Zeitschrift: Schweizerische Gesellschaft für Wirtschafts- und Sozialgeschichte = Société Suisse d'Histoire Economique et Sociale

Herausgeber: Schweizerische Gesellschaft für Wirtschafts- und Sozialgeschichte

Band: 12 (1994)

Artikel: Die Auswertung von Textquellen und quantifizierbaren Daten in einem textorientierten Datenverarbeitungskonzept

Autor: Hagnauer, Stephan

DOI: https://doi.org/10.5169/seals-871683

Nutzungsbedingungen

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Mehr erfahren

Conditions d'utilisation

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. En savoir plus

Terms of use

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. Find out more

Download PDF: 05.07.2025

ETH-Bibliothek Zürich, E-Periodica, https://www.e-periodica.ch

Stephan Hagnauer

Die Auswertung von Textquellen und quantifizierbaren Daten in einem textorientierten Datenverarbeitungskonzept

Dass es heute möglich und empfohlen ist, zahlreiche alltägliche geistes- und sozialwissenschaftliche Arbeitsschritte am Personalcomputer digital nachzuvollziehen und in undeterminierender, flexibler und komfortabler Form auch an umfangreichen Informationsmengen durchzuführen, soll am nachfolgend skizzierten Datenverarbeitungskonzept gezeigt werden.

Für das Nationalfonds-Projekt zur Erforschung der Geschichte der bernischen Staatsfinanzen suchte die Projektgruppe um Martin Körner (Bern) ein Verfahren zur individuellen Erfassung und intersubjektiven Auswertung Hunderter verschiedenartiger staatlicher Jahresrechnungen mit insgesamt Zehntausenden von Buchungssätzen. Es galt insbesondere, möglichst alle für spätere Fragestellungen relevanten Quelleninformationen zu erfassen und eine Reduktion oder interpretatorische Veränderung der Quellen bei der Aufnahme zu vermeiden, damit die nachfolgenden qualitativen und quantitativen Analysen nicht prädeterminiert werden.

Ausgehend von einer Anregung Markus Küngs (Bern) wurde deshalb der bemerkenswert naheliegende Versuch unternommen, die Textquellen zur Auswertung nicht – wie oft üblich – sofort in eine Datenbank überzuführen, sondern tatsächlich als Texte zu erfassen und so lange wie möglich in den Textstrukturen weiterzuverarbeiten.

Um Informationen in Textquellen erschnuppern, markieren und auswerten zu können, war es erforderlich, die Stärken moderner Computerprogramme (wie Word, Asksam und Access für Windows) durch spezielle Zusatzprogramme zu kombinieren, die während meiner Lizentiatsarbeit[1] und in intensiver Rücksprache mit Niklaus Bartlome[2] (Bern) als sogenannte Makroprogramme konzipiert wurden.

Das entwickelte Datenverarbeitungskonzept wird überblickartig und anhand ausgewählter Verfahrensschritte vorgestellt, weil es so oder ähnlich in verschiedenen sozial- und geisteswissenschaftlichen Gebieten zur qualitativen und quantitativen Quellenanalyse eingesetzt werden kann, ohne hohe Anforderungen an die Informatikkenntnisse der Benutzer zu stellen.

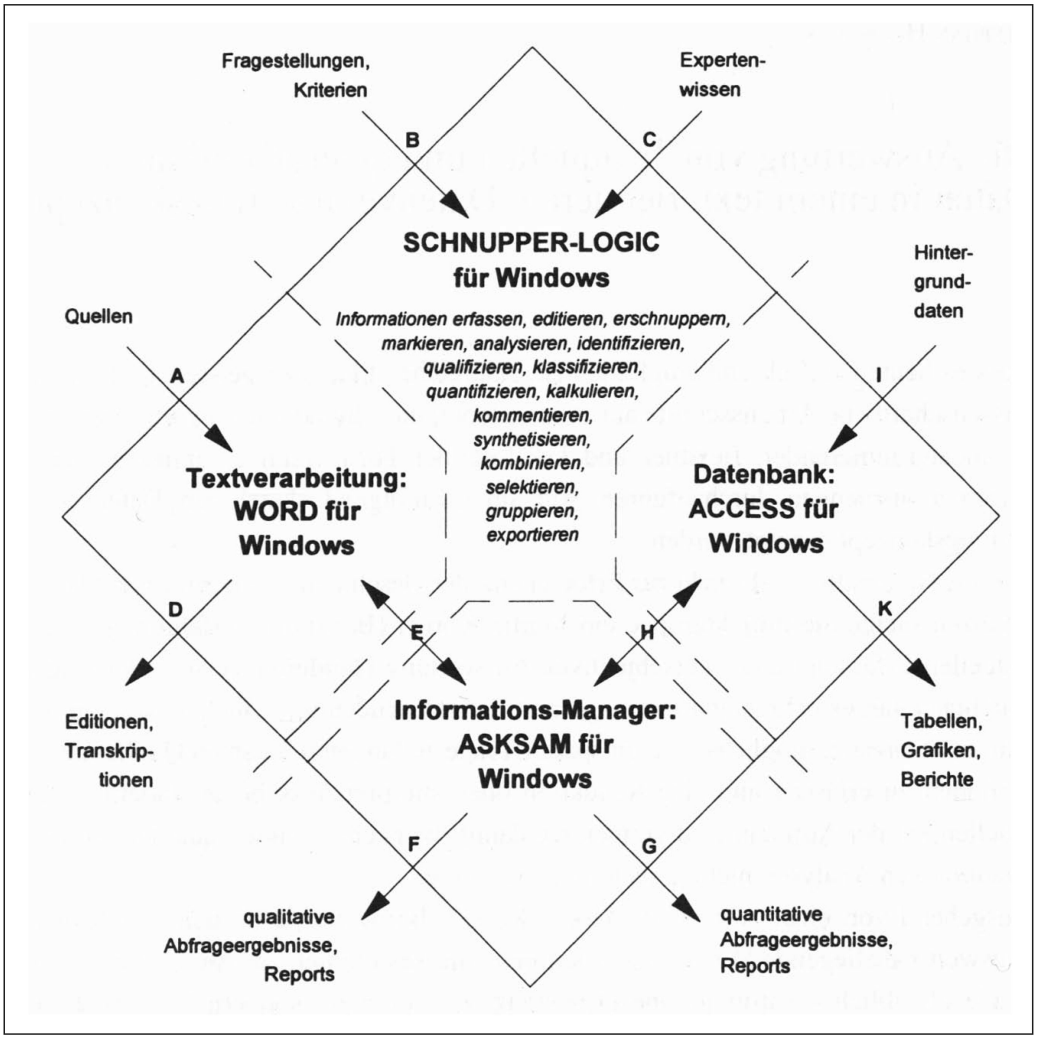
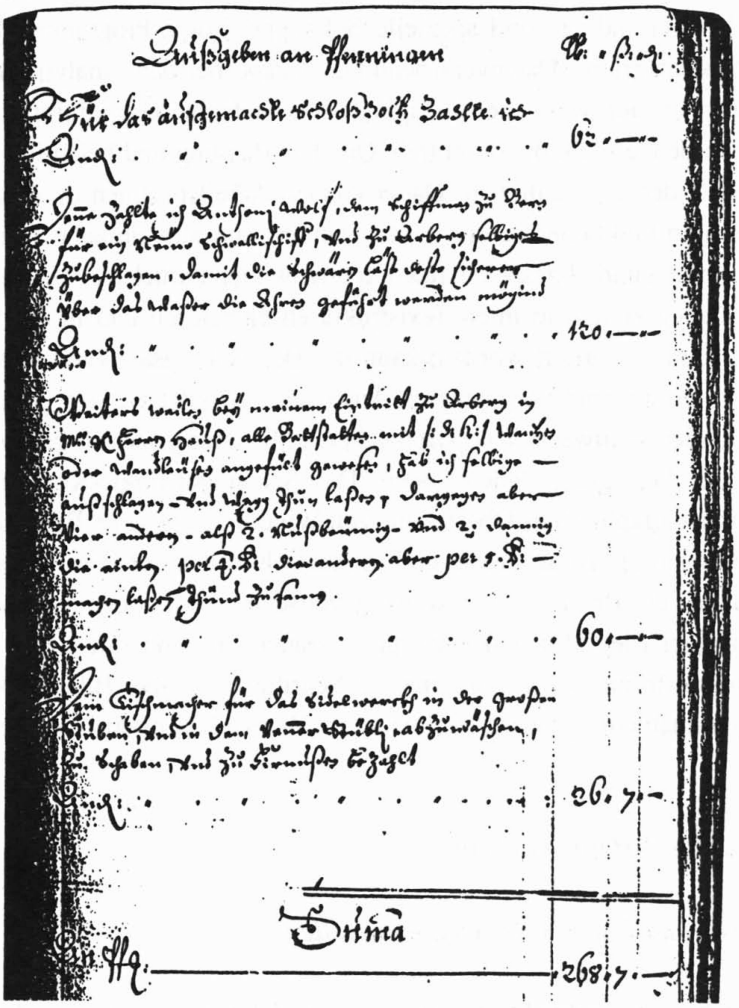


Abb. 1: *Übersicht zum textorientierten, integrierten Datenverarbeitungskonzept.*

1. Das Datenverarbeitungskonzept im Überblick

Das *Schnupper-Logic-Konzept* beruht auf dem Grundsatz, die Quellendaten in ihren originalen Textstrukturen in eine Computer-Textdatei überzuführen (vgl. Abb. 1, A) und die Informationen dann in den Texten zu erschnuppern und zu markieren, damit die so rekonstruierten Textstrukturen durch Informationsmanagement-Programme und durch relationale, tabellenartige Datenbankprogramme ausgewertet werden können, wobei der Rückgriff auf den originalen Datenkontext gewährleistet bleibt.



A | Transkription oder Scanning mit OCR

Vgl. Abb. 2b

Abb. 2a–2h: *Die allmähliche, flexible und beliebig intensive Verarbeitung von Informationen auf der Grundlage weitgehend originalgetreu erfasster Textquellen.*

Abb. 2a: *Der Text in handschriftlicher Gestalt (Rechnung des Amtes Aarberg vom 22. 10. 1683 bis 21. 4. 1684, S. 8; Staatsarchiv Bern, B VII, 850; verkleinert).*

Textverarbeitung, Informationsmanager und Datenbank werden durch die gemeinsame Windows-Grundlage und spezielle Schnupper-Logic-Programm-Module verbunden.[3] Das integrierte Datenverarbeitungskonzept, bei dem analytische Kriterien erst nachträglich auf weitgehend originalgetreu erfasste Grundinformationen angewendet werden, erlaubt die dezentrale Quellenerfassung und beliebig ausführliche Auswertungen der Informationen. Dabei können Arbeitsroutinen durch Schnupper-Logic-Programmfunktionen automatisiert werden, die auf einfache Muster-Erkennungsalgorithmen und abgespeichertes Expertenwissen zurückgreifen.
Sind die Quellen einmal in ihren Textstrukturen eingelesen und in die Schnupper-Logic-Umgebung integriert worden, kann der Bearbeitungsaufwand von allen Verarbeitungsstufen aus und von weiteren Anwendern genutzt werden, wobei sich die erfassten Quellen wahlweise in elektronischer Aufbereitung oder als gedruckte Transkriptionen zur Verfügung stellen lassen. Das Schnupper-Logic-Konzept nutzt die Fähigkeiten der graphischen Oberfläche von Windows, die dem Benutzer die Programmfunktionen als Bildsymbole und in Auswahlmenüs auf dem Bildschirm präsentiert, so dass der Benutzer die jeweils gewünschten Funktionen nur noch mit dem Bildschirmzeiger (der «Maus») anzutippen braucht. Es entbehrt somit nicht einer etwas sperrigen 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[4] noch 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 Transkriptionsaufwand zu minimieren:
– Bei Originalen, die nicht (mit sogenanntem Down-Loading) aus Text-Datenbanken[5] importiert oder mit dem Scanner eingelesen werden können, empfiehlt es sich, bei der Transkription wiederkehrende Textstrukturen als sogenannte Textbausteine abzuspeichern, damit diese später mit Kopierbefehlen wiederverwendet werden können.

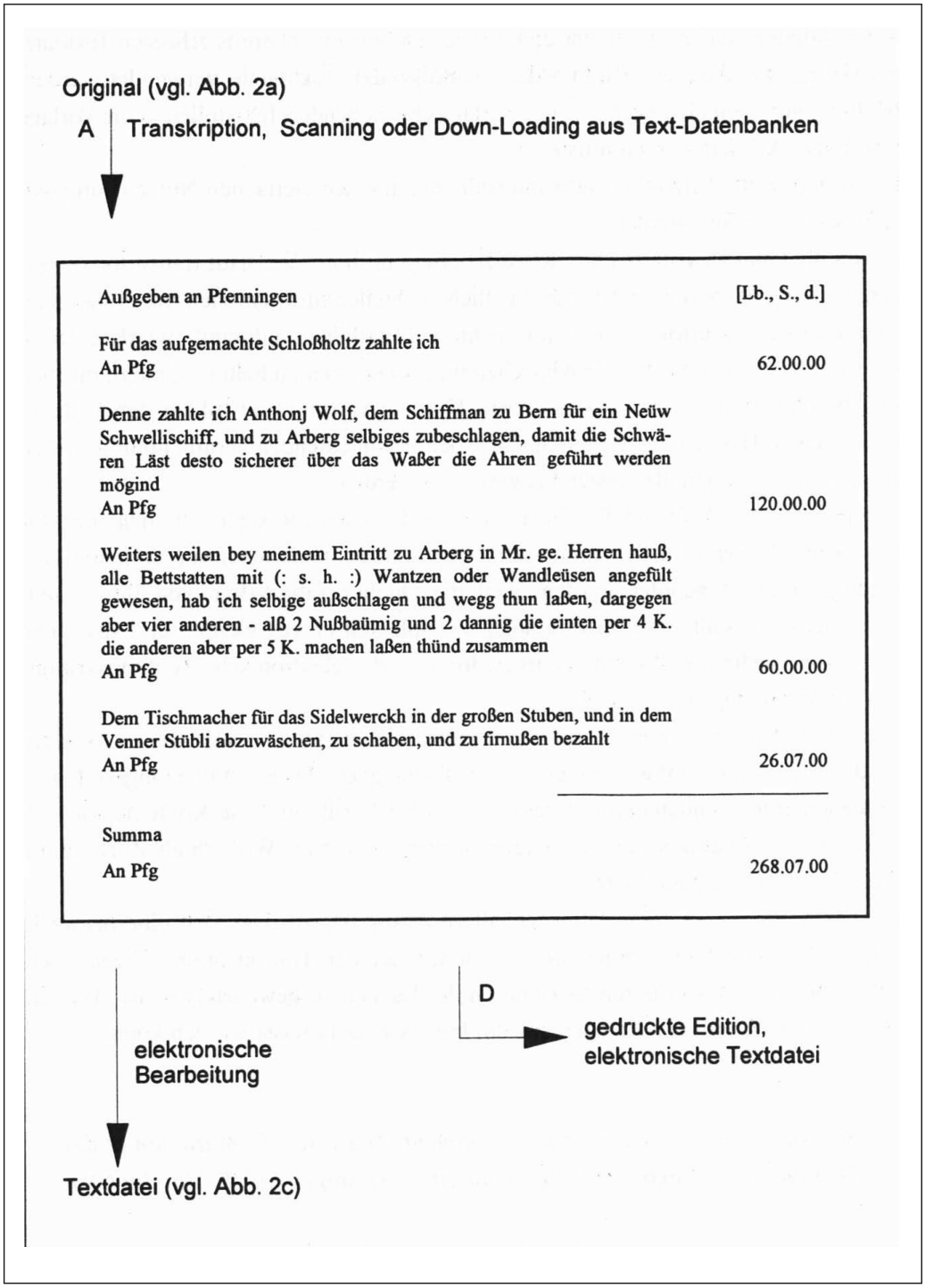


Abb. 2b: *Der Text in (wie hier) vollständig oder selektiv-partiell transkribierter Form. (Vgl. zu den Initialen der Pfeile die Übersicht in Abb. 1.)*

– Bei seriellen Quellen kann eine elektronische Kopie einer bereits erfassten Textdatei als Vorlage für die Transkription der nachfolgenden Textquelle verwendet werden, wodurch nur noch die gegenüber der Vorlage abweichenden Textstellen in die Vorlage hinein transkribiert werden müssen.[6]

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 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.

– Der einmalige Aufwand für die Erfassung der Quellentexte dient nicht nur zur Beantwortung der aktuellen Fragestellungen, sondern kann auch von weiteren Forschungsprojekten genutzt werden, 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.

Textabschnitte, deren Informationsgehalt zu gering ist, als dass sich eine manuelle Volltexterfassung rechtfertigen liesse, können bei der Transkription vorerst übergangen werden, wobei durch Deklaration der Lücken zu gewährleisten ist, dass die ausgesparten Textabschnitte bei späterem Interesse nacherfasst werden können.

2.2. Markieren, Verstichworten und Kommentieren von Informationen in den Textquellen («elektronische Leuchtstifte», «Flaggen» und «Randnotizen»)

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 sogenannten Maus

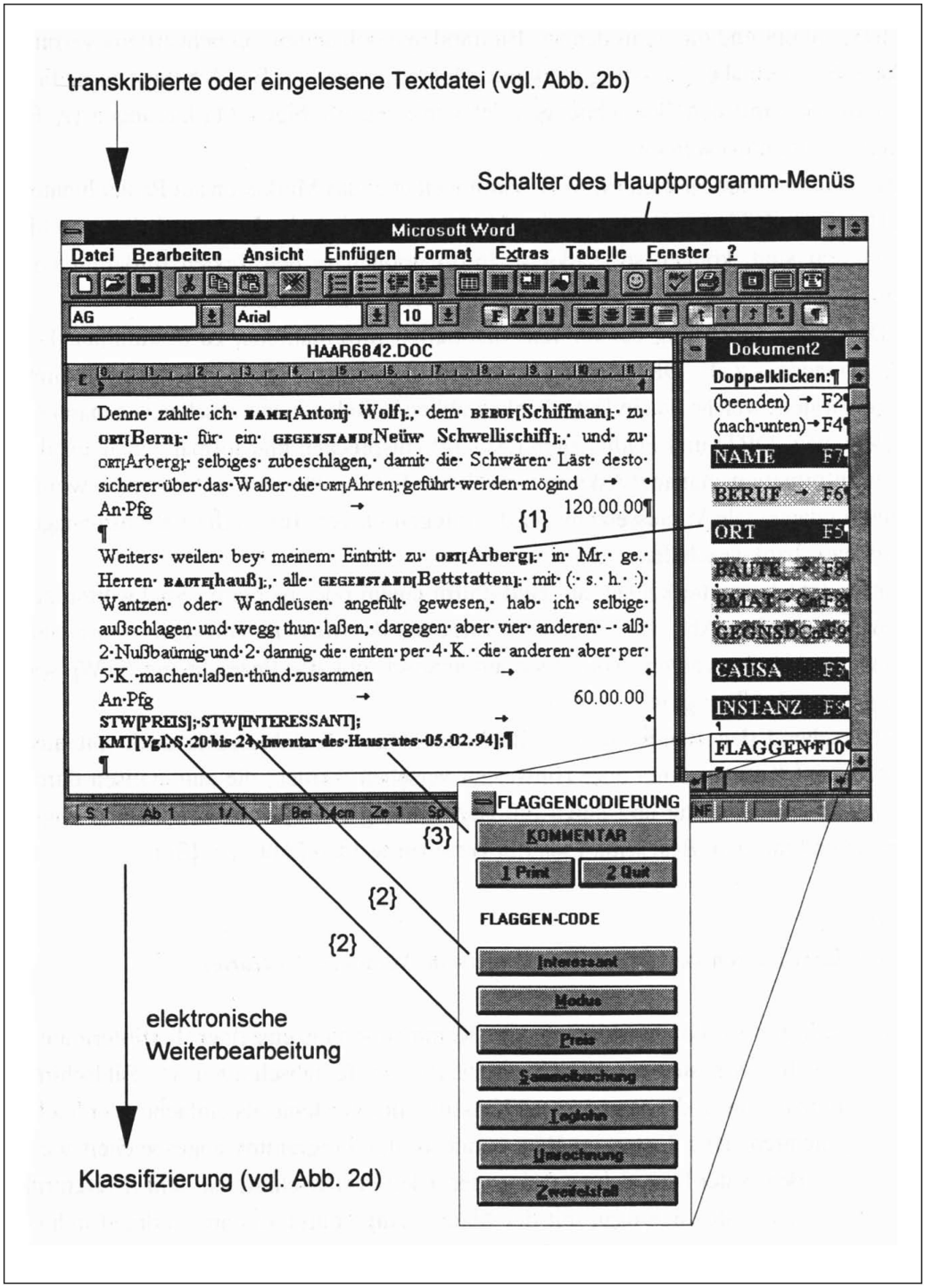


Abb. 2c: *Der Text während der (wahlweise farbigen) Markierung {1}, Verstichwortung {2} und Kommentierung {3} von Textinformationen (Flaggenfenster montiert).*

zuerst antippt und dann mit den am Bildrand symbolisierten «Leuchtstiften» verbindet, wobei nun aber die farbig 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 imText markierten Informationen zu bestimmten Bezeichnungen – z. B. ORT[Arberg] – werden diejenigen wichtigen Grundstrukturen rekonstruiert, welche den in Datenbanken üblichen Relationen zwischen sogenannten Feldnamen (ORT) und Feldinhalten (Arberg) entsprechen und 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, worauf sie automatisch mit der Bezeichnung STW[wie Flaggen in den Text gesetzt werden.
Entsprechend den Randnotizen in Büchern, mögen Textabschnitte auch mit ausführlichen[7] 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}).

2.3. Klassifizieren von Daten (systematische Verschlagwortung)

Zusätzlich zur Ad-hoc-Markierung und Kommentierung einzelner Textinformationen ist auch das systematische Klassifizieren vonTextabschnitten am Bildschirm durchführbar, soweit die gewünschten Klassifikationssysteme als einfache oder hochgradig hierarchisierte Listen im Wissensschatz des Programms abgespeichert werden.[8] 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 derAbschnitt automatisch mit dem entsprechenden Klassifikationscode gekennzeichnet wird.

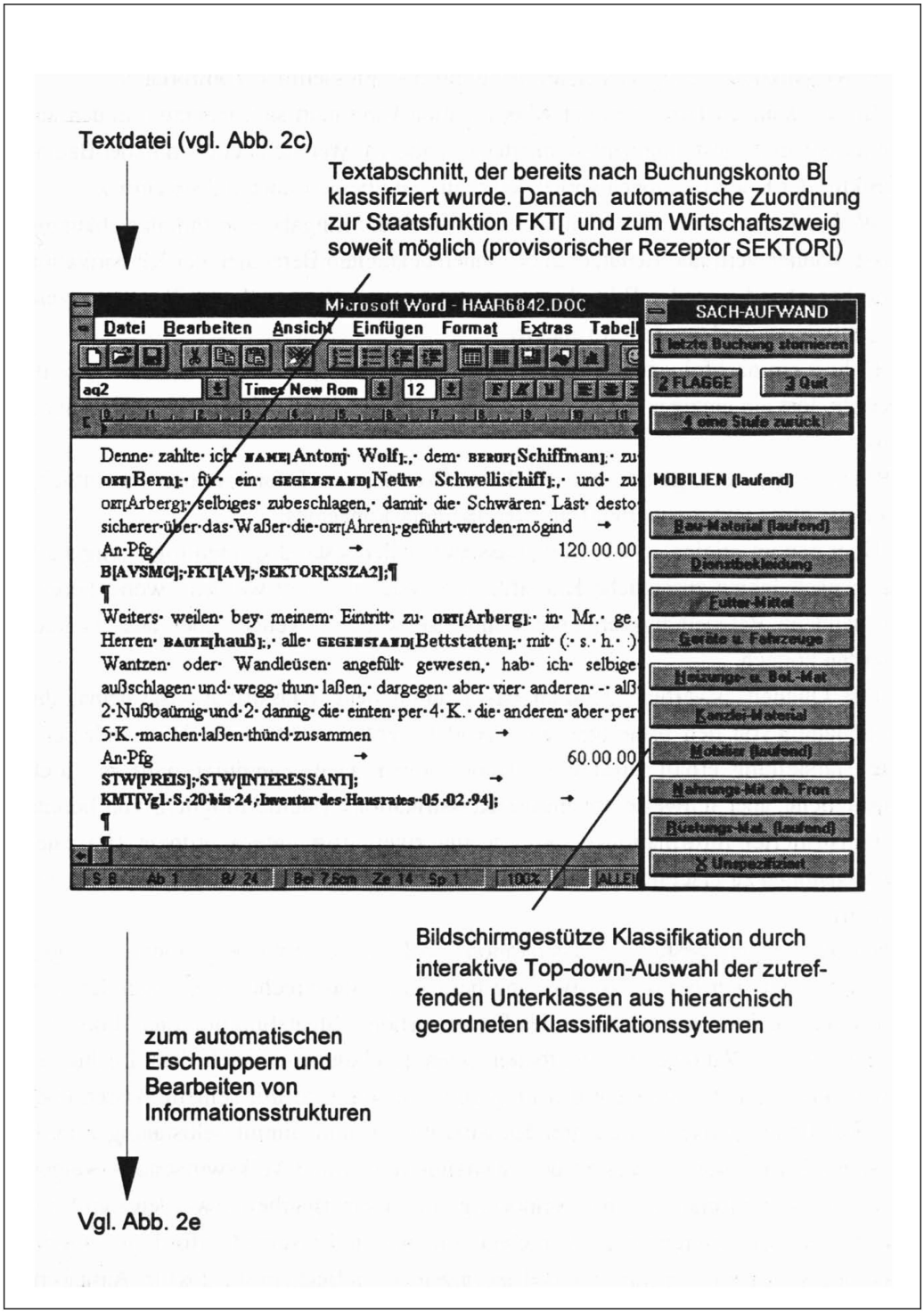


Abb. 2d: *Der Text während der Klassifikation von Textabschnitten am Bildschirm.*

Das Klassifizieren gestaltet sich am Bildschirm somit ziemlich komfortabel:
– In unbekannten Bereichen 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 die Eingabe von Initialen abkürzen lässt, können vertraute Benutzer in den ihnen bekannten Bereichen der Klassifikation vorübergehend von der Bildschirmunterstützung absehen und mit Tastaturkürzeln klassifizieren, und zwar molto allegro.
– 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 Datenverarbeitungskonzeptes:
– Über den im originalen Wortlaut erfassten Quellenbestand können folienartig auch noch nach Jahren zusätzliche Klassifikationssysteme gelegt werden,[9] 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 Freiheitsgrade, und die Darstellung erhält zusätzliche Dimensionen. Demgegenüber wird es durch zusätzliche, aber teilweise abhängige Klassifikationssysteme möglich, von bereits rekonstruierten Informationsstrukturen aus zweigartig untergeordnete (fraktale) Dimensionen zu erschliessen, also die Verästelung der Quellenauswertung zu verfeinern.
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 dem 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 rechtseindeutig sind. Dem Bearbeiter werden somit nur noch jene Zuordnungsfragen vorgelegt, die seinen Entscheid erfordern.[10] 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 circa 30% Buchungssätzen, die er noch präziser einer Unterklasse zuzuordnen hat.

2.4. Automatisches Erschnuppern und Bearbeiten von Informationsstrukturen durch einfache, wissensbasierte Muster-Erkennungsalgorithmen

In den als Texten aufgenommenen Quellen lassen sich mit Muster-Erkennungsalgorithmen ausführlich Informationen erschnuppern, wobei wiederkehrende Muster im Expertenwissen des Programms abgespeichert und Routinearbeiten automatisiert werden können (vgl. Abb. 2e).
Hinterlassen die automatisierten Programmfunktionen verborgen formatierte Bearbeitungsvermerke (mit Datum) in der Textdatei, erlaubt dies, die bisher durchgeführte

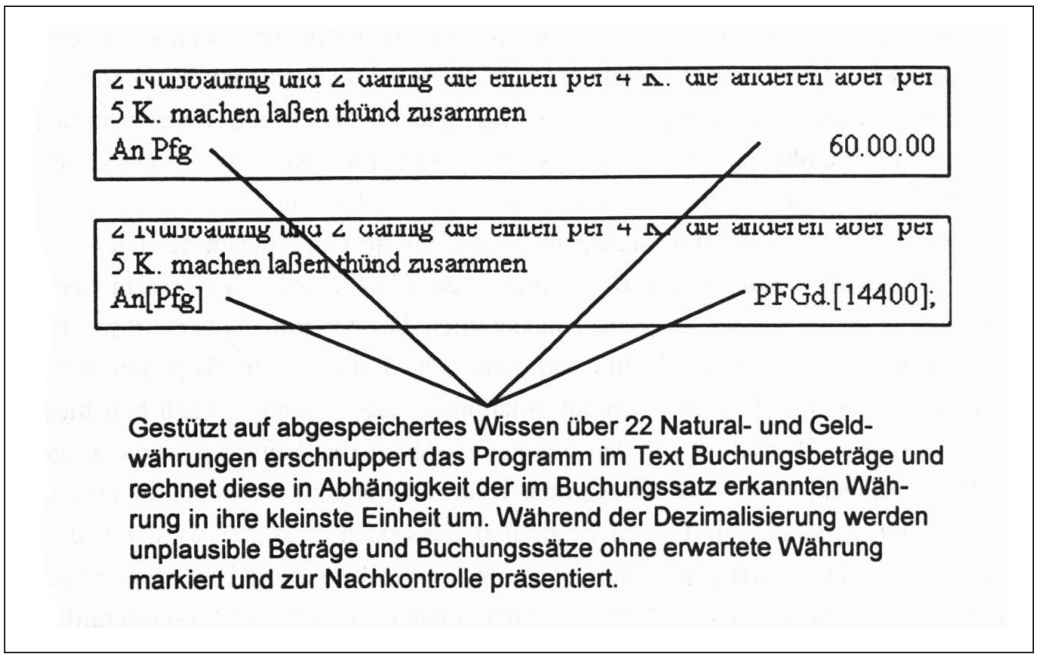


Abb. 2e: *Beispiel für das automatische Erschnuppern und Bearbeiten von Texten.*

Quellenbearbeitung automatisch zu rekonstruieren und den weiteren Ablauf der Bearbeitung durch Autopilotfunktionen zu steuern.
Um die nachfolgende Überführung des Textes in den Informationsmanager vorzubereiten, wird jeder Textabschnitt mit den in der Kopfzeile der Quelle erschnupperten bibliographischen Merkmalen sowie mit der Nummer der Seite und des Textabschnittes versehen, damit jeder Abschnitt auch nach der Fragmentierung der Quelle auf elektronische Karteikarten identifizierbar und zitierbar bleibt (vgl. Abb. 2f).

2.5. Fragmentieren des Textes auf «elektronische Karteikarten» des Informationsmanagers

Die Texte liegen im Informationsmanager Asksam nach wie vor in den originalen Satzstrukturen vor, sind nun aber auf einzelne «elektronische Karteikarten», sogenannte Dokumente, fragmentiert (im vorliegenden Fall nach Buchungssätzen), womit der in der Textverarbeitung betriebene Aufwand intensiv genutzt werden kann.
Von den zahlreichen Funktionen des Informationsmanagers seien hier nur einige derjenigen aufgezählt, die für sozial- und geisteswissenschaftliche Arbeiten äussert wirkungsvolle Werkzeuge darstellen: Hyper-, Voll- und Kontextsuchen mit kombiniertem boolschem Abfragen; Selektion nach Worten oder Feldinhalten; Report und Druck 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, es sei denn, man bringe die Geduld auf, sämtliche Karten zu fotokopieren. Demgegenüber ist es im elektronischen Informationsmanager möglich, nach beliebigen Kriterien (wie z. B. nach den in den vorangegangenen Bearbeitungen durchgeführten Markierungen, Verstichwortungen oder Klassifizierungen) reversibel die Daten – isoliert oder zusammen mit ihren originalen Kontexten – immer wieder neu zu selektieren, zu gruppieren, zu subsumieren sowie als elektronische Dateien abzuspeichern, als speziell geordnete Textauswahlen, Listen, Berichte und Zitatsammlungen auszudrucken oder in Darstellungen einzubauen.
Im Gegensatz zu den weiter verbreiteten, tabellenartig aufgebauten Datenbanken (wie

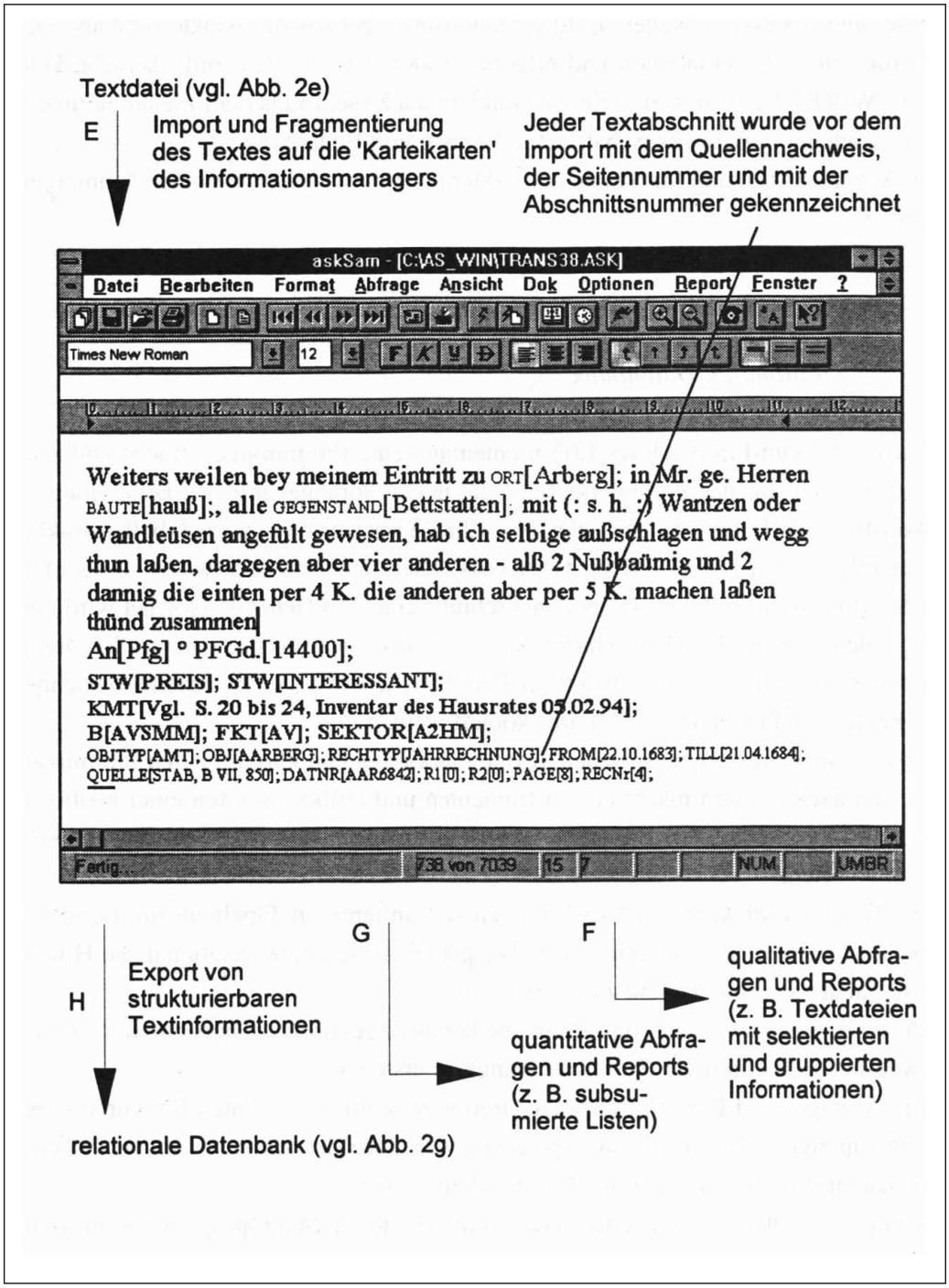


Abb. 2f: *Der Text nach der automatischen Fragmentierung in Textabschnitte und Überführung in «elektronische Karteikarten» des Informationsmanagers. Gezeigt ist die 738. von 7039 «Karteikarten», auf die circa 1200 Seiten Text kopiert wurden.*

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 zum voraus definiert zu werden.

2.6. Überführen von Textinformationen in die Tabellenstruktur der relationalen Datenbank

Obwohl Asksam-für-Windows 1.02 momentan keine Programmiersprache enthält,[11] können durch sogenannte Reportbefehle die in den vorangegangenen Bearbeitungen rekonstruierten 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 eröffnet wird und die Daten – z. B. BERUF[Schiffmann] – von der «Karteikarte» in vordefinierter Reihenfolge (und im sogenannten Rich-Text-Format) auf die entsprechend bezeichneten Felder der Tabellenzeile (Record) kopiert werden.[12]
Damit können die vorgängig in den Texten markierten und klassifizierten Informationen nun auch mit den mächtigen Instrumenten und Hilfsassistenten 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 Kreuz- und Pivottabellen aussagekräftige Übersichten über ausgewählte Daten und deren Zusammenhänge zu erstellen;
– die Abfrage- und Darstellungsprozeduren vorerst im sogenannten Entwurfsmodus zu konzipieren, zu testen und abzuspeichern, womit sie sich zukünftig auf wechselnde Datensampels anwenden und weiterentwickeln lassen;
– komplexe Datenanalysen oder -bearbeitungen durch Makroprogramme zu automatisieren.

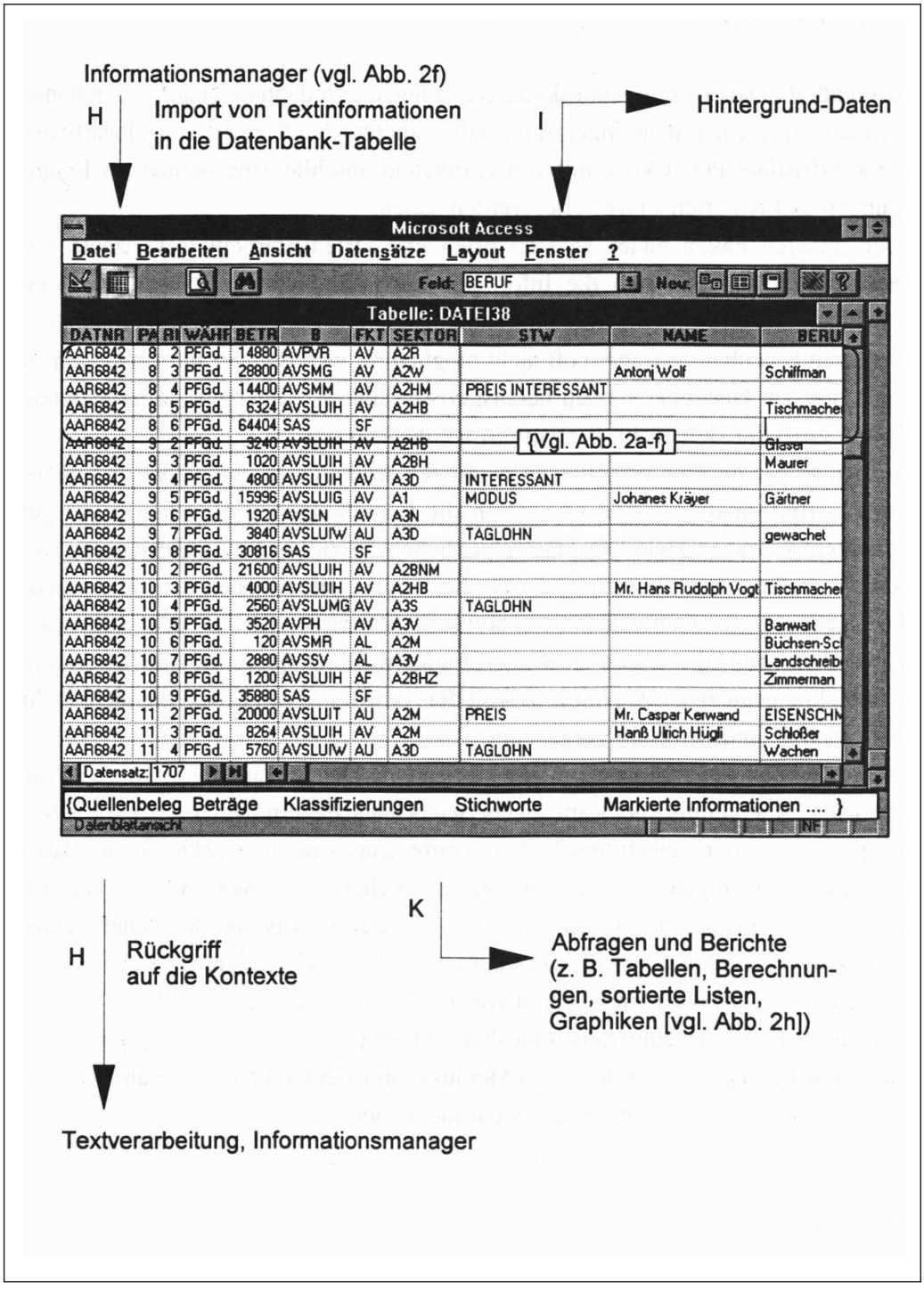


Abb. 2g: *Die strukturierten Textinformationen (Ausschnitt) nach der automatischen Überführung in die Tabellenstruktur der relationalen Datenbank.*

3. Schlussbemerkungen

Textorientierte Datenverarbeitungskonzepte 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 sich textorientierte Datenverarbeitungskonzepte vor allem dadurch an, dass sich die Informationserfassung von den anschliessenden Klassifikationen und Interpretationen trennen lässt, was demgegenüber bei tabellenorientierten Datenbankverfahren oft nicht möglich ist, weil dort die Informationen zu Beginn aus den Quellenkontexten herausgerissen und in vorgängig zu definierende Feldstrukturen der Datenbank übergeführt werden müssen.
Textorientierte Konzepte erschliessen heterogen strukturierte qualitative und quantitative Textinformationen und erleichtern die Bemühungen, die nach wie vor gut unterhaltenen Gräben zwischen unterschiedlichen Disziplinen zu überbrücken. Der modulare Aufbau des Konzeptes erlaubt, die Programmfunktionen auf unterschiedliche Quellen anzupassen und neue analytische Kriterien alternativ oder kumulativ anzuwenden. Weil alle Daten stets am Originalwortlaut überprüft und langfristig intersubjektiv genutzt werden können, wird der Anhäufung von Daten, die nicht mehr kollationierbar sind, wirksam vorgebeugt.
Es sollte deutlich geworden sein, dass textorientierte Datenverarbeitungskonzepte die Begrenzungen herkömmlicher Karteien überwinden und das ermöglichen, wovon schon lange gesprochen wird: elektronische Textverarbeitung – im eigentlichen Sinne. Dabei übernimmt der Computer in sehr benutzerfreundlicher, reversibler und automatisierbarer Form Funktionen der Schreibmaschine, des Leuchtstifts und des Zettelkastens, wie sie z. B. in der Wissenschaft, in Dokumentationen und Registraturen üblich sind, um Quellen zu analysieren und die Flut von Textinformationen zu verwalten.
Ein textorientiertes Datenverarbeitungskonzept 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.

Anmerkungen

1 Stephan Hagnauer, Analysen der Rechnungen der Ämter Aarberg, Büren, Erlach und Nidau 1631–1635 und 1681–1685. Elemente zur Geschichte der bernischen Staatsfinanzen (Arbeitstitel), Lizentiatsarbeit, Universität Bern.

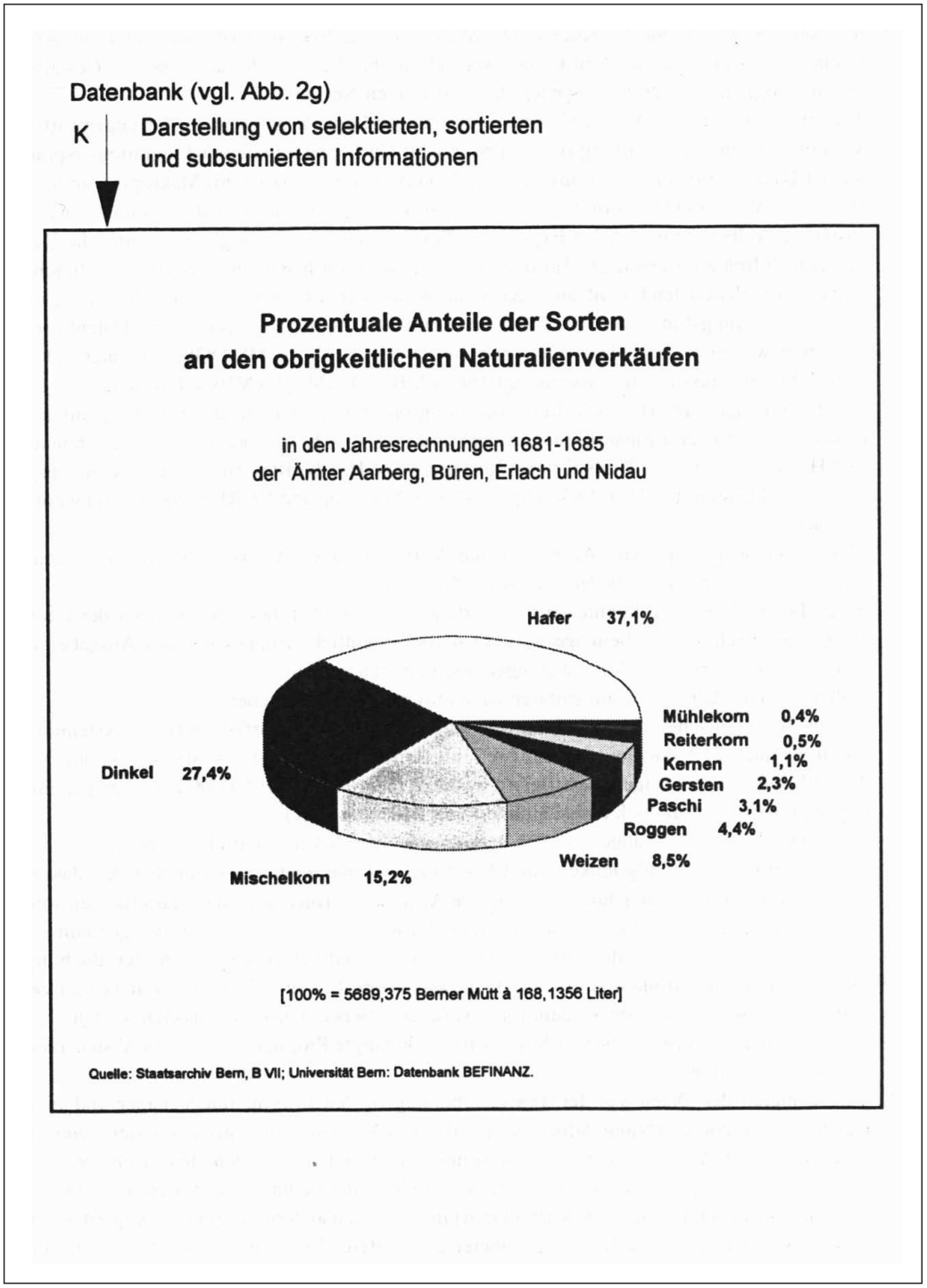


Abb. 2h: *Beispiel einer quantitativen Auswertung: Grafik auf der Grundlage von Daten, die in ihren Textstrukturen erfasst und aufbereitet wurden.*

2 Ich danke Prof. Dr. Martin Körner, Dr. Markus Küng, Niklaus Bartlome und Hans Anton Ebener für Anregungen und Kritik sowie speziell für ihre Vor- und Mitarbeit bei der Gestaltung des ins Programm integrierten Kontenplans zur Frühen Neuzeit.

3 Technisch gesehen handelt es sich beim Schnupper-Logic-Programm in der gegenwärtigen Version 2.1 (und in Erwartung der von Microsoft Corp. angekündigten Programmiersprache Visual-Basic-Application-Edition) um circa 220 kByte (circa 120 Seiten) Makroprogramme für Word-für-Windows (Microsoft Corp.), welche als Pre-Editing- und Transfermodule 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 exportiert werden können. Das Programm wurde entwickelt für MS-DOS-Computer mit folgender Minimalausstattung: Prozessor 80386, 4 MByte RAM, 100 MByte Harddisk.

4 Die Bemühungen um Handschriften-Erkennungsprogramme nähren die Hoffnung auf automatische Digitalisierung historischer Manuskripte: vgl. Eric L. Helsper, Lambert R. Schomaker und Hans-Leo Teulings, Tools for the Recognition of Handwritten Historical Documents, in: History and Computing 5/2 (1993), (Special Issue: Scanning and OCR), S. 88–93, mit weiterer Literatur.

5 Eine zunehmende Zahl von Agenturen und Verlagen bietet Druckerzeugnisse nun auch in elektronischer Form an (z. B. Neue Zürcher Zeitung).

6 Nach diesem Verfahren konnten bei der Erfassung bernischer Jahresrechnungen der Grundaufbau der Rechnungen (Seitengestaltung) sowie ordentliche Einnahmen und Ausgaben aus den bereits transkribierten Vorrechnungen übernommen werden.

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

8 Die in den bernischen Jahresrechnungen untersuchten Finanzvorfälle werden systematisch vorläufig 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 Bedeutung sind.

9 Was weitreichende Wirkungen auf den hermeneutischen Zirkel zu entfalten scheint.

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

11 Der Hersteller verweist auf die von Microsoft angekündigte Programmiersprache Visual-Basic-Application-Edition.

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