

# "Versteckte" Fakten- informationen in Textdatenbanken

**GENIOS bietet mit seinem Produkt "ChartLibrary" in einer Metadatenbank Graphiken und Tabellen aus diversen Quellen zur Ausgabe in Graphik- bzw. Faksimileformaten (vor allem PDF und GIF) an. Hiermit wird ein zielgenauer Zugriff auf einzelne Fakteninformationen möglich. ChartLib gestattet das Retrieval nach "versteckten" Fakten, die in Zeitungsartikeln oder in wissenschaftlicher Literatur eingebunden sind und deshalb bislang nur schwierig oder gar nicht zu finden waren.**

Es ist für viele Online-Archive derzeit noch nicht selbstverständlich, dass graphisches Material in den Dokumenten - auch bei Volltextdatenbanken - abfragbar online vorläge. Angeboten wird vielmehr in der Regel ausschließlich der Fließtext im ASCII-Format. Für den Nutzer besonders sichtbar wird dieses Manko durch die (durchaus vorhandenen) Bildüberschriften sowie den Hinweis "Graphik entfernt". Man ist hier auf Dokumentliefersdienste angewiesen, die den Artikel - diesmal mit den Graphiken - als Kopie zur Verfügung stellen. Hier geht GENIOS neue Wege: Graphiken werden nicht mehr ignoriert, sondern sowohl in der ChartLib als auch in der Ursprungsquelle - soweit möglich - aufgenommen. Ziel der ChartLib ist - in den Worten von GENIOS - "die Versorgung der Kunden aus Marktforschung, Beratung und Research mit direkt nutzbaren 'Schaubildern'".

Die ChartLib steckt noch in den Anfängen: Gut 40.000 Datensätze (die alten ASCII-Tabellen der TABL eingeschlossen) sind derzeit vorhanden. Wir halten das Konzept für richtungweisend. Deshalb haben wir Ende August 1999 die ChartLib sowohl in der World Wide Web-Variante ("WEB-SEARCH") als auch mittels der Client-Software ("InMedia") ein-

gehend getestet. Wir analysierten die Quellenlage, die Recherchemöglichkeiten sowie den Bezug zu den Texten, denen die Charts entstammen.

## Unterschiedliche Fragetypen, unterschiedlicher Informationsbedarf: CIN und POIN

Nutzer suchen unterschiedliche Informationen aus unterschiedlichen Informationstypen aus völlig unterschiedlichen Anlässen zur Deckung unterschiedlicher Bedarfe. In einer Grobeinteilung können wir aus zwei Fragetypen zwei Grundformen von Informationsbedarf ableiten. Betrachten wir zum Beispiel einige Informationsbedarfe, die sich in folgenden Fragen wiederfinden:

- Fragetyp A
  - Welches Marktvolumen hat der deutsche Maschinenbau 1998?
  - Was kostet ein Hektograd Chablis (in FF) im August 1999?

### CIN

1. Thematische Grenzen sind klar abgesteckt.
2. Die Suchfrageformulierung ist durch exakte Terme ausdrückbar.
3. Eine Fakteninformation reicht aus, um den Bedarf zu decken.
4. Mit der Übermittlung der Fakteninformation ist das Informationsproblem erledigt.

### ● Fragetyp B

- Welche Methoden der fuzzy logic lassen sich beim Data Mining einsetzen?
- Wie beschreiben Marktforscher und Wirtschaftswissenschaftler Konsumklima und Absatzchancen für ausländischen Wein in Ungarn?

Fragetyp A zielt auf die Übermittlung einer Fakteninformation ab. Der zugrundeliegende Informationsbedarf ist konkret. Valery Frants, Jacob Shapiro und Vladimir G. Voiskunskii benennen in ihrer Analyse des Informationsbedarfs (Automated Information Retrieval. San Diego 1997, 38) diesen Typ als "Concrete Information Need" (CIN). Fragetyp B läßt sich nicht durch die Angabe eines Faktums befriedigen. Das Informationsproblem wird hier durch die Übermittlung einer mehr oder minder großen Sammlung von Texten erfüllt. Frants et al. sprechen beim Typ B vom "Problem Oriented Information Need" (POIN). CIN und POIN lassen sich mittels einiger Charakteristika anschaulich vergleichen:

### POIN

1. Thematische Grenzen sind nicht exakt bestimmbar.
2. Die Suchfrageformulierung läßt mehrere terminologische Varianten zu.
3. In der Regel müssen diverse Dokumente beschafft werden. Ob der Informationsbedarf damit abschließend gedeckt ist, bleibt offen.
4. Mit der Übermittlung der Literaturinformationen wird ggf. das Informationsproblem modifiziert, oder ein neuer Bedarf entdeckt.

Die Relevanz einer Information zur Befriedigung eines CIN bzw. zur Beantwortung einer Frage vom Typ A ist exakt zu bestimmen. Entweder beantwortet der gefundene Datensatz unsere Frage oder nicht. Etwas anderes liegt beim POIN bzw. bei der Antwort auf eine Typ-B-Frage vor. Die Antwortmenge sammelt in der Regel Dokumente (oder Hinweise darauf) aus literaturbezogenen Datenbanken, also bibliographischen Datenbanken und Volltextdatenbanken. Hier haben wir mehrere "Treffer", die - mehr oder weniger - Aspekte dessen beantworten, was gefragt wurde. Die Relevanzeinschätzung beim POIN dürfte nicht so eindeutig ausfallen wie beim CIN.

Nun gibt es durchaus Bedarf an Fakten, also eigentlich CIN, der aber nur indirekt problemorientiert befriedigt werden kann. Dies ist immer dann der Fall, wenn Fakten in Texten "versteckt" sind, etwa eine Tabelle in einem Zeitungsartikel oder eine Graphik in einem wissenschaftlichen Aufsatz. Hier muß umständlich ein POIN konstruiert werden, um sich an das Faktum heranzutasten. Ist bei der entsprechenden Tabellen- und Abbildungsüberschrift nicht terminologisch sauber getextet worden oder übergeht ein Abstract (bei bibliographischen Datenbanken) die Graphik, so werden wir sie überhaupt nicht finden.

GENIOS verspricht für alle CIN Abhilfe. Die Suche nach Fakteninformation soll zielgenau - und ohne den Umweg über POIN - erledigt werden. Hier wird ein Weg weiterverfolgt, der von GENIOS bereits mit der Datenbank TABL (oder von Infratest Burke mit der FAKT) eingeschlagen worden ist. GENIOS hat hierzu eine Meta-Datenbank ("Library") angelegt, die alle Graphiken aus diversen Datenbanken zusammenfaßt: die ChartLibrary. Im Online-Bereich dürfte GENIOS mit diesem Angebot derzeit wenig Konkurrenz haben - einzig Profound verfügt über ein vergleichbares Produkt.

Was bringt dieses Angebot dem Nutzer? Wie ist der Zugriff auf die Fakten zu bewerkstelligen? Sprechen die nunmehr von ihrem Text isolierten Fakten für sich? Ist die Quellenlage zufriedenstellend?

## ChartLibrary: Die Quellen

Die ChartLibrary enthält Graphiken, Tabellen und Bilder aus diversen Datenbanken, so etwa aus verlagseigenen Zeitungen und Zeitschriften (darunter natürlich "Handelsblatt" und "WirtschaftsWoche") sowie aus anderen Zeitungen (z.B. "Computerwoche"). Um einen ungefähren Überblick über die Verteilung auf die unterschiedlichen GENIOS-Quellen zu bekommen, die die ChartLib speisen, wurde für

die letzten zwölf Monate (31. August 1998 bis 31. August 1999) eine Rangordnung der Quellen nach der Anzahl der Datensätze erstellt (Abbildung 1). Demnach sind fast 6.000 Charts aus dem "Handelsblatt" sowie rund 1.400 Charts aus der "WirtschaftsWoche" vorhanden. 56% aller Graphiken entstammen diesen beiden Organen des Handelsblatt-Verlages. Auf den nächsten Rangplätzen liegen Abbildungen aus "Computerwoche", "Werben und Verkaufen", "Horizont" und "Lebensmittel Praxis".

Für den gesamten Jahrgang 1998 liegen 9.137 Datensätze vor, im Jahr 1997 waren es 6.728. Frühere Jahrgänge werden fast ausschließlich durch die alte TABL gespeist, die zwar Tabellen, diese aber im ASCII- und nicht in Graphik- oder Faksimileformaten speichert. Insgesamt (mit den TABL-Datensätzen) umfaßt die ChartLib knapp über 40.000 Dokumentationseinheiten. Die Qualität und der Informationsgehalt der einzelnen Graphiken schwankt beträchtlich zwischen wenigen Zahlen und umfassenden, detaillierten Angaben.

Noch nicht alle Volltextquellen, die bei GENIOS aufliegen und die Graphiken oder Tabellen enthalten, sind in der ChartLib vertreten. "Prominentes" Beispiel ist die Datenbank "Password", die leider nur die Fließtexte der Artikel abbildet. Der Online-Leser dieses Artikels wird also das Ranking aus Abbildung 1 nicht sehen können. Wie der Leser der Printversion einschätzen kann, entgehen dem "Onliner" doch einige nicht uninteressante Informationen.

Bei GENIOS enthalten mehr Datenbanken Graphiken, als in ChartLib vertreten sind, z.B. das "Deutsche Marken-Suchsystem" (DEMAS). Das Quellenangebot mit Graphiken soll - nach Angaben von GENIOS - kontinuierlich ausgebaut werden. Auch die ChartLib wird noch erweitert.

## Suche nach Charts über die Textdatenbanken

Das Retrieval in den Textdateien läuft in zwei Formen ab. Einmal ist es bei Datenbanken, die Charts enthalten, möglich, direkt in der Suchmaske im Feld "Grafik/Tabelle" gezielt nach Wörtern aus dem Graphik- bzw. Tabellentitel zu suchen. Zum andern ist eine "normale" Suche, egal in welchen Feldern, möglich. Hier erfährt man jedoch erst aus der Titelliste, dass zum Fließtext eine Chart gehört. Diese Titelliste enthält pro Datensatz ● das Datenbankkürzel ● das Erscheinungsdatum ● den Artikeltitle ● Angaben zur Länge des Artikels und ● sofern eine Chart vorhanden ist: das Graphiksymbol.

Nach Aufruf des Volltextes, der nach wie vor ohne Graphiken ausschließlich den Fließtext anzeigt, erscheint ein Fenster mit Informationen zur Chart: ● Überschrift der Graphik oder Tabelle ● Dateiname ● Dateigröße und ● Preis.

Durch Doppelklick auf das Fenster wird die Datei übertragen (je nach Größe kann dies einige Zeit dauern), der passende Viewer geladen (bei PDF-Dateien also Adobe Acrobat) und die Chart angezeigt.

Die Ausgabeformate sind vielfältig, sie reichen von Graphik- (gif, tif, jpg) über Tabellen- (xls) bis zu Faksimileformaten (pdf), je nach

Datenbank	Anzahl Datensätze pro Jahr (*)
Handelsblatt	5871
WirtschaftsWoche	1401
Computerwoche	828
w&v Werben und Verkaufen	642
Horizont	621
Lebensmittel Praxis	529
Telecom Handel	305
Euro	230
ESV Zahlenbilder	200
fww international	180
Finanztest	179
Erdöl/Energie-Inf.	163
Z f. d. ge Kreditwesens	157
Management Berater	149
New Media Report	144
Der langfristige Kredit	140
Media und Marketing	133
Convenience Shop	126
com!	124
Bank und Markt	120
Internet World	119
Weinwirtschaft	110
Absatzwirtschaft	102
Network & Communications	80
Flugrevue	72
Cards - Karten - Carts	67
Europa Chemie	63
Die Zeit	51
ESV Altlastenspektrum	47
Außer-Haus-Konsum	41
Das Rathaus	40
ESV Immissionsschutz	27

(\*) 31.8.1998 bis 31.8.1999; einzelne Datensätze enthalten mehr als eine Graphik; N = 13.061

Abb. 1: GENIOS-Quellen der ChartLib nach Häufigkeit

Datenbank unterschiedlich. Die häufigsten Formate sind PDF und GIF. Alle Charts entstammen dem jeweiligen elektronischen Satzsystem; von einer hohen Qualität der Graphiken kann demnach ausgegangen werden.

Die Typen der graphischen Darstellungen haben eine große Bandbreite: Sie reichen von Marktprognosen über Rankings bis hin zu "Schaubildern". Ein Vorteil für den Nutzer kann sein, dass er die Charts direkt in eigene Präsentationen oder Veröffentlichungen einbauen kann.

## Suche nach vormals "versteckten" Fakten über die ChartLib

Sucht man (unabhängig von einer bestimmten Datenbank) nach den vormals "versteckten" Fakteninformationen in den Textdatenbanken, so steht als Metadatenbank die ChartLib zur Verfügung. Folgende Suchfelder sind bei InMedia bzw. bei WEB-SEARCH vorgesehen:

- freie Suche (Retrieval über alle Inhalte aller Felder)
- Titel (gemeint ist: Titel des Artikels)
- Titel Graphik/Tabelle
- Datum (von - bis)
- Dokumentnummer des Volltextes
- Quelle (via Auswahlfenster).

Bei InMedia hat der Nutzer darüber hinaus die Möglichkeit, die Suchmaske individuell zu erweitern (etwa durch die Felder "Firma/Institution", "Thema", "Branche", "SIC" usw.).

Der sinnvollste Sucheinstieg ist über das Feld Graphik bzw. Tabelle. Suchbar sind die Wörter der jeweiligen Überschriften. Bei den Quellen hat der Nutzer die Wahl, entweder alle Quellen auszuwählen oder genau eine. Nicht suchbar ist der Urheber einer Abbildung (z.B. "Globus Bilderdienst"); dies erfährt man erst durch die Copyright-Angabe in der Graphik.

Suchergebnis ist eine Titelliste (Titel der Artikel). Die "Volltext"-Anzeige unterscheidet sich jedoch von der gewohnten (echten) Volltextanzeige der einzelnen Datenbanken. Angezeigt werden die bibliographischen Angaben des Quellartikels, soweit vorhanden die Inhaltserschließung des Artikels sowie formale Angaben (wie die Dokumentnummer der Quelle). **Abbildung 2** zeigt zwei Beispiele.

Jeweils unterhalb der "Volltexte" erscheinen die uns schon bekannten Fenster mit den Informationen über die Chart

Betrachten wir in **Abbildung 2** das erste Beispiel, so fällt die (durchaus zufriedenstellende) Inhaltserschließung auf. Allerdings ist die Sacherschließung problematisch, wertet sie doch den dazu gehörigen Artikel aus und nicht die Graphik oder Tabelle. Die Graphik berichtet nämlich über den Umsatz mit Kaugummi in Deutschland zwischen 1993 und 1997.

Im zweiten Beispiel fehlt jede Inhaltserschließung; der Artikel-Titel ist völlig belanglos. Hier vermischen wir noch mehr als im Beispiel 1 eine detaillierte Inhaltserschließung der Chart, dass es um Senioren in Deutschland geht, erfährt man erst aus der Überschrift der Abbildung.

## Das Zusammenführen von Chart und Text

Sucht man in einer Volltextdatenbank, so hat der Nutzer stets den Fließtext und die darin enthaltenen Charts in einem Kontext gefunden. Anders ist dies bei der ChartLib; hier sind die Graphiken und Tabellen vom Text losgelöst. Da gewisse Abbildungen oder Zahlenwerke sich ohne den begleitenden Text nur schlecht oder gar nicht interpretieren las-

sen, wäre ein direkter Link zum Volltext wünschenswert. (Lexis-Nexis beispielsweise hat in diversen Dokumenten mittels "=Nr." solche Links eingebaut.) GENIOS kann dies derzeit nicht; an einer Lösung wird jedoch gearbeitet. Vorgesehen sind bei GENIOS (nicht nur in der ChartLib) Hyperlink

Momentan muß hier eine zweite Suche in der entsprechenden Textdatenbank gestartet werden. Als Suchstrategien bieten sich zwei Varianten an:

- Recherche nach der Dokumentnummer des Volltextes (geht nicht immer) oder
- Recherche nach sinntragenden Termen des Artikel-titel

Der Nutzer muß die Angaben jeweils notieren und neu eingeben. Sicherergestellt ist, dass der Nutzer die Textinformationen stets findet.

## Fazit

Auch andere Online-Archive bieten Fakten in Form von Tabellen oder Abbildungen an. Aber die meisten nicht in einer Datenbank. Ausnahme ist Profound, das die Tabellen und Graphiken seiner Marktforschungsberichte (meist als PDF-Datei) unter einer Oberfläche offeriert. Natürlich hat z.B. Lexis-Nexis Patentzeichnungen oder Warenzeichen

### Beispiel 1

CHARTLIB - Die GENIOS Grafikbibliothek  
 LEBENSMITTEL PRAXIS NR. 004 VOM 27.01.1998 SEITE 054  
 Markt Süßwaren  
 Wrigley / Neue Vertriebskanäle  
 Unternehmen: Wrigley  
 Namen: Pittelkow, Frouke  
 Laender: C4WGE Bundesrepublik Deutschland C4DEU C4BRD  
 Themen: Süßwaren; Kaugummi; Umsatz; Absatz; Vertriebstypen  
*Den Originaltext zur Grafik/Tabelle finden Sie in Datenbank: LP Dokumentnummer: 5092134980227*  
 Datenbank CHARTLIB  
 Dokumentnummer: LP5092134980227

### Beispiel 2:

CHARTLIB - Die GENIOS Grafikbibliothek  
 das Rathaus Nr. 04 vom 01.04.1999 Seite 101  
 das Thema  
*Den Originaltext zur Grafik/Tabelle finden Sie in Datenbank: RATHAUS Dokumentnummer: 0499001029*  
 Datenbank CHARTLIB  
 Dokumentnummer: RATHAUS0499001029

Abb. 2: Trefferanzeige bei der ChartLib



als Graphiken, aber es hat nicht die Abbildungen aus seinen News-Quellen. Das Angebot von GENIOS ist in der Tat innovativ.

Unsere Befürchtung, dass die vom Text losgelöste Graphik durch diese Isolation ihren Sinn verlieren könnte, erwies sich als unzutreffend. Es kann - allerdings derzeit recht umständlich - jederzeit vom Bild zum Text gewechselt werden, wenn dies denn zur Interpretation erforderlich ist.

Der Bedienungskomfort wäre mit einer einheitlichen Indexierung aller Graphiken beträchtlich zu steigern. Derzeit wird überhaupt nicht inhaltlich erschlossen, man läßt die Bild- und Tabellenüberschriften für sich sprechen. Es verwundert, dass GENIOS sein eigenes Klassifikationssystem, das etwa für Handelsblatt-Artikel eingesetzt wird, nicht auch auf die einzelnen Charts anwendet. Man könnte auch über eine Indexierung nach dem Standard-Thesaurus Wirtschaft nachdenken. Bei so wenigen Datensätzen (es sind ja nur gut 10.000 pro Jahr - zugegebenermaßen mit stark steigender Tendenz) wäre eine intellektuelle Indexierung doch wohl drin. Angesichts dieser Möglichkeiten, die GENIOS noch ausschöpfen kann, vergeben wir in der Rubrik Bedienungskomfort nur zwei Punkte (von fünf möglichen).

Die Qualität der Inhalte hängt stark von den Bemühungen der Datenbankproduzenten und weniger von GENIOS ab. Zum Teil ärgern Inhalte, wenn ein ganzer Aufsatz als PDF-Datei geliefert wird und nicht die einzelnen Abbildungen darinnen. Große Fehler konnten wir jedoch nicht finden. Beim Umfang der Quellen dürfen die jetzigen Aktivitäten nur als Anfang verstanden werden, fehlen doch z.B. alle Abbildungen aus deutschen wirtschaftswissenschaftlichen Zeitschriften. Dies haben zwar vorrangig die Informationsproduzenten zu verantworten, aber GENIOS könnte ggf. Ratschläge geben. (Ein Beispiel: Die "ifo Literaturdatenbank" bei GENIOS enthält die Abstracts aller Artikel des "ifo Schnelldienstes", Volltexte einiger Artikel liegen

als Fließtext kostenlos im World Wide Web; die Abbildungen, die für diese Zeitschrift wesentlich sind, sind elektronisch überhaupt nicht erreichbar. Optimal wäre es, wenn Abstract, Fließtext und Abbildungen innerhalb eines Systems aufliegen könnten.) Die von GENIOS grundsätzlich eingeschlagene Richtung stimmt aber: vier Punkte für die Qualität der Inhalte.

Der Preis erscheint zunächst recht hoch: 7,20 DM für ein "Bildchen" aus dem Handelsblatt bei der WEB-SEARCH könnte abschrecken. Bedenkt man aber Aufwand und Kosten, wenn die Originale via Online-Ordering bei einer Bibliothek beschafft werden müßten, erscheinen die Preise durchaus marktgerecht. (dass ich die Preise des gesamten Online-Marktes für zu teuer einstufe, sei hier nicht berücksichtigt.) Für diese Dimension erscheinen vier Punkte als angemessen.

Letztlich erweist sich ChartLib als interessantes Produkt, das vom Nutzer sicherlich gebührend wahrgenommen werden sollte. Findet das Produkt seinen Markt, muß GENIOS das Angebot erweitern und methodisch optimieren. ■

Wolfgang G. Stock

## Weiterführende Informationen

GENIOS Wirtschaftsdatenbanken / Kundenservice

Kasernenstr. 67  
40213 Düsseldorf

Tel.: 0211/887-1525

(Karola Peppinghaus)

oder -1524, -1529

Fax: 0211/887-1520

E-Mail: kundenservice@geniode  
oder

GENIOS Wirtschaftsdatenbanken  
Gärtnerweg 4-8

60322 Frankfurt

Tel.: 069/955108-13 (Sabine Klehm)

Fax: 069/5963434

E-Mail: klehm@geniode

Homepage: <http://www.geniode>

## Literaturhinweise

*Grafiken in GENIOS über InMedia. Suche in der Chartlib. Grafiken in GENIOS über WebSearch.* - Düsseldorf: GENIOS Wirtschaftsdatenbanken / Kundenservice, 1999.

**Eckhard Rahlenbeck:** Genios InMedia, der persönliche Recherche-Assistent. - In: *Password* Nr. 5 (1996), I-VI

### Auf dem ersten Blick

## ChartLib von GENIOS

Preis-Leistungsverhältnis: ○○○○

Bedienungskomfort: ○○

Qualität der Inhalte: ○○○○