

IZ-Arbeitsbericht Nr. 12

**Text-Fakten-Integration in ELVIRA**

Jürgen Krause, Thomas Mandl, Maximilian Stempfhuber

Dezember 1997

Das Projekt ELVIRA II wird vom  
Bundesministerium für Wirtschaft gefördert.  
Aktenzeichen IV C2-003060/22



InformationsZentrum  
Sozialwissenschaften

Lennéstraße 30  
D-53113 Bonn  
Tel.: 0228/2281-0  
Fax.: 0228/2281-120  
email: krause@bonn.iz-soz.de  
      mandl@bonn.iz-soz.de  
      stempfhuber@bonn.iz-soz.de  
Internet: <http://www.social-science-geis.de>

ISSN: 1431-6943  
Herausgeber: Informationszentrum Sozialwissenschaften der Arbeits-  
                  gemeinschaft Sozialwissenschaftlicher Institute e.V.  
                  (ASI)  
Druck u. Vertrieb: Informationszentrum Sozialwissenschaften, Bonn  
                          Printed in Germany

Das IZ ist Mitglied der Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen e.V. (GESIS),  
einer Einrichtung der Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz (WGL)

## Inhalt

<b>1 Einleitung</b>	4
<b>2 Datengrundlage</b>	4
<b>2.1 Textkorpora</b>	5
<b>2.2 Fakten: Zeitreihen</b>	6
<b>2.3 Zusatzfakten</b>	6
<b>2.4 Aufbereitete Tabellen und Grafiken</b>	7
<b>2.5 Internet-Links</b>	7
<b>2.6 Betriebswirtschaftliche Kennzahlen</b>	8
<b>2.7 Weitere Faktendatenbestände</b>	8
<b>3 Szenarien für die Integration Text-Fakten</b>	9
<b>3.1 Reinformen</b>	10
<b>3.2 Text-Fakten-Integration durch Transformationen</b>	13
3.2.1 Probleme bei Transformationen	16
3.2.2 Transformationen von Anfragen	17
3.2.3 Transformationen in iterativen Prozessen	20
<b>3.3 "Echte" Integration</b>	22
<b>3.4 Ansätze Schalenmodell</b>	23
<b>4 Vorgehensweise</b>	25
<b>4.1 Prototypische Realisierung</b>	25
<b>4.2 Evaluierung</b>	25
<b>5 Fazit</b>	26
<b>6 Literaturverzeichnis</b>	27

# 1 Einleitung

Dieser Arbeitsbericht beschreibt die Text-Fakten-Integration in ELVIRA (Elektronisches Verbandsinformations-, Recherche- und Analysesystem) aus Sicht der dabei auftretenden informationellen Prozesse. Dazu wird kurz der Datenumfang vorgestellt. Das zweite Kapitel entwickelt aus der Theorie des Information Retrieval (IR) ein Modell, das bereits durch die ersten empirischen Ergebnisse geprägt ist.

## 2 Datengrundlage

In den ersten Befragungen von Verbandsmitarbeitern wurde versucht, die verschiedenen Typen von Daten zu identifizieren, die potentiell für ELVIRA interessant sind. Die vorgestellte Menge ist damit als Maximalansatz zu verstehen, die sich in den ersten Prototypen nicht unbedingt widerspiegelt. Für die Konzeption sollen aber möglichst viele Datentypen berücksichtigt werden.

Die Unterscheidung zwischen Texten und Fakten ist zunächst intuitiv klar. Texte sind natürlichsprachliche Aufzeichnungen, während Fakten aus Zahlen in Tabellen bestehen. Diese einfache Unterscheidung wirft aber Probleme auf. Texte enthalten oft Fakten, wie etwa der Satz: "Adenauer wurde 1876 geboren" Die gleiche Information könnte ein Nutzer einer relationalen Tabelle entnehmen. Tabellen bestehen andererseits nicht nur aus Zahlen sondern auch aus Textbestandteilen.

Entscheidend ist in einem Informationssystem die Nutzersicht. Für den Nutzer ist es gleichgültig, ob er eine gewünschte Information einem natürlichsprachlichen Text oder einer Tabelle entnimmt, oder ob ihm für sein Problem ein Experte genannt wird, bei dem er die Information erhalten kann. Der Mensch kann Wissen aus Texten extrahieren und in Tabellenform darstellen und umgekehrt. Konzeptuell stehen Texte und Fakten im Information Retrieval also auf einer Stufe. Der Stand der Forschung und Technik erlaubt eine Extraktion von Daten/Fakten aus beliebigen natürlichsprachlichen Texten bei großen Datenmengen noch nicht. Die Inhaltserschließung benutzt daher verschiedene Verfahren. Die intuitive Definition reicht für die Arbeit innerhalb von ELVIRA

aus. Wichtig ist für die folgenden Betrachtungen v.a., daß Texte und Fakten bei den heute vorliegenden Datenbeständen verschieden indexiert sind.

## 2.1 Textkorpora

Nach den ersten Untersuchungen werden folgende Textbestände in der Projektphase berücksichtigt:

- eigene Publikationen der Verbände VDMA (Verband Deutscher Maschinen- und Anlagenbau), ZVEI (Zentralverband der Elektrotechnik und Elektronikindustrie) und HVB (Hauptverband der Deutschen Bauindustrie), z.B. ZVEI-Mitteilungen, ZVEI-Quartalsbericht
- Textdatenbank der BfAI (Bundesstelle für Außenhandelsinformationen) Diese Datenbank besteht zum einen aus Volltexten, die von Länderkorrespondenten der BfAI erstellt werden und zum anderen aus Referenzen auf Veröffentlichungen der BfAI und verschiedener Stellen im Ausland (z.B. Zollregelungen). Daneben enthält die von der BfAI vertriebene CD-ROM noch Adressen und Projekte, die in Abschnitt 2.7 noch erwähnt werden.
- Nachrichtendatenbank Reuters (teilweise englisch)
- nach Möglichkeit Zeitungskorpora (z.B. FAZ, Handelsblatt)
- Publikationen der Forschungsinstitute (ifo-Länderberichte, IW-Trends)
- weitere externe Textdatenbanken werden in die Überlegungen mit einbezogen

Bei verbandsinternen Publikationen bietet sich eine intellektuelle Verschlagwortung nach den Thesauri des Faktenzugangs an. Dadurch werden Texte, die Zeitreihen interpretieren, mit diesen verknüpft. Beispiele hierfür sind Texte von Fachverbänden (Untereinheiten, die ein bestimmtes Produktspektrum betreuen), in denen Besonderheiten bei der Entwicklung einzelner Produkte analysiert werden. Auch für Zeitreihen wie *”die Entwicklung des Maschinenbaus allgemein”* gibt es solche interpretierenden Texte.

Bei den externen Textkorpora ist wohl nur eine automatische Indexierung möglich. Im Gesamtkonzept existieren aber einzelne Korpora, die intellektuell indexiert sind und für die Thesauri gepflegt werden. Dazu gehören etwa die

Datenbestände des VCI (Verband der Chemischen Industrie) und die Texte der BfAI. Diese Informationen sollten soweit wie möglich genutzt werden.

## **2.2 Fakten: Zeitreihen**

In ELVIRA II wird auf die gleichen statistischen Zeitreihen zugegriffen, die schon in ELVIRA I verfügbar waren. Diese Datenbasis ist von Verband zu Verband unterschiedlich, die Strukturen sind jedoch identisch. Einen Überblick über den momentan erreichten Stand bietet Mandl/Stempfhuber 1997).

Die Zeitreihen werden von den Verbänden gepflegt und intellektuell nach vorgegebenen Nomenklaturen verschlagwortet. In ELVIRA I werden die Zeitreihen anhand dieser Deskriptoren recherchiert. Dies stellt einen Spezialfall der Faktenrecherche dar, bei dem die im Textretrieval üblichen Deskriptoren verwendet werden (cf. Krause/Mandl/Stempfhuber 1997).

## **2.3 Zusatzfakten**

Die Texte in den oben genannten Korpora enthalten typischerweise auch Fakten. Beispiele hierfür sind die Texte der BfAI, in denen auch Tabellen enthalten sind. Die Daten stammen aus den verschiedensten Quellen und haben in der Regel nicht den Stellenwert und die hohe Qualität der Zeitreihen der Verbände. Es handelt sich jedoch oft um andere Daten, die im Zeitreihenkorpus nicht vorkommen. Deshalb sind sie für den Benutzer in spezifischen Situationen interessant.

Erste Befragungen haben gezeigt, daß diese Fakten auch unabhängig von den Texten einen Informationswert bieten und von den Verbänden durchaus ohne den Kontext des Textdokumentes an Informationssuchende weitergegeben werden. Deshalb soll in ELVIRA II versucht werden, diese Informationen bei der Indexierung soweit wie möglich automatisch zu erkennen und als zusätzlichen Faktendatenbestand anzubieten. Diese aus Texten gewonnenen Fakten sollen Zusatzfakten genannt werden, um sie terminologisch von den Zeitreihen abzugrenzen.

## 2.4 Aufbereitete Tabellen und Grafiken

Zusatzfakten sind auch in von den Verbänden selbst erstellten Texten enthalten. Daneben erstellen die Verbände Tabellen und Grafiken auch unabhängig von Texten. Der VDMA etwa erstellt Grafiken zu aktuellen wirtschaftspolitischen Themen, die sich besonders für Folien zu Vorträgen eignen. Auch diese können den Firmen zur Verfügung gestellt werden. Besonders die Grafiken bieten durch die Auswahl und Aufbereitung der Daten einen Mehrwert gegenüber Rohdaten in Zeitreihenform

Tabellen und Grafiken, denen Daten aus den Zeitreihen zugrunde liegen, können diesen Zeitreihen intellektuell und ohne großen Mehraufwand zugeordnet werden. Dadurch steht dem Benutzer nicht nur die aufbereitete Sicht auf die Daten zur Verfügung, sondern er kann die entsprechenden Zeitreihen recherchieren und in seiner Auswertung Parameter wie z.B. den Zeitausschnitt ändern oder die Daten mit eigenen Zeitreihen verbinden.

## 2.5 Internet-Links

Der VDMA bietet als zusätzlichen Datenbestand ein Internet-Clearinghouse an. Ein oder mehrere Mitarbeiter, die häufig im Internet recherchieren, erstellen dieses Clearinghouse aus ihren Bookmarks und pflegen es. Zu jedem Link verfassen sie eine Kurzbeschreibung von einigen Zeilen. Dieser Kurztext kann dann automatisch indexiert werden. Beim VCI liegt ein solches Clearinghouse im Intranet bereits vor.

Dadurch wird zusätzlich der Informationsraum Internet für die ELVIRA-Nutzer erschlossen. Ist der Nutzer mit den Recherche-Ergebnissen aus den Datenbanken nicht zufrieden, kann er anhand der Links im Internet weitersuchen. Er muß also nicht den mühsamen Weg über Suchmaschinen und Hypertext-Links beschreiten, sondern erhält durch Vorauswahl, Qualitätsüberwachung und Inhaltserschließung des Verbandes bereits einen Ausschnitt des Internets, in dem eine weitere Suche sinnvoll erscheint.

Neben einer Referenzdatenbank von Links ist auch eine zweite Möglichkeit der Erschließung von Internet-Informationen in ELVIRA denkbar. Dabei würden automatisch Seiten aus dem Internet gespeichert und als eigene Text-

Dokumente abgelegt. Diese Dokumente wären über die Indexierung direkt recherchierbar, ohne daß der Benutzer Zugriff zum Internet braucht.

## **2.6 Betriebswirtschaftliche Kennzahlen**

Die betriebswirtschaftlichen Kennzahlen sind ein Fakten-Datenbestand des VDMA, der als Print-Produkt veröffentlicht wird. Die Kennzahlen resultieren aus einer regelmäßig unter Mitgliedsfirmen durchgeführten Umfrage, wobei die Ergebnisse über den gesamten Maschinenbau gemittelt werden. Sie informieren z.B. darüber, wie hoch die Investitionsquote im Maschinenbau ist und wieviel Prozent des Preises typischerweise für Vertriebsrisiken einkalkuliert werden. Ein ähnlicher Datenbestand für den HVB findet sich in einer Publikation des Fraunhofer Informationszentrums Raum und Bau.

Die betriebswirtschaftlichen Kennzahlen sind zwar Zeitreihen, der zeitliche Aspekt ist jedoch nachgeordnet. Die historische Entwicklung ist weniger wichtig als der Vergleich oder Überblick über mehrere aktuelle Kennzahlen. Der Drei-Facetten-Zugang ist für ihre Beschreibung gut geeignet, so daß sie im gleichen Stil wie die Zeitreihen recherchiert werden können. Zur Darstellung sollte aber nicht die Zeitreihen-Tabelle benutzt werden, sondern an die Print-Form angelehnte Tabellen, in denen zusammengehörende Kennzahlen nebeneinander stehen. Ein wichtiger Aspekt der Kennzahlen sind die dazugehörigen textuellen Definitionen, die in der gedruckten Version etwa die Hälfte der Platzes beanspruchen.

## **2.7 Weitere Faktendatenbestände**

Ein für die Firmen wichtiger Datenbestand sind Messen. Die Messedatenbank der AUMA (Ausstellungs- und Messe-Ausschuß der Deutschen Wirtschaft e.V.) stellt eine interessante Mischung von Text und Fakten dar. Zum einen besteht sie aus typischen tabellarischen Fakten wie Ausstellungsfläche, Besucherzahl und Datum der Messe. Daneben enthält die Beschreibung bezüglich Branchenschwerpunkt und Angebotsschwerpunkt nur Texte. Zusätzlich können die Adressen der Veranstalter abgefragt werden.

Adressen spielen auch außerhalb der Messen eine Rolle. Die BfAI unterhält eine Adressdatenbank zu ausländischen Stellen und Anwälten. Auch Firmenadressen können bei der Suche nach Herstellern bestimmter Produkte interessant sein.

### **3 Szenarien für die Integration Text-Fakten**

Bei der Text-Fakten-Integration werden drei Stufen unterschieden:

#### **A. Fakten- und Textrecherche in Reinform**

In integrierten Informationssystemen, die Text- und Faktenretrieval erlauben, wird es immer Benutzer geben, die nur nach Texten oder nur nach Fakten suchen, die also ihr Informationsbedürfnis hinsichtlich des Typs der Ergebnisdokumente einschränken. Deshalb müssen in ELVIRA II sowohl Text- als auch Faktensuche in Reinform möglich sein. Die Faktensuche ist bereits in ELVIRA I realisiert und wird übernommen (cf. Krause/Mandl/Stempfhuber 1997). Für die Textrecherche wird ein neuer Zugang realisiert, der in der Grundkonzeption an das Textretrieval in sozialwissenschaftlichen Datenbanken (Projekt GESINE, cf. Marx et al. 1995) und an die experimentelle Realisierung KONEXIS (cf. Schaefer 1997) angelehnt ist.

#### **B. Transformationen zwischen den Suchtypen**

Auf den Reinformen für Texte und Fakten bauen die Transformationen auf. Der Benutzer kann das für einen Modus bereits formalisiertes Informationsbedürfnis auf einen anderen Modus anwenden, ohne die Anfrage neu formulieren zu müssen. So kann das System z.B. nach einer Faktenrecherche dem Benutzer zusätzlich Texte zu den Fakten liefern.

#### **C. "Echte" Integration**

Die "höchste" Stufe der Integration geht nur vom Informationsbedürfnis aus. Der Benutzer hat ein Problem und legt nicht fest, ob er Texte oder Fakten als Ergebnisdokumente erhalten will. Das System erkennt unabhängig vom Modus, welche Dokumente relevant sind.

Im folgenden werden diese drei Stufen näher erläutert. Dabei beschränkt sich die Darstellung hauptsächlich auf die Integration von Texten und statistischen Zeitreihen.

### **3.1 Reinformen**

Die Prozesse bei den Reinformen der Recherche lassen sich folgendermaßen beschreiben:

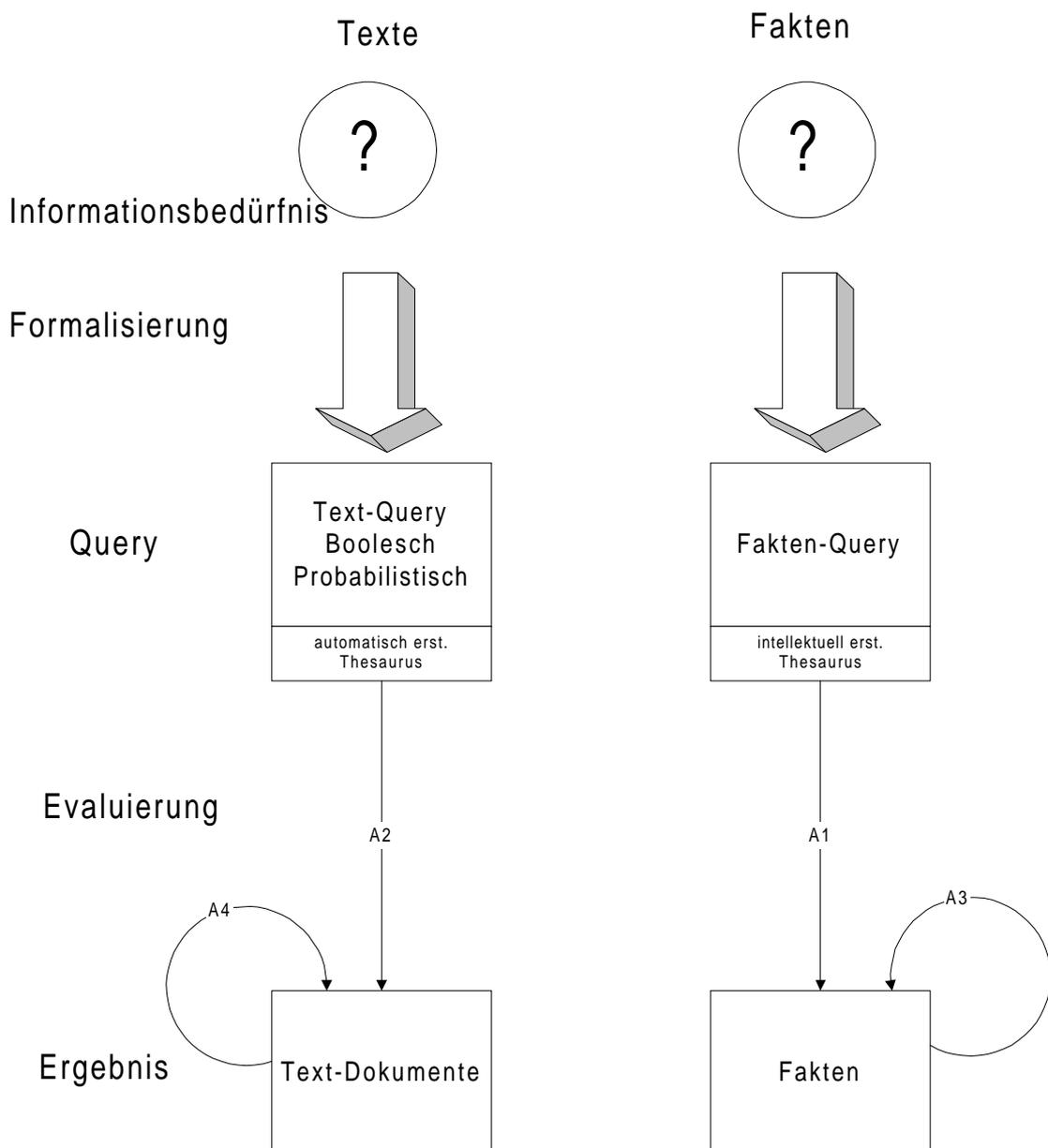
- A1 Fakten-Anfrage - > Fakten-Ergebnis
- A2 Text-Anfrage - > Text-Ergebnis

Die bisherige ELVIRA-Faktenanfrage entspricht dem Typ A1. A2 stellt eine Standardrecherche in einem Text-Retrieval-Objekt dar. Diese greift auf automatisch oder intellektuell indexierte Texte zu. Auch bei dieser Reinform ergeben sich erste Synergie-Effekte aus der Kombination von Text- und Faktensuche in einem System.

In der Praxis zeigt sich, daß Informationsprozesse meist iterativ ablaufen. Ein Benutzer stellt mehrere Anfragen und bewertet mehrere Ergebnisse bevor er zu seinem Endergebnis kommt. Dies kann so ablaufen, daß der Benutzer mit dem Ergebnis unzufrieden ist und eine neue Anfrage formuliert, weil seine Suchbegriffe im Sinne der Indexierungssprache sein Informationsbedürfnis nicht gut abbilden. Häufig verändert sich im Laufe der Interaktion das ursprüngliche Informationsbedürfnis. Eine spezifische Art des iterativen Retrievals ist relevance feedback, wie es in probabilistischen Retrievalmodellen angelegt ist. Dabei markiert der Benutzer Dokumente oder Teile von Dokumenten als für ihn relevant. Das System erstellt auf Basis der alten Anfrage eine neue, bei der die in den relevanten Dokumenten vorkommenden Terme höher gewichtet sind feedback (cf. Womser-Hacker 1997).

Daraus ergeben sich für die Reinformen folgende zusätzliche Arten von Retrievalprozessen:

- A3 Fakten-Ergebnis - > Fakten-Ergebnis
- A4 Text- Ergebnis - > Text-Ergebnis



**Abb. 3.1: Reinformen der Recherche**

Für den Fall A3 wird eine Erweiterung für den Faktenzugang von ELVIRA getestet. Dabei kann der Benutzer in der Grafik die Browser einblenden und Parameter ändern. Dadurch kann er z.B. zu einer Zeitreihe im internationalen Außenhandel zusätzliche Länderkombinationen einblenden oder zu einer Konjunktur-Zeitreihe zur Nachrichtentechnik zusätzlich den Kurvenverlauf zur übergeordneten Branche Elektrotechnik betrachten. Dies bedeutet, daß die Zustandsanzeige auch in der Grafik manipulierbar bleibt (cf. Mandl/Stempfhuber 1997).

Der Fall des Textretrievals (A4) wird z.B. durch Hypertextlinks auf Texte vom gleichen Autor realisiert. Dabei ist die Verknüpfung eindeutig deduktiv ableitbar. Markiert der Benutzer im Sinne von relevance feedback Dokumente und löst dann eine modifizierte Suche aus, so kann die Beziehung zwischen diesen Dokumenten und dem neuen Ergebnis nur vage bestimmt werden. Die im relevance feedback eingesetzten Verfahren müssen auf ihre Eignung im ELVIRA-Kontext untersucht werden. Die Extraktion von Deskriptoren aus markierten Textpassagen ist ebenfalls ein Standard-Verfahren, das bereits in vielen experimentellen (z.B. FAPSY) und kommerziellen Systemen eingesetzt wird.

Wie bei den Texten gibt es auch bei den Fakten sowohl deduktiv eindeutig ableitbare als auch vage Beziehungen. Der oben geschilderte Fall von A3 nutzt eindeutig ableitbare Zusammenhänge. Daneben gibt es zahlreiche Beziehungen zwischen Faktendaten, die vager Natur sind. So beeinflussen manche Zeitreihen andere und dienen daher für Prognosen. Maßzahlen hierfür können mit Korrelationsanalysen oder anderen statistischen Verfahren berechnet werden. Erste empirische Untersuchungen im Rahmen von ELVIRA ergaben jedoch keine entsprechenden Informationsbedürfnisse (cf. Mandl/Stempfhuber 1997). Möglicherweise lassen sich solche Beziehungen sinnvoll über die Kookurrenzen der Fakten-Deskriptoren in den Text-Korpora herleiten.

Die Reinformen treten immer dann auf, wenn ein Benutzer sich bewußt auf einen Modus beschränkt und nur Dokumente dieses Typs in der Ergebnismenge will. Selbst bei den Reinformen kann es jedoch sinnvoll sein, dem Benutzer entgegen seiner Entscheidung für Fakten zusätzlich einen Text zu liefern. Der Text kann für das Verständnis der gelieferten Fakten wichtig sein, er kann die Qualitätsbeurteilung beeinflussen oder auf eine aktuelle Entwicklung verweisen, die für die Interpretation der Zahlen wichtig ist.

**Beispiel:** Teilweise fragen Benutzer nach Daten, die nicht vorliegen. Anstatt einer Nullantwort, sollte das System idealerweise flexibel reagieren und Parameter gezielt abschwächen, um so ein Ergebnis vorlegen zu können. Folgende Anfrage ergibt eine Nullantwort:

- aktuelle Arbeitskosten im Maschinenbau in Südostasien

Im System sind aber folgende Daten vorhanden:

- Arbeitskosten von 1993 im Maschinenbau in Südostasien

- aktuelle Arbeitskosten in der *verarbeitenden Industrie* in *westlichen Industrieländer*

Dabei handelt es sich um Zusatzfakten, die nicht in der Zeitreihendatenbank enthalten sind, sondern in einer Tabelle in einem Text vorkommen. Bei der Indexierung soll ELVIRA erkennen, daß dieser Teil des Textes eine Tabelle mit Fakten enthält und soll die darin vorkommenden Begriffe benutzen.

- aktuelle *Bruttostundenverdienste* im Maschinenbau in Südostasien
- ein Text, der berichtet, daß die Arbeitskosten in Südostasien in den letzten Jahren stark gestiegen sind

Die drei ersten Dokumente bestehen aus Fakten, in denen Bedingungen der ursprünglichen Suche abgeschwächt sind. Je nach Aufgabenstellung kann der Benutzer entscheiden, welche Parameter für ihn wichtig sind und eines der Faktendokumente auswählen. Analog dem Vorgehen eines menschlichen Informationsvermittlers sollte ihm das erste Faktendokument aber nicht ohne das letzte Dokument, einen Text vorgelegt werden, selbst, wenn er die Suche auf Fakten beschränkt hat. Denn der Text ist für die Aussagekraft der Fakten entscheidend, er schränkt sie stark ein. Der Text stellt in diesem Fall einen Mehrwert dar, auch wenn der Benutzer eigentlich nur an Fakten dachte.

Unter Einbeziehung der erzwungenen Integration ergeben sich zusätzlich folgende Prozesse:

- A5 Fakten-Anfrage - > Text-Ergebnis
- A6 Text-Anfrage - > Fakten-Ergebnis

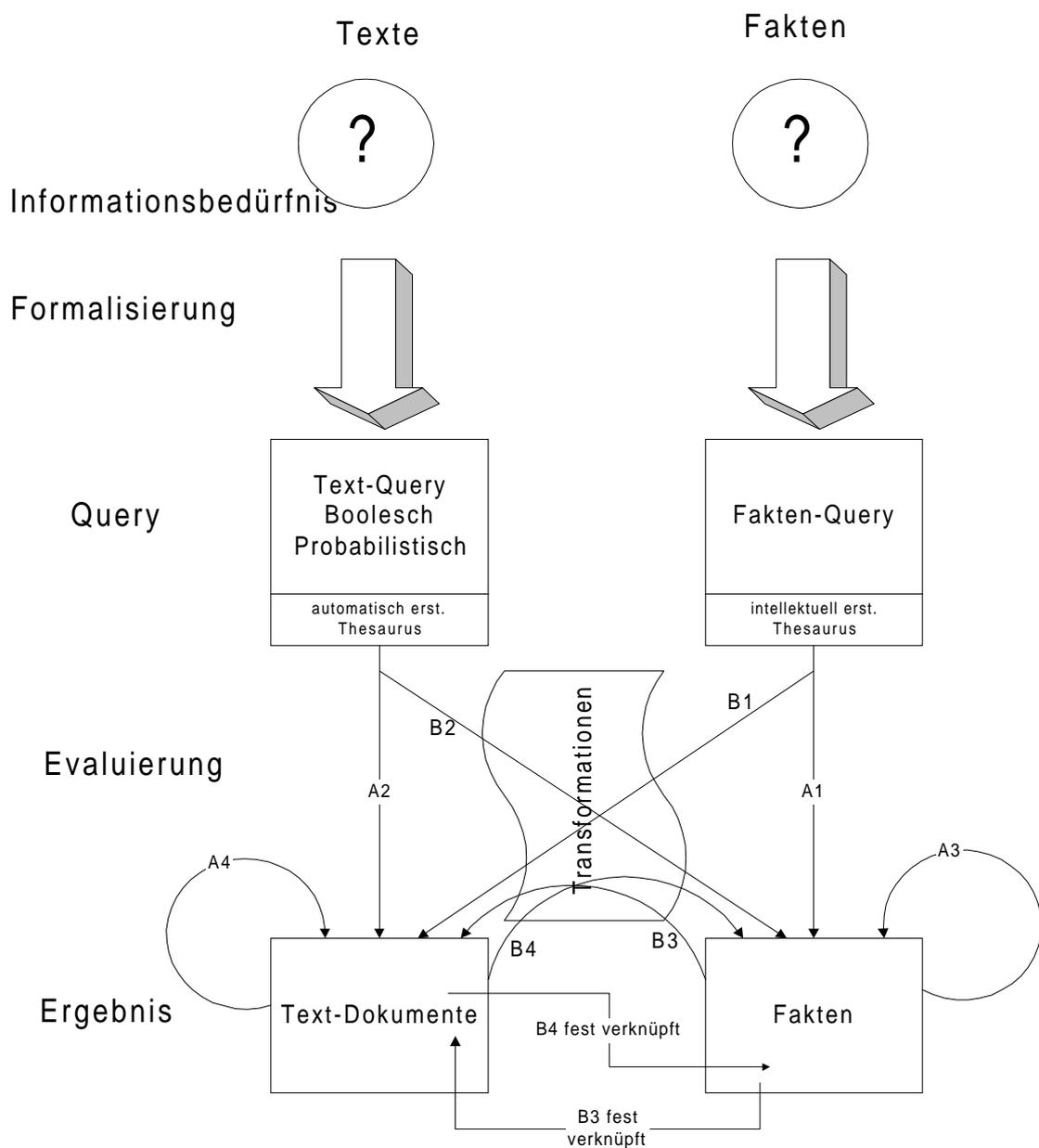
Dieser Typ steht zwischen Reinform und Transformation. Aus Nutzersicht scheint es sinnvoll, ihn unter Reinform zu behandeln, wenn er auch auf zahlreiche Konzepte der Transformationen zurückgreift.

## **3.2 Text-Fakten-Integration durch Transformationen**

Bei den Transformationen führen Anfragen aus einem Modus zu Ergebnissen in einem anderen. Solche Informationsbedürfnisse können auch im iterativen Retrieval vorkommen. Dies führt zu folgenden Fällen:

- B1 Fakten-Anfrage - > (Fakten- und) Text-Ergebnis
- B2 Text-Anfrage - > (Text- und) Fakten-Ergebnis
- B3 Fakten-Ergebnis - > Text-Ergebnis
- B4 Text- Ergebnis - > Fakten-Ergebnis

Die folgende Grafik gibt einen Überblick über die Reinformen und die Transformationen:



**Abb. 3.2: Reinformen und Transformationen**

Der Mehrwert bei den Transformationen besteht darin, daß der Benutzer eine bereits formalisierte Anfrage nicht neu formalisieren muß, um Dokumente in einem anderen Modus zu erhalten. Die Formalisierung eines Informationsbedürfnisses zu einer Anfrage in einem Informationssystem ist nach der Theorie der informationellen Prozesse ("cognitive viewpoint", cf. Ingwersen 1992) immer mit einem Verlust verbunden. Diese Formalisierungsarbeit soll der Benutzer nur einmal leisten müssen. Das System übernimmt die Anfrage und transformiert sie in eine formale Anfrage für einen anderen Modus. Dazu nutzt das System Wissen über die Zusammenhänge zwischen den beiden Repräsentationsmechanismen, im Fall von ELVIRA, den Indexierungssprachen.

Durch iterative Prozesse mit Ergebnisevaluierungen kann ein Benutzer eventuell seine Query so verändern, daß sie eher seinem ursprünglichen Informationsbedürfnis entspricht. Eine wichtige Rolle spielt dabei relevance feedback. Der Benutzer markiert Dokumente oder Teile davon als relevant und das System nutzt diese Information um daraus eine neue Query zu konstruieren. Relevance feedback kann in der Praxis sehr gute Ergebnisse bringen, wie die TREC-Studien zur Qualität von Retrieval-Verfahren gezeigt haben (cf. Womser-Hacker 1997). Der Benutzer hat durch seine Formalisierung des Informationsbedürfnisses für ihn relevante Dokumente gefunden. Diese Dokumente, die sein Informationsbedürfnis teilweise erfüllen, beeinflussen die neue Formalisierung des Informationsbedürfnisses. Die kognitive Last bei der Produktion der Anfrage ist wesentlich geringer, da Beispiele vorgegeben werden, und der Benutzer nicht in einer grafischen oder formalen Anfragesprache auf abstraktem Niveau sein Informationsbedürfnis beschreiben muß. Dem System bleibt nun der letzte Schritt der Abstraktion überlassen. Es benutzt die formale Repräsentation der relevanten Dokumente für die Konstruktion der neuen Query.

Hierbei wird der Mehrwert der Transformationen besonders deutlich. Relevance feedback ist zwar sehr erfolgversprechend, wird aber von Benutzern selten genutzt, da es mit Aufwand verbunden ist. Dieser Aufwand wird bei Transformationen nur einmal fällig. Ist der Benutzer durch relevance feedback einmal bei einer guten Formalisierung seines Informationsbedürfnisses angelangt, so ist der Aufwand dafür im anderen Modus nicht mehr notwendig.

### 3.2.1 Probleme bei Transformationen

Es ist zu erwarten, daß es nicht genügt eine einfache Kopie der Anfrage im jeweils anderen Suchobjekt einzusetzen. Dieser naive Ansatz trifft auf folgende Probleme:

- **Formale Unverträglichkeit:** Der Faktenzugang von ELVIRA arbeitet mit einem spezifischen Ausschnitt der Booleschen Logik, der die Aufteilung der Suchbegriffe in drei Facetten ausnutzt (cf. Krause/Mandl/Stempfhuber 1997). In dem Suchobjekt für Texte kann der Benutzer die Boolesche Logik mit größerer Freiheit benutzen, so daß Textanfragen entstehen, die in der ELVIRA-Faktensuche nicht dargestellt werden können. Weder bei Boolescher noch bei probabilistischer Text-Query hat das System Informationen darüber, in welche Facette des Faktenzugangs ein Suchbegriff gehört.
- **Anderes Indexierungsverfahren:** Die Deskriptoren, die Zeitreihen beschreiben sind intellektuell vergeben. Sie stehen innerhalb ihrer Nomenklatur in hierarchischen Beziehungen. Um die Beziehungen von Deskriptoren zwischen einzelnen Nomenklaturen und zu anderen gängigen Fachbegriffen zu bestimmen stehen Thesauri zur Verfügung. Dagegen werden die Textbestände in der Regel automatisch erschlossen und höchstens in Ausschnitten intellektuell indexiert.
- **Anderes Indexierungsvokabular:** Aus den verschiedenen Verfahren und der Natur der Bestände ergeben sich zwangsläufig verschiedene Indexierungssprachen. Bei intellektuell erschlossenen Beständen ist die Anzahl der Deskriptoren in der Regel niedriger als bei automatischer Erschließung. Somit haben selbst gleiche Deskriptoren unterschiedliche Differenzierungsfähigkeiten. Desweiteren scheinen sich die Texte in dem Anwendungsgebiet nicht so spezifisch auf einzelne Produkte zu beziehen, wie dies bei den tiefgegliederten Zeitreihen der Fall ist. Z.B. gibt es zu einer Zeitreihe auf unterster Produktebene nicht immer einen Text, sondern eher zu einer höheren Aggregationsstufe. Dies muß bei der Formulierung einer erfolgversprechenden Text-Query berücksichtigt werden.

Daraus ergibt sich, daß bei den Transformationen eine formale Anfrage in eine andere Anfragesprache übersetzt werden muß.

Für die Transformationen sollen in ELVIRA zum einen deduktive und zum anderen vage Verfahren eingesetzt werden. Besonders erfolgversprechend dürften kombinierte Methoden sein. Deduktive Verfahren nutzen z.B. die hierarchischen Beziehungen im Thesaurus aus. Wo dies nicht zu Erfolg führt, können vage Methoden etwa aus dem Bereich der Statistik oder Mustererkennung die Transformation leisten.

### 3.2.2 Transformationen von Anfragen

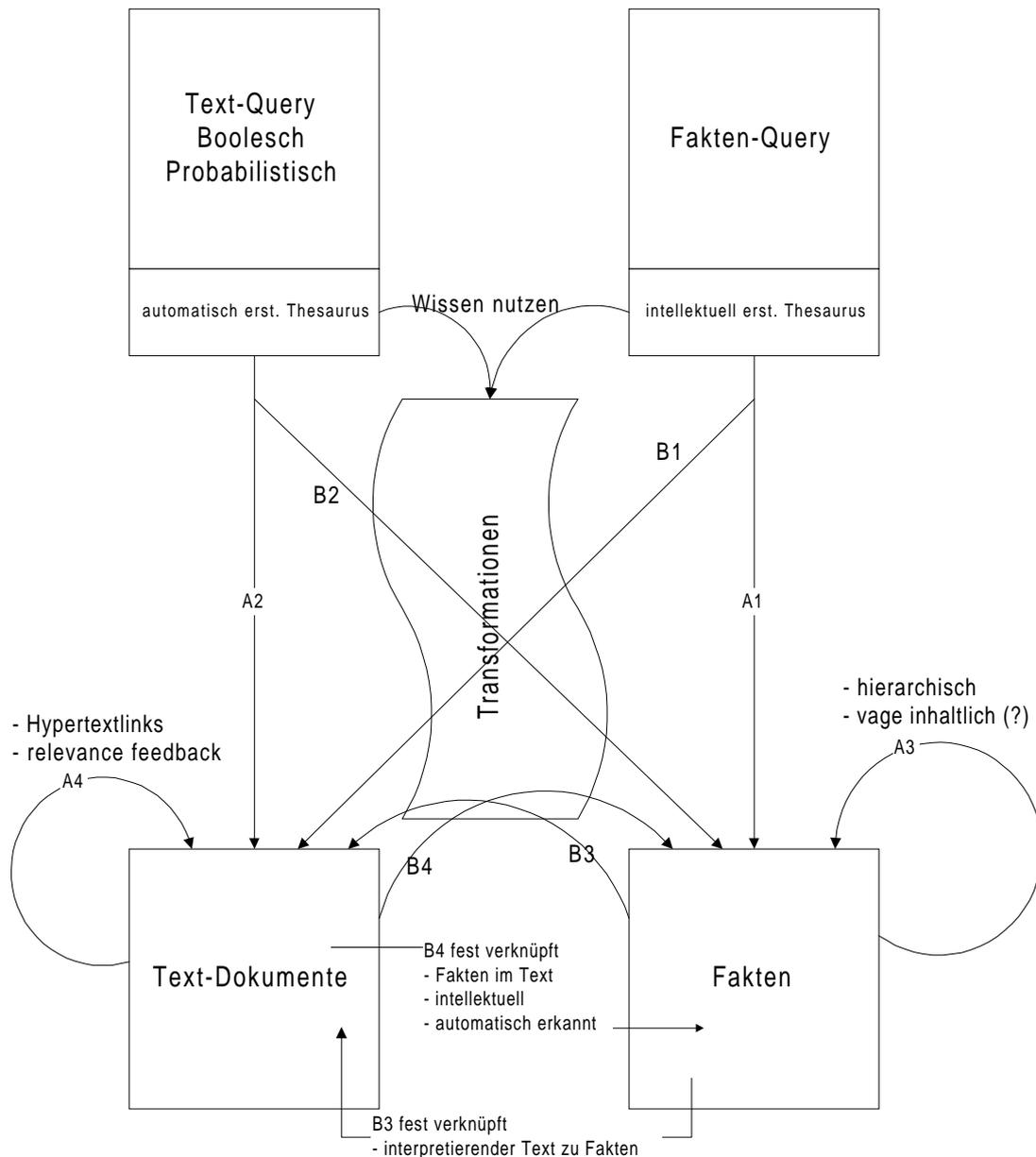
Im folgenden werden die verschiedenen Fälle B1 bis B4 näher erläutert.

Im Fall B1 wird eine formalisierte Fakten-Anfrage in eine Text-Anfrage umgesetzt. Dabei muß zunächst festgelegt werden, ob eine boolesche oder eine probabilistische Anfrage gestellt wird. Dann setzt das System die Deskriptoren aus den drei Facetten der Faktenanfrage in sinnvolle Text-Deskriptoren um, wobei je nach der Entscheidung boolesch oder probabilistisch eventuell verschiedene Deskriptoren sinnvoll sind.

Dabei werden deduktive Verfahren, wie Erweiterung aus intellektuell erstellten Thesauri und probabilistische Verfahren, wie Erweiterung durch automatisch erstellten Text-Thesaurus (Term-Dokument-Matrix) eingesetzt. (z.B. Zeitreihen-Branche zu spezifisch, dazu kein Text, sondern nur zu einer Ebene höher).

**Beispiel:** Folgende Anfrage wurde beim VDMA gestellt: Bietet Kolumbien Marktpotentiale für den Einstieg in den lateinamerikanischen Markt? Nach Interpretation der Sachbearbeiterin ist dies in erster Linie eine Faktenanfrage. Antwort ist die Zeitreihe "Thema: Export, Import; Branche: Maschinenbau insgesamt; Land: Summe Berichtsländer aus/nach Kolumbien". Der Markt berechnet sich aus Produktion, Import und Export. Da Produktionsdaten für Kolumbien in der Zeitreihendatenbank momentan nicht zur Verfügung stehen, können nur zwei Komponenten recherchiert werden.

Als Ergänzung bieten sich Texte aus der BfAI an. Folgende intellektuell erstellte Query mit der BfAI CD-ROM bringt gute Ergebnisse:  
Eingabedatum: "960102" bis "970718" / Land: Kolumbien / Freitext: Maschinen\* und (Import oder Export)



**Abb.3.2.2 : Transformationen im Detail**

Das Thema der Zeitreihe, bzw. der Zeitreihen-Anfrage ist *Import*, ein Unterthema des Außenhandels. Daraus wird zum einen die Auswahl der Datenbank abgeleitet, zum anderen kann *Import* im Freitext als Suchwort verwendet werden. Da der BfAI Datenbestand ohnehin nur Dokumente zum Außenhandel enthält, bewirkt das Weglassen von *Import* eine Steigerung des Recalls; die Precision sinkt nicht sehr stark. Die zusätzliche Verwendung des Begriffs *Markt* bringt in diesem Fall keine Veränderung,

während die Erweiterung mit den Synonymen *Einfuhr* und *Ausfuhr* ca. 30% mehr Treffer bringt.

Die Festlegung des Eingabedatums auf das obige Intervall geschah rein heuristisch und ist dadurch motiviert, daß die Nutzer immer sehr aktuelle Texte präferieren.

Aus der Facette *Land* wurde das Partnerland *Kolumbien* in das Feld für Länder bei der Textsuche übernommen. *Summe Berichtsländer* entfällt als für die BfAI ungeeignetes Suchwort (entspricht bei Zeitreihen annähernd der gesamten Welt). Eine Tabelle oder ein Text, der über Importe Kolumbiens berichtet, ohne ein weiteres Bezugsland anzugeben, setzt ebenfalls die gesamte Welt als Bezug. Im BfAI Datenbestand führt das einfache Weglassen zu keiner Verschlechterung, da die Texte meist die Sicht des betreffenden Landes oder die Beziehungen zu Deutschland beschreiben. Die Zeitreihen dagegen erlauben die Abfrage der Datenströme zwischen den Berichtsländern und allen anderen Ländern. Bei den Texten ist dies nur sehr eingeschränkt der Fall. Trotzdem verbessert sich das Ergebnis erheblich, wenn Kolumbien nicht im Freitext, sondern im Länderfeld eingesetzt wird, das intellektuell verschlagwortet ist.

In der Facette *Branche* wurde *Maschinenbau insgesamt* zu *Maschinen\** expandiert, da es in diesem Bestand kein guter Freitext-Deskriptor ist.

**Beispiel:** Ein Benutzer hat Zeitreihen zu *Speicherprogrammierbaren Steuerungen* recherchiert. Laut Erläuterungen zur GP 95 (Güterverzeichnis für Produktionsstatistiken) gehören hierzu u.a. auch *Prozeßrechner*. Dies müßte in einem Thesaurus festgehalten sein. Da aufgrund der Zeitreihen-Suche nicht klar ist, welche Produkte die Firma produziert, sollten bei einer Transformation zu einer Text-Suche mit allen Synonymen und Unterbegriffen erweitert werden. Die Erweiterung mit *Prozeßrechner* führt im BfAI-Datenbestand zu einer höheren Trefferanzahl.

## B.2 Text-Anfrage - > (Text- und) Fakten-Ergebnis

Eine formale boolesche oder probabilistische Text-Anfrage muß in eine boolesche Faktenanfrage umgesetzt werden. Dabei ist zu versuchen, aus den beliebigen Suchbegriffen die richtigen Begriffe für die Faktensuche zu formen. Ein

Problem ist dabei die Umsetzung von umgangssprachlichen Produkt- und Branchenbegriffen in Begriffe der amtlichen Statistik. Bei einer Faktensuche müssen im aktuellen System alle Facetten belegt sein. Einzelne Facetten sind eventuell aus Filtereinstellungen abzuleiten, wenn sie in der Suche nicht explizit vorkommen.

**Beispiel:** Waschmaschinen in Ungarn

Ein Benutzer hat mit diesem Thema eine Recherche im BfAI Textbestand durchgeführt. Die Query umfaßte *Ungarn* als gebundenen Deskriptor für Land und den Begriff *Waschmaschinen*. Mit dem Ergebnis ist er zunächst zufrieden, er will aber nun genauere statistische Angaben zu diesem Thema. ELVIRA sollte nun die Zeitreihen zu den *Waschvollautomaten* finden, wie die Waschmaschinen auf Produktebene in der WA (Warenverzeichnis für die Außenhandelsstatistik) genannt werden. Als Themen sollte *Import* und *Export* gewählt werden, was in der Text-Anfrage so nicht vorkommt. Da der Benutzer aber im BfAI-Datenbestand gesucht hat, sollte ELVIRA diese Themen aufgrund der Filtereinstellung richtig setzen. Als Land muß *Ungarn* gewählt werden. Da Ungarn kein Berichtsland ist, kann es nur als Partnerland gewählt werden. Als Berichtsland sollte *Deutschland* und *Summe Berichtsländer* gewählt werden. Dies müßte als heuristisches Wissen über den Datenbestand festgelegt werden.

### 3.2.3 Transformationen in iterativen Prozessen

Bei den iterativen Prozessen liegt dem System mehr Information über das Informationsbedürfnis vor. Der Benutzer hat aus der Ergebnismenge bereits Text- oder Fakten-Dokumente (oder Ausschnitte wie einzelne Kurvenpunkte oder Text-Passagen) gewählt und will von diesen ausgehend weitersuchen. Dies entspricht dem relevance feedback mit Wechsel der Modalität. Der einfachste Fall sind jeweils intellektuell vergebene Verknüpfungen.

#### B.3: Fakten-Ergebnis - > Text-Ergebnis

Falls es zu einer Zeitreihe einen Text gibt, der dieser intellektuell zugeordnet wurde, ist dies in der Ergebnisdarstellung sichtbar. Der Benutzer kann diesen Text dann über einen Hypertext-Link anspringen. Für B4 gilt dies analog, die Sprungrichtung ist dann vom Text zur Zeitreihe.

Will der Benutzer darüber hinaus Texte zu einer oder mehreren für ihn relevanten Zeitreihen erhalten, so kann das System eine oder mehrere Faktenanfragen konstruieren, die genau zu diesen Zeitreihen führen. Diese können dann im Sinne von Fall B1 transformiert und abgearbeitet werden.

Darüber hinaus kann das System versuchen, aus dem gewählten Ausschnitt des Faktenergebnisses weiteres Wissen zu gewinnen. Wurden zwei Zeitreihen aus verschiedenen Datenbeständen ausgewählt, so will der Benutzer eine Beziehung herstellen. Z.B. könnte ihn interessieren, ob es Texte gibt, die zwischen den Zeitreihen eine Abhängigkeit feststellen. Markiert der Benutzer einen Bereich einer Zeitreihe als für ihn relevant und will nun hierzu Texte, so kann das System daraus sein Informationsbedürfnis konkretisieren. Handelt es sich etwa um auffällige Steigungen, so werden entsprechende zusätzliche Deskriptoren für die Textsuche eingesetzt. Ansonsten können die Jahreszahlen mit als Suchbegriffe eingehen.

#### **B4: Text- Ergebnis - > Fakten-Ergebnis**

Hierzu gehört die bei B3 schon erwähnte intellektuelle Verknüpfung von Zeitreihen und Texten. Zusätzlich gibt es für diese Richtung weitere Möglichkeiten zur festen Verknüpfung. Texte enthalten häufig größere Fakteneinheiten in Form von Tabellen oder Grafiken. Die Verbände könnten solchen Texten intellektuell Links zu den Faktendaten hinzufügen. Sind solche Links nicht vorhanden, kann das System versuchen, anhand von formalen und inhaltlichen Merkmalen der Texte oder Textpassagen auf die enthaltenen Daten zu schließen.

**Beispiel:** In der verbandseigenen Publikation "Bauindustrie aktuell" ist eine Tabelle zu Baugenehmigungen von 1990 bis 1997 enthalten. Der Benutzer möchte diese Reihe weiter zurückverfolgen und sie in eine eigene Grafik einbauen.

Der Verband hat zu dieser Veröffentlichung Link auf die der Zeitreihen definiert, die in ihm interpretiert und teilweise dargestellt ist. Der Benutzer kommt über diesen Link an die Fakten und kann sie im vollen Umfang einsehen und für die Eigennutzung exportieren.

Durch die Integration von Text- und Faktenretrieval ergibt sich für den Fall A3 (iteratives Faktenretrieval) eine interessante Variante. Bisher ist hier von Testpersonen die Ausweitung auf Zeitreihen mit inhaltlichem Bezug eher abgelehnt

worden. Durch die automatische Indexierung der Texte steht mit der Term-Dokument-Matrix ein automatisch erstellter Thesaurus für den Anwendungsfall zur Verfügung. Damit können zu Zeitreihen neue Zeitreihen gefunden werden, deren Deskriptoren häufig gemeinsam in Texten vorkommen.

### 3.3 "Echte" Integration

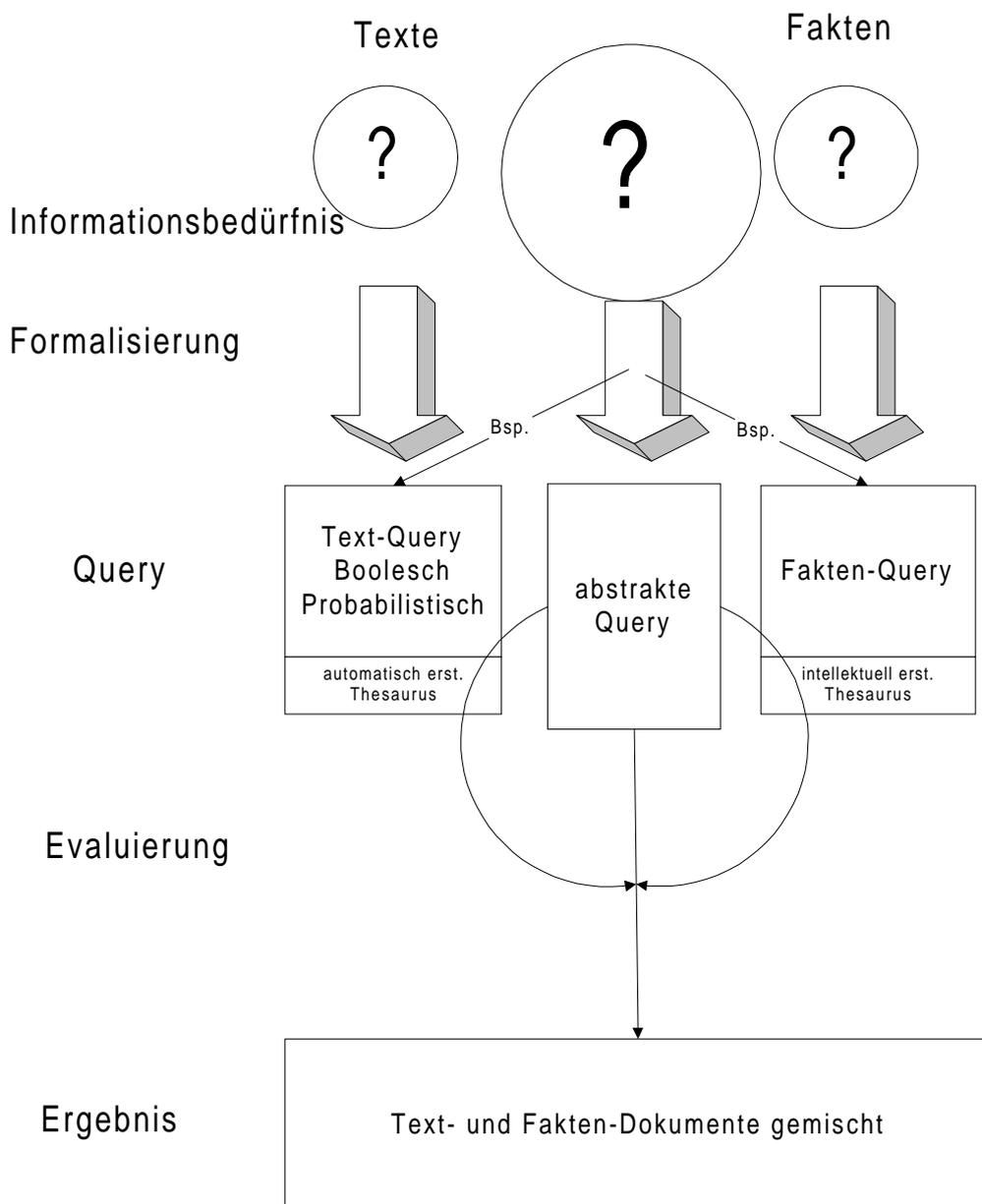
Echte Integration liegt dann vor, wenn der Benutzer sein Informationsbedürfnis völlig abstrakt formuliert und es nicht auf Texte, Fakten oder eine andere Modalität einschränkt.

#### **C: Infobedürfnis - > ... -> Text-Ergebnis + Fakten-Ergebnis**

Die ersten empirischen Untersuchungen haben gezeigt, daß solche Bedürfnisse im Anwendungsfall von ELVIRA vorkommen. Ob Benutzer ihre Suchen auch so konzeptualisieren, ist dagegen noch nicht klar. Es ist möglich, daß bei der Suchformulierung bereits in Beispielen gedacht wird und damit das abstrakte Problem auf der Ebene der Suchformulierung keine Entsprechung mehr hat.

**Beispiel:** Standardpakete. Ein Benutzer will "alles" zu einem Land wissen. Er erhält in diesem Fall von den Informationsvermittlern der Verbände ein Paket mit verschiedensten Informationen. Dabei schränken die Informationsvermittler die Anfrage mit oder ohne Rücksprache meist auf das Produktspektrum der Firma ein.

Ob diese Art der Integration für ELVIRA II sinnvoll ist, können erst weitergehende empirische Untersuchungen zeigen. In der ersten Phase müssen entsprechende Tests durchgeführt werden.



**Abb. 3.3: "Echte" Integration (Stufe 3)**

### 3.4 Ansätze Schalenmodell

Wie bei der Darstellung der Datengrundlage deutlich wurde, gibt es zwischen einzelnen Datenbeständen Qualitätsunterschiede bezüglich der Inhalte und der Inhaltserschließung.

Vom Verband aufbereitete Daten besitzen inhaltlich die höchste Qualität. Die Verbandszeitreihen sind intellektuell indexiert und im ELVIRA-Faktenzugang eindeutig recherchierbar. Dadurch sind sie von den Zusatzfakten klar abgegrenzt. Die von den Verbänden erstellten Texte haben vom Inhalt her eine höhere Qualität als etwa die BfAI-Nachrichten. Jedoch steht für die Verbandstexte kein intellektuell gepflegter Thesaurus wie bei BfAI zur Verfügung, der die Erschließungsqualität verbessern kann.

Diese Beispiele zeigen, daß im Falle von ELVIRA kein homogener, einheitlich erschlossener Datenbestand vorliegt, wie er vielen Informationssystemen zugrunde liegt. Diese Problematik wird im Schalenmodell (cf. Krause 1996) thematisiert. Das Schalenmodell sieht vor, daß in einem Datenpool verschiedene Schalen von Dokumenten existieren, die mit verschiedener Qualität erschlossen wurden. Dies ist z.B. sinnvoll, wenn aus Kostengründen nur ein Kernbereich von Dokumenten intellektuell erschlossen wird und eine weitere Schale automatisch indexiert wird.

## **4 Vorgehensweise**

Bei den weiteren Arbeiten im Rahmen der Text-Fakten-Integration wird weiterhin ein starker Praxisbezug angestrebt. Dies sollte früh zu evaluierbaren Prototypen führen.

### **4.1 Prototypische Realisierung**

Es erscheint sinnvoll, nach der Realisierung des Textzugangs zunächst einfache Transformationen zu realisieren und diese in die Benutzungsoberfläche zu integrieren. Hierzu zählen die fest verknüpften Varianten von B1 bis B4, also Umstiege von Fakten zu Texten und umgekehrt sowohl aus Anfrage als auch iterativ aus Ergebnissen. In diesem Rahmen sollte weiterhin an B1 und B2 für den nicht fest verknüpften Fall gearbeitet werden. Daneben gehören auch die Zusatzfakten in den Bereich der einfachen Transformationen. Bei den Tests dieser ersten Prototypen ist mit weiteren Anforderungen zu rechnen.

Die Untersuchung und Realisierung komplexerer Transformationen sollten einer zweiten Phase vorbehalten sein. Darin können B1 und B2 aufgrund der ersten empirischen Ergebnisse ausgebaut werden und die Verfahren auf die Fälle B3 und B4 übertragen werden, also auf die Umstiege ausgehend von Ergebnissen. Weiterhin soll versucht werden Teile von Dokumenten als Ausgangspunkte von Transformationen B3 und B4 zuzulassen.

Im gesamten Zeitraum müssen weiterhin empirische Hinweise auf Text-Fakten-Integration nach Stufe C gesucht werden.

### **4.2 Evaluierung**

Wie der vorhergehende Abschnitt bereits zeigt, muß die Entwicklung eines innovativen Systems durch eingehende empirische Untersuchungen begleitet werden. Im Kontext der Text-Fakten-Integration stellt sich wie bei allen Information Retrieval Systemen besonders die Frage nach der Evaluierung der Recherche-Ergebnisse. Wie langwierig eine eingehende statistische IR-Evaluierung

ist, zeigt in den letzten Jahren besonders die TREC-Initiative (für einen Überblick cf. Womser-Hacker 1996).

Dieser Aufwand ist im Rahmen von ELVIRA nicht möglich. Zudem müßte für eine innovative Technik wie die Text-Fakten-Integration erst geprüft werden, ob das gebräuchliche Evaluierungsinstrumentarium geeignet ist. Zudem vernachlässigt die statistische Auswertung oft qualitative Eigenheiten der informationellen Prozesse.

Die Evaluierung der Recherche-Ergebnisse aus der Text-Fakten-Integration muß daher im wesentlichen wie die bisherigen empirischen Untersuchungen durchgeführt werden. Dazu werden mit Experten an wenigen Beispielen Recherche-Ergebnisse intensiv besprochen. Durch diese Art der kognitiven Benutzertests lassen sich bessere qualitative Ergebnisse als durch umfangreiche statistische Untersuchungen.

## 5 Fazit

Im Rahmen der Text-Fakten-Integration werden sich zahlreiche neue Herausforderungen ergeben. Dieser Arbeitsbericht kann nur ein erster Einstieg in die Thematik sein. Eine besondere Rolle von werden Vagheit und dynamische Anpassung spielen. Da Transformationen häufig nicht auf deduktiv ableitbarem Wissen aufbauen, muß vage und annäherungsweise geschlossen werden. Dazu sind eingehende Kenntnisse über die vagen Prozesse im IR nötig. Als besonders erfolgversprechende Modellierungstechniken bieten sich vage Methoden aus der Informatik an. Einen ersten Einstieg in diese Problematik bietet Mandl (1997). Um die Transformationen sinnvoll in die Benutzungsoberfläche zu integrieren, müssen neue Konzepte der dynamischen Anpassung untersucht werden. Hier ist es entscheidend, für den Benutzer nachvollziehbare Anpassungen zwischen den einzelnen Zugängen so zu definieren und das Systemverhalten jederzeit vorhersehbar ist. Zusätzlich gilt, es den Faktenzugang so zu flexibilisieren, daß möglichst viele der neuen Fakten-Datenbestände integrierbar sind.

## 6 Literaturverzeichnis

Ingwersen, Peter (1992): Information Retrieval Interaction. London 1992.

Krause, Jürgen (1996): Informationserschließung und -bereitstellung zwischen Deregulation, Kommerzialisierung und weltweiter Vernetzung - Schalenmodell. IZ-Arbeitsbericht Nr. 6, IZ Sozialwissenschaften, Bonn.

Krause, Jürgen; Mandl, Thomas; Stempfhuber, Maximilian (1996): Design des ersten Prototypen des ZVEI-Verbandsinformationssystems IZ-Arbeitsbericht Nr. 4, IZ Sozialwissenschaften, Bonn.

Mandl, Thomas (1997): Vages Information Retrieval und die Verarbeitung unsicheren Wissens. ELVIRA-Arbeitsbericht Nr. 7, IZ Sozialwissenschaften, Bonn.

Mandl, Thomas; Stempfhuber, Maximilian (1997): Ergebnisse aus der Praxisphase von ELVIRA. ELVIRA-Arbeitsbericht Nr. 14, IZ Sozialwissenschaften, Bonn.

Marx, Jutta; Mutschke, Peter; Schommler, Marcus (1995): Möglichkeiten der intelligenten Integration heterogener Datenbestände: das Projekt GESINE. IZ-Arbeitsbericht Nr. 2, IZ Sozialwissenschaften, Bonn.

Schaefer, André (1997): Entwicklung eines prototypischen Informationssystems für das Projekt KO-NEXIS. Diplomarbeit. Universität Koblenz-Landau, Fachbereich Informatik.

Womser-Hacker, Christa (1996): Das MIMOR-Modell. Mehrfachindexierung zur dynamischen Methoden-Objekt-Relationierung im Information Retrieval. Habilitationsschrift. Universität Regensburg, Informationswissenschaft.

## Liste der Arbeitsberichte im Projekt ELVIRA

### ELVIRA I

- |   |                                     |  |
|---|-------------------------------------|--|
| 1 | Mandl                               | Anforderungen und Benutzereigenschaften für das ZVEI-Verbandsinformationssystem (Juni 1995)  |
| 3 | Krause/Stempfhuber/ Mandl           | Design des ersten Prototypen des ZVEI-Verbandsinformationssystems ELVIRA (April 1996)<br>2. Vers.: Dez. 96; entspricht IZ-Arbeitsbericht 4 |
| 4 | Stempfhuber                         | Realisierung der Benutzungsoberfläche von ELVIRA mit PowerBuilder<br>ersetzt durch ELVIRA-Arbeitsbericht 8                                 |
| 5 | Krause                              | Das WOB-Modell. Zur Gestaltung objektorientierter, grafischer Benutzungsoberflächen (Dezember 1995)<br>entspricht IZ-Arbeitsbericht 1      |
| 6 | Mandl/Stempfhuber                   | Die Benutzertests zur ersten Version von ELVIRA (Dezember 1996)  |
| 7 | Mandl                               | Vages Information Retrieval und die Verarbeitung unsicheren Wissens (Januar 1997)<br>gekürzt als Kapitel im Endbericht                     |
| 8 | Stempfhuber                         | Generalisierung des 3-Facettenzugangs (1997)   |
| 9 | Stempfhuber/Mandl/<br>Gräßer/Hermes | Beschreibung der Benutzungsoberfläche von ELVIRA (Dezember 1996) 2. Vers.: Juli 1997; entspricht Handbuch; auch Anhang Endbericht          |

### ELVIRA II

- |    |                                |  |
|----|--------------------------------|--|
| 11 | Krause/Mandl/<br>Stempfhuber   | Das Verbandsinformationssystem ELVIRA II. Projektskizze (April 1997)<br>aus Projektantrag; entspricht IZ Arbeitsbericht 11 |
| 12 | Krause/Mandl/<br>Stempfhuber   | Text-Fakten-Integration in ELVIRA (Dez. 1997)<br>entspricht IZ Arbeitsbericht 12   |
| 14 | Mandl/Stempfhuber              | Ergebnisse aus der Praxisphase von ELVIRA (Dezember 1997)  |
| 15 | Mandl/Schaefer/<br>Stempfhuber | Exemplarische Transformationen für die Text-Fakten-Integration (Februar 1998)  |
| 16 | Krause/Schaefer                | Benutzungsoberfläche für Textretrieval [Arbeitstitel] (1998) geplant   |
| 17 | Krause/Mutschke                | Indexierung und Fulcrum Evaluierung [Arbeitstitel] (1998) geplant  |
| 18 | Stempfhuber                    | ELVIRA II im Internet [Arbeitstitel] (1998)<br>geplant   |

In der Reihe **IZ-Arbeitsberichte** sind bisher erschienen:

- 1 Krause, J.: Das WOB-Modell. Zur Gestaltung objektorientierter, graphischer Benutzungsoberflächen (Dezember 1995)
- 2 Marx, J.; Mutschke, P.; Schommler, M.: Möglichkeiten der intelligenten Integration heterogener Datenbestände: das Projekt GESINE (Dezember 1995)
- 3 Krause, J.: Visualisierung und graphische Benutzungsoberflächen (Mai 1996)
- 4 Krause, J.; Mandl, T.; Stempfhuber, M.: Design des ersten Prototypen des ZVEI-Verbandsinformationssystems ELVIRA (April 1996, vergriffen)
- 5 Mutschke, P.: Relationale Datenbanksysteme im Vergleich: Eine Zwischenbilanz (Dezember 1995)
- 6 Krause, J.: Informationserschließung und -bereitstellung zwischen Deregulation, Kommerzialisierung und weltweiter Vernetzung - Schalenmodell - (September 1996)
- 7 Weingärtner, M.: Computergestützte Gruppenarbeit: State of the Art (September 1996)
- 8 Binder, G., Herfurth, M., Krause, J., Marx, J., Stahl, M.: Nutzerforschung, Gestaltung informationeller Abläufe und Softwareentwicklung am Informationszentrum Sozialwissenschaften (Dezember 1996)
- 9 Riege, U.; Schomisch, S.; Schommler, M.: COGET - Computergestützte Erstellung von Themendokumentationen (Dezember 1996)
- 10 Frisch, E.; Kluck, M.: Pretest zum Projekt German Indexing and Retrieval Test-database (GIRT) unter Anwendung der Retrievalsysteme Messenger und freeWAISsf (Juli 1997/Nachauflage Oktober 1997)
- 11 Krause, J.; Mandl, T.; Stempfhuber, M.: Das Verbandsinformationssystem ELVIRA II. Projektskizze (April 1997)
- 12 Krause, J.; Mandl, T.; Stempfhuber, M.: Text-Fakten-Integration in ELVIRA (Dezember 1997)