Vorlesung CL

Endliche Automaten und ihre Verwendung in der morphologischen Verarbeitung

Hans Uszkoreit

WS 00/01

Automaten

Automaten in der weiteren Bedeutung des Wortes sind ein zentrales Konzept aber nicht formal definiertes Konzept in der Informationsverarbeitung.

Wenn wir durch eine Sequenz von Handlungen bestimmte Effekte auslösen, dann ist das nicht unbedingt Informationsverarbeitung. Normalerweise wird mit einer jeden Handlung kausal eine Wirkung erzeugt.

Beispiel: ich öffne eine Tür mit zwei Schlössern, indem ich jedes Schloß mit einem Schlüssel aufschließe.

Wenn wir aber ein System haben, das so gebaut ist, daß es in Abhängigkeit von meinen Handlungen „Entscheidungen“ über seine Handlungen trifft, dann liegt Informationsverarbeitung vor. In diesem Fall bewirken meine Handlungen die Handlungen der Maschine nicht direkt kausal.

Automaten sind Systeme, die in Abhängigkeit von meinen Handlungen, bestimmte Aktionen ausführt bzw auslöst.

Damit wir von Automaten sprechen, muß es mindestens einen Entscheidungspunkt geben.

Automaten

Automaten werden verwendet, um Symbol- oder Handlungsabfolgen zu überprüfen, zu analysieren oder abzuarbeiten.

Sie werden z.B. bei der Verarbeitung von Benutzereingaben verwendet.

Dabei wird geprüft,

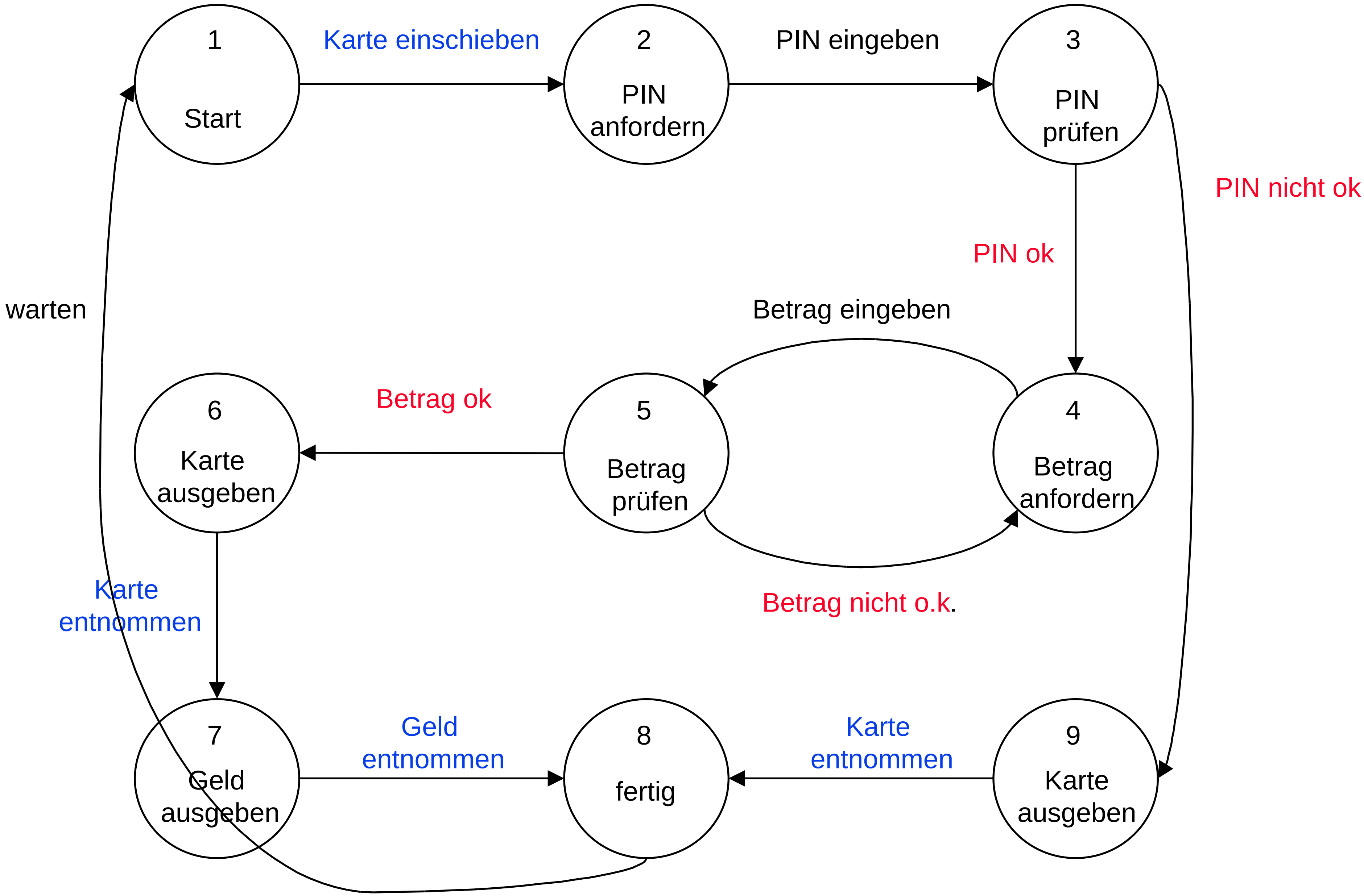
- ob Eingaben korrekt sind und
- in der richtigen Reihenfolge bzw. im richtigen Kontext erfolgen.

Automaten spielen auch eine Rolle bei der Überprüfung bzw. Übersetzung von Programmier- oder Spezifikationssprachen in die internen Sprachen des Computers.

In der Sprachverarbeitung werden Automaten an vielen Stellen eingesetzt, so z.B. bei der morphologischen Analyse von Wörtern.

„Endlicher Geld-Automat“

Das Beispiel ist leicht abgewandelt von Prof. Dr. Reinhard Völler (Hamburg) übernommen.



Bestandteile eines Automaten:

- endliche Menge von Zuständen
- Startzustand, Endzustand (auch mehrere)
- Eingaben: Worte einer Sprache
- Ausgaben: Worte einer Sprache
- Regeln für die Auswirkungen einer Eingabe
- Wichtig: diese Regeln müssen vollständig sein, d.h. für jede Eingabe muß eine Reaktion des Automaten spezifiziert sein.
- Notfalls wird ein Fehlerzustand definiert.

Endliche Automaten

Ein endlicher Automat   ist ein mathematisches Modell eines Systems mit Ein- und Ausgaben.

Ein solches System befindet sich immer in einer aus einer endlichen Anzahl möglicher interner Konfigurationen.

Man sagt auch: das System befindet sich in einem Zustand .

Beispiele:

Ein Schaltkreis mit n Gattern befindet sich in einem von möglichen Zuständen.

Texteditoren oder lexikalische Analysatoren von Compilern kann man als endliche Automaten modellieren.

Auch ein Computer ist ein endlicher Automat. Allerdings ist dieses Modell wegen der großen Anzahl möglicher Zustände nicht besonders hilfreich.

Deterministischer endlicher Automat (DFSA)

Ein deterministischer endlicher Automat ist ein Fünftupel A= (Z, E, δ, z_0, F)

Z	Menge der Zustände
E	Menge der Eingabesymbole
$\delta: Z \times E \rightarrow Z$	Zustandsübergangsfunktion
z_0 Z	Anfangszustand
F Z	Menge der Endzustände

DFSA als Akzeptor

Für eine Eingabekette $w = e_1, e_2, \ldots, e_n$ soll überprüft werden, ob sie durch einen Automaten A akzeptiert wird

Wir definieren uns zwei Variablen, q für den gegenwärtigen Zustand und e für das gegenwärtige Eingabesymbol

Wir setzen $q = z_0$ und wiederholen dann für jedes Symbol der Eingabekette e_i

$$q := \delta(q, e_i)$$

$$e := e_{i+1}$$

Wenn $e = e_n$ und q F dann ist die Eingabekette durch A akzeptiert, ansonsten gilt sie als zurückgewiesen.

Beispiel 1

Namensliste:

Peter Müller

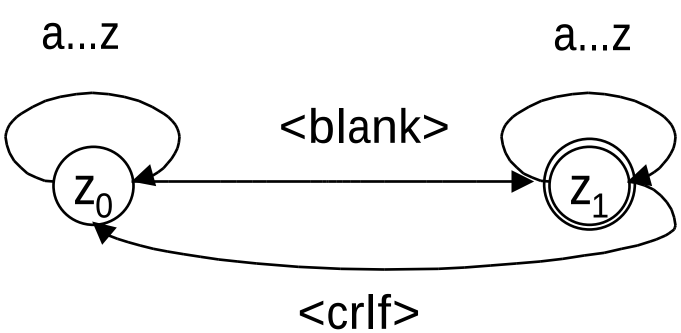
.

.

.

Doris Steckler

	z_0	z_1
a	z_0	z_1
.	z_0	z_1
.	z_0	z_1
.	z_0	z_1
z	z_0	z_1
<blank>	z_1	
<crlf>		z_0



	z_0	z_1
z_0	a...z	<blank>
z_1	<crlf>	a...z

Nichtdeterministischer endlicher Automat

Ein nichtdeterministischer endlicher Automat ist ein Fünftupel A= (Z, E, δ, z_0, F)

Z	Menge der Zustände
E	Menge der Eingabesymbole
δ: Z x E $\rightarrow 2^Z$	Zustandsübergangsfunktion
z_0 Z	Anfangszustand
F Z	Menge der Endzustände

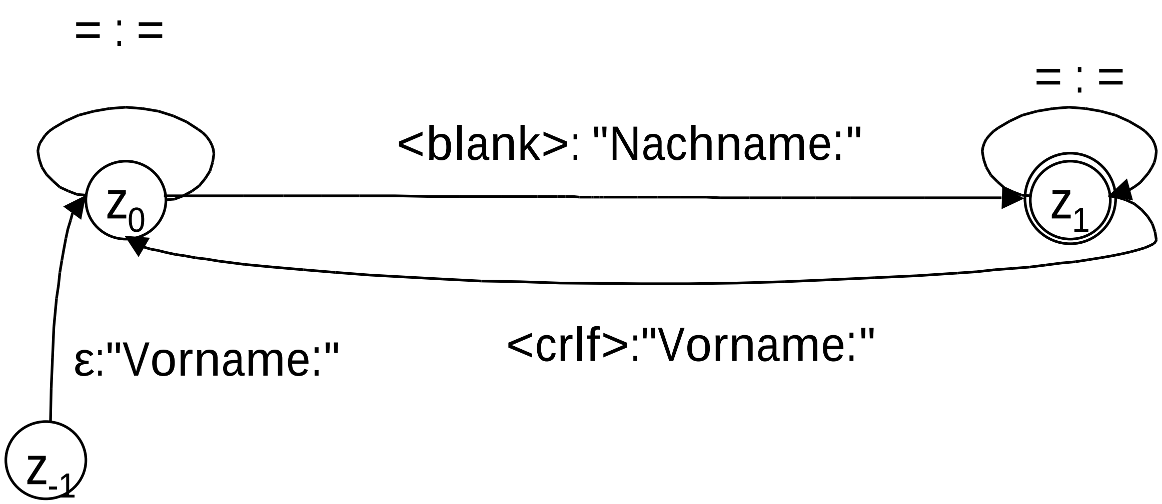
Automaten mit Ausgabe

Ein Mealey-Automat ist ein Sechstupel $A = (Z, E, A, \delta, z_0, \lambda)$

Z	Menge der Zustände
E	Menge der Eingabesymbole
A	Menge der Ausgabesymbole
$\delta: Z \times E \rightarrow 2^Z$	Zustandsübergangsfunktion
$\lambda: Z \times E \rightarrow A$	Ausgabefunktion
$z_0 \quad Z$	Anfangszustand

Beispiel 2

	z_{-1}	z_0	z_1
ε : **"Vorname:"**	z_0		
= : =		z_0	z_1
<blank>: "Nachname:"		z_1	
<crlf> : "Vorname:"			z_0



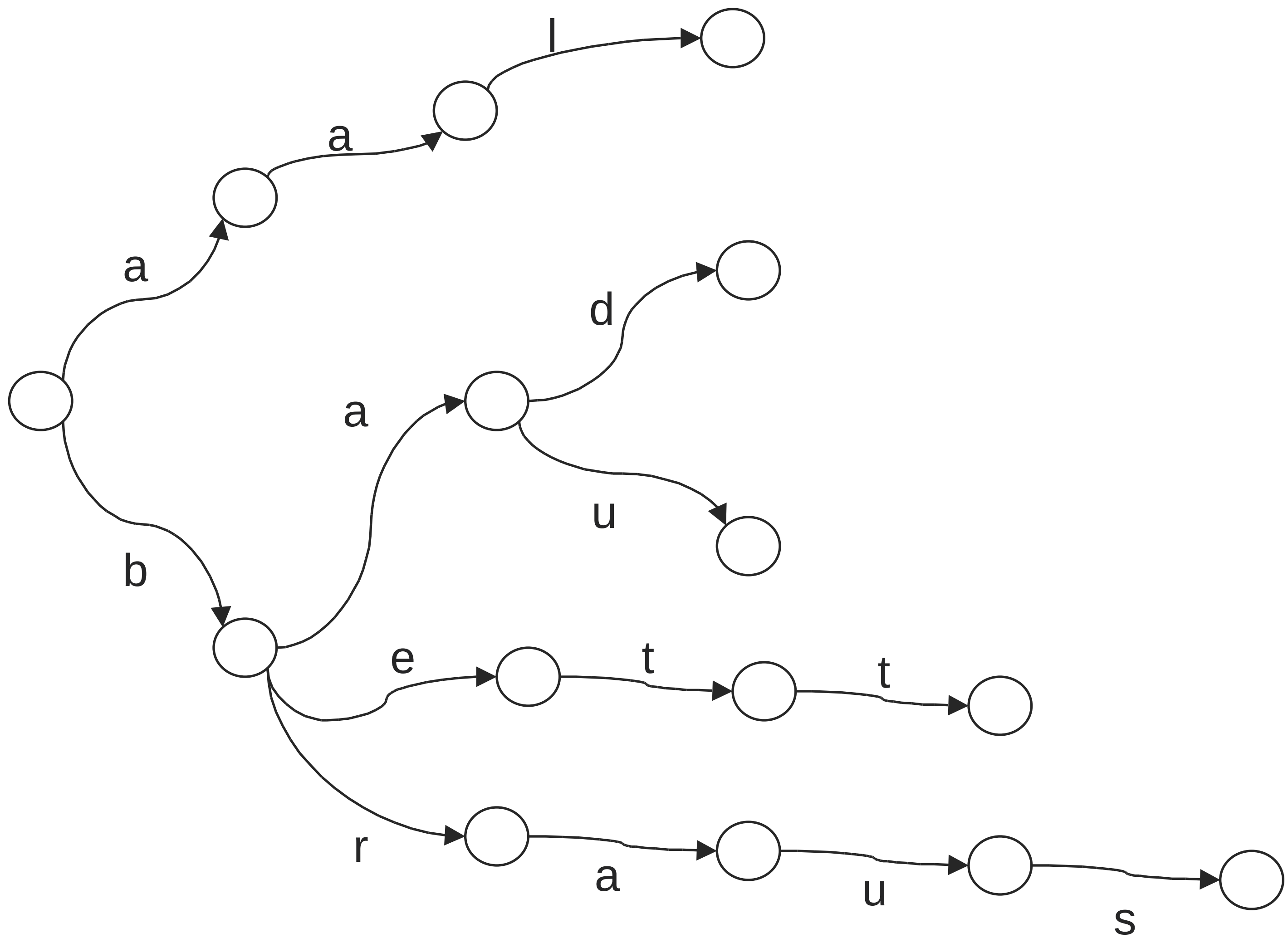
Beispiel 2

Vorname: Marion Nachname: Abbecker

Vorname: Klaus Nachname:Becker

Vorname: Günter Nachname:Bruck

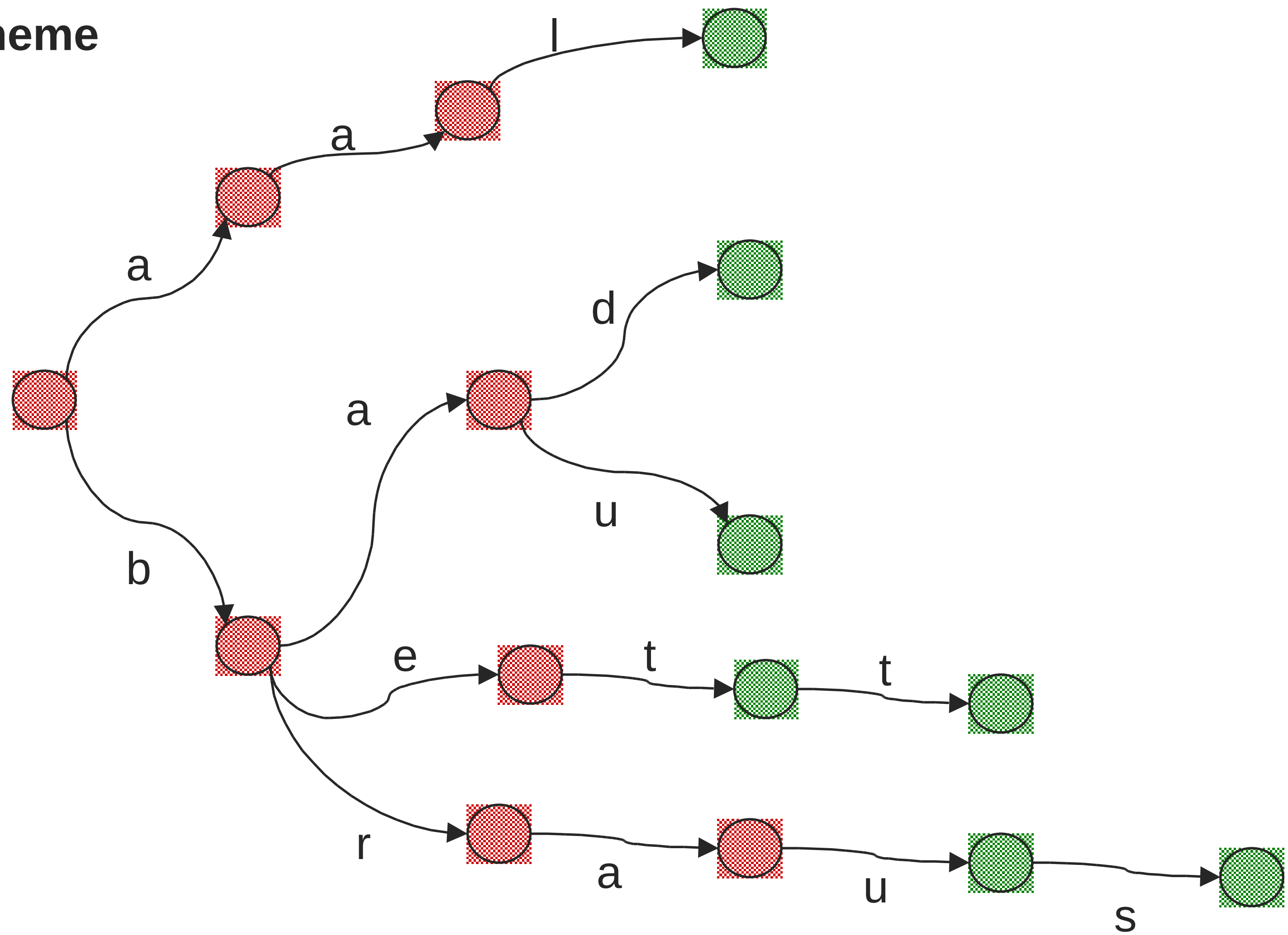
Buchstabenbaum



Buchstabenbaum

Verbstamm-Morpheme

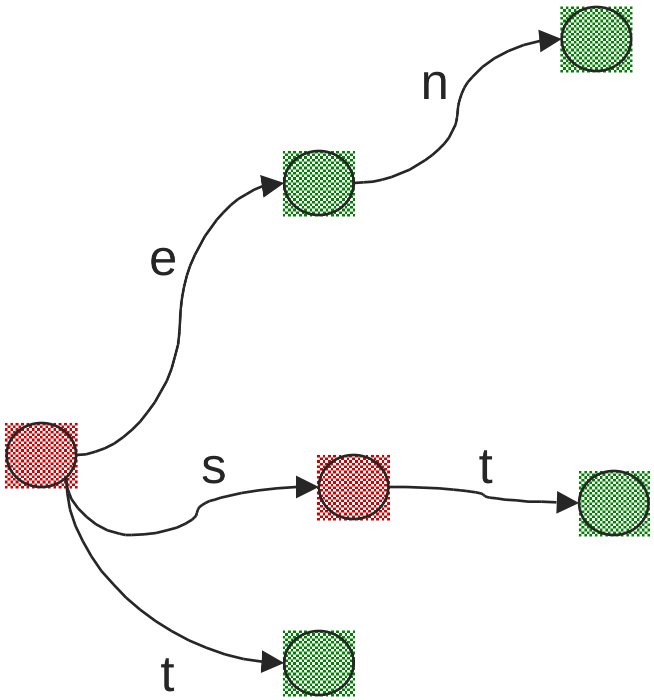
aal
bad
bau
bet
bett
brau
braus

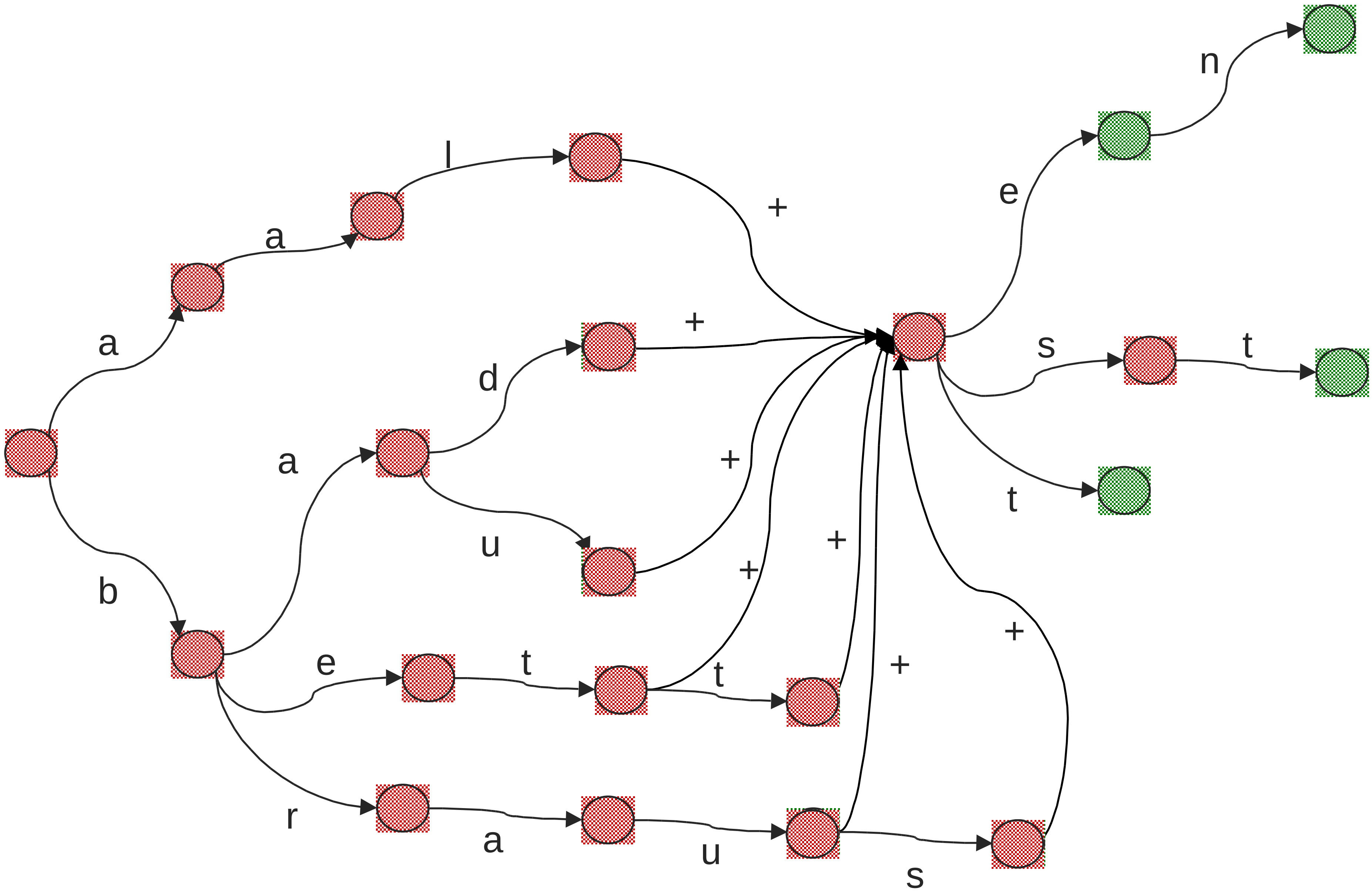


Buchstabenbaum

Verbsuffix-Morpheme

-e
-en
-st
-t



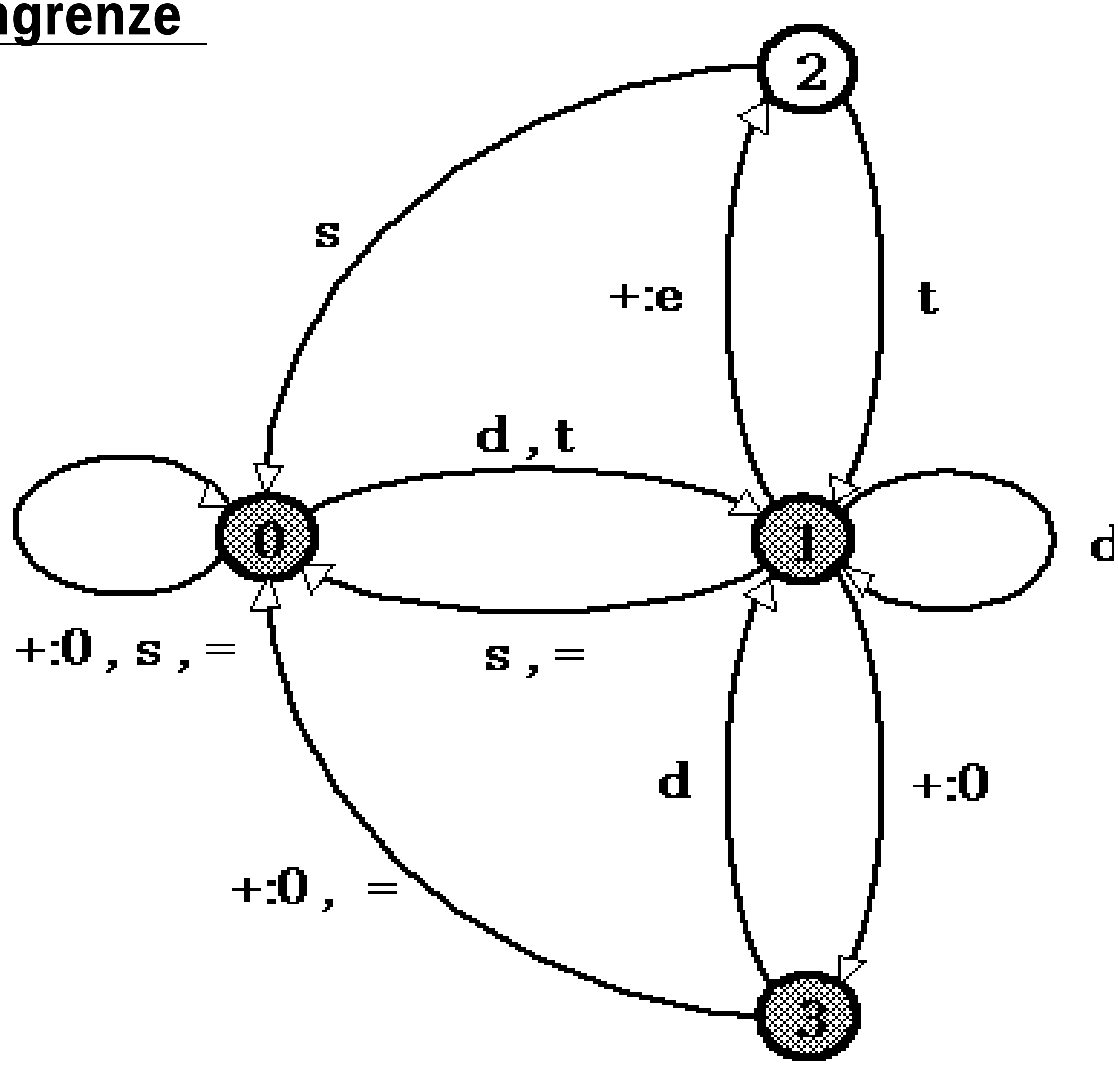


a
a
l
+
a
b
a
d
u
+
+
e
t
t
+
+
r
a
u
s
+
+
e
n
s
t
t

Zwei-Ebenen Morphologie

e-Epenthese an der Morphemgrenze (vereinfacht)

Regel: +:e <=> {d, t} _ {s, t}



Übergangstabelle:

	0	1	2	3
+:e	-	2	-	-
+:0	0	3	-	0
d:d	1	1	-	1
s:s	0	0	0	-
t:t	1	1	1	-
=:=	0	0	-	0

Abgewandeltes Beispiel von Prof. Dr. Harald Trost (Wien)