

10 stu,

DOKUMENTATION/INFORMATION

Schriftenreihe des Instituts für Informationswissenschaft, Erfindungswesen und
Recht der Technischen Hochschule Ilmenau

Informationslogische Sprache MIZK

Heft 33

**IX. Kolloquium über Information und Dokumentation
vom 12. bis 14. November 1975**

Themenkreis 4b
Urheberrecht
und Sondervorträge zur Informationswissenschaft

Ilmenau 1977

ANDRZEJ TRYBULEC

Informationslogische Sprache MIZAR

I. Informationslogische Sprachen

Als Informationslogische Sprachen bezeichnen wir faktografische Sprachen, mit deren Hilfe wir sowohl Informationen aus dem gegebenen Gebiet als auch die zwischen ihnen auftretenden logischen Verbindungen darstellen können. Informationslogische Sprache kürzen wir mit IL Spr. ab.

II. Anwendung der IL Spr.

Informationslogische oder angenäherte Sprachen können in folgenden Gebieten Anwendung finden:

1. in informationslogischen Systemen ,d. h. in Systemen, die zur Wiederauffindung und Bearbeitung von Informationen logische Beziehungen nutzen, als Eingangssprachen;
2. bei der Maschinenübersetzung, insbesondere bei der Übersetzung wissenschaftlicher Literatur, als vermittelnde Sprachen;
3. bei der Automatisierung des Redigierens wissenschaftlicher Artikel; die Eingangssprache ist eine IL Spr., die Ausgangssprache ist die natürliche Sprache. Infolge von Schwierigkeiten bei der Erzielung effektiver Systeme zur syntaktischen Analyse der natürlichen Sprachen kann die Übersetzung aus einer IL Spr. in einige natürliche Sprachen bei der Überwindung der Sprachbarriere in der wissenschaftlichen Information hilfreich sein;
4. in Systemen zur Kontrolle der Richtigkeit von Entscheidungen, als Eingangssprache. Bei diesen Anwendungen muß der Translator IL Spr. die Möglichkeit besitzen, die Richtigkeit des Vorgehens zu kontrollieren und auch mit der Informationsbank zusammenzuarbeiten, die den Kompetenzbereich, die empirischen Angaben u. ä. in Abhängigkeit von dem Gebiet umfaßt, in dem das System angewandt wird. Als typische Anwendungsgebiete kann man erwähnen:
 - Administration,
 - Medizin (Diagnostik und Auswahl der Therapie),
 - Gerichtswesen,
 - Wissenschaft (vor allem Mathematik — in diesem Falle wäre folgende Benennung zutreffend: System zur Kontrolle der Richtigkeit des denkmäßigen Vorgehens);

5. in der Didaktik, vor allem beim programmierten Lernen. Für die Projektanten der IL Spr können diese Anwendungen besonders attraktiv sein, da sie die Einführung relativ kleiner Fragmente der Sprache ermöglichen und zwar mit einem Nutzen für die Nutzer, während andere Anwendungen die Einführung großer Systeme erfordern.
6. in den Untersuchungen zur künstlichen Intelligenz.

Alle diese Anwendungsgebiete erfordern IL Spr. mit spezifischen Eigenschaften. So ist z. B. bei der Automatisierung des Redigierens die Unterbringung von Anmerkungen des Redakteurs unbedingt notwendig. In den IL Spr. für die Automatisierung der Lehre muß man besonders auf die genaue Fehlermeldung achten und bei dem programmierten Lernen auf die genaue Diagnostik der Fehlerursachen. Davon unabhängig erscheint der Aufbau von universellen IL Spr. zweckmäßig, die nach einer eventuellen Ergänzung für alle oder zumindest für die Mehrheit der Anwendungszwecke geeignet sind. Das ermöglicht eine komplexe Lösung der Probleme, die mit der Nutzung der IL Spr. in Zusammenhang stehen.

Das anzustrebende Ziel ist z. B. die folgende Nutzung der IL Spr. für die Mathematik:

1. Der Mathematiker bereitet seinen Artikel in einer IL Spr. vor.
2. Er „setzt den Artikel in Bewegung“ in Zusammenarbeit mit dem Computer, der die synthetischen Fehler und einige semantische Fehler aufzeigt.
3. Das System kontrolliert die Richtigkeit des denkmäßigen Vorgehens; es bereitet das Dokument für den Redakteur vor, wenn Sprünge im Vorgehen auftreten, mit denen das System nicht zurechtkommt.
4. In Zusammenarbeit mit der Datenbank, die Gesetze, Definitionen und bibliographische Angaben enthält, überprüft der Computer, ob die erzielten Ergebnisse neu sind, ob die Verweise richtig sind usw.
5. Nach ihrer Akzeption werden die im Artikel enthaltenen Informationen automatisch in die Datenbank überführt.
6. Das System übersetzt den Artikel in die vorgegebenen natürlichen Sprachen und redigiert ihn (ordnet die bibliographischen Angaben, numeriert die Gesetze usw.) nach entsprechenden Hinweisen.
7. Nach Erlangung zusätzlicher Informationen (Publikationsort) übergibt das System Angaben, die den Artikel betreffen, an dokumentengrafische Systeme sowie an Systeme von Typ SCI.
8. Das System schafft Artikel auf maschinellen lesbaren Trägern, die an die Druckerei weitergeleitet werden können.

Obwohl sich die vorgestellte Vision an der Grenze zur science fiction befindet, zeigt sie nach Meinung des Autors doch auf die richtige Forschungsrichtung.

III. Prinzipien zum Aufbau von IL Spr

Man kann allgemeine Anforderungen aufstellen, die durch die IL Spr erfüllt werden sollten:

1. Die IL Spr sollte das klassische logische Rechnen beinhalten, was an sich aus der Definition hervorgeht. Die Auswahl der Regeln für Schlußfolgerungen muß ausreichend breit sein, um ein freies denkmäßiges Vorgehen zu ermöglichen. Man kann sich also nicht mit sparsamen Systemen zufrieden geben, die z. B. nur den „modus ponendo ponens“ nutzen. Es ist auch notwendig, Beweisführungen über Annahmen zuzulassen.
2. Es ist zweckmäßig, das die IL Spr die Arithmetik der ganzen Zahlen enthält und vielleicht auch die Arithmetik der vollen Zahlen.
3. Infolge der immer größeren Verbreitung der mengen-theoretischen Techniken wäre es zweckmäßig, wenn die IL Spr ein ausreichend reiches Fragment der Mengentheorie enthält.
4. Es müssen die Mittel vorgesehen sein, die ein freies Organisieren des denkmäßigen Vorgehens erlauben.

Diese Anforderungen betreffen praktisch beliebige IL Spr, unabhängig vom Anwendungsgebiet. Ein attraktives Ziel ist der Aufbau einer universellen IL Spr, die die Logik und eventuell ausgewählte Gebiete der Mathematik enthält. Darüber hinaus müßten Varianten für verschiedene Anwendungsgebiete (Mathematik, Medizin, Recht) geschaffen werden, indem eine entsprechende Auswahl von spezifischen Termini getroffen wird, ähnlich, wie man verschiedene Varianten von Programmiersprachen schaffen kann, indem die universelle Sprache durch verschiedene Gruppen von spezifischen Prozeduren ergänzt wird. Man muß allerdings darauf hinweisen, daß man es bei Programmiersprachen mit einigen Dutzend spezifischen Prozeduren zu tun hat, während bei IL Spr einige Zehntausend spezifische Termini auftreten.

Außer den semantischen Anforderungen kann man auch auf bestimmte syntaktische Anforderungen weisen:

1. Die Sprache muß genügend reich sein, um das Schreiben von Texten in IL Spr zu ermöglichen, die hinsichtlich ihrer syntaktischen und logischen Struktur nur unbeträchtlich von Texten in natürlicher Sprache abweichen.
2. Gleichzeitig darf die Grammatik nicht zu kompliziert sein, da dies das Erlernen der Sprache erschweren würde.
3. Die ausgewählten Bezeichnungen und die Syntax sollten ähnlich wie in Programmiersprachen sein. Vor allem sollte die Sprache nach Möglichkeit nicht den Rahmen des Codes ISO-7 überschreiten.

IV. Die Sprache MIZAR

Ein erstes Projekt der Sprache MIZAR wurde von dem Autor im Institut für Bibliothekswissenschaft und wissenschaftliche Information der Warschauer Universität erarbeitet. Die weiterführenden Arbeiten zur Sprache MIZAR werden bei der Sektion für angewandte Mathematik der Plocker Wissenschaftlichen Gesellschaft geleistet.

Der Name MIZAR ist die Abkürzung für „mathematisch-informationswissenschaftlicher Apparat für das automatische Redigieren“, da anfänglich die Sprache MIZAR für das System des automatischen Redigierens vorgesehen war.

Gegenwärtig erscheint es auf Grund der schnellen Einführung am wichtigsten, die Sprache MIZAR für das programmierte Lernen einzusetzen. Daraus ergibt sich die Einstellung zur schnellen Inbetriebnahme seiner Fragmente. Da das Projekt noch nicht bis zu Ende bearbeitet ist (Juni 75), stelle ich hier nur ein Fragment vor, das die Rechnung mit Sätzen beinhaltet. Weiterführende Informationen über MIZAR kann bei folgender Adresse erhalten werden:

Plocker Wissenschaftliche Gesellschaft
 Sektion Angewandte Mathematik
 09 — 400 PLOCK (VR Polen)
 Pl. Narutowicza 8

Als Sprachmuster (bei der Auswahl der Bezeichnungen, Grammatik usw.) diene bei der Erarbeitung von MIZAR die Sprache Algol 60. Ähnlich wie in Algol wird eine endliche Menge von Abkürzungen verwandt, die sich dadurch unterscheiden, daß sie unterstrichen bzw. dicker gedruckt werden. Die Abkürzungen müssen aus der Sicht von MIZAR als nicht zerlegbar aufgefaßt werden. Es gibt folgende Abkürzungen:

- | | | |
|------------------|---|--|
| begin | } | — Klammern, ähnlich wie in Algol |
| end | | |
| proof | | — in bestimmten Kontexten vertritt sie begin |
| then | } | — wird für die Beschreibung der Satzbeurteilung genutzt |
| by | | |
| thus | | — steht vor Schlußfolgerungen des denkmäßigen Vorgehens |
| hence | | — wird anstelle von then thus benutzt |
| ref | | — bedeutet einen Verweis zur Literatur oder zur Datenbank |
| let | | — beginnt die Annahme und wird in den Konstruktionen let ... denote ... verwendet, die neue Bezeichnungen einführen, sowie in Konstruktionen let ... be ... , die neue Objekte einführen |
| denote | } | — bereits beschrieben |
| be | | |
| is | } | — wird in Primärsätzen verwendet |
| of | | |
| in | | |
| for | | — als Abkürzung von is element of , sowie being element of |
| for | | — allgemeiner Quantifikator |
| ex | | — Existentialquantifikator |
| (exist) | | |
| being | } | — wird in Quantifikatoren mit begrenztem Bereich verwendet |
| such that | | |
| def | | — beginnt eine Definition |

Vielleicht werden auch noch bestimmte andere Änderungen eingeführt.

Als Veränderliche u. ä. werden Identifikatoren genutzt, deren Definition die gleiche ist wie in Algol, d. h. der Identifikator ist eine Summe von Buchstaben und Ziffern, die mit einem Buchstaben beginnt.

Ein richtiger Text in der Sprache MIZAR bildet einen „Block“, der mit **begin** anfängt und mit **end** aufhört; er setzt sich aus einer bestimmten Anzahl von Sätzen zusammen, die durch Semikolon getrennt sind. Es gibt einige Satztypen; der wichtigste Typ sind Gesetze (statement). Außerdem noch Verweise (reference), Definitionen (definition) usw. Im Besonderen kann ein Satz ein Block sein.

V. MIZAR — PC

Ausführlich beschreibt der Autor ein Fragment, dessen Aufbau abgeschlossen ist: MIZAR — PC, das die Rechnung mit Sätzen beinhaltet. Es besitzt folgende Grammatik:

* (block) :: = **begin** {block tail}
 {assumption} :: = **let** {simple sentence}
 {block tail} :: = {sentence} **end** | {sentence} ; {block tail}
 {compound segment} :: = {conclusion} |
 {assumption} ; {compound segment} |
 {statement} ; {compound segment}
 {compound statement} :: = {label} : {compound
 statement} |
 {compound tail} :: = {compound segment} **end** |
 {compound segment} ; {compound tail}
 {conclusion} :: = {unlinked conclusion} |
 {linked conclusion}
 {conjunction} :: = {secondary sentence} |
 {conjunction} \wedge {secondary sentence}
 * {designator} :: = {identifier}
 {designator list} :: = {designator} |
 {designator} , {designator list}
 {disjunction} :: = {conjunction} |
 {disjunction} \wedge {conjunction}
 {implication} :: = {disjunction} |
 {implication} \supset {disjunction}
 * {justification} :: = {designator list} |
 {reference}
 {label} :: = {identifier}
 {linked conclusion} :: = **hence** {unlinked sentence}
 {linked sentence} :: = **then** {unlinked sentence}
 {negative sentence} :: = {positive sentence}
 * {positive sentence} :: = {identifier} |
 ({unlabelled simple})
 {proof} :: = {simple sentence} **proof**
 {compound tail}
 {reference} :: = {label} : {reference} |
 {unlabelled reference}
 * {reference identifier} :: = {identifier}
 {secondary sentence} :: = {positive sentence} |
 {negative sentence}

* <sentence>	:: =	<statement> <reference>
<simple sentence>	:: =	<unlabelled simple>
<label>	:	<simple sentence>
<simple statement>	:: =	<unlinked sentence>
<linked sentence>		
<statement>	:: =	<simple statement>
<compound statement> <proof>		
<unlabelled compound>	:: =	x begin <compound tail>
<unlinked conclusion>	:: =	thus <unlinked sentence>
<unlabelled reference>	:: =	ref <reference identifier>
<unlinked sentence>	:: =	<simple sentence>
<simple sentence>		by <justification>

Die durch * bezeichneten Definitionen sind in der Sprache MIZAR erweitert.

Wie man leicht sieht, ist der Satzbau dem Aufbau der Booleschen Ausdrücke im Algol ähnlich. MIZAR — PC benutzt nur einen Teil der erwähnten Abkürzungen, nämlich; **begin**, **proof**, **by**, **ref**, **then**, **thus**, **hence**, **end**.

Es folgen die zugelassenen Regeln für das Schlußfolgern und die Schreibweise ihrer Nutzung:

1. Einklinkung (linkage)

Die Abkürzungen **then** und **hence** wurden nur zur Vergrößerung der Lesbarkeit eingeführt; man kann sie stets mit Hilfe von **thus** und **by** eliminieren. Die Konstruktionen vom Typ:

x; **then** y
 x; **then** y **by** A
 x; **hence** y
 x; **hence** y **by** A

kann man immer ersetzen durch:

B: x; y **by** B
 B: x; y **by** A, B
 B: x; **thus** y **by** B
 B: x; **thus** y **by** A, B

2. Beweis proof

Die Konstruktion:

<simple sentence> **proof** <compound tail> bedeutet dasselbe wie:
begin <compound tail>

und kann durch diese ersetzt werden, wenn sie keine Schlußfolgerung darstellt.

3. Begründung (justification)

Die Konstruktion:

x **by** A, B (x **by** A),

bedeutet, daß man x aus Sätzen erhalten kann, die durch die Etiketten A und B bzw. A bezeichnet werden, und zwar mit Hilfe der zugelassenen Regel zum Schlußfolgern. Beispiele:

$$A : p \quad q; \dots; B : p; \dots; q \text{ by } A, B \\ p \wedge q; \text{ then } p$$

Der Apparat der Regeln zum Schlußfolgern wurde nicht endgültig festgelegt; zum Zwecke des programmierten Lernens kann er verändert werden. Es wurden jedoch folgende Begrenzungen angenommen:

1. Die Menge der Annahmen darf nicht größer als 2 sein (die Menge der Annahmen kann jedoch leer sein, d. h. Axiom – Schemata sind zugelassen)
2. Die Menge der auftretenden Veränderlichen darf nicht größer als 3 sein (also folgende Regel ist nicht zugelassen:

$$p \supset q, p \wedge \neg q \vee s$$

3. In der Schlußfolgerung und in den Annahmen darf höchstens ein 2-Argumente-Funktor verwendet werden:

4. Zusammengesetzte Gesetze

Der Inhalt eines zusammengesetzten Gesetzes hängt allein von den in ihm auftretenden Annahmen (assumption) und Schlußfolgerungen (conclusion) ab. Formal kann man das Erhalten eines einfachen Satzes mit dem gleichen Inhalt wie folgt beschreiben:

Nach den Annahmen stellen wir eine Implikation und nach den Schlußfolgerungen (mit Ausnahme der letzten) eine Konjunktion; die Klammern gruppieren wir von rechts.

Der Inhalt des Satzes

begin let a; ...; thus b; ...; let c, ... thus end

ist der Inhalt des Satzes

$$a \supset (b \wedge (c \supset d)).$$

Ein zusammengesetzter Satz kann in der Begründung genauso wie ein einfacher Satz mit gleichem Inhalt genutzt werden.

Im Beweis mit der Form:

(simple sentence) **proof** (compound tail) muß der Inhalt

(simple sentence) der gleiche sein wie der Inhalt des Gesetzes **begin** (compound tail). Man sollte darauf achten, daß man diese Anforderung nicht mit Hilfe einer kontextlosen Grammatik beschreiben kann.

VI. Realisierung auf der EDVA Odra – 1204

Gegenwärtig werden Arbeiten zur Programmierung von MIZAR – PC und weiteren Teilen der MIZAR auf der EDVA Odra – 1204 durchgeführt. Die Arbeiten werden durch die Plocker Wissenschaftliche Gesellschaft finanziert.

Bis Juni 1975 wurde durch Rudnicki, Matuszewski und den Autor die syntaktische Analyse realisiert.

Die Meldung der syntaktischen Fehler hat den Translator Algol 1204 zum Vorbild.

Beispiele:

integer UNEXPECTED (Gebrauch einer nicht in der Sprache MIZAR auftretenden Abkürzung)

INCOMPLETE end end (unvollständiger Text, zur Fehlerbeseitigung muß **end end** addiert werden)

begin UNSIM then (wenn der Text folgendermaßen begann:

INCORRECT begin p v \neg p then ...)

Bei Fehlern dieses Typs werden leicht zu entschlüsselnde Abkürzungen verwendet: UNSIM — (unlabelled simple), COMTA — (compound tail) u. ä.

Literatur

K. Jerzykiewicz, J. Szczepkowicz — Algol 1204, PWN, Warszawa 1973

Übersetzt von Reinhard Schramm, TH Ilmenau, Institut INER

In der Schriftenreihe „DOKUMENTATION/INFORMATION“ erschienen bisher:

Heft 1 bis 8 siehe an derselben Stelle in den Heften 9 bis 25

- Heft 9 V. Kolloquium des Instituts für Dokumentation, Patentwesen und Recht der TH Ilmenau vom 15. bis 17. November 1967, Themenkreis 1: Dokumentation und Forschungsplanung
- Heft 10 V. Kolloquium des Instituts für Dokumentation, Patentwesen und Recht der TH Ilmenau vom 15. bis 17. November 1967, Themenkreis 2: Aktuelle und perspektivische Aufgaben der Dokumentation
- Heft 11 V. Kolloquium des Instituts für Dokumentation, Patentwesen und Recht der TH Ilmenau vom 15. bis 17. November 1967, Themenkreis 3: Information und Schutzrechtspolitik
- Heft 12 Ansprachen und Vorträge anlässlich der Immatrikulation der ersten Studierenden der Fachrichtung Informations- und Dokumentationswissenschaft an der TH Ilmenau am 31. Oktober 1968
- Heft 13 VI. Kolloquium des Instituts für Informationswissenschaft, Erfindungswesen und Recht der TH Ilmenau vom 19. bis 21. November 1969, Themenkreis 1: Gegenstand und besondere Aspekte der Informations- und Dokumentationswissenschaft
- Heft 14 VI. Kolloquium des Instituts für Informationswissenschaft, Erfindungswesen und Recht der TH Ilmenau vom 19. bis 21. November 1969, Themenkreis 2: Praktische Fragen der Information und Dokumentation
- Heft 15 VI. Kolloquium des Instituts für Informationswissenschaft, Erfindungswesen und Recht der TH Ilmenau vom 19. bis 21. November 1969, Themenkreis 3: Erfindungs- und Patentwesen
- Heft 16 Informationsartikel über Arbeiten von Studierenden der 1. Matrikel des postgradualen Fernstudiums der Informations- und Dokumentationswissenschaft am Institut INER der TH Ilmenau
- Heft 17 VII. Kolloquium über Information und Dokumentation vom 17. bis 19. November 1971, Themenkreis 1: Ausbildung in der Informations- und Dokumentationswissenschaft
- Heft 18 VII. Kolloquium über Information und Dokumentation vom 17. bis 19. November 1971, Themenkreis 2: Schutz wissenschaftlicher Erkenntnisse
- Heft 19 VII. Kolloquium über Information und Dokumentation vom 17. bis 19. November 1971, Themenkreis 3: Theorie und Praxis der Information und Dokumentation (1. Teil)
- Heft 20 VII. Kolloquium über Information und Dokumentation vom 17. bis 19. November 1971, Themenkreis 3: Theorie und Praxis der Information und Dokumentation (2. Teil)

Fortsetzung Seite 3 des Umschlages