

Einsatz heuristischer Suchverfahren zur Erzeugung eines Akrostichons

Masterarbeit SS2012

Christof Bräutigam

`christof.braeutigam@uni-weimar.de`

Bauhaus-Universität Weimar

27. September 2012

Akrostichon

- ▶ Text, dessen Zeilenanfänge einen eigenen Sinn ergeben

Akrostichon

- ▶ Text, dessen Zeilenanfänge einen eigenen Sinn ergeben

For some time now I have lamented the fact that major issues are overlooked while many unnecessary bills come to me for consideration. Water reform, prison reform, and health care are major issues my Administration has brought to the table, but the Legislature just kicks the can down the alley.

Yet another legislative year has come and gone without the major reforms Californians overwhelmingly deserve. In light of this, and after careful consideration, I believe it is unnecessary to sign this measure at this time.

Aus einem Brief von A. Schwarzenegger (Okt. 2009)

Akrostichon

- ▶ Text, dessen Zeilenanfänge einen eigenen Sinn ergeben

For some time now I have lamented the fact that major issues are overlooked while many **u**nnecessary bills come to me for consideration. Water reform, prison reform, and health **c**are are major issues my Administration has brought to the table, but the Legislature just **k**icks the can down the alley.

Yet another legislative year has come and gone without the major reforms Californians **o**verwhelmingly deserve. In light of this, and after careful consideration, I believe it is **u**nnecessary to sign this measure at this time.

Aus einem Brief von A. Schwarzenegger (Okt. 2009)

Problemstellung

- ▶ gegeben einen Text t und ein Zielakrostichon a
- ▶ Erzeuge a in t

Problemstellung

- ▶ gegeben einen Text t und ein Zielakrostichon a
- ▶ Erzeuge a in t
- ▶ Beispiel: Erzeuge Akrostichon „Hello“ im folgenden Text
Hello world is an easy example program to learn a new \leftrightarrow
programming language before writing other, more \leftrightarrow
complex programs.

Problemstellung

- ▶ gegeben einen Text t und ein Zielakrostichon a
- ▶ Erzeuge a in t

- ▶ Beispiel: Erzeuge Akrostichon „Hello“ im folgenden Text

Hello world is an easy
example program to
learn a new programming
language before writing
other, more complex programs.

Problemstellung

- ▶ gegeben einen Text t und ein Zielakrostichon a
- ▶ Erzeuge a in t
- ▶ Beispiel: Erzeuge Akrostichon „Hello“ im folgenden Text

Hello world is an easy
example program to
learn a new programming
language before writing
other, more complex programs.

- ▶ Lösungsansatz
 1. Erzeuge Varianten des Originaltextes (Suchraum)
 2. Suche nach einer Variante mit Akrostichon

Problemlösung durch Suche

- ▶ „Finden oder Konstruieren eines Objekts mit bestimmten Eigenschaften.“ [Pearl 1984]
- ▶ Formelle Problembeschreibung
 - ▶ *Datenbasis*: Codierung von Objekten im Suchraum
 - ▶ *Operator*: Mechanismus zur Transformation von Objekten
 - ▶ *Steuerungsstrategie*: Reihenfolge der Transformationen
- ▶ Systematische Suche
 - ▶ *Vollständigkeit*: Alle Objekte im Suchraum betrachten
 - ▶ *Effizienz*: Jedes Objekt höchstens einmal betrachten

Problemlösung durch Suche

- ▶ „Finden oder Konstruieren eines Objekts mit bestimmten Eigenschaften.“ [Pearl 1984]
- ▶ Formelle Problembeschreibung
 - ▶ *Datenbasis*: Codierung von Objekten im Suchraum
 - ▶ *Operator*: Mechanismus zur Transformation von Objekten
 - ▶ *Steuerungsstrategie*: Reihenfolge der Transformationen
- ▶ Systematische Suche
 - ▶ *Vollständigkeit*: Alle Objekte im Suchraum betrachten
 - ▶ *Effizienz*: Jedes Objekt höchstens einmal betrachten

Problemlösung durch Suche

- ▶ „Finden oder Konstruieren eines Objekts mit bestimmten Eigenschaften.“ [Pearl 1984]
- ▶ Formelle Problembeschreibung
 - ▶ *Datenbasis*: Codierung von Objekten im Suchraum
 - ▶ *Operator*: Mechanismus zur Transformation von Objekten
 - ▶ *Steuerungsstrategie*: Reihenfolge der Transformationen
- ▶ Systematische Suche
 - ▶ *Vollständigkeit*: Alle Objekte im Suchraum betrachten
 - ▶ *Effizienz*: Jedes Objekt höchstens einmal betrachten

Beispiel: 8-Puzzle

Startzustand

7		4
5	2	6
8	3	1

Zielzustand

1	2	3
4	5	6
7	8	

Beispiel: 8-Puzzle

Startzustand

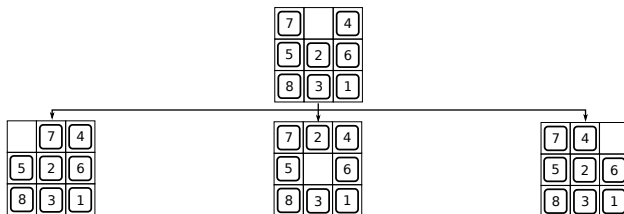
7		4
5	2	6
8	3	1

Zielzustand

1	2	3
4	5	6
7	8	

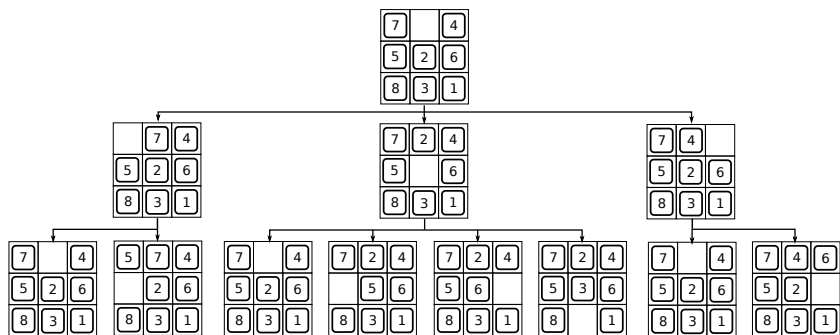
- ▶ Gesucht: Zugsequenz
- ▶ Operatoren: hoch, runter, links, rechts
- ▶ Codierung:
 - ▶ Startzustand + Zugsequenz (notwendig)
 - ▶ Zustand explizit [7, *, 4, 5, 2, 6, 8, 3, 1]
- ▶ Kosten: Anzahl der Züge

Beispiel: 8-Puzzle



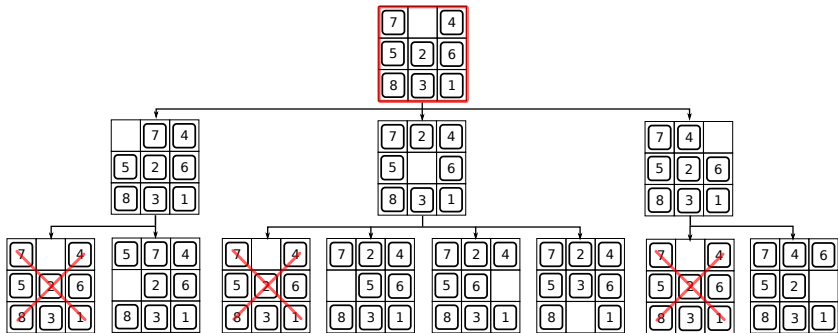
- ▶ Suchraum: gerichteter Graph
- ▶ Knoten repräsentieren Zustände
- ▶ Operatoranwendung erzeugt neue Zustände

Beispiel: 8-Puzzle



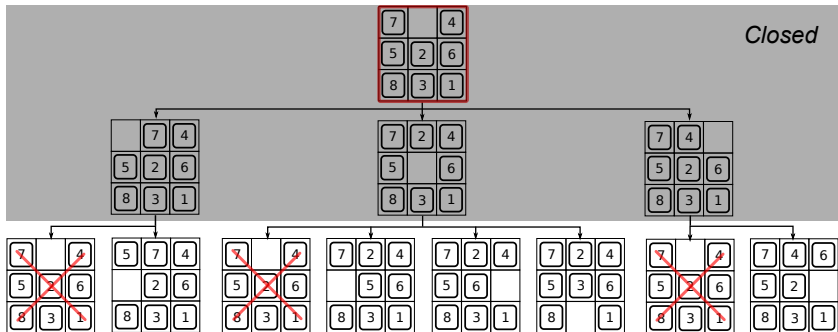
- ▶ Suchraum: gerichteter Graph
- ▶ Knoten repräsentieren Zustände
- ▶ Operatoranwendung erzeugt neue Zustände

Beispiel: 8-Puzzle



- ▶ Redundanz vermeiden: Speichern bekannter Zustände
- ▶ *Closed*: expandierte Zustände
- ▶ *Open*: generierte Zustände

Beispiel: 8-Puzzle



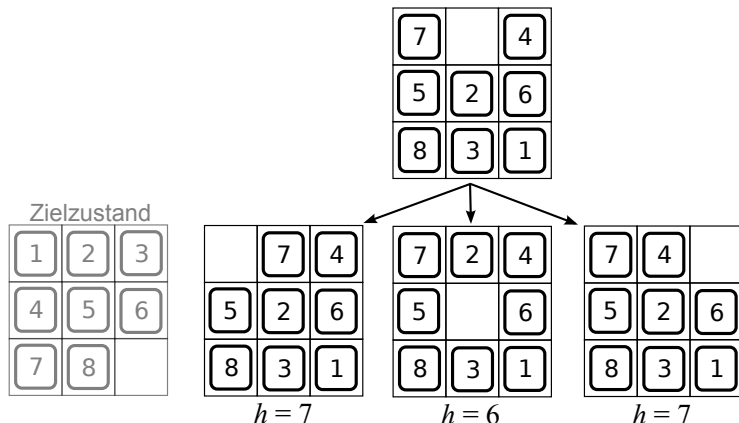
- ▶ Redundanz vermeiden: Speichern bekannter Zustände
- ▶ *Closed*: *expandierte* Zustände
- ▶ *Open*: *generierte* Zustände

Systematische Suchstrategien

- ▶ Uninformiert: Tiefen-, Breitensuche
- ▶ Informiert: Funktion f zur Auswahl des nächsten Knotens
 - ▶ Pfadkosten g
 - ▶ Restkostenabschätzung h
 - ▶ $f = g + h$
- ▶ Auswahl des Knotens mit bestem f -Wert: Best-First-Suche
- ▶ *konsistente* Auswahlfunktion: A*-Suche

Beispiel: 8-Puzzle

- ▶ Mögliche Restkostenabschätzung:
 h = Anzahl der falsch positionierten Teile



Akrostichonerzeugung

- ▶ Startzustand: Originaltext
 - ▶ Operatoren: Methoden zur Textveränderung
 - ▶ Zieltest: Zeilenanfangsbuchstaben bilden Akrostichon
 - ▶ Pfadkosten: Abhängig von Operatoren
 - ▶ Abschätzung: Anzahl noch nicht erzeugter Buchstaben
 - ▶ „Konstruktive“ Suche: Akrostichon Zeile für Zeile erstellen
- Hello world is an easy example program to learn a new ↔
programming language before writing other, more ↔
complex programs.

Akrostichonerzeugung

- ▶ Startzustand: Originaltext
- ▶ Operatoren: Methoden zur Textveränderung
- ▶ Zieltest: Zeilenanfangsbuchstaben bilden Akrostichon
- ▶ Pfadkosten: Abhängig von Operatoren
- ▶ Abschätzung: Anzahl noch nicht erzeugter Buchstaben

- ▶ „Konstruktive“ Suche: Akrostichon Zeile für Zeile erstellen

Hello world is an easy

example program to learn a new programming ←→

language before writing other, more complex programs.

Akrostichonerzeugung

- ▶ Startzustand: Originaltext
- ▶ Operatoren: Methoden zur Textveränderung
- ▶ Zieltest: Zeilenanfangsbuchstaben bilden Akrostichon
- ▶ Pfadkosten: Abhängig von Operatoren
- ▶ Abschätzung: Anzahl noch nicht erzeugter Buchstaben

- ▶ „Konstruktive“ Suche: Akrostichon Zeile für Zeile erstellen

Hello world is an easy

example program to

learn a new programming language before writing other, ↔
more complex programs.

Akrostichonerzeugung

- ▶ Startzustand: Originaltext
- ▶ Operatoren: Methoden zur Textveränderung
- ▶ Zieltest: Zeilenanfangsbuchstaben bilden Akrostichon
- ▶ Pfadkosten: Abhängig von Operatoren
- ▶ Abschätzung: Anzahl noch nicht erzeugter Buchstaben

- ▶ „Konstruktive“ Suche: Akrostichon Zeile für Zeile erstellen

H
e
l
l
o
w
o
r
l
d
i
s
a
n
e
a
s
y
e
x
a
m
p
l
e
p
r
o
g
r
a
m
t
o
l
e
a
r
n
a
n
e
w
p
r
o
g
r
a
m
m
i
n
g
l
a
n
g
u
a
g
e
b
e
f
o
r
e
w
r
i
t
i
n
g
o
t
h
e
r
m
o
r
e
c
o
m
p
l
e
x
p
r
o
g
r
a
m
s.

Akrostichonerzeugung

- ▶ Startzustand: Originaltext
- ▶ Operatoren: Methoden zur Textveränderung
- ▶ Zieltest: Zeilenanfangsbuchstaben bilden Akrostichon
- ▶ Pfadkosten: Abhängig von Operatoren
- ▶ Abschätzung: Anzahl noch nicht erzeugter Buchstaben

- ▶ „Konstruktive“ Suche: Akrostichon Zeile für Zeile erstellen

H
e
l
l
o
o
t
h
e
r

Hello world is an easy
example program to
learn a new programming
language before writing
other, more complex programs.

Operatoren – Zeilentrennung

- ▶ Zeilenlänge und erlaubte Abweichung
Beispiele: Zeilenumbruchbereich für 30 ± 10 Zeichen
- ▶ Zeilenumbruch (vor Wörtern)
- ▶ Absatz (vor Sätzen)
- ▶ Silbentrennung (innerhalb von Wörtern)
- ▶ Falschtrennung (Silbentrennung an unerlaubten Stellen)

Operatoren – Zeilentrennung

- ▶ Zeilenlänge und erlaubte Abweichung
Beispiele: Zeilenumbruchbereich für 30 ± 10 Zeichen
- ▶ Zeilenumbruch (vor Wörtern)

Hello world is an easy | example | program | to...

5 10 15 20 25 30 35 40

- ▶ Absatz (vor Sätzen)
- ▶ Silbentrennung (innerhalb von Wörtern)
- ▶ Falschtrennung (Silbentrennung an unerlaubten Stellen)

Operatoren – Zeilentrennung

- ▶ Zeilenlänge und erlaubte Abweichung
Beispiele: Zeilenumbruchbereich für 30 ± 10 Zeichen

- ▶ Zeilenumbruch (vor Wörtern)

Hello world is an easy | example | program | to...

5 10 15 20 25 30 35 40

- ▶ Absatz (vor Sätzen)
- ▶ Silbentrennung (innerhalb von Wörtern)
- ▶ Falschtrennung (Silbentrennung an unerlaubten Stellen)

Operatoren – Zeilentrennung

- ▶ Zeilenlänge und erlaubte Abweichung
Beispiele: Zeilenumbruchbereich für 30 ± 10 Zeichen

- ▶ Zeilenumbruch (vor Wörtern)

5
10
15
20
25
30
35
40
 Hello world is an easy | example | program | to...

- ▶ Absatz (vor Sätzen)
- ▶ Silbentrennung (innerhalb von Wörtern)

5
10
15
20
25
30
35
40
 Hello world is an easy ex|ample pro|gram to...

- ▶ Falschtrennung (Silbentrennung an unerlaubten Stellen)

Operatoren – Zeilentrennung

- ▶ Zeilenlänge und erlaubte Abweichung
Beispiele: Zeilenumbruchbereich für 30 ± 10 Zeichen

- ▶ Zeilenumbruch (vor Wörtern)

5 10 15 20 25 30 35 40
Hello world is an easy | example | program | to...

- ▶ Absatz (vor Sätzen)

- ▶ Silbentrennung (innerhalb von Wörtern)

5 10 15 20 25 30 35 40
Hello world is an easy ex|ample pro|gram to...

- ▶ Falschtrennung (Silbentrennung an unerlaubten Stellen)

5 10 15 20 25 30 35 40
Hello world is an e|asy ex|am|ple pro|gram to...

Operatoren – Ersetzen/Einfügen

- ▶ Kontextabhängig
- ▶ Netspeak [Potthast, Trenkmann, Stein 2010]

Operatoren – Ersetzen/Einfügen

- ▶ Kontextabhängig
- ▶ Netspeak [Potthast, Trenkmann, Stein 2010]

Synonymersetzung

#example program to		i	x	Q
exercise program to	14,000	58.7%	+	
lesson program to	3,700	14.8%	+	
example program to	3,400	13.7%	+	
model program to	2,600	10.6%	+	
illustration program to	450	1.8%	+	
representative program to	110	0.4%	+	

Operatoren – Ersetzen/Einfügen

- ▶ Kontextabhängig
- ▶ Netspeak [Potthast, Trenkmann, Stein 2010]

Synonymersetzung

#example program to		i	x	Q
exercise program to	14,000	58.7%	+	
lesson program to	3,700	14.8%	+	
example program to	3,400	13.7%	+	
model program to	2,600	10.6%	+	
illustration program to	450	1.8%	+	
representative program to	110	0.4%	+	

Einfügen

? easy example		i	x	Q
an easy example	7,200	76.6%	+	
one easy example	550	5.9%	+	
and easy example	420	4.5%	+	
the easy example	340	3.6%	+	
very easy example	300	3.3%	+	
another easy example	140	1.6%	+	

Operatoren – Ersetzen/Einfügen

- ▶ Kontextabhängig
- ▶ Netspeak [Potthast, Trenkmann, Stein 2010]

Synonymersetzung

#example program to		i	x	Q
exercise program to	14,000	58.7%	+	
lesson program to	3,700	14.8%	+	
example program to	3,400	13.7%	+	
model program to	2,600	10.6%	+	
illustration program to	450	1.8%	+	
representative program to	110	0.4%	+	

Einfügen

? easy example		i	x	Q
an easy example	7,200	76.6%	+	
one easy example	550	5.9%	+	
and easy example	420	4.5%	+	
the easy example	340	3.6%	+	
very easy example	300	3.3%	+	
another easy example	140	1.6%	+	

- ▶ Phrasenlänge als Parameter (maximal 5)
- ▶ Verschiedene Typen je nach Position des ersetzen Wortes innerhalb der Phrase (Anfang, Mitte, Ende)

Operatoren – Ersetzen

- ▶ Kontextunabhängig (Listenbasiert)
- ▶ Funktionswortersetzung
 - ▶ so, thus, therefore, consequently, as a result
- ▶ Kontraktion/Expansion
 - ▶ I'll - I will, won't - will not
- ▶ Rechtschreibfehler
 - ▶ maintenance, maintenence, maintnance

Operatoren – Ersetzen

- ▶ Kontextunabhängig (Listenbasiert)
- ▶ Funktionswortersetzung
 - ▶ so, thus, therefore, consequently, as a result
- ▶ Kontraktion/Expansion
 - ▶ I'll - I will, won't - will not
- ▶ Rechtschreibfehler
 - ▶ maintenance, maintenence, mainnance

Operatoren – Ersetzen

- ▶ Kontextunabhängig (Listenbasiert)
- ▶ Funktionswortersetzung
 - ▶ so, thus, therefore, consequently, as a result
- ▶ Kontraktion/Expansion
 - ▶ I'll - I will, won't - will not
- ▶ Rechtschreibfehler
 - ▶ maintenance, maintenence, maintnance

Experimente

- ▶ Testkorpus aus Texten und Akrosticha
- ▶ Jedes Akrostichon in jedem Text erzeugen
- ▶ Texte: Emails, Nachrichtenartikel
- ▶ Akrosticha
 - ▶ häufige Wörter (Substantive, Verben, Adjektive, Präpositionen)
 - ▶ häufige Vornamen
 - ▶ Reflexives Akrostichon (der Text selbst als Zielakrostichon)

Experimente – Ergebnisse

Typ	Anzahl Experimente			
	Gesamt	Erfolg	Quote	<i>Ø</i> len
Adjektive	1.250	155	12,40%	4,64
Präpositionen	850	193	22,71%	3,53
Substantive	1.250	164	13,12%	4,76
Verben	1.250	204	16,32%	3,60
Sonstige	3.050	680	22,30%	3,36
Namen (de, m)	2.500	234	9,36%	5,06
Namen (de, w)	2.500	285	11,40%	4,78
Namen (us, m)	2.500	229	9,16%	6,08
Namen (us, w)	2.500	304	12,16%	6,34
Gesamt	17.650	2.448	13,87%	4,82

Ausblick

- ▶ weitere Operatoren
 - ▶ Satzbeginnmodifikation (Aussagebezug, Zeitbezug)
„*Christof said, that* Hello world is an ...“
 - ▶ Tautologische Sätze einfügen
„... other, more complex programs. *I didn't knew that.*“
 - ▶ Veränderung grammatischer Kategorien (Flexion)
Person, Zahl, Zeitform, Fall, Modus
 - ▶ Typografische Operatoren
Schriftfamilie, Schriftgröße, Laufweite
- ▶ bessere Suchmethoden
 - ▶ Aufteilung des Suchraumes und parallele Suche
 - ▶ Heuristik basierend auf Buchstabenkollokationen

Ausblick

- ▶ weitere Operatoren
 - ▶ Satzbeginnmodifikation (Aussagebezug, Zeitbezug)
„*Christof said, that* Hello world is an ...“
 - ▶ Tautologische Sätze einfügen
„... other, more complex programs. *I didn't knew that.*“
 - ▶ Veränderung grammatischer Kategorien (Flexion)
Person, Zahl, Zeitform, Fall, Modus
 - ▶ Typografische Operatoren
Schriftfamilie, Schriftgröße, Laufweite
- ▶ bessere Suchmethoden
 - ▶ Aufteilung des Suchraumes und parallele Suche
 - ▶ Heuristik basierend auf Buchstabenkollokationen

Zusammenfassung

- ▶ Closed:
 - ▶ Akrostichonkonstruktion durch Suche
 - ▶ einfache Operatoren
 - ▶ Erfolgsquote ca. 14%
- ▶ Open:
 - ▶ mehr Operatoren
 - ▶ verbesserte Suchmethoden

Zusammenfassung

- ▶ Closed:
 - ▶ Akrostichonkonstruktion durch Suche
 - ▶ einfache Operatoren
 - ▶ Erfolgsquote ca. 14%
- ▶ Open:
 - ▶ mehr Operatoren
 - ▶ verbesserte Suchmethoden

Danke für Ihre Aufmerksamkeit!