzhaushalte der bernischen Ämter Aarberg, Büren, Erlach und Nidau in den Jahren 1631-1635 und 1681-1685. Elemente zur Geschichte der bernischen Staatsfinanzen*. Unveröffentlichte Lizentiatsarbeit Universität Bern. Bern 1995.

nen verborgen. Aus der Rechnung des Amtes Aarberg von 1683-1684 erfahren wir beispielsweise, dass damals auch ein Landvogt – wie hier der Patrizier Viktor von Erlach – nicht vor Wanzen und Wandläusen sicher war.<sup>4</sup> Aus anderen Rechnungen lassen sich Speisezettel<sup>5</sup> rekonstruieren oder kann Verlauf und Art der Kriminalität<sup>6</sup> verfolgt werden.

In diesen Quellen sollten nun nicht nur die quantitativen Angaben (Finanzielles), sondern auch die qualitativen Informationen erfasst werden. Da diese qualitativen Angaben aber sehr unterschiedlich und kaum oder gar nicht strukturiert sind, drängte sich bei der Datenaufnahme die vollständige Transkription der Texte auf. Bei herkömmlichen seriellen Untersuchungen wird dagegen oft schon im Archiv eine bestimmte Datenreihe isoliert und direkt aufgenommen. Demgegenüber bewahrt eine Abschrift der Quelle den originalen Kontext sämtlicher Informationen. Das Problem des Umgangs mit verschiedenartigen Informationen lässt sich somit zunächst noch hinausschieben, denn die vollständige Abschrift der Texte erlaubt auch die Trennung der Datenerfassung (Transkription) von der Klassifikation der Informationen.

Zur Erfassung und Verarbeitung der Daten wurde ein Verfahren entwickelt, das auf einem normalen PC angewendet werden kann und kein Grosssystem notwendig macht. Damit ist für jeden Mitarbeiter ein individuelles Arbeiten – unabhängig von Zugangsrestriktionen oder Ortsbeschränkungen – möglich. Ausserdem baut das Verfahren auf weit verbreiteter Standardsoftware auf.

## 1. Das Datenverarbeitungs-Konzept im Überblick

Das Datenverarbeitungs-Konzept beruht auf dem Grundsatz, die Quelldaten in ihren originalen Textstrukturen in eine Computer-Textdatei überzuführen. Erst in einem zweiten Schritt werden die Daten klassifiziert und die darin enthaltenen Informationen markiert. Für die Auswertung werden die bearbeiteten Texte darauf in ein Informations-Management-Programm (free-form-Datenbank) und schliesslich in eine relationale Datenbank exportiert.

---

4 Staatsarchiv Bern, B VII, 850.

5 Eine solche Untersuchung an Hand von Rechnungen bietet Rippmann, Dorothee: «Dem Schlossherrn in die Küche geschaut: Zur Ernährung im Spätmittelalter und in der Frühen Neuzeit». In: *Geschichte 2001*. Mitteilungen der Forschungsstelle Baselbieter Geschichte Nr. 15 (Beilage zu den Baselbieter Heimatblättern), Liestal 1994, S. 1-12.

6 Vgl. dazu etwa Bartlome, Niklaus: «Zur Bussenpraxis in der Landvogtei Willisau im 17. Jahrhundert». In: *Jahrbuch der Historischen Gesellschaft Luzern* 11, 1993, S. 2-15.

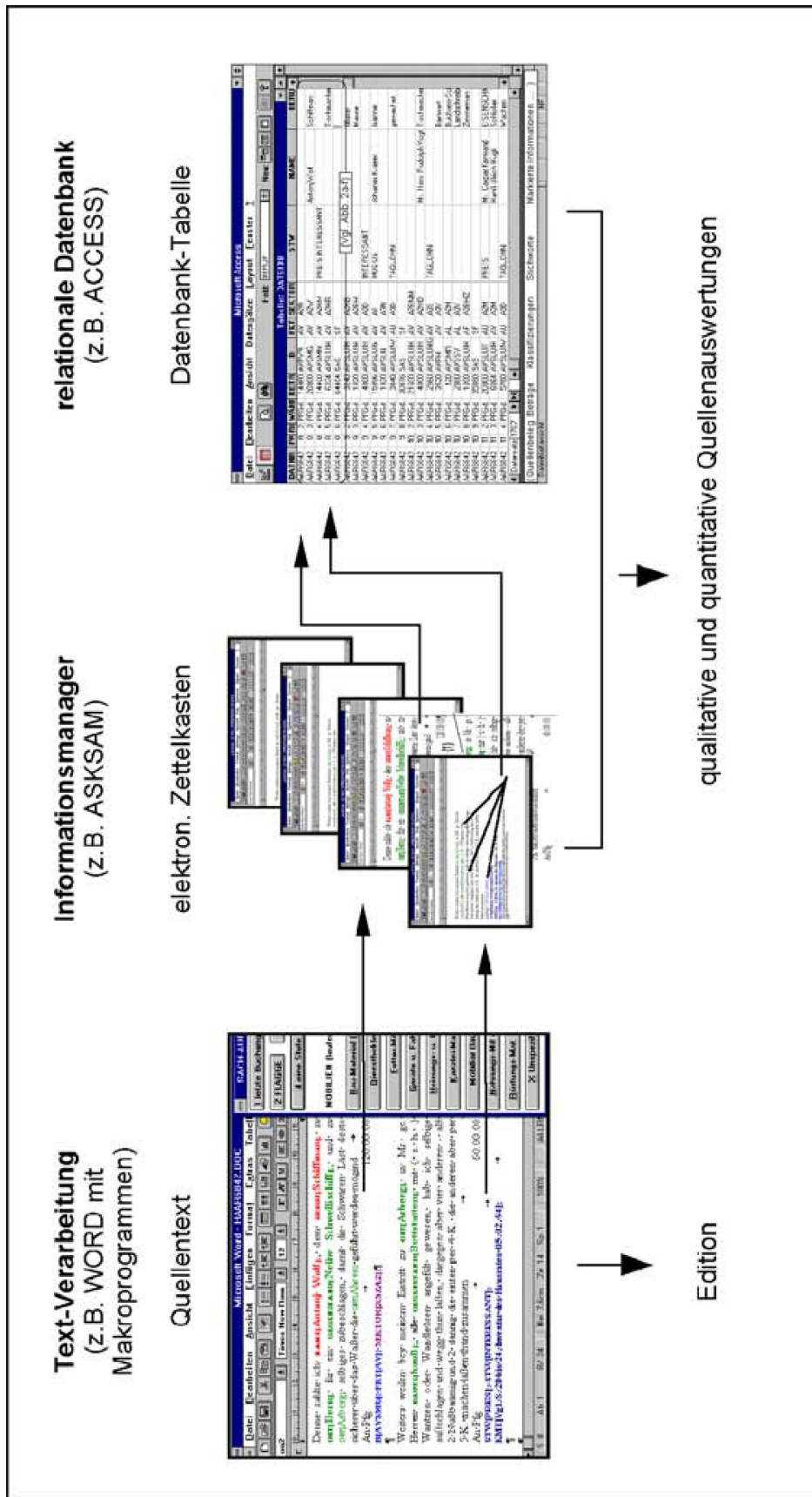


Abb. 1: Das Datenverarbeitungs-Konzept in seinen Grundzügen: Die Informationen aus dem Quellentext werden über den elektronischen Zettelkasten in die relationale Struktur der Datenbank übergeführt.

Ein Bearbeiter kann somit nicht nur alle statistischen Auswertungen vornehmen, die Datenbanken üblicherweise erlauben, er kann auch jederzeit auf den originalen Datenkontext zurückgreifen (vgl. Abb. 1). Die bearbeiteten Quellen können zudem von weiteren Anwendern (Forschungsprojekte, Unterricht) genutzt werden, wobei sie sich wahlweise in elektronischer Aufbereitung oder als gedruckte Transkriptionen zur Verfügung stellen lassen.

Textverarbeitung, Informationsmanager und Datenbank sind durch die gemeinsame WINDOWS-Grundlage und spezielle Programm-Module verbunden.<sup>7</sup> Das integrierte Datenverarbeitungs-Konzept, bei dem analytische Kriterien erst nachträglich auf weitgehend originalgetreu erfasste Grundinformationen angewendet werden, erlaubt den dezentralen Einsatz von Mitarbeitern und Mitarbeiterinnen bei der Quellenerfassung und beliebig ausführliche Auswertungen der Informationen. Wiederkehrende Arbeitsroutinen lassen sich mit einfachen Muster-Erkennungs-Algorithmen und abgespeichertem Expertenwissen automatisieren.

Die Programm-Module nutzen die Fähigkeiten der graphischen Oberfläche von WINDOWS, die dem Benutzer die Programmfunktionen als Bildsymbole und in Auswahlmenüs auf dem Bildschirm präsentiert, so dass er die jeweils gewünschten Funktionen nur noch mit der Computermaus anzutippen braucht und nicht genötigt ist, sich durch umfangreiche Lektüre von Programmhandbüchern mit unterschiedlichen Befehlscodes vertraut zu machen. Es entbehrt somit nicht einer gewissen Ironie, wenn nun einzelne Schritte eines über Bildsymbole gesteuerten Programmablaufs dennoch rudimentär in Worten beschrieben werden.

## **2. Ausgewählte Verfahrensschritte**

### *2.1. Erfassen und Edieren der Originalquellen*

Während sich heute gedruckte Quellen mittels elektronischer Lesegeräte (Scanner) und Optical-Character-Recognition (OCR) in Textdateien überführen lassen, wird man bei handschriftlichen Quellen vorläufig<sup>8</sup> noch

---

7 Technisch gesehen handelt es sich bei diesen Modulen um ca. 220 kByte (ca. 120 Seiten) Makro-Programme für WORD-FÜR-WINDOWS (MICROSOFT Corp.), welche die Rekonstruktion, Aufbereitung und den Export von Textinformationen ermöglichen, damit die Daten mit dem Informationsmanager ASKSAM-FÜR-WINDOWS (NORTH AMERICAN SOFTWARE, München) und mit der relationalen Datenbank ACCESS-FÜR-WINDOWS (MICROSOFT) weiterbearbeitet und über die Konvertierungsfunktionen der WINDOWS-Applikationen in zahlreiche andere Datenformate weiterexportiert werden können. Das Programm wurde entwickelt für MS-DOS-Computer mit folgender Minimalausstattung: Prozessor 80386, 4 MByte RAM, 100 MByte Harddisk.

8 Die Bemühungen um Handschriften-Erkennungsprogramme nähren die Hoffnung auf automatische Digitalisierung historischer Manuskripte: vgl. Helsper, Eric L.; Schomaker, Lambert R.; Teulings,

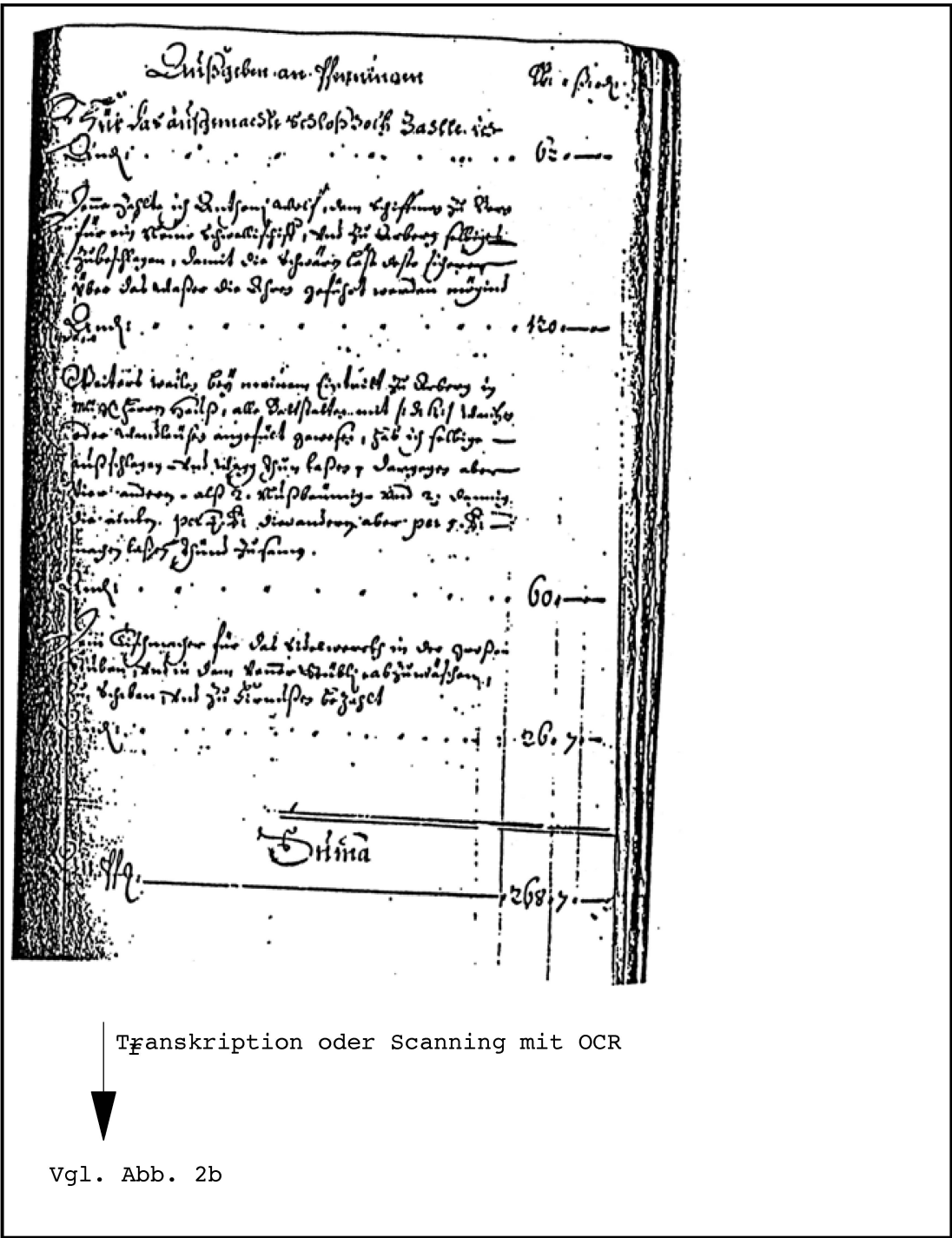
kaum darum herumkommen, die Texte unter Einsatz komfortabler Textverarbeitungsprogramme zu transkribieren, also z.B. im Archiv abzutippen. Damit zeigt sich, dass bei einem textorientierten Datenerfassungskonzept gerade dieser erste Verfahrensschritt den grössten Anteil am Arbeitsaufwand haben kann. Es gilt somit, den anfänglichen Transkriptionsaufwand zu minimieren: Wiederkehrende Textstrukturen lassen sich als Textbausteine abspeichern, damit diese später mit Kopierbefehlen wiederverwendet werden können. Bei seriellen Quellen empfiehlt es sich zuweilen sogar, eine vollständige Kopie einer bereits erfassten Datei als Vorlage für die Transkription des nachfolgenden Texts zu benützen, womit nur noch die abweichenden Stellen abzuändern sind.

Der so reduzierte Erfassungsaufwand sollte durch einen vielfachen Nutzen aufgewogen und übertroffen werden:

- Weil die Quellen vorerst ohne Klassifizierung in ihrem Wortlaut transkribiert werden, kann die Erfassung der handschriftlichen Quelleninformationen mit einfachsten Textprogrammen erfolgen und auf mehrere Mitarbeiter/-innen aufgeteilt werden. Diese brauchen sich nicht a priori an Klassifizierungsrichtlinien zu halten, sondern machen sich bei der Transkription erst mit dem Material vertraut und bilden dabei Fragestellungen und analytische Kriterien, die in den anschliessenden Verarbeitungsschritten auf den gesamten Quellenbestand angewendet werden.
- Die einmalige, aufwendige Erfassung der Quellentexte ermöglicht spätere Mehrfachnutzungen. Nicht nur die gerade aktuellen Fragestellungen, sondern auch spätere Forschungsprojekte können diese Vorarbeiten weiterverwenden, weil die Quellen (abgesehen von ihrer handschriftlichen Gestalt) in ihrer originalen Komplexität erfasst werden und sich ohne interpretatorische Veränderungen als gedruckte oder elektronische Texttranskriptionen zur Verfügung stellen lassen.
- Da ja die Einzeldaten nicht aus den originalen Kontexten herausgerissen, sondern mit diesen zusammen erfasst werden, erlaubt das integrierte Datenverarbeitungsverfahren stets auch unter wandelnden Perspektiven den Rückgriff auf diese Kontexte, die sich zudem im Informationsmanager bequem sortieren und nach Wunsch als Zitate in die Darstellungen einbauen lassen.

---

Hans-Leo: «Tools for the Recognition of Handwritten Historical Documents». In: *History and Computing*, Vol. 5, No. 2, 1993, (Special Issue: Scanning and OCR), S. 88-93; mit weiterer Literatur.



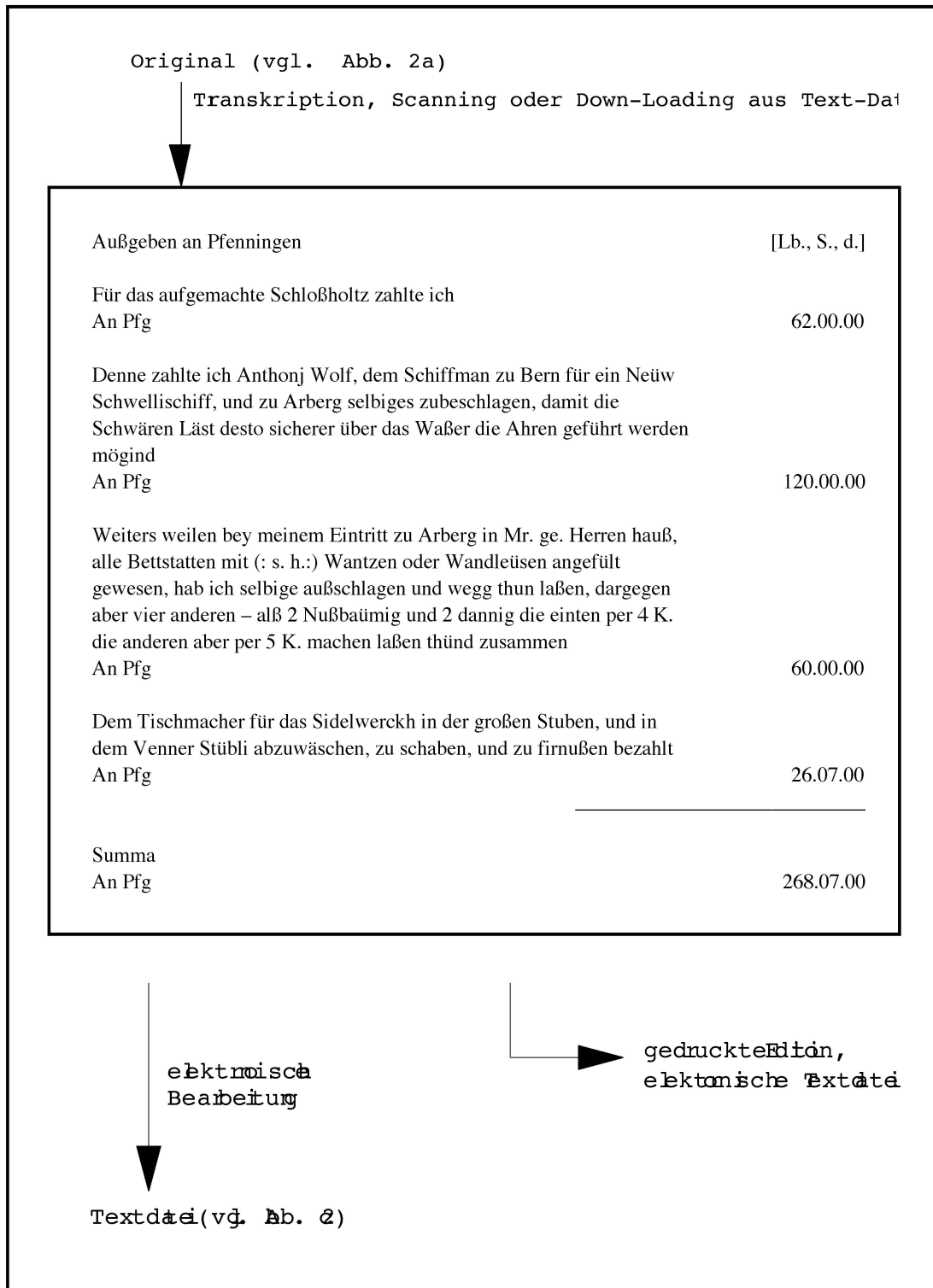


Abb. 2b: Der Text in (wie hier) vollständig oder selektiv-partiell transkribierter Form. (Quelle: vgl. Abb. 2a).



## 2.2. Markieren, Verstichworten und Kommentieren von Informationen in den Textquellen

Ähnlich wie beim weit verbreiteten Markieren von Textstellen mit Leuchtstiften auf Papier, lassen sich auch am Bildschirm in den zuvor abgespeicherten Textdateien Informationen markieren (Abb. 2c: {1}), indem man diese mit der Maus zuerst antippt und dann mit den am Bildrand symbolisierten «elektronischen Leuchtstiften» verbindet, wobei nun aber die farblich markierten Informationen (z.B. «Arberg») zusätzlich automatisch mit den Bezeichnungen der verschiedenfarbigen Markierungen (z.B. `ORT[...]`) verbunden werden.

Das Markieren am Bildschirm weist damit weit über das Markieren auf Papier hinaus:

- Da die am Bildschirm entstehenden Markierungen digitaler Natur und also maschinenlesbar sind, können so markierte Informationen in den Texten fürderhin von Programmen automatisch erkannt und weiterverarbeitet werden.
- Durch die Zuordnung der im Text markierten Informationen zu bestimmten Bezeichnungen – z.B. `ORT[Arberg]` – werden diejenigen wichtigen Grundstrukturen rekonstruiert, welche den in Datenbanken üblichen Relationen zwischen Feldnamen (`ORT`) und Feldinhalten (Arberg) entsprechen und welche formal genau auf die vom Informationsmanager ASKSAM verwalteten Feldstrukturen angepasst sind, womit eine fundamentale Voraussetzung für die Integration von Text, Informationsmanager und Datenbank geschaffen wird.

Ganze Textabschnitte können am Bildschirm einem oder mehreren Stichworten zugeordnet werden (Abb. 2c: {2}), indem diese mit der Maus im Bildschirm-Auswahlmenü angeklickt werden. Ein Makroprogramm setzt darauf Stichworte in der Form «`STW[...]`» wie «Flaggen» oder «Etiketten» automatisch ans Absatzende.

Entsprechend den Randnotizen in Büchern, mögen Textabschnitte auch mit ausführlichen<sup>9</sup> Kommentaren oder Hinweisen versehen werden, die automatisch durch spezielle Formatierung und durch die Bezeichnung «`KMT[...]`» vom Originaltext abgegrenzt und mit dem Bearbeitungsdatum versehen werden (Abb. 2c: {3}).

---

9 Unbeschränkte Zahl von Kommentaren zu je maximal 32'000 Zeichen.

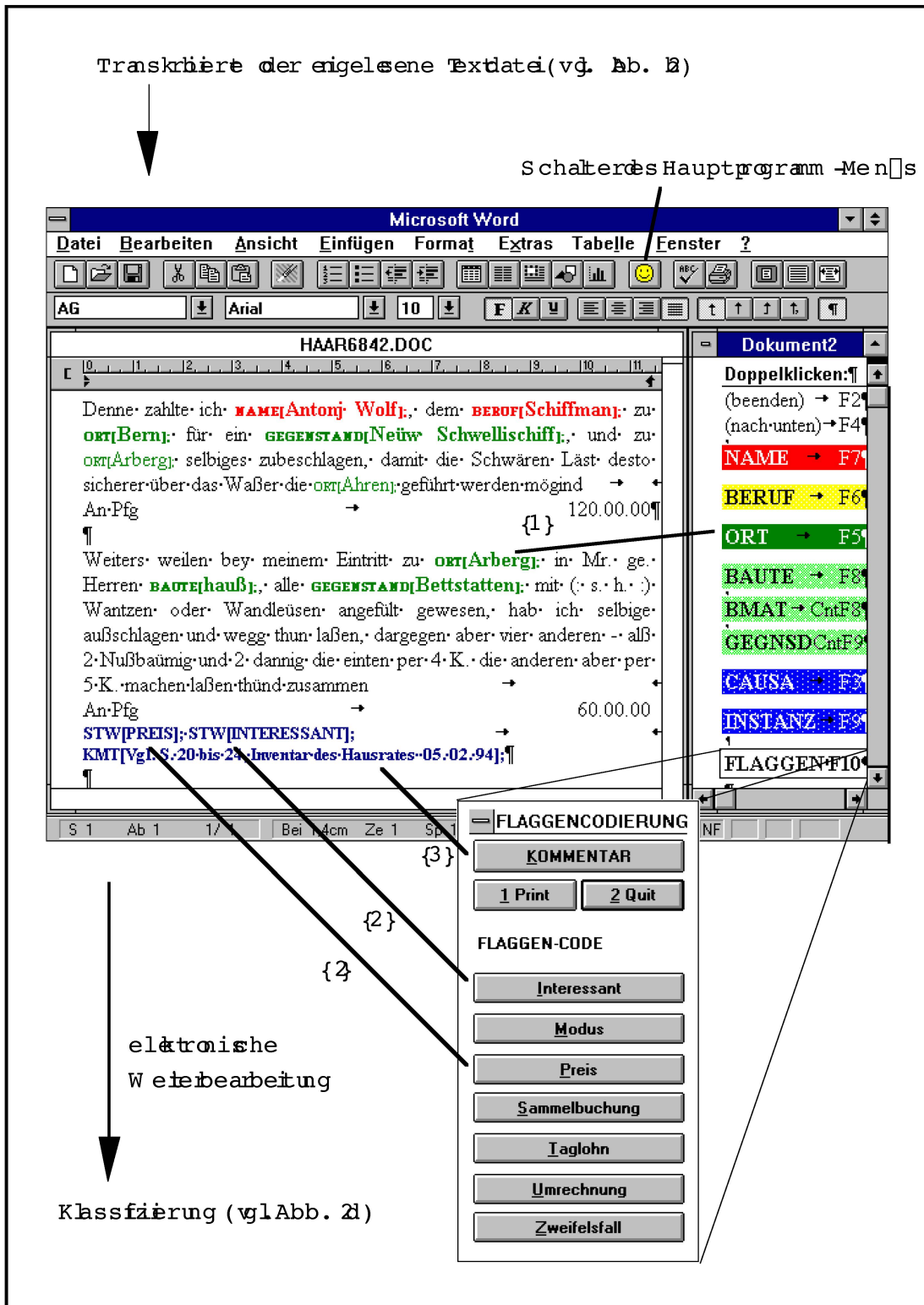


Abb. 2c: Der Text w&ahrend der (wahlweise farbigen) Markierung {1}, Verstichwortung {2} und Kommentierung {3} von Textinformationen. [Flaggenfenster montiert]

### 2.3. Klassifizieren von Daten

Zusätzlich zur Ad-hoc-Markierung und Kommentierung einzelner Textinformationen ist auch das konsequente, systematische Klassifizieren (Verschlagworten) von Textabschnitten am Bildschirm durchführbar, soweit die gewünschten Klassifikationssysteme als einfache oder hochgradig hierarchisierte Listen im Wissensschatz des Programms abgespeichert werden.<sup>10</sup> Der Bearbeiter lässt sich nacheinander jeden Textabschnitt auf dem Bildschirm darstellen und tastet sich dazu mit der Maus solange durch die am Bildrand in hierarchischer Abfolge präsentierten Ebenen (Verästelungen) der Klassifikation, bis der Textabschnitt in der feinsten Unterklasse eindeutig zugeordnet ist, worauf der Abschnitt automatisch mit dem entsprechenden Klassifikationscode gekennzeichnet wird.

Das Klassifizieren gestaltet sich am Bildschirm somit ziemlich komfortabel:

- In unbekanntem Bereich der Klassifikation kann man sich jederzeit an den auf dem Schirm dargestellten Untereinheiten orientieren. Wer sich dennoch in der Baumstruktur der Klassifikationen verirrt, kehrt auf die übergeordneten Äste zurück.
- Weil sich die Auswahl von Unterklassen durch Eingabe von Initialen abkürzen lässt, können vertraute Benutzer in den ihnen bekannten Bereichen der Klassifikation vorübergehend von der Bildschirmunterstützung absehen und sehr schnell mit Tastaturkürzel klassifizieren.
- Gehört ein nachfolgender Textabschnitt zu derselben Klasse wie der zuletzt bearbeitete, löst ein einziger spezieller Tastaturschlag die entsprechende Kennzeichnung aus.

Weiter zeigen sich gerade bei der Klassifikation von Informationen wesentliche Vorteile eines textorientierten Datenverarbeitungs-Konzeptes:

- Über den im originalen Wortlaut erfassten Quellenbestand können folienartig auch noch nach Jahren zusätzliche Klassifikationssysteme gelegt werden,<sup>11</sup> wobei bereits vorhandene Bearbeitungsvermerke wahlweise weiterverwendet oder ausgeblendet werden können.
- Die Quellenauswertung gewinnt dabei durch neue Klassifikationssysteme, die unabhängig von den bisherigen angewendet werden, weitere Möglichkeiten der Differenzierung, womit die Darstellung zusätz-

---

<sup>10</sup> Die in den bernischen Jahresrechnungen untersuchten Finanzvorfälle werden systematisch nach 212 Buchungskonten (in 7 hierarchischen Stufen), 14 Staatsfunktionen und 31 Volkswirtschaftszweigen klassifiziert, wodurch theoretisch über 50'000 Zuordnungen möglich sind, wovon vermutlich einige Hundert von wirklicher Bedeutung sind.

<sup>11</sup> Was weitreichende Wirkungen auf den hermeneutischen Zirkel zu entfalten scheint.

liche Dimensionen erhält. Zudem können zu bestehenden Kategorien jederzeit weitere, zusätzliche Untereinheiten hinzugefügt werden, womit die Verästelung der Quellenauswertung erhöht und verfeinert wird.

Textdatei (vgl. Abb. 2c)

Textabschnitt, der bereits nach Buchungskor klassifiziert wurde. Danach automatische Zu zur Staatsfunktion FKT[ ] und zum Wirtschaft soweit möglich (provisorischer Rezeptor SEK

The screenshot shows a Microsoft Word window titled 'Microsoft Word - HAAR6842.DOC'. The main text area contains several lines of text with classification codes in various colors: red for 'NAME|Antoni Wolf;', blue for 'BERUF|Schiffman;', green for 'ORT|Bern;', yellow for 'GEGENSTAND|Neuw Schwellischiff;', and purple for 'AVSMG|FKT|SEKTOR|XSZA2;'. The text is partially obscured by a diagonal line. To the right of the text is a sidebar titled 'SACH-AUFWAND' containing a list of categories: '1 letzte Buchung stornieren', '2 FLAGGE', '3 Quit', '4 eine Stufe zurück', 'MOBILIEN (laufend)', 'Bau-Material (laufend)', 'Dienstbekleidung', 'Futter-Mittel', 'Geräte u. Fahrzeuge', 'Heizungs- u. Bel-Mat', 'Kanzlei-Material', 'Mobilier (laufend)', 'Nahrungs-Mit oh. Fron', 'Rüstungs-Mat. (laufend)', and 'X Unspezifiziert'. Below the screenshot, there is a vertical arrow pointing down from the text 'zum automatischen Erschuppen und Bearbeiten von Informationsstrukturen' to the text 'Vgl. Abb. 2e'. To the right of this arrow, there is a block of text: 'Bildschirmgestützte Klassifikation durch interaktive Top-down-Auswahl der zugehörigen Unterklassen aus hierarchisch geordneten Klassifikationssystemen'.

zum automatischen Erschuppen und Bearbeiten von Informationsstrukturen

Vgl. Abb. 2e

Bildschirmgestützte Klassifikation durch interaktive Top-down-Auswahl der zugehörigen Unterklassen aus hierarchisch geordneten Klassifikationssystemen

Abb. 2d: Der Text während der Klassifikation von Textabschnitten am Bildschirm.

Dass die computergestützte Klassifikation von Texten erhebliche Rationalisierungen erlaubt, zeigt sich bei der Auswertung bernischer Staatsrechnungen nach den drei Kriterien Buchungskonto (Natur des Finanzvorfalls), Staatsfunktion und Volkswirtschaftszweig: Während der Bearbeiter einen Buchungssatz nach Buchungskonto klassifiziert, vergleicht das Programm die Eingabe mit seinem Wissen über die Relationen zwischen den drei Klassifikationen und nimmt selbständig Zuordnungen des Buchungssatzes zu den Staatsfunktionen und Volkswirtschaftszweigen vor, soweit die Relationen eindeutig sind. Dem Bearbeiter werden somit nur noch jene Zuordnungsfragen vorgelegt, die seinen Entscheid erfordern.<sup>12</sup> Greift beispielsweise das Programm aufgrund des eingegebenen Buchungskontos für <Ausgaben/Ausgaben der Verbrauchsrechnung/Sachaufwand/Mobilien/Geräte u. Fahrzeuge (mit Staatsfunktion [Verwaltung]) auf die Annahme zurück, dass es sich um einen Kapitalfluss in den 2. Sektor handelt, aber keine eindeutige Zuordnung zu untergeordneten Wirtschaftszweigen möglich ist, wird automatisch der provisorische Code <SEKTOR[XSZA2] als Rezeptor in den Text gesetzt. Anhand dieser Rezeptoren steuert das Programm den Bearbeiter bei der nachfolgenden volkswirtschaftlichen Klassifikation bloss zu jenen ca. 30% der Buchungssätze, die er noch präziser einer Unterklasse zuzuordnen hat.

#### *2.4. Automatisches Entdecken und Bearbeiten von Informationsstrukturen durch einfache, wissensbasierte Muster-Erkennungsalgorithmen*

In den als Texten aufgenommenen Quellen lassen sich mit Muster-Erkennungsalgorithmen Informationen erschnuppern, wobei wiederkehrende Muster im Expertenwissen des Programms abgespeichert und Routinearbeiten automatisiert werden können (Vgl. Abb. 2e).

---

12 Von der aufwendigen Möglichkeit, die Klassifikation derart weiter zu automatisieren, dass das Programm vollautomatisch durch semantische Analyse der Buchungssätze klassifizieren sollte, wurde abgesehen, weil eine ohne jegliche Mitwirkung des Bearbeiters vorgenommene Klassifikation aufgrund der orthographischen und inhaltlichen Variabilität der Buchungssätze hätte nachkontrolliert werden müssen, was vermutlich mit einem höheren Zeitaufwand verbunden gewesen wäre, als bei dem im Text beschriebenen halbautomatischen Verfahren.

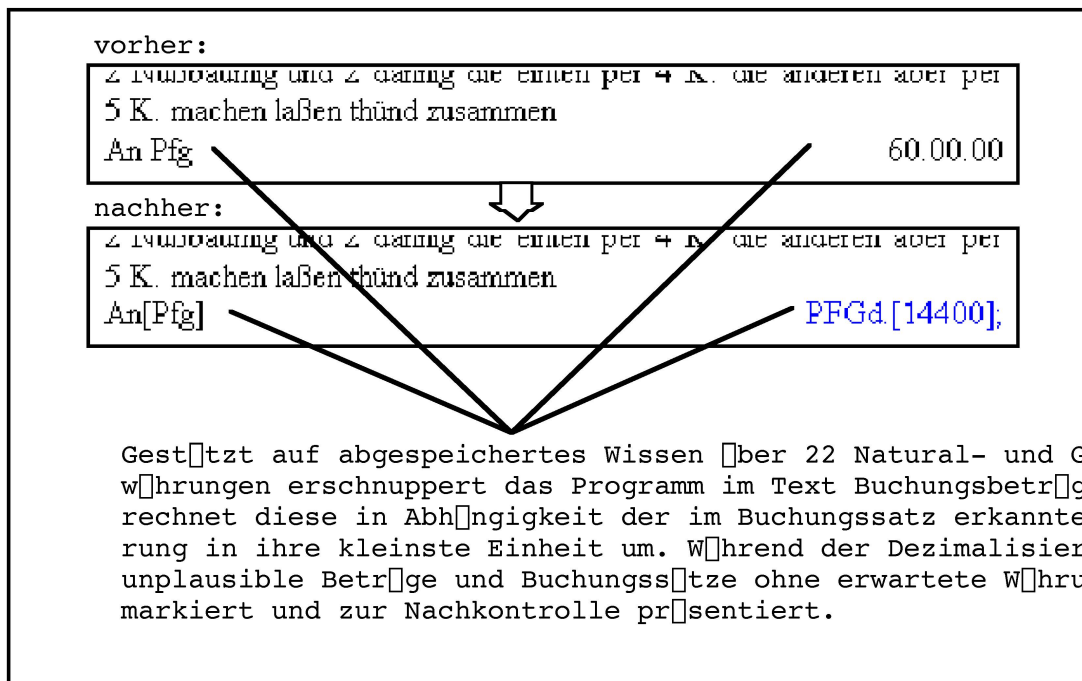


Abb. 2e: Beispiel für das automatische Entdecken und Bearbeiten von Texten.

Weil die automatisierten Programmfunktionen verborgen formatierte Bearbeitungsvermerke (mit Datum) in der Textdatei hinterlassen, lassen sich die bisher durchgeführten Quellenbearbeitungen automatisch rekonstruieren und der weitere Ablauf der Bearbeitung durch Autopilotfunktionen steuern.

Um die nachfolgende Überführung des Textes in den Informations-Manager vorzubereiten, wird jeder Textabschnitt mit den in der Kopfzeile der Quelle festgehaltenen bibliographischen Merkmalen sowie mit der Nummer der Seite und des Textabschnittes versehen, damit jeder Abschnitt auch nach der Fragmentierung der Quelle auf elektronische Karteikarten identifizier- und zitierbar bleibt (vgl. Abb. 2f).

### 2.5. Fragmentieren des Textes auf <elektronische Karteikarten> des Informationsmanagers

Die Texte werden bei der Übernahme in den Informationsmanager ASKSAM in ihren originalen Satzstrukturen belassen, aber automatisch nach Abschnitten oder individuellen Trennmarken auf einzelne <elektronische Karteikarten> (sog. «Dokumente») fragmentiert, womit der in der Textverarbeitung betriebene Aufwand intensiv genutzt werden kann.

Von den zahlreichen Funktionen des Informationsmanagers seien hier nur diejenigen aufgezählt, die für sozial- und geisteswissenschaftliche Arbeiten äusserst wirkungsvolle Werkzeuge darstellen:

- Hyper-, Voll- und Kontextsuchen mit kombinierten booleschen Abfragen;
- Selektion nach Worten oder Feldinhalten;
- Ausgabe von selektierten, gruppierten oder aggregierten Daten und Textabschnitten;
- Export von Feldinhalten allein oder zusammen mit ihren Kontexten nach kombinierten Kriterien sortiert und summiert in Texte oder in Tabellen etc.

Die Vorteile des Informationsmanagers gegenüber herkömmlichen Karteien sind dabei augenfällig: Sollen die Karten einer klassischen Kartei nach einem neuen Kriterium geordnet werden, muss dazu bekanntlich die alte Ordnung zerstört werden. Demgegenüber ist es im elektronischen Informationsmanager möglich, nach beliebigen Kriterien (wie z.B. nach den vorangegangenen Markierungen oder Klassifizierungen) reversibel die Daten – isoliert oder zusammen mit ihren originalen Kontexten – immer wieder neu auszuwählen, zu gruppieren, zu summieren sowie als Dateien abzuspeichern oder auszudrucken.

Im Gegensatz zu den weiter verbreiteten, tabellenartig aufgebauten Datenbanken (wie DBASE oder ACCESS) verwaltet der Informationsmanager sowohl strukturierte als auch unstrukturierte Informationen und erlaubt, je <Karteikarte> (Record) dasselbe Feld (z.B. WIRKUNG[ . . . ]) nicht nur einmal, sondern auch mehrmals oder nie aufzuführen, was gerade im sozialwissenschaftlichen Bereich von Bedeutung ist. Die Anzahl, Typen und Merkmale der Felder brauchen nicht im voraus definiert zu werden.

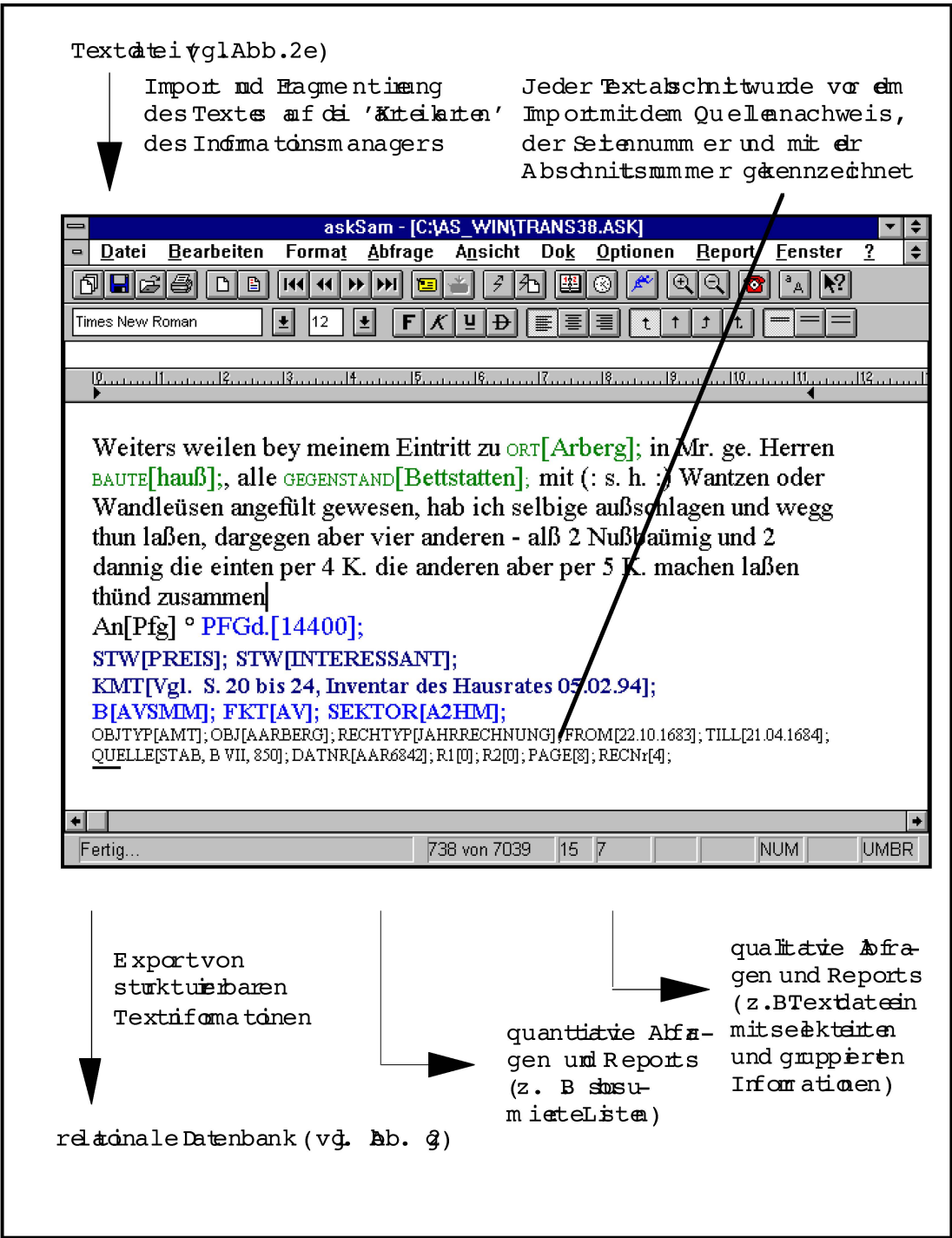


Abb. 2f: Der Text nach der automatischen Fragmentierung in Textabschnitte und Überführung in «elektronische Karteikarten» des Informationsmanagers. Gezeigt ist die 738. von 7039 «Karteikarten», auf die ca. 1200 Seiten Text kopiert wurden.



## 2.6. Überführen von Textinformationen in die Tabellenstruktur der relationalen Datenbank

In ASKSAM-FÜR-WINDOWS können durch sogenannte Reportbefehle die vorher bearbeiteten Daten von den <elektronischen Karteikarten> in die Tabellenstruktur einer relationalen Datenbank (z.B. ACCESS) übertragen werden, wobei je <Karteikarte> (und somit indirekt je Quellenabschnitt) eine Tabellenzeile (Record) eröffnet wird.<sup>13</sup>

Damit können die vorgängig in den Texten markierten und klassifizierten Informationen nun auch mit den mächtigen Instrumenten einer relationalen Datenbank ausgewertet werden, wobei hier speziell die Möglichkeiten hervorgehoben seien,

- die Daten zu selektieren, zu sortieren, zu subsumieren, in Tabellenform, Berichten oder als Grafiken darzustellen sowie über gemeinsame Felder relational mit Hintergrunddaten zu verbinden und zu vergleichen;
- durch sogenannte Kreuztabellen aussagekräftige Übersichten über ausgewählte Daten und deren Zusammenhänge zu erstellen;
- komplexe Datenanalysen oder -bearbeitungen durch Makro-Programme zu automatisieren.

---

13 Das Kopieren der Daten aus der Textverarbeitung in den Informationsmanager und in die Datenbank ist Voraussetzung dafür, dass die spezifischen Vorteile und Instrumente der jeweiligen Programme zur Auswertung genutzt werden können, verstößt aber gegen den Grundsatz, dass Informationen möglichst nur einmal vorliegen sollen. Nachträgliche Änderungen an den Daten (z.B. Korrekturen) müssen somit in die beiden ändern Programme kopiert werden, wenn deren Daten auch auf den früheren und späteren Bearbeitungsstufen à jour gehalten werden sollen. Es ist deshalb zu wünschen, dass künftige Informationsmanager ihre Text-, Datenbank- und Programmierfunktionen so weit ausbauen, dass das hier dargestellte Konzept auf einem Programm allein aufbauen könnte.

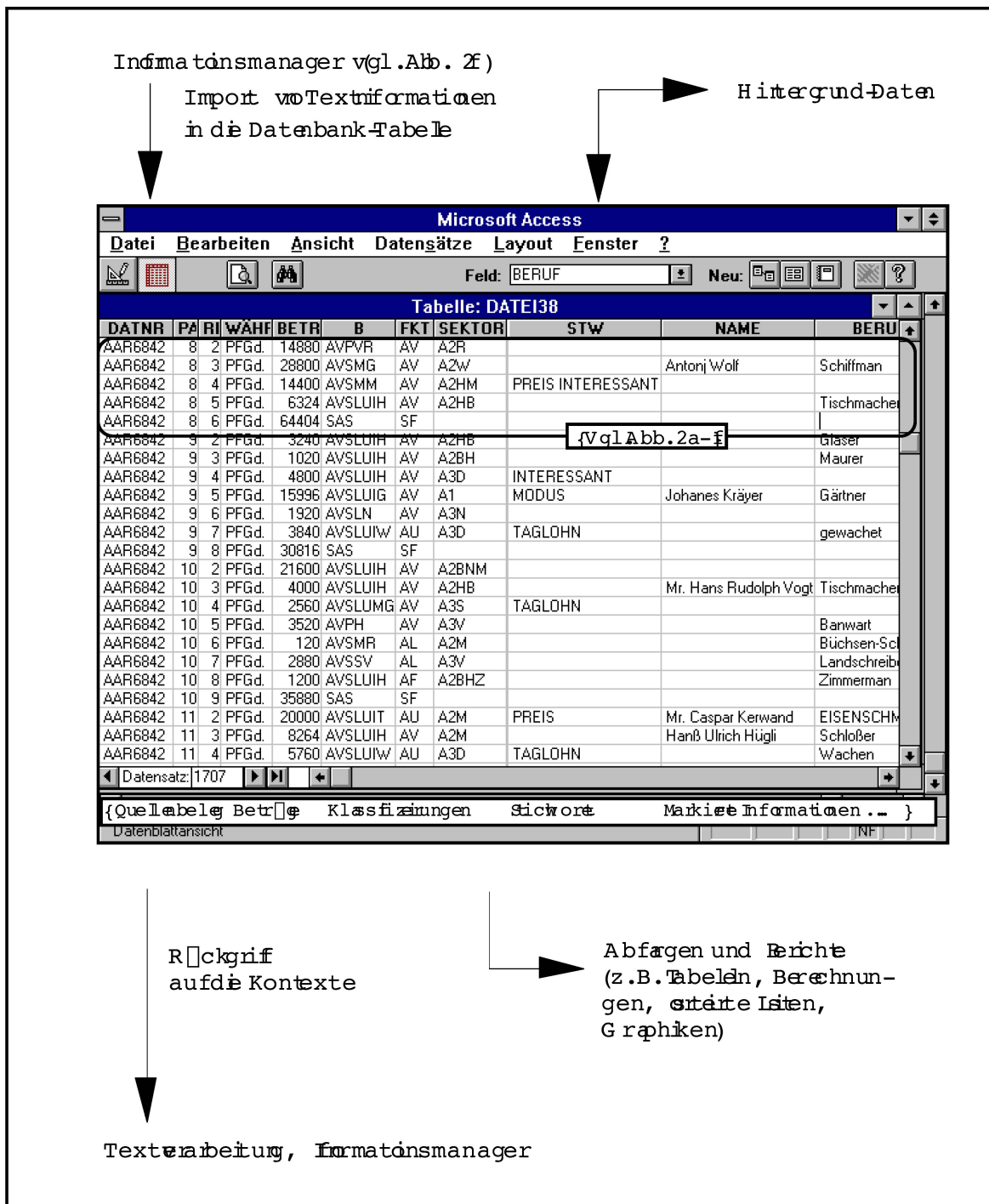


Abb. 2g: Die strukturierten Textinformationen (Ausschnitt) nach der automatischen Überführung in die Tabellenstruktur der relationalen Datenbank.

### 3. Schlussbemerkungen

Textorientierte Datenverarbeitungs-Konzepte erschliessen heterogen strukturierte qualitative und quantitative Textinformationen. Sie drängen sich dann nicht auf, wenn handschriftliche Quellen mit geringer Informationsdichte durch einen Einzelbearbeiter unter kurzfristiger Perspektive mit von vornherein abschliessend definierten Fragestellungen und Kriterien ausgewertet werden sollen.

In den anderen Fällen bieten sie sich insbesondere deswegen an, weil sich die Informationserfassung von der anschliessenden Klassifikation und Interpretation trennen lässt, was bei tabellenorientierten Datenbankverfahren oft nicht möglich ist, weil dort die Informationen zu Beginn aus dem Quellenkontext herausgerissen und in vorgängig zu definierende Feldstrukturen der Datenbank übergeführt werden müssen. Der modulare Aufbau des oben dargestellten Konzepts erlaubt es dagegen, die Programmfunktionen an unterschiedliche Quellen anzupassen und jederzeit neue analytische Kriterien alternativ oder kumulativ anzuwenden, falls dies revidierte oder neue Fragestellungen erfordern sollten. Weil die Texte als Transkription vorliegen, können alle Informationen stets am Originalwortlaut überprüft werden. Zudem ist deswegen der Datenbestand mehrfach verwendbar und somit langfristig auch für neue Projekte nutzbar.

Mit der Verbindung von quantitativer mit qualitativer Analyse kombiniert das oben vorgestellte Datenverarbeitungs-Konzept die Stärken von Textverarbeitung und Datenbank und übernimmt damit in benutzerfreundlicher Form Funktionen der Schreibmaschine, des Leuchtstifts und des Zettelkastens zugleich.

Ein textorientiertes Datenverarbeitungs-Konzept empfiehlt sich somit gerade auch jenen, die der gelegentlich geäusserten Meinung sind, der Computer löse auf elegante Weise Schwierigkeiten, die es ohne ihn gar nicht gäbe.