

WILIE

Ein wissensbasiertes Literaturerfassungssystem

Schirra, J.R., Brach, U., Wahlster, W., Woll, W.

Universität des Saarlandes

FR 10.2 Informatik IV

Im Stadtwald 15

6600 Saarbrücken 11

Abstract: Da konventionelle Systeme zur interaktiven Literaturerfassung ausschließlich mit vollständigen Literaturangaben in einem strengen Eingabeformat arbeiten können, werden sie von Wissenschaftlern, die damit ihre individuelle und fachbezogene Bibliographie aufbauen wollen, als zu aufwendig abgelehnt. Ein für diesen Zweck akzeptables Literaturerfassungssystem darf deshalb dem Benutzer keine unnötige Formatierarbeit aufzwingen und muß ihm erlauben, seine Literaturhinweise so zu verkürzen, als würde das System das entsprechende Fachgebiet kennen. Es wird das KI-System WILIE beschrieben, das aus knappen Benutzerangaben mithilfe seiner Wissensbasis vollständige Einträge in eine Literaturdatenbank erzeugt. Die Wissensbasis selbst wird sukzessive erweitert, indem relevante Information aus neu eingegebenen Literaturhinweisen automatisch extrahiert wird. Nach einer Analyse der Aufgabenstellung und der Einsatzbedingungen des WILIE-Systems wird ein unter ingenieurmäßigen Gesichtspunkten implementierter Prototyp vorgestellt. Das System besteht aus einer Komponente zur strukturellen Analyse, teilweise als ATN kodiert und einer mithilfe von Inferenzregeln arbeitenden Auswertungskomponente. Diese greift auf eine framestrukturierte Wissensbasis zu, die bibliographisches, geographisches und biographisches Wissen enthält.

1. Motivation

Bisherige Systeme zur interaktiven Titelerfassung in Hochschulen und anderen Forschungseinrichtungen arbeiten in der Regel zentral und indirekt. Zwischen den Wissenschaftler, der seine neu eingegangenen Dokumente (z.B. Berichte, Bücher) erfaßt haben möchte (im folgenden 'Benutzer' genannt) und den Rechner treten vermittelnd Formalerfasser (Bibliothekare oder Dokumentare), indem sie aus den Dokumenten der Benutzer systemgerechte Eingaben formulieren und die zum Ansetzen des Titels notwendigen Informationen extrahieren. Diese Tätigkeit setzt umfangreiches Erfahrungswissen aus der dokumentarischen Praxis sowie fachbezogenes Faktenwissen voraus. Vielen Wissenschaftlern ist diese Vorgehensweise (Abb. 1) zu aufwendig, wenn sie ihre individuelle und fachbezogene Bibliographie aufbauen wollen; sie wünschen sich ein System auf einem Kleinrechner, das an ihrem Arbeitsplatz zur Verfügung steht und ihre Literaturangaben direkt verarbeitet. Ein wissensbasiertes Literaturerfassungssystem (WILIE) soll deshalb die Kenntnisse eines Dokumentars für ein eingeschränktes Fachgebiet mit der Funktionalität eines konventionellen Titelerfassungssystems vereinigen (Abb. 2). WILIE gehört somit zur Klasse der kooperativen Zugangssysteme (vgl. Wahlster 84), die einen breiteren und effizienteren Einsatz konventioneller DV-Systeme (hier: Literaturdatenbanken) ermöglichen sollen.

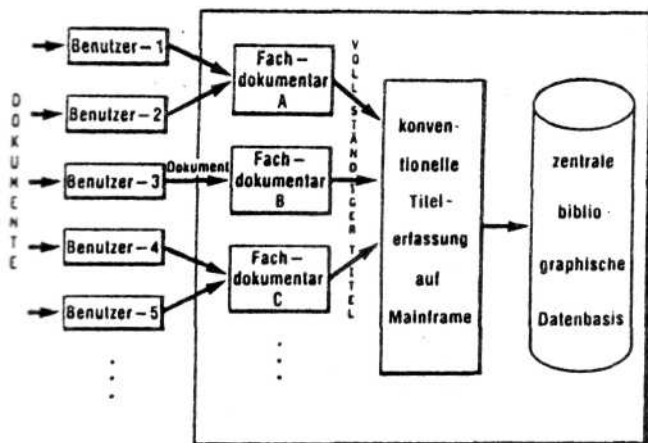


Abb. 1: Zentrale Titelerfassung in wissenschaftlichen Einrichtungen; Ein Fachdokumentar wird als Erfassungsspezialist eingeschaltet



Abb. 2: Direkte wissensbasierte Titelerfassung

Bisherige Titelerfassungssysteme erwarten vollständige Literaturhinweise in einem strengen Eingabeformat.

Beispiel 1: Vollständige Literaturangabe

McDermott, John (1983):

Extracting Knowledge from Expert Systems;

in: Proceedings of the Eighth International Joint Conference on Artificial Intelligence, Karlsruhe

Abweichungen von der vorgegebenen Reihenfolge werden ebensowenig akzeptiert wie inkonsistent gesetzte Trennzeichen. Vollständig muß die Literaturangabe sein, weil diese Systeme zu einem Dokument nur die Information kennen, die in der Eingabe enthalten war. Unvollständige Hinweise werden nicht in die interne Darstellung übergeführt.

Der Zugang zur Datenbasis eines konventionellen Titelerfassungssystems wird also durch eine *Formatschranke* und die *Vollständigkeitsforderung* sehr stark behindert. Demgegenüber soll WILIE mit seiner Wissensbasis durch *strukturelle Analyse* und *automatische Vervollständigung* den Zugang für alle Benutzer weit öffnen (Abb. 3).

Dazu muß ein System zur direkten Titelerfassung den folgenden beiden Forderungen des Benutzers genügen.

- Der Benutzer möchte sich bei der Eingabe nicht nach einer vorgegebenen Reihenfolge richten müssen, sondern erwartet, daß das System flexibel und robust auf die aktuelle Eingabestruktur reagiert.
- Er will wie einem Fachkollegen gegenüber auch knappe Literaturhinweise verwenden. Das System soll mithilfe seines Wissens die fehlenden Teile ergänzen.

Beispiel 2: Knappe Literaturangabe

McDermott; Extracting Knowledge from Expert Systems,

IJCAI8.

Der Vorname des Autors fehlt hier ebenso wie Erscheinungsort und -datum. Der Verweis auf den Tagungsband ist auf die Kurzangabe der entsprechenden Konferenz reduziert.

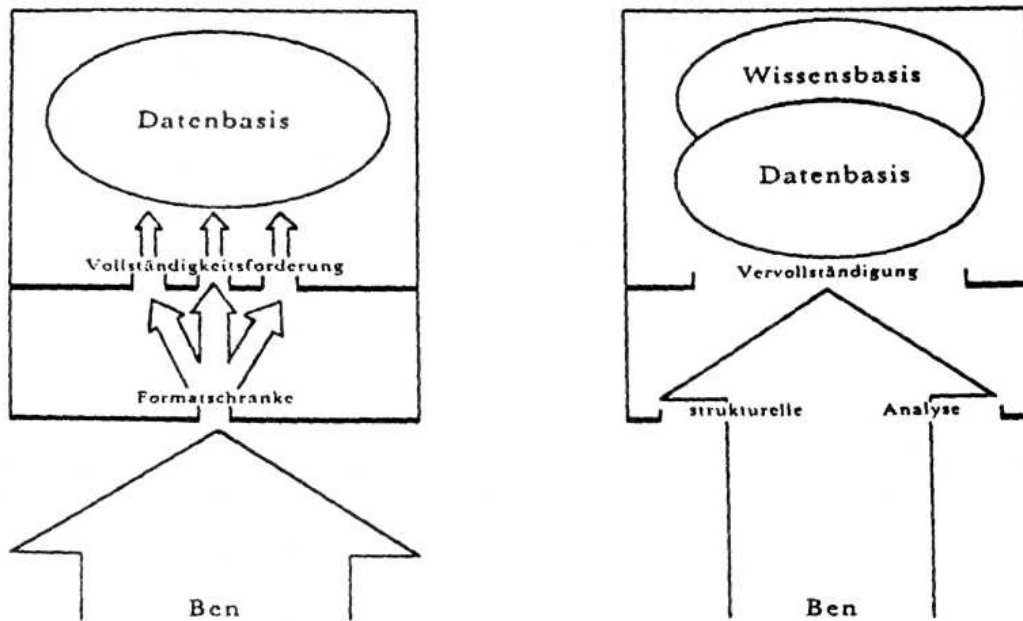


Abb. 3: Abstrakter Schnitt durch ein

Konventionelles Erfassungssystem wissensbasiertes

Zusammengefaßt ergibt sich als Hauptaufgabe von WILIE, aus Literaturhinweisen wissensbasiert eine bibliographisch möglichst vollständige Datenbasis aufzubauen. Sekundäres Ziel ist es, die Wissensbasis sukzessive zu erweitern, indem aus den Benutzereingaben automatisch neues bibliographisches, geographisches und biographisches Objektwissen extrahiert wird. Aus den Forderungen des Benutzers nach relativer Formatfreiheit und Unvollständigkeit der Eingabe kann eine Zweiteilung des Systems abgeleitet werden in eine Komponente zur *strukturellen Analyse* des Eingabestrings, die den Inhalt des Literaturhinweises in eine normalisierte Form überführt, und eine *Auswertungskomponente*, die diese vervollständigt (Abb. 4).

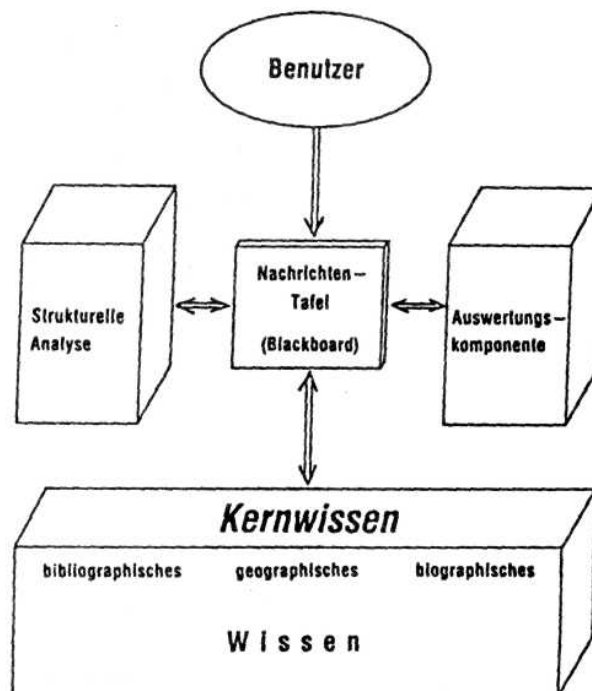


Abb. 4: Erster Entwurf der Architektur von WILIE

Das Ergebnis des Versuchs, durch ingenieurmäßiges Vorgehen mit bestehenden KI-Werkzeugen, wie ATNs, FRAME-Sprachen und Regelinterpretierern einen Prototypen des WILIE-Systems zu schaffen, wird in den folgenden Kapiteln im einzelnen vorgestellt. Der Entwicklungsschwerpunkt in dieser *rapid prototyping*-Phase lag demnach weniger auf innovativen Repräsentations- und Programmierkonzepten als auf einer möglichst umfassenden und effizienten Problemlösung.

2. Die Komponente zur strukturellen Analyse

Im WILIE-Prototypen sind die beiden Komponenten zur strukturellen Analyse und zur Auswertung hintereinandergeschaltet. Der strukturellen Analyse stehen die folgenden beiden Wissensquellen zur Verfügung:

- eine Sammlung bibliographischer Abkürzungen und Strukturpartikeln (Abb. 5)

| gefundene Abkürzung oder Strukturpartikel | Interpretation | Systemerwartung |
|--|----------------|-----------------------|
| -Proc. | Proceedings | Konferenznamen |
| -pp. | Pages | zwei Seitenangaben |
| -eds. | Herausgeber | mehrere Personennamen |
| -Press | Verlag | Verlagsnamen |

Abb. 5: Auszug aus der Liste der Abkürzungen und Strukturpartikeln

- eine Beschreibung der Struktur von Personennamenlisten bzw. von Titeln, wobei unter anderem die Reihenfolge der Satzzeichen besonders berücksichtigt wird. Dadurch ist die Sicht dieser Parsing-Komponente auf den Eingabestring sehr stark eingeschränkt (Abb. 6).

```

McDermott: Extracting Knowledge from Expert Systems, IJCAI-8.
⇒ @: @, @. ( Eingabe des Parsers )      N - Nachname
                                          T - Titel
                                          R - Rest
                                          @ - Block
⇒ [N, T, R] ( Ausgabe des Parsers )

```

Abb. 6: Beispiel für eine stark reduzierte Titelangabe ohne Strukturpartikeln

Die strukturelle Analyse ist ein dreistufiges Verfahren, das auf die erwähnten Wissensquellen zugreift, die als Listen bzw. als ATN (Bates 78) dargestellt sind (siehe Abb. 11 am Ende des Berichtes). Die erste interne Darstellung des Eingabestrings wird vom *Scanner* aufgebaut; es ist eine flache Liste von LISP-Punktpaaren, die jeweils einen Block (Teilstring zwischen zwei Trennzeichen) und das darauffolgende Trennzeichen enthalten. Diese Primärdarstellung ist isomorph zum Eingabestring. Anschließend reduziert der Vorparser die Komplexität dieser ersten Darstellungsform für die weitere Verarbeitung, indem er die enthaltenen bibliographischen Abkürzungen und Strukturpartikeln samt den zugehörigen Blöcken aus der Liste extrahiert und durch die Markierung (% . n) ersetzt. Die extrahierte Information wird in der zweiten internen Darstellungsform, einem Frame vom Typ *Titel* in Slots mit entsprechender Bezeichnung %n abgelegt.

Die vereinfachte Primärstruktur wird an den ATN-Parser (vgl. Abb. 7) übermittelt, der daraus mithilfe des Wissens über die Struktur von Personennamen und Titeln diese Teile der Eingabe ermittelt und entsprechend strukturiert in der *Titel*-Instanz abspeichert. Neben der Reihenfolge der Trennzeichen zum Erkennen der Personennamen wird eine Teilmenge des Strukturwortschatzes der gewählten Sprache dazu benutzt, die Titelblöcke zu markieren.

Beispiel 3: Varianten für die Angabe eines Personennamens

| | |
|-------------------|-----------------|
| John von Neumann | J. v. Neumann |
| von Neumann, John | v. Neumann J. |
| J. von Neumann | John v. Neumann |

Beispiel 4: inkonsistente Personenliste

von Neumann, John, Schmidt, R.; P. Wegener:

Beispiel 5: Titel mit Markierungswörtern

Planning in the world of the air traffic Controller

In einigen Fällen kann der Parser keine eindeutige Interpretation finden; zum Beispiel:

Schlageter, Stucky; versus *Schlageter, Gunter;*

Wenn ein führender Nachname von Komma gefolgt wird, versucht der Parser, das anschließende Wort als Familiennamen einer weiteren Person aufzufassen. Gelingt das, wird dieses Paar von Personennamen mit einer Unsicherheitsmarkierung versehen, da es sich möglicherweise beim zweiten Block um einen Vornamen handelt.

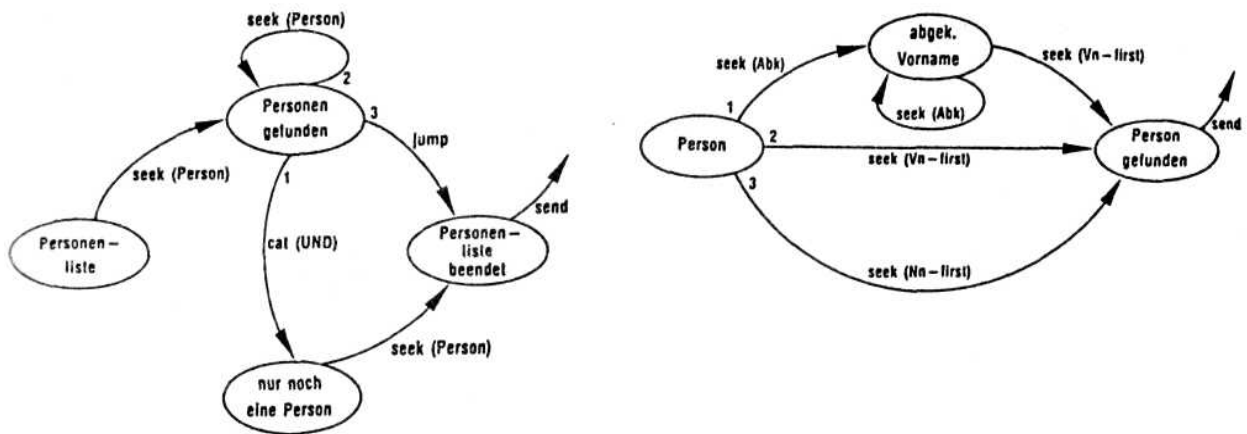


Abb. 7: Ausschnitt aus dem ATN-Teilnetz zur Analyse von Personennamen

3. Die Auswertungskomponente

Die Auswertungskomponente soll die Teile des Strings identifizieren, die von der strukturellen Analyse schon erkannt wurden, und eine Zuordnung zu den Objekten in der Datenbasis herstellen. Außerdem muß die Funktion der noch nicht erkannten Teile der Eingabekette und deren Identität bestimmt werden. Letztlich soll eine endgültige Darstellung des Inhalts der Literaturangabe aufgebaut und unter Umständen die Wissensbasis erweitert werden. Es ist offensichtlich, daß dazu vielfach auf das bio-, geo- und bibliographische Kernwissen zugegriffen werden muß. Implementiert wurde diese

Komponente als eine Sequenz von vier *Diskriminationsnetzen* in der Form von IF-THEN-ELSE-Regeln. Diese Teilnetze sind gemäß der Schwierigkeit ihrer Aufgaben hintereinander gestaffelt. Das erste identifiziert die gefundenen Personen und den Titel, das zweite die im Literaturhinweise enthaltenen bibliographischen Abkürzungen und Strukturpartikel. Dann werden in der dritten Phase die noch immer nicht erkannten Teile in der primären Liste verarbeitet und zuletzt überprüft, ob alle Objekte zusammenpassen. Während den ersten drei Phasen werden gegebenenfalls neue *individual-Frames* instantiiert. Das führt insbesondere in den Phasen 2 und 3 zu Erweiterungen des Kernwissens. Die Vervollständigung einer knappen Literaturangabe ergibt sich automatisch als Folge der Zuordnung zu der internen Darstellung der entsprechenden Objekte. Sammelstelle aller beteiligten Teilstrukturen bleibt die *Titel-Instanz* in der Art einer zentralen Nachrichtentafel (vgl. Erman et al. 80).

Um die Personen zu identifizieren, sucht WILIE in Phase 1 zunächst in der Wissensbasis nach Personenobjekten mit entsprechenden Namen. Diese Assoziation kann mehrdeutig sein. Falls das Kernwissen zu einem Namen kein Objekt enthält, muß eine neue Personeninstanz kreiert werden. Die *Parser- Unsicherheiten* können an dieser Stelle aufgelöst werden, wenn entsprechende Einträge vorliegen. Sonst wird durch einen Klärungsdialog die richtige Interpretation vom Benutzer erfragt, um zu aufwendige Schlußfolgerungsprozesse zu vermeiden.

Auch zur Identifikation der Blöcke mit bibliographischen Abkürzungen oder Strukturpartikeln wird zuerst im Kern wissen nach Strukturen des erwarteten Typs mit gleichem Namen gesucht. Falls auf diese Weise eine solche Beziehung hergestellt wurde, beginnt die dem Inselemparsing (Bates 78) analoge *Inselanalyse*: Ausgehend von den nun identifizierten Blockinseln wird versucht noch nicht erkannte Stringteile in diese Interpretation einzubeziehen, wozu die Platzhalter der Blöcke mit bibliographischen Abkürzungen und Strukturpartikeln benötigt werden.

Beispiel 6: Es sei ein Verlag durch die Partikel *press* markiert und vom Vorparser extrahiert worden.

```

...((Clarendon Press) . .) (Oxford . .) (England . .)
⇒ ...(% . 1) (Oxford . .) (England . .)
+ TITEL. %1: Clarendon Press
in der Wissensbasis:      VERLAG - 17
                          Name: Clarendon Press
                          Ort: Oxford
                          Staat: England

```

Die Wissensbasis enthalte das entsprechende Verlagsobjekt mit dem zugehörigen Sitz. Dann können die beiden in der Primärliste verbliebenen Blöcke neben dem Platzhalter leicht identifiziert werden. Alle übrig gebliebenen Blöcke der Primärstruktur werden in Phase 3 bearbeitet. Ausgehend vom Schema einer vollständigen Literaturangabe versucht WILIE heuristisch, die noch nicht besetzten Stellen aufzufüllen. Anfragen an den Benutzer unterstützen gegebenenfalls diesen Prozeß, der ebenfalls durch *Inselanalyse* abgerundet wird.

Zuletzt wird überprüft, ob die Interpretationen konsistent sind, wobei insbesondere die Unsicherheiten und Mehrdeutigkeiten der Assoziationen aufgelöst werden. Besonders die biographischen Daten werden dabei zu Rate gezogen.

Beispiel 7: Die Angabe: "McDermott: *Extracting Knowledge from Expert Systems*" soll verarbeitet werden; in der Wissensbasis seien zwei Personenobjekte des Nachnamens *McDermott* enthalten:

John McDermott
D. McDermott

Thema: Expert Systems
Thema: Spatial Reasoning

Dann wird durch Vergleich der angegebenen Arbeitsgebiete mit dem Resultat einer Themaextraktion aus dem Titel *John McDermott* als Autor gewählt.

Die erwähnten Klärungsdialoge werden von einer Interaktionskomponente initiiert. Neben einfachen Entscheidungsfragen der Form: "Ist *Stucky* Vorname von *Schlageter*?", womit z.B. Parser-Unsicherheiten aufgelöst werden können, werden zur Unterstützung des Wissenserwerbs auch kompliziertere Fragen nach der Klassenzugehörigkeit von Teilstrings gestellt:

Sys: Bitte Verlagsnamen markieren!
(mit Zeichen ungleich *space*)
Addison Wesley Press at Reading
Ben: xxxxxxxxxxxxxxxxxxxxxx

Abbildung 8a zeigt eine der Regeln, welche die Diskriminationsnetze definieren; die Regel mit Nummer 29 ist Teil der *Inselanalyse*:

Falls der Sitz des gefundenen Verlags bereits in der Wissensbasis eingetragen ist, dann versuche diese Ortsbezeichnungen oder Synonyme davon in den umliegenden Blöcken zu finden. Andernfalls versuche, den noch unbekanntem Sitz aus der Literaturangabe mithilfe des geographischen Wissens zu ermitteln."

Jeder *Ausgang* einer Regel ist mit einer Nachfolgeregeln markiert. Abbildung 8b stellt einen Ausschnitt des Netzes zu Phase 2 dar. Neben Rekursion sind auch verallgemeinerte Regeln mit mehr als zweiwertigen Entscheidungen zugelassen.

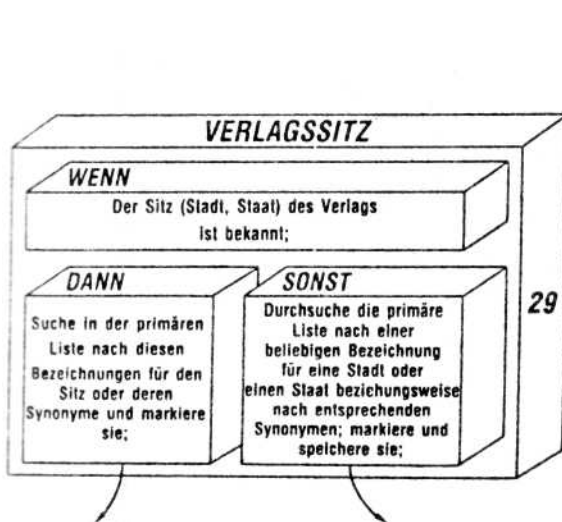


Abb. 8a: Beispiel für eine Regel

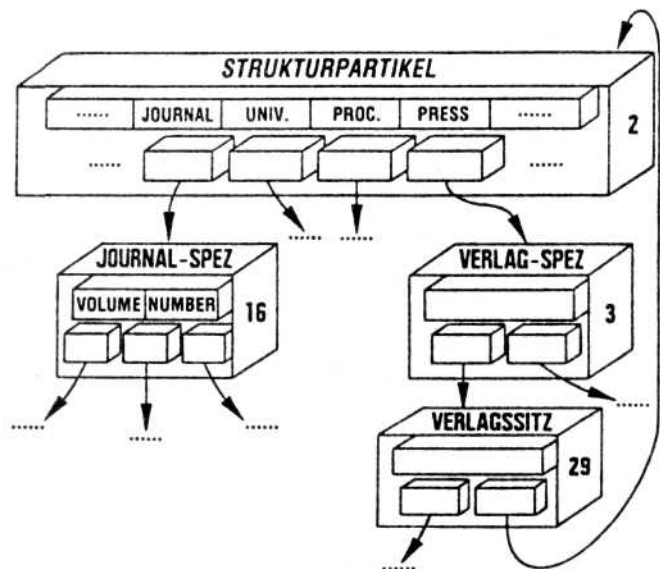


Abb. 8b: Ausschnitt aus dem Diskriminationsnetz

4. Das Kernwissen

Dieser wesentliche Teil der Wissensbasis von WILIE enthält das Wissen über die Objekte des Publikationswesens und ist in drei Wissensquellen aufgeteilt: das bibliographische, das geographische und das biographische Wissen. Im folgenden sind dazu einige Wissensseinheiten beispielhaft aufgeführt.

- bibliographisches Wissen

| | | | |
|-----------|--|----------|---|
| Springern | ⇒ Instanz von Körperschaft vom Typ Verlag Name: Springer Verlag Ort: Berlin, New York, London, Tokio spez-Veröff: Informatik Fachberichte | IJCAI-83 | ⇒ Instanz von Konferenz vom Typ IJCAI Ort: Karlsruhe Staat: FRG Jahr : 1983 Nummer 8 |
|-----------|--|----------|---|

| | | |
|-------|---|-------------------------|
| IJCAI | ⇒ Instanz von Konferenz Name: International Joint Conference on Artificial Intelligence Thema: KI Period: 2 | - biographisches Wissen |
|-------|---|-------------------------|

| | |
|-----------|---|
| McDermott | ⇒ Instanz vom Person Name McDermott Vorname: John Thema: Expert Systems 1-uni: CMU 1-time 1983 |
|-----------|---|

- geographisches Wissen

| | |
|----------|---|
| New York | ⇒ Instanz von Geo-Einheit vom Typ Ort Staat: New York |
|----------|---|

| | |
|--|--|
| | ⇒ Instanz von Geo-Einheit vom Typ Staat Kapitale: Albany Orte: New York Staat: USA |
|--|--|

| | |
|--|---|
| | ⇒ Instanz von Person Name McDermott Vorname D. Thema: Spatial Reasoning 1-uni: Yale-University 1-time 1980 |
|--|---|

Wie die Beispiele zeigen, kann es beim Anwenden dieser Wissensquellen zu Mehrdeutigkeiten kommen. Diese Ambiguitäten sind mithilfe des Kontextes jedoch im allgemeinen auflösbar, wozu z.B. die Frameslots *Thema*, *1-uni* und *1-time* (1 - last) in den Personenobjekten wesentlich beitragen können. An die 1-Slots ist das Wissen des Systems über die letzte Arbeitsstelle der Person gebunden.

WILIE benutzt als Repräsentationssprache eine Untermenge von FRL-Konstrukten (Roberts/Goldstein 77) und modelliert mit dem Frame-Netz (Abb. 9) die Relationen des Publikationswesens. Prototypische Frames fassen Informationen zu größeren Einheiten zusammen, was infolge der Vererbungsmechanismen eine sehr ökonomische Wissensrepräsentation erlaubt. Die Möglichkeit, Frames nur teilweise zu instantiieren, eröffnet den wichtigen Zugang zu unvollständigem Wissen und Default-Werten. Durch die Dämonen der *procedural attachments* schließlich kann auf Datenbankobjekte so zugegriffen werden, als seien es *individual-Frames*.

Da es sich beim Kernwissen offensichtlich um Massendaten handelt, ergab sich die Notwendigkeit, einen Teil der Frames außerhalb des Hauptspeichers zu halten. Deshalb werden alle individuellen Instanzen in die relationale Datenbank INGRES (Woodfill et al. 83) ausgelagert. Auf sie wird mithilfe des generischen Frame-Netzes, das als erweitertes Datenbankschema aufgefaßt wird, durch Dämon-Prozeduren (vgl. Roberts/Goldstein 77) über die LISP-INGRES-Schnittstelle (Bock 84) zugegriffen.

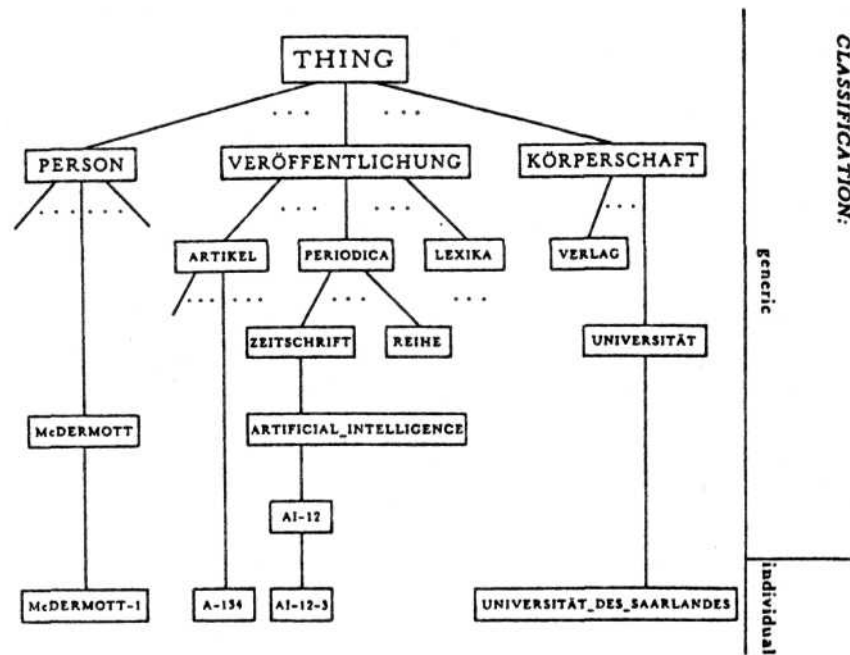


Abb. 9: Ausschnitt aus dem Framenetz: nur ako- / instance-Verweise sind eingetragen

Das Extrahieren der Daten aus der Datenbank und ihre Instantiierung als Frame geschieht also für die zugreifende Komponente unsichtbar; die Aufspaltung in aktives und inaktives Wissen ist nur innerhalb der Verwaltung des Kernwissens relevant.

5. Der Formatierer

Nachdem von der Auswertungskomponente eine endgültige Darstellung des Inhalts der Literaturangabe als *individual*-Instanz im Framenetz des Kernwissens erstellt wurde, gibt WILIE zur Kontrolle und externen Weiterverarbeitung den vollständigen Hinweis in einem wählbaren Format aus. Der dafür zuständige Formatierer soll ferner dem Benutzer erlauben, selbst neue Formate gemäß den Konventionen verschiedener Verlage zu definieren (zur Zeit noch nicht implementiert). Die Formatspezifikationen bilden kontextfreie Grammatiken, wodurch mehrere Ebenen des Auflösungsgrades und eine gute Verständlichkeit ermöglicht werden. Außerdem lassen sich durch Einfügen neuer Ableitungsregeln sehr leicht Formatvarianten erzeugen.

Beispiel 8: Einige Ableitungsregeln

Literaturhinweis-1 \Rightarrow *perslist* Jahr : Titel, NL TAB Res

perslist \Rightarrow *perslist-1*

Jahr \Rightarrow *Jahr-1*

Jahr-J \Rightarrow ([Jahr])

perslist-1 \Rightarrow { *person* ; }*

[XYZ] - Pseudoterminal; druckt entsprechenden Teil XYZ der Literaturangabe aus;

Implementiert sind diese Grammatikregeln als Instanzen des prototypischen Frames *Schema*. Der Name einer *Schema*-Instanz zu einer Regel $X \Rightarrow YZ \dots$ entspricht der linken Seite X ; die rechte Seite der Regel $YZ \dots$ wird als flache Liste im Slot *RES* abgelegt.

Die Behandlung der terminalen Symbole ist einer der Vorteile dieser Darstellung; Terminale werden als Slotnamen aufgefaßt. Der dort abgelegte Wert veranlaßt den Interpretierer, die entsprechende Zeichenkette auszugeben. Da das System im wesentlichen für alle Formatvarianten die gleichen Terminale verwendet, können die entsprechenden Slots im Prototypen *Schema* definiert werden und sind dann in jeder *individual*-Instanz durch Vererbung verfügbar.

Um die Flexibilität der Spezifikationen zu erweitern, können Nonterminale *lokal* zu einer Regel definiert werden. Diese *zwischen geschalteten Nonterminalsymbole* werden nicht direkt als Namen einer *Schema*-Instanz aufgefaßt, sondern als Bezeichnung eines Slots in der Darstellung der aktuellen Regel. Erst der Wert dieses Slots ist dann ein echtes Nonterminal (Abb. 10).

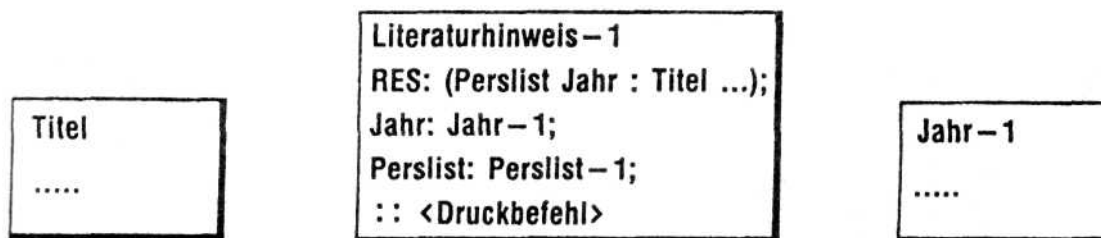


Abb. 10: Beispiel für Schemadarstellung

6. Technische Daten, Kritik und Ausblick

Der vorgestellte Prototyp eines WILIE-Systems ist auf einer VAX 11/780 unter Berkeley UNIX 4.2 implementiert. Programmiersprache ist FRANZ-LISP op. 38/79, das um FRL-artige Framekonstrukte (Winston/Horn 81) erweitert wurde. Der vom System benötigte Gesamtspeicherplatz FRANZ-LISP, INGRES-Verwaltung und LISP-Code beträgt etwa 2.5 MByte, wobei sich der Quellcode von WILIE aus ca. 420 LISP-Funktionen zusammensetzt. Die Verarbeitungszeit beläuft sich bei vollständigem Verarbeitungstrace im compilierten Modus auf 2-20 CPU-Sekunden. Dabei ist die Laufzeit des Datenbanksystems nicht berücksichtigt.

Die Implementierung des Prototypen wurde von fünf Studenten im Laufe von vier Monaten durchgeführt (für eine detaillierte Beschreibung vgl. Beiche et al. 85).

Zur Zeit wird an einem Re-Design gearbeitet, das die folgenden Kritikpunkte und Erweiterungsmöglichkeiten berücksichtigt:

- Die in einigen Wissensquellen verwendete prozedurale Darstellung muß durch eine deklarative ersetzt werden, damit die unten erwähnten Erweiterungen auf einer abstrakten Ebene durchgeführt werden können; das betrifft insbesondere den ATN-Parser und die Inferenzregeln der Auswertungskomponente.
- Das System soll vom Benutzer an andere Sprachen angepaßt werden können - das bisherige System war im wesentlichen am Englischen orientiert; u.a. müssen die bibliographischen Abkürzungen und Strukturpartikeln sowie die Liste der Titelspezifischen Worte ergänzt werden können.

- Der Benutzer soll das System auf ein anderes Fachgebiet einstellen können; das impliziert u.U. eine Änderung des Kernwissens, z.B. neue Personentypen wie "Übersetzer" im philologischen Bereich.
- Die Interaktionskomponente sollte in Richtung auf eine Erklärungskomponente ausgebaut werden.
- Ein Editor für die Formatspezifikationen ist zu integrieren.

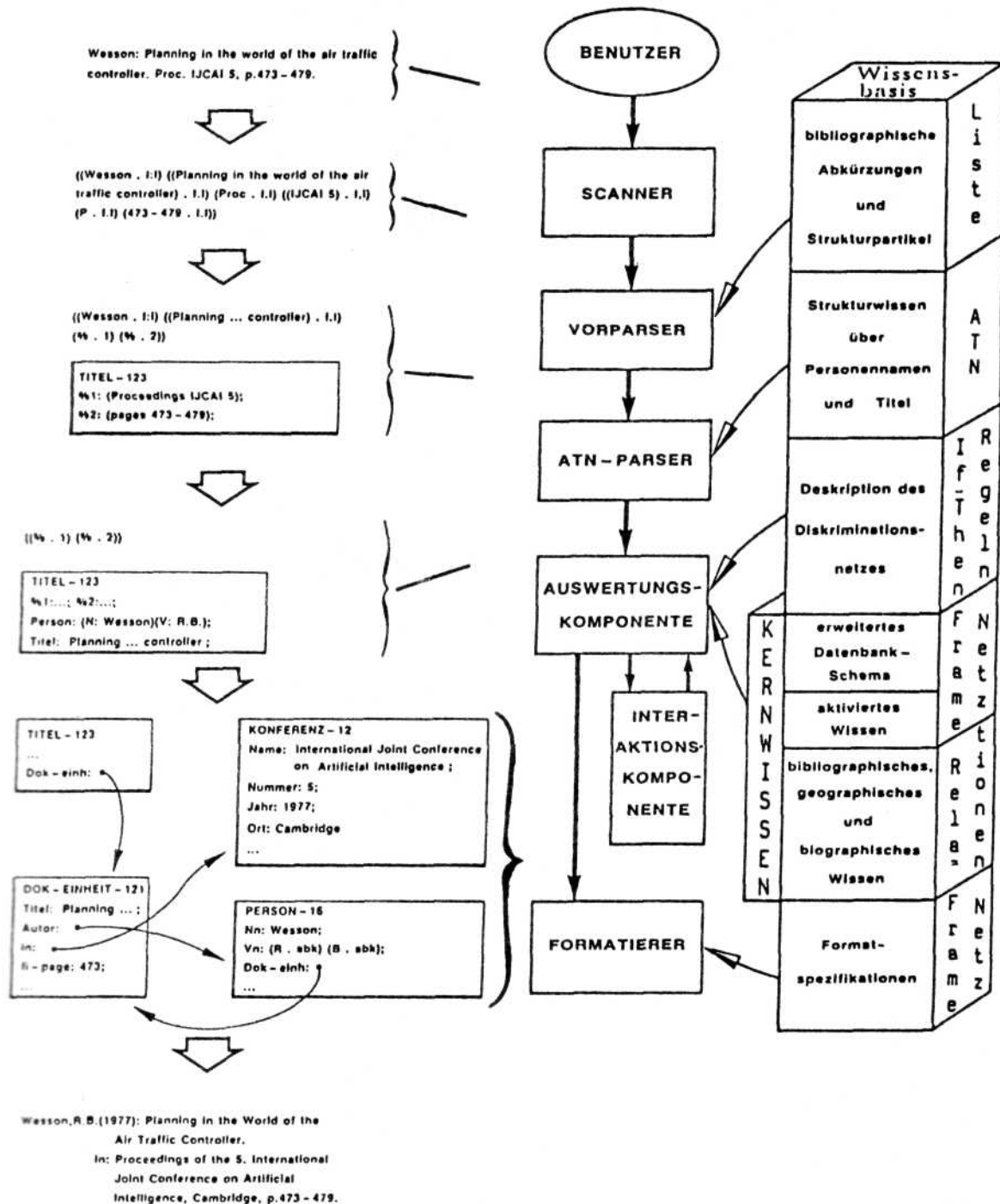


Abb. 11: Systemarchitektur, Wissensquellen des Systems und Verarbeitungsbeispiel

LITERATUR:

(Bates 78)

Bates, M. (1978): The Theory and Practice of Augmented Transition Network Grammars. In: Bolc, L. (ed.): *Natural Language Communication with Computers*. New York: Springer, pp. 191-260.

(Beiche et al. 85)

Beiche, H. P., Brach, U., Kalmes, J., Profitlich, H. J., Woll, W. (1985): SYSTEM VERLAG- Ein System zur wissensbasierten Vervollständigung von Literaturangaben. Internes Memo, FR10.2 Informatik IV, Univ. des Saarlandes, Saarbrücken.

(Bock 84)

Bock, R. (1984): Zugangssystem von FRANZ-LISP zum relationalen DB-System INGRES. Internes Memo, FR10.2 Informatik IV, Univ. des Saarlandes, Saarbrücken.

(Erman et al. 80)

Erman, L., Hayes-Roth, F., Lesser, V., Reddy, D. (1980): The HEARSAY-II Speech-understanding System: Integrating Knowledge to Resolve Uncertainty. In: *Computing Surveys*, Vol. 12, No. 2, pp. 213-253.

(Foderaro/Sklower 82)

Foderaro, J., Sklower, K. (1982): The FRANZ-LISP Manual. Univ. of Calif., Berkeley.

(Hayes-Roth et al. 83)

Hayes-Roth, F., Waterman, D., Lenat, D. (eds.) (1983): *Building Expert Systems*. Reading, Mass.: Addison-Wesley.

(Roberts/Goldstein 77)

Roberts, J., Goldstein, I. (1977): The FRL Manual. Memo No. 409, Artificial Intelligence Laboratory, MIT, Cambridge, Mass.

(Schirra/Hochgesand 84)

Schirra, J., Hochgesand, U. (1984): Bericht der Projektgruppe WILIE. Internes Memo, FR10.2 Informatik IV, Univ. des Saarlandes, Saarbrücken.

(Wahlster 84)

Wahlster, W. (1984): Cooperative Access Systems. In: *Future Generations Computer Systems*, Vol. 1, No. 2, pp. 103-111.

(Wimton/Horn 81)

Winston, P. H., Horn, B. (1981): *LISP*. Reading, Mass.: Addison-Wesley.

(Woodfill et al. 83)

Woodfill, J., Siegal, P., Ranstrom, J. Meyer, M., Allman, E. (1983): *The INGRES Reference Manual. Version 7*. Univ. of Calif., Berkeley.