Informatik-Fachberichte 307

Herausgeber: W. Brauer
im Auftrag der Gesellschaft für Informatik (GI)

Subreihe Künstliche Intelligenz

Mitherausgeber: C. Freksa
in Zusammenarbeit mit dem Fachbereich 1
„Künstliche Intelligenz" der GI

G. Klose E. Lang Th. Pirlein (Hrsg.)

Ontologie und Axiomatik der Wissensbasis von LILOG

Wissensmodellierung im
IBM Deutschland LILOG-Projekt

Springer-Verlag
Berlin Heidelberg New York London Paris
Tokyo Hong Kong Barcelona Budapest

Herausgeber

Gudrun Klose
TU Berlin, Fachbereich Informatik (20)
Franklinstraße 28/29, W-1000 Berlin 10

Ewald Lang
Universität Wuppertal, SFB 282 „Theorie des Lexikons"
Gauß-Straße 20, W-5600 Wuppertal

Thomas Pirlein
IBM Deutschland GmbH
Wissenschaftliches Zentrum, WT-IWBS 7000-75
Schloßstraße 70, W-7000 Stuttgart 1

CR Subject Classification (1991): I.2, I.2.4

ISBN-13: 978-3-540-55306-9 e-ISBN-13: 978-3-642-77387-7
DOI: 10.1007/978-3-642-77387-7

Dieses Werk ist urheberrechtlich geschützt. Die dadurch begründeten Rechte, insbesondere die der Übersetzung, des Nachdrucks, des Vortrags, der Entnahme von Abbildungen und Tabellen, der Funksendung, der Mikroverfilmung oder der Vervielfältigung auf anderen Wegen und der Speicherung in Datenverarbeitungsanlagen, bleiben, bei auch nur auszugsweiser Verwertung, vorbehalten. Eine Vervielfältigung dieses Werkes oder von Teilen dieses Werkes ist auch im Einzelfall nur in den Grenzen der gesetzlichen Bestimmungen des Urheberrechtsgesetzes der Bundesrepublik Deutschland vom 9. September 1965 in der jeweils geltenden Fassung zulässig. Sie ist grundsätzlich vergütungspflichtig. Zuwiderhandlungen unterliegen den Strafbestimmungen des Urheberrechtsgesetzes.

© Springer-Verlag Berlin Heidelberg 1992

33/3140-543210 – Gedruckt auf säurefreiem Papier

Vorwort

Die *Wissensbasis von LILOG* stand in der über fünfjährigen Laufzeit des LILOG-Projekts stets im Brennpunkt des Interesses - als Integrationsbereich der diversen Subprojekte und Verteilerknoten für kooperativ zu lösende Aufgaben; als Quelle praktischer Erfordernisse und Testfeld theoretischer Erwägungen stand sie folglich auch im Schnittpunkt vielfältiger Diskussionen. Der vorliegende Band will den nach Abschluß des Projekts erreichten Stand dieser Diskussionen dokumentieren und damit zur Erarbeitung von *erfahrungsgestützten und theoriegeleiteten Entwurfsprinzipien* für Wissensbasen beitragen.

Die Beiträge des Bandes beziehen sich vornehmlich auf die Wissensbasis von LEU/2, die zweite LILOG-Experimentier-Umgebung, die im Juli 1990 im Wissenschaftlichen Zentrum der IBM Deutschland GmbH als prototypische Implementierung eines textverstehenden Systems der internationalen Öffentlichkeit vorgestellt wurde. Der für diesen Band gewählte spezielle Blickwinkel läßt sich durch einen kurzen Blick auf die Projekt-Geschichte wie folgt verdeutlichen. Das übergreifende Gesamtziel des 1985 begonnenen Projekts LILOG (= '**Li**nguistic and **log**ic methods for the computational understanding of German') wird in dem von O. Herzog und C.-R. Rollinger herausgegebenen Abschlußberichtsband[1] so formuliert:

> "The various reasons for IBM to perform LILOG ... resulted in the practical goal to develop a system capable of understanding German texts, [i.e. a system] able to extract knowledge from these texts and to represent it in such a way that enables further processing." [p. 3]

Zur Erreichung dieses ehrgeizigen Ziels wurde fünfeinhalb Jahre lang großflächige Grundlagenforschung in fast allen einschlägigen Disziplinen und Bereichen geleistet, mit einer durchschnittlichen Anzahl von ca. 60 (insgesamt etwa 200) Wissenschaftlern und - zum Zeitpunkt der Entwicklung von LEU/2 - sechs arbeitsteilig in Subprojekten mit LILOG kooperierenden Universitäten (Hamburg, Stuttgart, Osnabrück, Tübingen, Saarbrücken und Trier). Der als Referenz-Buch für LILOG unerläßlich zu konsultierende Berichtsband von [Herzog, Rollinger (Eds.)] präsentiert das gesamte Spektrum von Inhalten, Strukturen und Ergebnissen des LILOG-Projekts im Überblick, während der vorliegende Band ergänzend dazu konzentriert ist auf eine detaillierte Rekonstruktion und Problemanalyse der Erfahrungen, die beim Aufbau der Wissensbasis von LILOG gesammelt werden konnten.

Die Implementierung des ersten Prototypen im Jahre 1988 brachte eine erste Konkretisierung von bis dahin erreichten Projektergebnissen. Relevant für die Entwicklung waren die danach präziser stellbaren Anfragen und schärfer konturierbaren Einsichten zum *Inhalt* der zu kodierenden Wissensbasis: um dem gesteckten Ziel nach Schaffung eines wissensbasierten Textverstehenssystems näher zu kommen, mußten beträchtliche Anteile sowohl von Alltagswissen ("common sense knowledge" oder "world knowledge") als auch von linguistisch aufbereitetem Sprachwissen ("word knowledge") repräsentiert und für verschiedene Systemkomponenten verfügbar gemacht werden. Die Wissensbasis wurde in dem innerhalb des Projekts entwickelten Formalismus L_{LILOG} geschrieben und mit der dazugehörigen Inferenzmaschine getestet - vgl. hierzu Kap. 2.1 in [Herzog, Rollinger (Eds.)] und den Beitrag von Bollinger (in diesem Band). Angesichts

---

1 O. Herzog, C.-R. Rollinger (Eds.): Text Understanding in LILOG: Integrating Computational Linguistics and Artificial Intelligence. Final Report on the IBM Germany LILOG Project. (= Lecture Notes in Artificial Intelligence 546). Springer-Verlag, Berlin-Heidelberg-New York etc. 1991

des zu erstellenden Frage-Antwort-Systems ergab sich für die Wissensbasis zudem die Anforderung, daß das benötigte Hintergrundwissen sowohl die Repräsentation der textuellen Information im Sinne *expliziten Textwissens* unterstützt als auch weitergehende Inferenzprozesse bezüglich des *impliziten Textwissens* erlaubt - vgl. hierzu die Beiträge von Gängler & Wachsmuth und von Börkel (in diesem Band). Nach der Fertigstellung von LEU/2 im Sommer 1990 wurde den an der Implementierung Beteiligten eine etwa einjährige Phase intensiver Auswertung ermöglicht, in deren Verlauf Ontologie und Axiomatik der Wissensbasis zum Thema eines von den Herausgebern im Oktober 1990 veranstalteten Workshops wurden, dessen kontroverse und kreative Diskussion zur Konzipierung dieses Bandes geführt hat. Die hier abgedruckten Beiträge sind jedoch inzwischen so weit überarbeitet und aktualisiert worden, daß sie bereits die nächste Diskussionsrunde repräsentieren.

Anlage und Problemauswahl des vorliegenden Bandes sind durch mehrere ineinandergreifende Motive bestimmt. Zunächst einmal ist es angesichts der eindrucksvollen Größenordnung und Bandbreite des LILOG-Projekts von aktuellem Interesse, den letztlich funktionsfähig implementierten Kernbereich der Wissensbasis in den Fokus einer kritischen und weiterführenden Analyse "post festum" zu rücken. Sodann dürfte es die Wirkung von LILOG im internationalen Bereich nachhaltig unterstützen, wenn die hier beim Knowledge Engineering gesammelten Erfahrungen in detaillierter und reflektierter Form mitgeteilt und zur Debatte gestellt werden. Es gibt in der Literatur zur Wissenrepräsentation und zur Architektur von Wissensbasen bisher kaum Studien, die auf einen vergleichbar reichhaltigen Erfahrungsfundus zurückzugreifen willens oder in der Lage wären. Nun liegt mit der ausgezeichneten Monographie von Ernest Davis[2] erstmals auch ein Lehrbuch zur Wissensrepräsentation vor, das verschiedene Ansätze und Formalisierungsvarianten aus der Literatur referierend aufarbeitet und als ergänzende Lektüre für den vorliegenden Band unbedingt zu empfehlen ist, aber ersetzen kann sie ihn nicht. Zum einen bekennt Davis im Vorwort seines Buches die Ausblendung von "representations of knowledge based on linguistic considerations" als "important omission", während unser Band genau diesem Aspekt besonders viel Aufmerksamkeit widmet. Zum anderen hat die vorliegende Aufsatzsammlung eine zu einem Lehrbuch komplementäre Konzeption. Statt um die Darstellung mehr oder minder exemplarischer Beispiele geht es uns primär darum, bei der Entwicklung der Wissensbasis von LILOG getroffene oder erwogene Modellierungsentscheidungen zu dokumentieren und in ihrem Zusammenhang mit der gesamten Implementierung zu diskutieren.

Die geläufige Tatsache, daß Inhalt und Technologie von Wissensbasen auf Fachkonferenzen ebenso wie in der einschlägigen Literatur eher verschwiegen als öffentlich ausgebreitet werden, hat uns ermutigt, am Beispiel von LEU/2 einmal wirklich in die Tiefe zu gehen und einen gerade abgeschlossenen Abschnitt der eigenen Forschung gründlich und schonungslos auszuleuchten. Gewisse Fehlentscheidungen oder Notlösungen werden somit nicht vertuscht, sondern ausdrücklich in die Auswertung einbezogen.

Im Zentrum der Diskussion steht das Hintergrundwissen, das seit der Fertigstellung des zweiten Prototypen einen integralern Bestandteil des Systems bildet. Zu diesem Zeitpunkt umfaßt die Wissensbasis von LEU/2 etwa 700 Konzepte, ca. 300 Axiome ("Regeln") sowie im Durchschnitt 20 Attribute pro Konzept - zu Details vgl. Börkel und Gerstl (in diesem Band). Die Aufteilung der Konstrukte in Ontologie

---

2 Ernest Davis: Representations of Commonsense Knowledge, Morgan Kaufmann Publishers, San Mateo, CA 1990

einerseits und Axiomatik andererseits wird vom Formalismus nahegelegt, aber von mehreren Autoren des Bandes problematisiert. Die vielfältigen Querverbindungen, mit Hilfe derer die Wissensbasis als Integrationsbereich und Verteilerknoten mit den übrigen Komponenten des Systems interagiert, finden in der Anlage des Bandes ihren direkten Niederschlag durch abgestimmte wechselseitige Bezugnahmen unter den Autoren, durch die z. T. mehrfache Kommentierung von Hauptbeiträgen und durch die Gliederung des Bandes in vier inhaltlich eng verbundene thematische Teile.

Teil I: **Aspekte der Wissensmodellierung in natürlichsprachlichen Systemen** spannt zunächst die einschlägigen Problemfelder aus verschiedenen Blickwinkeln auf. Als Einführung dient ein pragmatisch orientierter kritischer Rückblick auf den Verlauf der Implementierung der LILOG-Wissensbasis. Es folgen ein Plädoyer aus linguistischer Sicht für die repräsentationelle Differenzierung von sprachlichem und nicht-sprachlichem Wissen anhand von LEU/2 und ein Grundsatzartikel aus KI-Sicht mit (durch Fallstudie illustrierten) Reflexionen über die Nutzung naiver Theorien in der Wissensrepräsentation.

Teil II: **Wissensmodellierung und sprachliche Oberfläche** diskutiert die unterschiedlichen Ansprüche, denen die Wissensmodellierung eines textverstehenden Systems unter dem Aspekt ihrer kognitiven Adäquatheit und unter den Anforderungen der Generierung gerecht werden muß. Dabei wird die in Teil I aufgeworfene Frage der Differenzierung von sprachlichem und nicht-sprachlichem Wissen aus der jeweiligen Perspektive von Analyse und Generierung aufgegriffen und weitergeführt.

Teil III: **Wissensmodellierung und Inferenzen** beginnt mit einer methodologisch akzentuierten Studie zum "Ontological Engineering", die am Beispiel des Wissens über räumliche Objekte auf die Rolle empirischer Theorien (z.B. aus der linguistischen Semantik) für die Wissensmodellierung hinweist und die Schritte ihrer Überführung in die LEU/2-Wissensrepräsentationen vorführt. Es folgt ein Erfahrungsbericht, der am Beispiel des temporalen Bereichs den engen Zusammenhang von Modellierungs- und Inferenzaspekten thematisiert. Dieser Zusammenhang wird in den folgenden Aufsätzen weiter aufgefächert - in einer kritisch-konstruktiven Bestandsaufnahme der aktuellen Implementierung von L_{LILOG} und in einem Beitrag, der die Modellierung des für sprachnahe Inferenzen benötigten regelhaften Hintergrundwissens diskutiert.

Teil IV: **Entwicklung und Verwaltung der Wissensbasis** bietet zunächst einen Beitrag, der positive und negative Erfahrungen mit der Verwaltung der Wissensbasis in praktikable Richtlinien zur Projektorganisation umzusetzen versucht. Den Abschluß bildet eine Retrospektive auf die LILOG-Wissensbasen vom ersten Prototypen bis zu LEU/2 und kennzeichnet die dabei gewonnenen Erfahrungen.

Die **Anhänge** dokumentieren Struktur und Konzeptinventar der Wissensbasis, die Syntax von L_{LILOG} sowie das zugrunde gelegte Textkorpus.

Zentriert auf die Wissensbasis geben die hier abgedruckten Beiträge und Kommentare, indem sie über konkrete Entscheidungen im Alltag des Knowledge Engineering informieren, zugleich auch einen erheblichen Teil der in LILOG geleisteten Arbeit zur Integration von Wissensbasis und Nachbarkomponenten wie Syntax, Semantik und Lexikon wieder. Jedes dieser Anschlußstücke für sich wie auch ihre Abstimmung untereinander wären sicher einen eigenen Band wert. Ausführlichere Darstellungen dieser Komponenten finden sich in den Kap. 1.2 - 1.4 in [Herzog, Rollinger (Eds.)].

Was in diesem Band eingestandenermaßen zu kurz kommt, aber an den Leser zumindest als Anregung weitergegeben werden sollte, ist eine *Gegenüberstellung* der Wissensbasis von LILOG mit den Wissensbasen vergleichbarer Projekte. Ein erster Rundblick in der KI-Landschaft ergibt folgendes Bild:

Im deutschen Sprachraum ist derzeit keine vergleichbar große implementierte Wissensbasis für Zwecke der Textverarbeitung oder ähnliche Zielsetzungen auszumachen, aber das kann sich ja rasch ändern. Im anglo-amerikanischen Bereich sind zumindest vier einschlägige Ansätze zu nennen, die für einen Systemvergleich in Frage kommen, aber doch wesentliche - hier nur in Stichworten angedeutete - Unterschiede zu LILOG aufweisen. So verfügt das CYC Project des MCC in Austin, Texas[3] zwar über eine deutlich größere Wissensbasis als LEU/2, jedoch wurde die geplante natürlichsprachliche Anbindung wegen mangelnder Eignung der Wissensbasis vorerst wieder verworfen [Jim Barnett, pers. Mit. vom Oktober 1991]. Das Penman-Upper-Model[3] ist insbesondere für Sprachgenerierung eingerichtet und unterstützt weder Belange der Analyse noch Inferenzprozesse - vgl. den Beitrag von Novak (in diesem Band). Die von K. Dahlgren und J. McDowell[4] erstellte Wissensbasis des Los Angeles Scientific Center der IBM erlaubt lediglich die flache Verarbeitung von relativ einfachen sprachlichen Konstruktionen - so können weder Adjektive analysiert noch Inferenzen gezogen werden. Das von J. Hobbs am SRI geleitete TACITUS Project [5] verdient Beachtung als Vergleichsprojekt vor allem wegen seiner Behandlung von "abductive inferences" durch ein pragmatisches Modul.

Wir verstehen diesen Band, der anhand der Wissensbasis von LEU/2 in die Tiefe zu gehen versucht, zugleich als Aufforderung, unter den hier entwickelten Prämissen in die Breite des verfügbaren Angebots an alternativen Konzeptionen zu gehen. Erfahrung zahlt sich aus erst im Vergleich.

Abschließend möchten wir uns bei den Autoren für ihre Mitarbeit, bei den Gutachtern für wertvolle Hinweise und bei den Herausgebern der Informatik-Fachberichte für die Aufnahme des Bandes in diese Reihe herzlich bedanken.

November 1991 Gudrun Klose, Ewald Lang, Thomas Pirlein

---

3 Guha, R. V. & Lenat, D. B.: CYC: A Midterm Report. *AI Magazine* 11, 3.33-59, Fall 1990

3 Bateman, J.A.: Upper Modelling: Organizing Knowledge for Natural Language Processing, in: Proceedings of the 5th International Workshop on Language Generation, Pittsburgh, PA, 1990, 54-61

4 Dahlgren, K. & McDowell, J. & Stabler, E. : Knowledge Representation for Common Sense Reasoning with Text. *Computational Linguistics* 15, 3.149-170, 1989

5 Hobbs, J. R., Croft, W. *et al.* : The TACITUS Commonsense Knowledge Base, Technical Report, SRI, Menlo Park, CA 1988

Inhalt

Vorwort

I. Aspekte der Wissensmodellierung in natürlichsprachlichen Systemen
Gudrun Klose, Thomas Pirlein
Die Modellierung der LEU/2-Wissensbasis im Überblick 3
Ewald Lang
Linguistische vs. konzeptuelle Aspekte der LILOG-Ontologie 23
Geoff Simmons
Kommentar zu E. Lang 46
Kai von Luck
Wissensrepräsentation, naive Theorien und sprachverstehende Systeme 55

II. Wissensmodellierung und sprachliche Oberfläche
Claudia Maienborn
Motivation einer Ontologie aus linguistischer Perspektive – 3 Fallstudien aus LEU/2 73
Bart Geurts
Linguistic Knowledge and Knowledge of the World:
Comments on C. Maienborn's Paper 91
Peter Gerstl
Kommentar zu C. Maienborn 95
Ewald Lang
Kommentar zu den Beiträgen von C. Maienborn und B. Geurts 100
Hans-Joachim Novak
Die LILOG-Ontologie aus Generierungssicht 106
Zuzana Krifka Dobes
Linking Linguistic and Extra-Linguistic Information in Generation 113

III. Wissensmodellierung und Inferenzen
Geoff Simmons
Empirical Methods for "Ontological Engineering" – Case Study: Objects 125
Toni Bollinger
Verarbeitung von L_{LILOG} – Theorie und Praxis oder
Was kann man noch besser machen bei der Verarbeitung von L_{LILOG}? 155
Ralf Röhrig, Elisabeth Neugebauer
Kommentar zu T. Bollinger 167
Sven Lorenz
Erfahrungen mit der Zeitontologie und -axiomatik 171
Barbara Gängler, Ipke Wachsmuth
Antwortgenerierung, flexible Wortwahl und elaborative Inferenzen –
ein Regelinventar für LEU/2 179
Thomas Pirlein
Kommentar zu B. Gängler und I. Wachsmuth 196

IV. Entwicklung und Verwaltung der Wissensbasis
Peter Gerstl
Praxisgerechte Verwaltung der Ontologie 201
Gudrun Klose, Martin Mezger, Adrian Müller
Kommentar zu P. Gerstl 213
Manfred Börkel
Vom ersten LILOG-Prototypen zu LEU/2: Zur Entwicklung der Wissensbasen 217

Anhang
Struktur der Wissensbasis in LEU/2 233
Dokumentation der Texte 245
Syntax von L_{LILOG} 249

Anschriften der Autoren 255