

INHALT

| | |
|--|----|
| 0. Einleitung | 9 |
| 1. Kommunikationstechnologie und Sprache | 13 |
| 1.1 Die Rolle der natürlichen Sprache in der vermittelten Kommunikation | 13 |
| 1.2 Darstellungsformen von Sprache und Text | 17 |
| 1.3 Informationserschließung | 19 |
| 1.4 Maschinelle Sprachübersetzung (MÜ) | 22 |
| 1.4.1 Die heutige Bedeutung der MÜ | 22 |
| 1.4.2 Hauptprobleme der MÜ | 22 |
| 1.4.3 Ablauf maschineller Übersetzungsprozesse | 23 |
| 1.5 Natürlichsprachliche Kommunikation mit Computern | 28 |
| 2. Linguistische Grundlagen | 31 |
| 2.1 Linguistik – eine empirische Wissenschaft | 31 |
| 2.1.1 Der Korpusgedanke | 31 |
| 2.1.2 Prozeduren | 33 |
| 2.2 LDV als explizite Beschreibung von Sprache | 35 |
| 2.2.1 Das Problem der sprachlichen Einheit | 36 |
| 2.2.2 Strukturbeschreibung | 43 |
| 2.3 Sprache und Wissen | 47 |
| 3. Verarbeitungstechniken | 51 |
| 3.1 Algorithmen | 51 |
| 3.1.1 Begriffsbestimmung | 51 |
| 3.1.2 Darstellungsformen von Algorithmen | 52 |
| 3.2 Sprachliche Zeichen und Codes | 57 |
| 3.2.1 Erkennen von Sprache allgemein | 57 |
| 3.2.2 Erkennen von Sprache durch Computer | 58 |
| 3.2.3 Schriftsysteme | 59 |
| 3.2.4 Binärcodes | 59 |
| 3.2.5 Optische Datenerfassung durch Computer | 62 |
| 3.3 Sortieren | 63 |
| 3.4 Suchen und Abbilden | 64 |

| | | |
|---------|--|-----|
| 3.5 | Isolierung von sprachlichen Einheiten | 66 |
| 3.5.1 | Ebene der Grapheme | 66 |
| 3.5.2 | Ebene der Morpheme | 67 |
| 3.5.3 | Ebene der Sätze und Satzteile | 70 |
| 3.5.4 | Ebene der Silben | 71 |
| 3.6 | Darstellung von Information: Listen, Bäume, Netze | 73 |
| 3.6.1 | Einfache Listen | 73 |
| 3.6.2 | Strukturierte Listen und Bäume | 74 |
| 3.6.3 | Netze | 76 |
| 3.6.4 | Darstellung und Verarbeitung von Listen | 78 |
| 4. | Typische Algorithmen | 81 |
| 4.1 | Texterfassung und -segmentierung | 81 |
| 4.1.1 | Zum Begriff der Kodierung | 81 |
| 4.1.2 | Kodierungssysteme | 82 |
| 4.1.3 | Verarbeitung von Kodierungen | 85 |
| 4.1.4 | Segmentierung | 86 |
| 4.2 | Register, Index, Konkordanz | 88 |
| 4.2.1 | Wortformenregister | 90 |
| 4.2.2 | Index und Konkordanz | 90 |
| 4.2.3 | Problematik ausschließlich wortformenbasierter Register und Indizes | 93 |
| 4.3 | Morphologische Analyse | 94 |
| 4.3.1 | Kombination und Dekombination | 95 |
| 4.3.2 | Lemmatisierung | 97 |
| 4.3.2.1 | Voraussetzungen | 97 |
| 4.3.2.2 | Verfahren | 98 |
| 4.4 | Deskriptive Grammatiken | 100 |
| 4.4.1 | Allgemeine Grundlagen | 100 |
| 4.4.2 | Fragment einer deskriptiven Grammatik | 101 |
| 4.4.3 | Nutzen deskriptiver Grammatiken | 105 |
| 4.5 | Parsing-Techniken | 106 |
| 4.5.1 | Begriffsbestimmung und allgemeine Voraussetzungen | 106 |
| 4.5.2 | Parser und Parsing-Strategien | 107 |
| 5. | Praktische Umsetzung von Algorithmen in Programme | 113 |

| | | |
|---------|--|-----|
| 5.1 | Möglichkeiten der Programmiersprache PL/1 | 113 |
| 5.2 | Zeichenkettenverarbeitung | 115 |
| 5.2.1 | Datenein- und -ausgabe | 115 |
| 5.2.1.1 | Zeichenorientierter Modus | 115 |
| 5.2.1.2 | Satzorientierter Modus | 117 |
| 5.2.2 | Druckgestaltung | 119 |
| 5.2.3 | Manipulation von Zeichenketten | 123 |
| 5.2.3.1 | Von PL/1 bereitgestellte Möglichkeiten | 123 |
| 5.2.3.2 | Erweiterungsmöglichkeiten | 128 |
| 5.3 | Verarbeitung von Texten | 130 |
| 5.3.1 | Segmentierung | 131 |
| 5.3.2 | Kodierung und Umkodierung | 134 |
| 5.3.3 | Register | 138 |
| 5.3.4 | Lexikonvergleich | 142 |
| 5.3.5 | Index und Konkordanz | 146 |
| 5.3.5.1 | KWIC | 146 |
| 5.3.5.2 | Wortstellenindex | 150 |
| 5.3.6 | Suche nach Zeichenketten | 152 |
| 5.3.7 | Statistische Untersuchungen | 156 |
| 5.4 | Textanalyse | 162 |
| 5.4.1 | Morphologische Analysen | 163 |
| 5.4.1.1 | Verbformenbestimmung | 163 |
| 5.4.1.2 | Substantivsegmentierung | 168 |
| 5.4.1.3 | Flexionsformengenerierung | 173 |
| 5.4.2 | Parser | 176 |
| 5.5 | Erläuterungen zu den Aufgaben aus 5.2 – 5.4 | 180 |
| 5.6 | Unterschiede beim Einsatz von Microcomputer-PL/1 | 182 |
| | Literaturverzeichnis | 189 |
| | Sachregister | 197 |