



Text-Fakten-Integration in Informationssystemen

Jürgen Krause, Thomas Mandl, André Schaefer,
Maximilian Stempfhuber

IZ Sozialwissenschaften, Lennéstr. 30, D-53113 Bonn, st@bonn.iz-soz.de
Universität Koblenz-Landau, Institut für Informatik

Inhalt

- 1 Text-Fakten-Integration
- 2 Das Verbandsinformationssystem ELVIRA
- 3 Konzept für die Text-Fakten-Integration in ELVIRA
 - 3.1 Text- und Faktensuche in Reinform
 - 3.2 Transformationen zwischen den Suchtypen
 - 3.3 Modalitätsunabhängige Suche
 - 3.4 Zusammenfassung und weiteres Vorgehen
- 4 Textrecherche mit integrierter Faktensuche
- Literatur

Zusammenfassung

Der Beitrag stellt den Prototypen eines integrierten Informationssystems vor, das Daten verschiedener Typen (Modalitäten) enthält. Das zugrundeliegende Modell entstand aus empirischen Untersuchungen für Texte und Fakten, eignet sich aufgrund der gewählten Abstraktionsstufe jedoch auch für andere Datentypen wie z.B. Grafiken oder Internet-Seiten. Als Anwendungsfall dient das Elektronische Verbandsinformations-, Recherche- und Analysesystem ELVIRA.

1 Text-Fakten-Integration

Heterogenitätsbehandlung im Information Retrieval ist eine der großen Herausforderungen für zukünftige Informationssysteme. Ein Spezialfall hiervon ist die integrierte Suche nach Texten und Fakten, die bei zahlreichen Dokumenttypen sinnvoll sein kann (z.B. bibliographische Daten, Faktendatenbanken, Datenbanken für Forschungsprojekte, Dienstleistungsangebote, Rechner, Förderprogramme). Bei der Text-Faktenintegration können drei Typen von Suchen vorkommen:



a) Kombinierte Suchen

Bei kombinierten Suchen werden Bedingungen für Texte und Fakten und ihre Zusammenhänge in einer Query formuliert, wodurch die Beziehungen zwischen den Dokumenten besser ausgenutzt werden. Ein Beispiel ist die Faktensuche bei der man sich primär auf die Dokumentattribute (i.d.R. bibliographische Angaben) bezieht. Dies ist zum einen zur Realisierung bestimmter Suchstrategien im Rahmen einer Dokumentrecherche notwendig, etwa wenn man nach Autoren, Zeitschriften oder Konferenzen zu einem bestimmten Thema sucht, um mit diesem Ergebnis die Suche voranzutreiben. Zum anderen sind die Fakten selbst häufig das Ziel einer Suche, wenn man z.B. nach Autoren als Experten auf einem bestimmten Gebiet sucht oder nach Institutionen, die als Kooperationspartner in Frage kommen.

b) Umstiege durch Transformationen

In vielen Situationen wollen Nutzer z.B. nach einer Textanfrage mehr zu einem Thema wissen und möchten nun auch Fakten und z.B. Internet-Seiten dazu sehen. Da die Inhaltserschließung bei diesen verschiedenen Modalitäten unterschiedlich ist und oft anderes Vokabular benutzt wird, müssen bereits formalisierte Informationsbedürfnisse in Anfragen der anderen Modalitäten transformiert werden. Transformationen sollten in allen Richtungen (von Fakten zu Texten, von Texten zu Fakten, usw.) und sowohl aus initialer Suchanfrage als auch aus dem Ergebnis heraus möglich sein.

c) Darstellungsform-unabhängige Anfragen

Informationsbedürfnisse können auch von vorne herein unabhängig von der Darstellungsform des gespeicherten Wissens sein, also sowohl durch Textdokumente als auch durch Faktendaten beantwortet werden. Z.B. kann eine Information über eine Chemikalie in einer Tabelle oder einem beschreibenden Text gefunden werden. Für solche Informationsbedürfnisse muß ein System die Formulierung einer generellen Anfrage zulassen, diese in alle modalitätsabhängigen Anfragen transformieren und die Ergebnisse kombinieren.

Obwohl das Bewußtsein, daß bei einer realen Entscheidungs- oder Problemlösungssituation verschiedene Dokumenttypen, vor allem auch unstrukturierte Dokumentbestände (Texte) neben strukturierten wichtig sind, weit verbreitet ist; beschränkt sich ihre Integration in der Regel auf die Fragen der technischen Verwaltung in verteilten Datenbanken und die technisch effiziente Verarbeitung von Anfragen (z.B. [Gu et al. 1993]). Darüberhinaus wird die Integration von Texten und Fakten oft als erster Schritt einer multimedialen Verknüpfung verstanden (cf. [Womser-Hacker 1997]).

Praxisbeispiele einer Integration von Text- und Fakten-Material sind die Anwendungen einer schweizerischen Großbank (cf. [Amstutz et. al. 1991]) und ein Kriminalinformationssystem der niederländischen Polizei (cf. [Hoogeveen et al. 1992]). Ein Konzept für kombinierte Suchen, bei denen sowohl Bedingungen für

Fakten als auch für Texte formuliert werden, bietet [Fuhr 1992]. Auf diesen Typ beschränken sich auch viele Anwendungen, wie etwa die bei der GMD-IPSI entwickelte experimentelle Forschungsdatenbank HYDRA (cf. [Gu et al. 1993]), die Fakten- und Textwissen über EU-Projekte beinhaltet. Die beschreibenden Textteile sind verbunden mit strukturierten Fakteninformationen wie dem Projektleiter, Förderstelle oder Projektlaufzeit.

Erste konzeptuelle Grundlagen für eine Integration von Zeitreihen und Texten enthält für die Sozialwissenschaften Marx et al. 1995 und für die wirtschaftlichen Zeitreihen des Verbandsinformationssystems ELVIRA Krause et al. 1997. Zu diesen Vorüberlegungen gehört auch das MIMOR-Modell von Womser-Hacker 1997, das die Text-Fakten-Integration in einem System der Mehrfachindexierung zur dynamischen Methoden-Objekt-Relationierung im Information Retrieval berücksichtigen möchte. Methoden für Transformationen diskutiert [Mandl 1998].

Das Konzept für die Text-Fakten-Integration im Projekt ELVIRA II (vom Bundesministerium für Wirtschaft gefördert unter Aktenzeichen IV C2-003060/22), das im folgenden Überblick dargestellt wird, basiert auf empirischen Untersuchungen mit Nutzern von ELVIRA I (cf. [Krause et al. 1997]). Hierbei wurde zusammen mit Verbandsmitarbeitern und Marktforschern in Unternehmungen evaluiert, ob eine Integration von Texten in das für die Recherche von Zeitreihen zu statistischen Erhebungen optimierte System ELVIRA I sinnvoll ist. Interviews und die Analyse von Anfragen von Firmen an die Verbände ergaben, daß textuelle Daten nicht nur explizit nachgefragt werden, sondern oft auch durch die Informationsvermittler nachgewiesen werden, wenn zu einem Thema keine Zeitreihen vorliegen oder die Texte zusätzliche Informationen enthalten. Aufbauend auf dieser Beobachtung wurden einige konkrete Beispiele im Detail analysiert und dienen als Grundlage für die allgemeine Modellbildung (cf. [Krause et al. 1997] und [Mandl et al. 1998]).

2 Das Verbandsinformationssystem ELVIRA

ELVIRA ist ein Client/Server-System zur Recherche und Weiterverarbeitung von Zeitreihen, die überwiegend vom Statistischen Bundesamt, den Industrieverbänden und Wirtschaftsforschungsinstituten erhoben und veröffentlicht werden. Das ursprünglich für den Zentralverband der Elektrotechnik- und Elektronikindustrie (ZVEI, Frankfurt) entwickelte System bieten inzwischen auch der Hauptverband der Deutschen Bauindustrie (HVB, Berlin) und der Verband Deutscher Maschinen- und Anlagenbau (VDMA, Frankfurt) ihren Mitgliedsunternehmen als Online-System an. Insgesamt ist ELVIRA I an etwa 200 Arbeitsplätzen in Unternehmen und Verbänden installiert und wird von Marktforschern bzw. Informationsvermittlern benutzt (cf. [Scheinost et al. 1998]).

Der Schwerpunkt bei der Entwicklung von ELVIRA lag auf der software-ergonomischen Gestaltung der Benutzungsoberfläche. Der Benutzer soll möglichst schnell und ohne die interne Struktur der Datenbestände zu kennen die für ihn relevanten Daten aus der großen Gesamtmenge (teilweise weit über 5 Mio. Zeitreihen pro Verband) recherchieren. Zu diesem Zweck liegen die Zeitreihen in

deskribierter Form vor, wobei jede Zeitreihe durch ein Deskriptorentripel beschrieben wird. Die Deskriptoren jedes Datenbestands (z.B. Produktion, Auftragseingang usw.) lassen sich in drei Facetten gliedern (Thema, Branche und Land) und hierarchisch ordnen. Das Suchobjekt für Zeitreihen (Abb. 1) stellt diese Deskriptorlisten dar, wobei die Liste mit der Facette *Land* aus Platzgründen verdeckt ist.

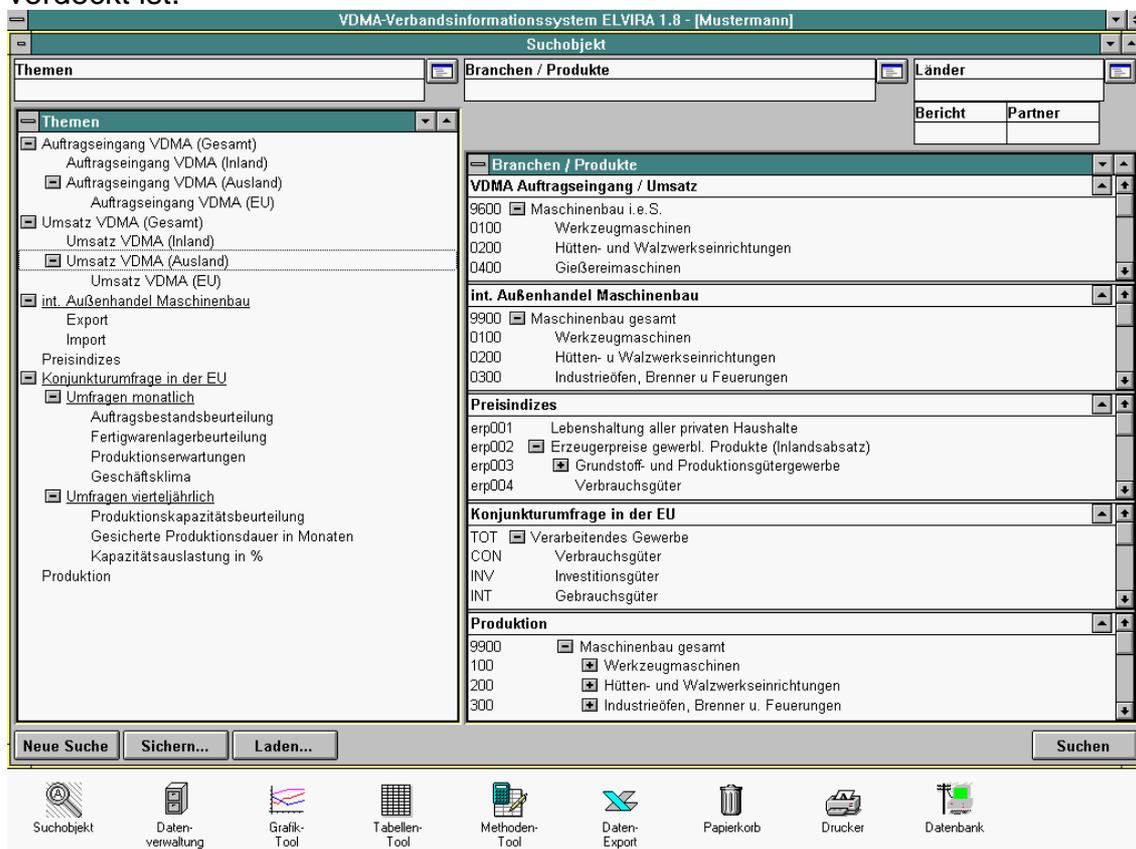


Abb. 1: ELVIRA Recherchebildschirm

Die Vorlageleistung des Systems wurde dahingehend optimiert, daß dem Benutzer zu jedem Zeitpunkt der Anfrageformulierung die Auswahl aus Deskriptorenlisten ermöglicht wird, da die Deskriptoren aufgrund ihrer Komplexität i.d.R. nicht fehlerfrei eingegeben werden können (z.B. „Gleichstrommotoren mit einer Leistung von 0,075 KW bis 0,75 KW“). Zusätzlich erlaubt das System die Selektion innerhalb der drei Facetten in beliebiger Reihenfolge, so daß der Anwender Fragestellungen wie „Die Produktion in 1000 DM von Produkt X in Deutschland“ oder „Zu Produkt X die Produktion in 1000 DM in Deutschland“ direkt formulieren kann und nicht an eine hierarchische Abfolge gebunden ist.

Um zu gewährleisten, daß der Benutzer in allen drei Facetten die Deskriptoren jeweils eines Bestandes kombiniert, paßt sich die Oberfläche dynamisch an und zeigt bzw. verdeckt die einzelnen Deskriptorenlisten (Abb. 2). Dieser Vorgang kann in jeder Facette ausgelöst werden und erlaubt die gleichzeitige Recherche in beliebig vielen Datenbeständen. Die selektierten Begriffe werden in einer Zustandsanzeige im oberen Bereich des Fensters angezeigt, wobei sich die Höhe der Zustandsanzeige und der Vorlagelisten gegenseitig beeinflusst.

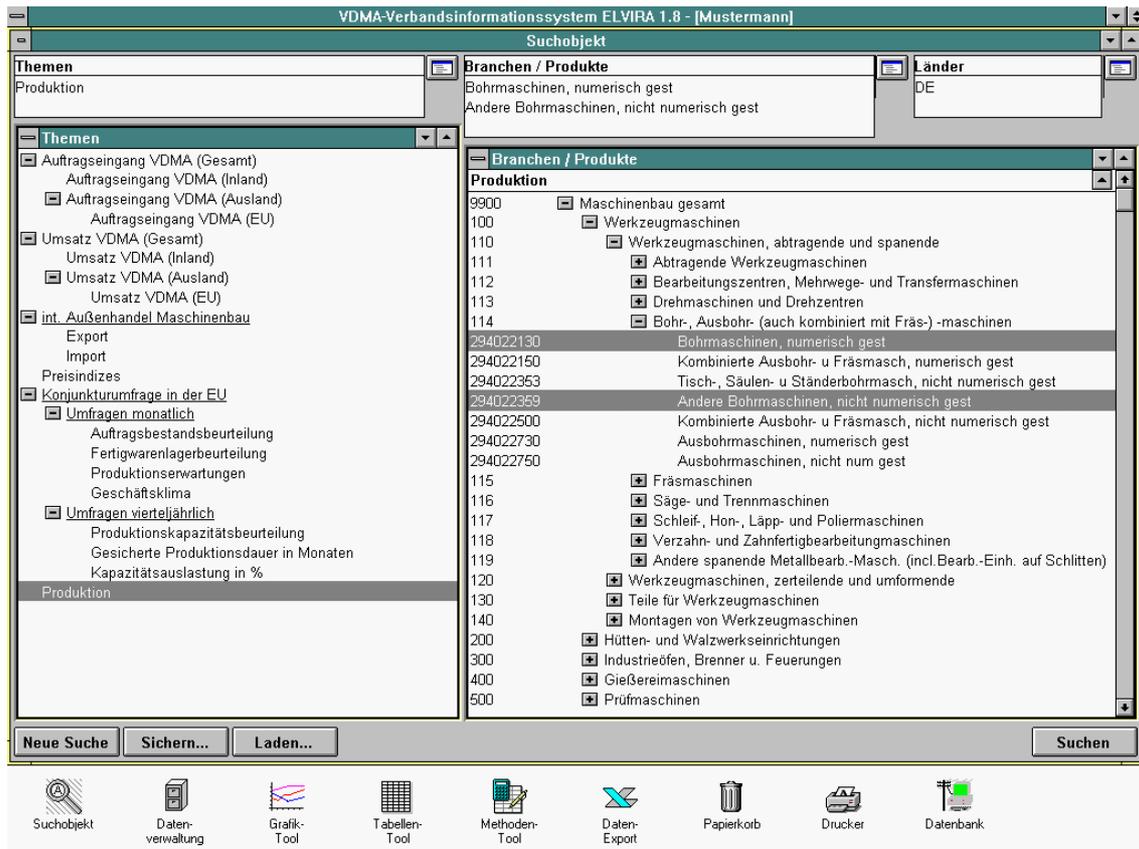


Abb. 2: Recherreformulierung bei der Zeitreihensuche

Ziel von ELVIRA II ist die Integration der Textrecherche in ELVIRA und die Gestaltung der Text-Fakten-Integration. Durch thesaurusbasierte und statistische Verfahren wird auf der Benutzungsoberfläche zum einen die Formulierung von Anfragen über heterogene Datentypen (z.B. Zeitreihen, Texte und Messetermine) ermöglicht und zum anderen die automatische Transformation zwischen unterschiedlichen Indexierungsvokabularien unterstützt.

3 Konzept für die Text-Fakten-Integration in ELVIRA

Die Modellbildung zur Text-Fakten-Integration in ELVIRA II stützt sich auf umfangreiche Interviews mit Nutzern von ELVIRA I und Mitarbeitern der Industrieverbände. Zusätzlich wurden über 100 Anfragen von Firmen an die Verbände ausgewertet, um sowohl konkrete Informationsbedürfnisse als auch Problemlösestrategien bei deren Beantwortung zu ermitteln. Die Analyse relevanter Text- und Datenbestände vervollständigte die Datengrundlage.

Als Ergebnis konnten drei Stufen der Text-Fakten-Integration formuliert werden (cf. [Krause et al. 1997]), die sowohl Benutzern mit spezifischen Vorerfahrungen aus ELVIRA I oder der Nutzung anderen Hosts gerecht werden als auch neue Konzepte der Informationssuche integrieren. Der im allgemeinen Modell postulierten kombinierten Suche entsprechen bei ELVIRA keine Informationsbedürfnisse. Die erste Stufe bilden die Reinformen der Recherche.

A. Text- und Faktensuche in Reinform

Auch in integrierten Informationssystemen treten immer Informationsbedürfnisse auf, die auf eine Modalität beschränkt sind, bei denen der Anwender als Ergebnis also nur Texte *oder* Fakten erhalten will. Dazu müssen Zugänge bereitstehen, die für diese Aufgabe optimiert sind.

B. Transformationen zwischen den Suchtypen

Transformationen unterstützen den Benutzer dabei, eine für eine Modalität formulierte Anfrage (z.B. nach Zeitreihen) in eine andere Modalität (z.B. nach erklärenden Texten) umzusetzen. Das System transformiert zwischen formalen Unterschieden in den Anfragemöglichkeiten und zwischen dem unterschiedlichen Indexierungsvokabular. Die Transformationen laufen hierbei teilweise transparent im Hintergrund ab oder geben dem Benutzer zusätzliche Möglichkeiten, Parameter der gewünschten zusätzlichen Modalitäten explizit zu setzen. Die spezialisierten Zugänge unterstützen die Anfrageformulierung durch eine dynamisch adaptierte Oberfläche, die z.B. im Zeitreihenzugang boole'sche Operatoren für die Suche nach Texten zulässt, sowie kontextsensitiv angepasste Eingabehilfen, wie z.B. Thesauri und Cross-Konkordanzen.

C. Modalitätsunabhängige Suche

Die "echte" Integration geht nur vom Informationsbedürfnis des Benutzers aus, ohne bei der Anfragformulierung auf eine bestimmte Modalität des Ergebnisses abzielen und durch die Oberflächengestaltung bereits einen Prototypen der Ergebnismenge im mentalen Modell des Benutzers zu implizieren.

Die Abb. 3 zeigt das Modell der Text-Fakten-Integration in ELVIRA II im Überblick und stellt neben den Reinformen (Stufe A) die möglichen Transformationen (Stufe B) schematisch dar. Die Transformationen auf Stufe B erfolgen hauptsächlich durch die automatische Übersetzung von Anfragen, wie in den folgenden Abschnitten erläutert. Teilweise werden sie aber auch durch feste, intellektuell erstellte Verknüpfungen zwischen inhaltlich zusammengehörenden Text- und Fakten-Dokumenten realisiert, die Experten z.B. bei der Erstellung eines Textes zu den darin verwendeten Fakten definieren.

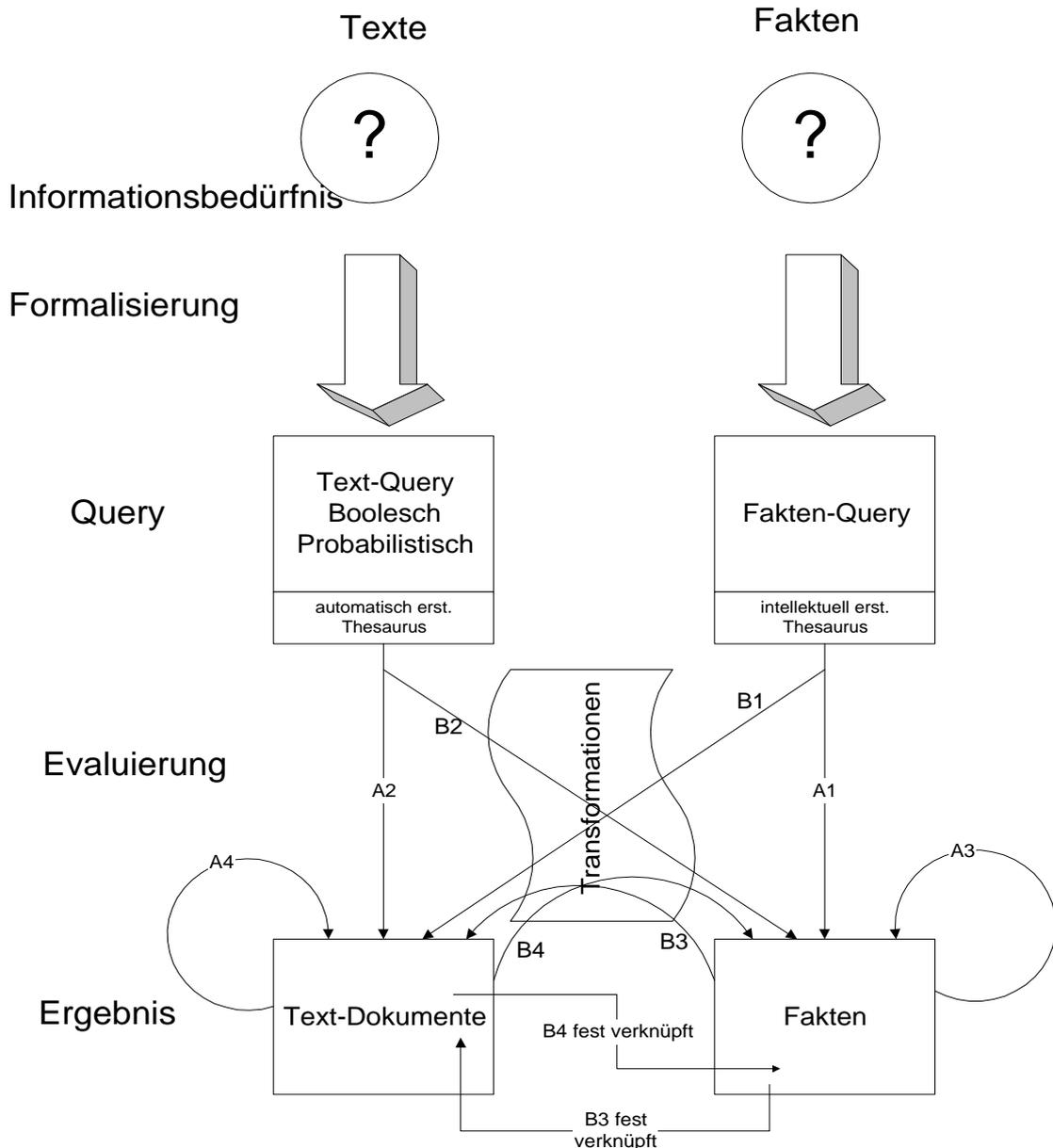


Abb. 3: Reinformen und Transformationen

3.1 Text- und Faktensuche in Reinform

Die Prozesse bei den Reinformen der Recherche lassen sich formal wie folgt beschreiben (vgl. Abb. 3):

A1: Fakten-Anfrage führt zu Fakten-Ergebnis

A2: Text-Anfrage führt zu Text-Ergebnis

Die bisherige ELVIRA-Faktenanfrage entspricht dem Typ A1 und die in Entwicklung befindliche Textretrieval-Komponente dem Typ A2.

In der Praxis laufen informationelle Prozesse iterativ ab, d.h. ein Anwender bewertet meist das aktuelle Ergebnis und formuliert aufgrund seiner Wertung die ursprüngliche Anfrage um. Erst nach evtl. mehreren Zyklen erhält er ein befriedigendes Ergebnis, das im Falle von ELVIRA nicht bereits dem gewünschten

Endergebnis entsprechen muß (oft schließen sich hier noch zusätzliche Verarbeitungsschritte an).

Eine spezifische Art des iterativen Retrievals ist das im Textretrieval verbreitete Relevance Feedback. Dabei markiert der Anwender in den Ergebnismengen die für ihn relevanten Dokumente. Das System nutzt diese Informationen zur Formulierung der neuen Anfrage. Der Benutzer gelangt mit dem aktuellen Ergebnis - oder Teilen davon - direkt zu einem neuen Ergebnis. Diese Technik ist auch im Faktenretrieval von ELVIRA sinnvoll, da sich durch die in Kapitel 2 erläuterte Verschlagwortung von Zeitreihen die Verfahren aus dem Textretrieval direkt übernehmen lassen. Somit ergeben sich zwei weitere Suchtypen (vgl. Abb. 3):

A3: Fakten-Ergebnis führt zu Fakten-Ergebnis

A4: Text-Ergebnis führt zu Text-Ergebnis

3.2 Transformationen zwischen den Suchtypen

Automatische Transformationen können Anfragen einer Modalität in andere Modalitäten übersetzen und den Benutzer entlasten. Mit einer Textanfrage ist so z.B. die gleichzeitige oder iterative Suche nach Texten und Zeitreihen möglich. Die Aufgabe der Transformationsfunktionen ist es hierbei, die Eigenheiten der Datenbestände möglichst gut aufeinander abzubilden. Strukturelle und semantische Unterschiede verhindern oft eine direkte Überführung, wobei sich die Probleme i.d.R. erst nach eingehender Analyse darstellen und zumindest für den gelegentlichen Benutzer nicht sofort erkennbar sind. Neben strukturellen Unterschieden im Aufbau verhindern vor allem unterschiedliche Indexierungsvokabulare den direkten Transfer. So weicht z.B. die Wortwahl in amtlichen Nomenklaturen zu Zeitreihendaten vom Sprachgebrauch in der Praxis ab. Dies hat zur Folge, daß z.B. mit dem Suchbegriff „Waschmaschine“ zwar Texte gefunden werden können, Faktendaten aber aufgrund der Verwendung des Begriffs „Waschvollautomat“ nicht nachgewiesen werden. Beim Transfer von Fakten zu Texten ergeben sich ähnlich gelagerte Probleme, wenn Attributausprägungen von Fakten(-beschreibungen) bei Textdaten im Volltext und zusätzlich als Schlagwort erscheinen können. Auch hier ist nach eingehender Analyse eine dedizierte Transformationsfunktion zu spezifizieren, welche die primäre Suchintention des Benutzers optimal widerspiegelt.

Unter Einbeziehung der iterativen Fälle ergeben sich aus Nutzersicht somit folgende Suchprozesse (vgl. Abb. 3):

B1: Fakten-Anfrage führt zu (Fakten- und) Text-Ergebnis

B2: Text-Anfrage führt zu (Text- und) Fakten-Ergebnis

B3: Fakten-Ergebnis führt zu Text-Ergebnis

B4: Text- Ergebnis führt zu Fakten-Ergebnis

3.3 Modalitätsunabhängige Suche

Auf der Stufe C der Integration fokussiert der Anwender nur noch auf sein Informationsbedürfnis und nicht auf die Modalität der Ergebnisdokumente (vgl. Abb. 4):

C: Informationsbedürfnis führt zu Text-Ergebnis + Fakten-Ergebnis

Dieses Suchbedürfnis kann der Anwender wohl nicht in einem für eine bestimmte Modalität optimierten Zugang formulieren. Vielmehr müßte ein Werkzeug für eine abstrakte Suchanfrage entwickelt werden, das nach Möglichkeit keinen spezifischen Ergebnistyp bereits bei der Anfrageformulierung impliziert.

Erste empirischen Untersuchungen haben gezeigt, daß solche Bedürfnisse im Anwendungsfall von ELVIRA vorkommen. Ob Benutzer die Suche auch so konzeptualisieren, ist dagegen noch nicht klar. Es ist möglich, daß der Suchformulierung bereits ein Prototyp des Ergebnisses zugrunde liegt und damit das abstrakte Problem auf der Ebene der Suchformulierung schon keine Entsprechung mehr hat. Hierzu ein Beispiel:

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 des anfragenden Unternehmens ein.

Ob diese Art der Integration für ELVIRA II sinnvoll ist, können erst weitergehende empirische Untersuchungen zeigen. In einer ersten Phase sollen hierzu entsprechende Tests bei ELVIRA-Nutzern durchgeführt werden.

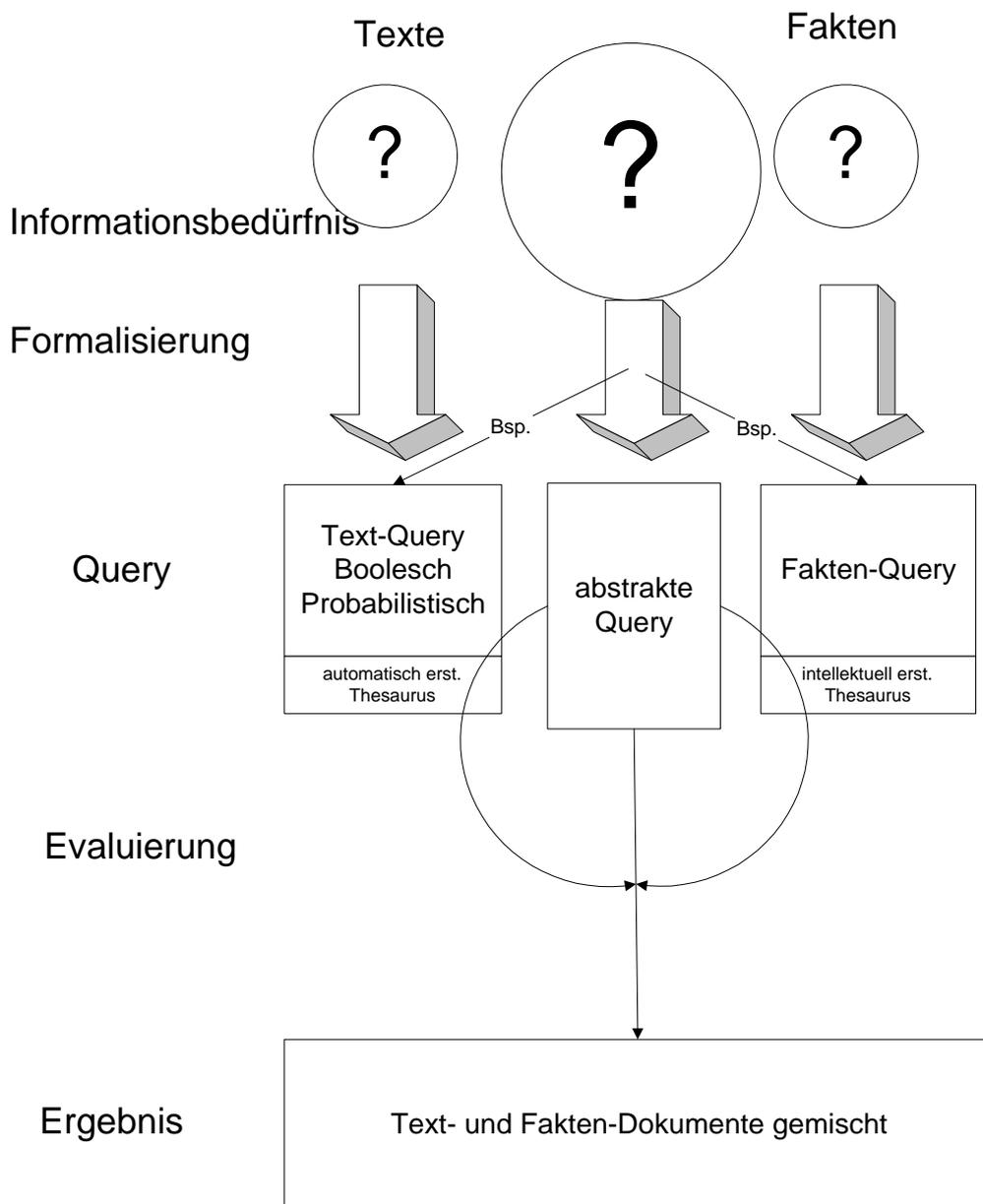


Abb. 4: Modalitätsunabhängige Suche

3.4 Zusammenfassung und weiteres Vorgehen

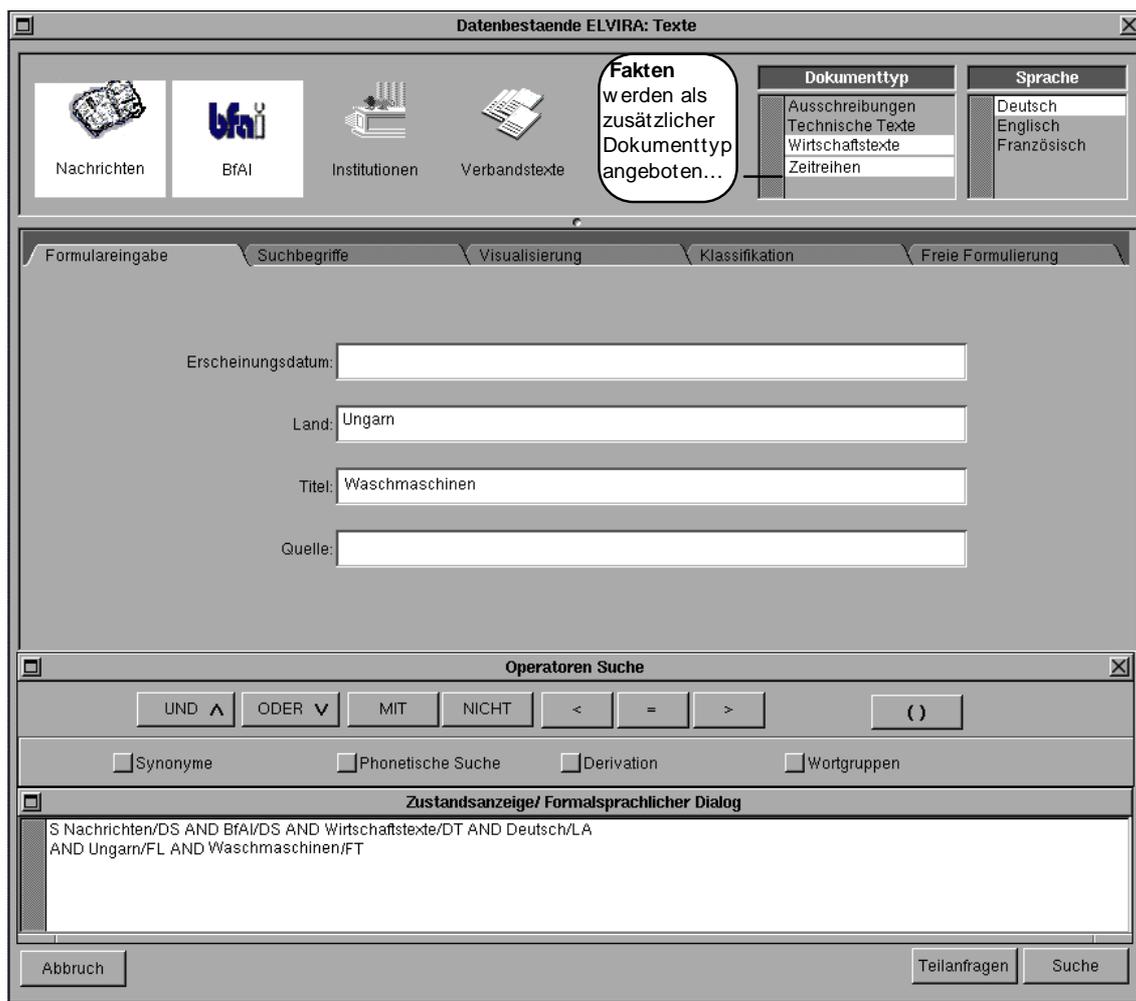
Empirische Untersuchungen im Anwendungsfall ELIVRA haben gezeigt, daß von den Integrationsstufen der kombinierten Text-Fakten-Recherche besonders die Transformationen auf der Stufe B für die Benutzer sinnvoll sind. Mit der Modellbildung gingen umfangreiche Analysen unterschiedlichster Datenarten und -quellen einher. Wie zu erwarten, zeigte sich, daß sowohl der formale Aufbau der Datenbestände als auch die Qualität der Erschließung erhebliche Unterschiede aufweist. Die Spannweite reicht von nach formalen Regeln erschlossenen und konsistent formatierten Daten (z.B. Zeitreihen, Messetermine, einzelne Textbestände) bis zu sehr grob bzw. ungegliederten Datentypen (z.B. Texte aus News-Tickern). Einige Felder aus Faktendaten finden ihre Entsprechung in

bibliographischen Angaben oder Schlagwörtern von Textbeständen, während andere Textbestände nur über automatische Indexierung erschlossen werden können. Dies bedingt den Einsatz unterschiedlichster Verfahren zur Transformation von Anfragen. Der Schwerpunkt der aktuellen Entwicklung zur Realisierung von Transformationen liegt auf den folgenden Bereichen: Über ein umfassendes Metamodell werden die heterogenen Datenbestände beschrieben und integriert, wobei neben dem formalen Aufbau auch Angaben über anwendbare Transformationsalgorithmen und verfügbare Hilfsmittel (z.B. Klassifikationen, Thesauri oder Konkordanzen) abgelegt sind. Thesaurusbasierte Verfahren, die dem Benutzer den Zugang über seine eigene Begrifflichkeiten ermöglichen. Hierzu befindet sich ein Thesaurus im Aufbau, der zunächst einzelne Teilbereiche aus dem Spektrum der beteiligten Verbände abdeckt und durch eine Konkordanz zu unterschiedlichen Datenbeständen vervollständigt wird. Vage Verfahren auf der Basis von Kookurenzen und neuronalen Netzen, die mit möglichst geringem Pflegeaufwand von Seiten der Verbände Suchanfragen zwischen dem unterschiedlichen Indexierungsvokabular der Datenbestände transformieren können (cf. [Mandl 1998]). Diese Ansätze extrahieren, die ihr Wissen in Form von Kookurenzen aus Doppelkorpora, die mit zwei Verfahren indexiert wurden. Ein erfolgversprechender statistischer Ansatz von [Sheridan/Ballerini 1996] wurde bereits für Multilinguales Retrieval eingesetzt und wird für ELVIRA adaptiert.

4 Textrecherche mit integrierter Faktensuche

Dieser Abschnitt veranschaulicht den Ablauf einer integrierten Suche anhand einer skizzenhaften Visualisierung der Interaktionsschritte. Das Szenario stammt aus [Mandl et. al. 1998], wo Beispiele für alle Typen der Integration gesammelt wurden. Sie wurden zusammen mit Verbandsmitarbeitern auf Plausibilität und Praxisrelevanz geprüft.

Die in Kapitel 3 aufgeführte Reinform zur Textrecherche wurde mittlerweile prototypisch realisiert und in ELVIRA II integriert (cf. [Krause/Schaefer 1998]). Im gewählten Beispiel sucht der Benutzer primär nach Texten und öffnet dementsprechend das Text-Suchobjekt. Daneben möchte er aber auch Zeitreihen erhalten, die relevant bzgl. seiner Problemstellung sind (Auswahl in der Liste Dokumenttyp; vgl. Abb. 5, oben rechts). Der Anwender aus dem Bereich der Elektroindustrie interessiert sich für den Markt für Waschmaschinen in Ungarn.



Ab

b. 5: Text-Suchobjekt (prototypischer Entwurf)

Beim Suchprozeß wird die Text-Anfrage intern in eine Zeitreihen-Anfrage transformiert. Die in Abb. 6 dargestellte Zustandsanzeige der Zeitreihensuche (siehe auch Abb. 2) macht dies deutlich.

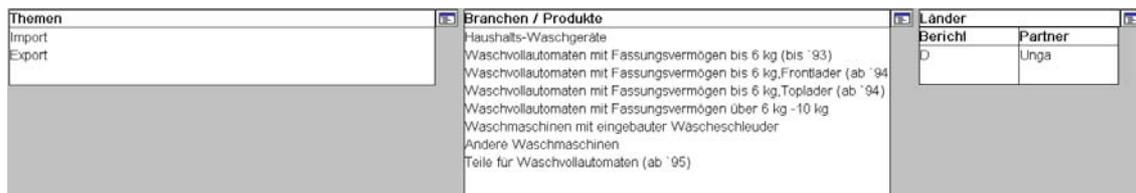


Abb. 6: Interne Zeitreihen-Anfrage

Die Leistung der Transformation besteht in diesem Fall in der Ausweitung von *Waschmaschinen* auf den Begriff *Waschvollautomaten*, der in der Nomenklatur verwendet wird. Dies kann durch eine Synonym-Relation im Thesaurus erreicht werden. Liegt kein Thesaurus vor, können aus dem IR bekannte Verfahren zur Termexpansion eingesetzt werden. Diese basieren meist auf Kookurrenzen und gehen davon aus, daß Begriffe, die häufig gemeinsam in Dokumenten vorkommen, ähnliche Bedeutung haben.

Die Auswahl der richtigen Themen *Export* und *Import* kann durch eine Regel zum Datenbestand BfAI (Bundesstelle für Außenhandelsinformationen) erreicht

werden, nach der sich diese Texte sich immer mit dem Außenhandel befassen. Aber auch diese Regel ist nicht unbedingt notwendig, da zu Ungarn momentan nur Zeitreihen zum Außenhandel vorliegen und das System die geeigneten Themen für die Zeitreihen-Suche auch aus dem Land-Deskriptor der Text-Suche erschließen kann. Das Ergebnis besteht aus einer gemischten Liste, die sowohl Texte aus dem BfAI Bestand als auch die entsprechenden Zeitreihen aus der Statistikdatenbank zum Außenhandel enthält (Abb. 7).

TITEL	LAND	QUELLE	EINGDATUM
Ausländische Produkte beherrschen den Markt	Ungarn	BfAI	960724
Mehr Waschmaschinen aus Ungarn	Ungarn	BfAI	951018
ZEITREIHEN ZUM THEMA			
Export (84501111) Waschvollautomaten mit Fassungsvermögen bis 6 kg,Frontlader (ab '94), 1000 DM Deutschland nach Ungarn			06.02.1997
Export (84501111) Waschvollautomaten mit Fassungsvermögen bis 6 kg,Frontlader (ab '94), Stück Deutschland nach Ungarn			06.02.1997
Export (84501119) Waschvollautomaten mit Fassungsvermögen bis 6 kg,Toplader (ab '94), 1000 DM Deutschland nach Ungarn			06.02.1997
Export (84501119) Waschvollautomaten mit Fassungsvermögen bis 6 kg,Toplader (ab '94), Stück Deutschland nach Ungarn			06.02.1997
Export (84509000) Teile für Waschvollautomaten (ab '95), 1000 DM Deutschland nach Ungarn			06.02.1997
Export (84509000) Teile für Waschvollautomaten (ab '95), Stück Deutschland nach Ungarn			06.02.1997
Import (84501111) Waschvollautomaten mit Fassungsvermögen bis 6 kg,Frontlader (ab '94), 1000 DM Deutschland aus Ungarn			06.02.1997
Import (84501111) Waschvollautomaten mit Fassungsvermögen bis 6 kg,Frontlader (ab '94), Stück Deutschland aus Ungarn			06.02.1997
Import (84509000) Teile für Waschvollautomaten (ab '95), 1000 DM Deutschland aus Ungarn			06.02.1997
Import (84509000) Teile für Waschvollautomaten (ab '95), Stück Deutschland aus Ungarn			06.02.1997

b. 7: Ergebnisliste mit Texten und Zeitreihen

Das Design der Benutzungsoberfläche für die Integration ist neben der Realisierung von Transformationen der Schwerpunkt der momentanen Entwicklung von ELVIRA. Das obige Beispiel zeigt die prototypische Implementierung der Transformationen von Texten zu Fakten. Eine ähnliche Funktionalität wird in das Zeitreihenrecherche-Werkzeug integriert. Diese Prototypen werden mit Benutzern getestet, wobei auch die Qualität der Transformationen untersucht wird.

Literatur

[Amstutz et.al. 1991]

Amstutz, H.; Holländer-Thönssen, B.: *Elektronische Ablage und Archivierung auf der Basis eines Database Management Information Retrieval Systems: Die Bedürfnisse - Das Angebot - Die Realität*. In: Fuhr, Norbert (Hrsg.) (1991): Information Retrieval. GI/GMD-Workshop Darmstadt. S. 78-93.

[Fuhr 1992]

Fuhr, Norbert: *Integration of Probabilistic Fact and Text Retrieval*. In: Belkin, Nicholas; Ingwersen, Peter; Pejtersen, Annelise (Hrsg.) (1992): SIGIR'92. Copenhagen, Denmark. June 21-24 1992. S. 211-222.

[Gu et al. 1993]

Gu, J.; Thiel, U.; Zhao, J.: *Efficient Retrieval of Complex Objects: Query Processing in a Hybrid DB and IR System*. In: Knorz, Gerhard; Krause, Jürgen; Womser-Hacker, Christa (Hrsg.) (1993): *Information Retrieval '93. Von der Modellierung zur Anwendung*. Proc. der 1.Tagung Information Retrieval. Konstanz. S. 67-81.

[Krause et al. 1997]

Krause, Jürgen; Mandl, Thomas; Stempfhuber, Maximilian: *Text-Fakten-Integration in ELVIRA*. ELVIRA-Arbeitsbericht 12, IZ Sozialwissenschaften, Bonn, 1997.

[Krause et al. 1998]

Krause, Jürgen; Mandl, Thomas; Schaefer, André; Stempfhuber, Maximilian: *Stand des Projektes ELVIRA II*. ELVIRA-Arbeitsbericht 19, IZ Sozialwissenschaften, Bonn, 1998.

[Krause/Schaefer 1998]

Krause, Jürgen; Schaefer, André: *Textrecherche-Oberfläche für ELVIRA II*. ELVIRA-Arbeitsbericht 16, IZ Sozialwissenschaften, Bonn, 1998.

[Mandl 1998]

Mandl, Thomas: *Vague Transformations in Information Retrieval*. In this volume.

[Mandl et al. 1998]

Mandl, Thomas; Schaefer, André; Stempfhuber, Maximilian: *Exemplarische Transformationen für die Text-Fakten-Integration*. ELVIRA-Arbeitsbericht 15, IZ Sozialwissenschaften, Bonn, 1998.

[Marx et al. 1995]

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

[Scheinost et al. 1998]

Scheinost, Ulrich; Haas, Hansjörg; Krause, Jürgen; Lindlbauer, Jürg (Hrsg.): *Marktanalyse und Marktprognose. Das ZVEI Verbandsinformationssystem ELVIRA*. IZ Sozialwissenschaften, Bonn, 1998.

[Sheridan/Ballerini 1996]

Sheridan, Páraic, Ballerini, Jean Paul: *Experiments in Multilingual Information Retrieval using the SPIDER System*. In: Frei et al. 1996. p. 58-65.

[Womser-Hacker 1997]

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