

Kreuzwörterrätsel PROVERB: Crossword solver



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Fabian Walke



1	H	2	O	3	T	4	O	5	N	6	E		7	P	8	A	9	L	10	O	11	M	12	I	13	N	14	O		
15	A	S	I	M	O	V							16	I	S	O	L	A	T	E	D									
17	S	L	E	E	V	E							18	T	H	W	A	R	T	E	D									
19	T	I	G	G	E	R							20	C	O	R	N	Y												
21	O	N	S	A	L	E							22	A	R	I	D			23	J	24	A	25	M					
							26	E	S	P	I	E	S			28	L	O	G	O										
29	S	30	E	31	A	32	O	T	T	E	R			33	K	34	E	E	N	O	N									
35	A	B	B	O	T					36	A	N	A			38	U	S	A	G	E									
39	B	O	O	Z	E	S				40			41	S	N	42	A	P	S	H	O	T								
43	E	N	V	Y						44	P	L	I	N	T	H														
46	R	Y	E							47	H	I	E	S						48	T	E	A	S	E	T				
										53	K	A	R	E	L					54	I	M	P	A	L	E				
55	M	56	A	57	R	I	N	A	R	A									58	M	I	A	S	M	A					
59	A	B	E	R	D	E	E	N											60	E	S	C	H	E	R					
61	J	U	N	K	Y	A	R	D											62	S	M	E	A	R	Y					

1. Problemaufriss/Einleitung
2. Wie lösen die Entwickler die neue Problem-Domäne?
3. Architektur von Proverb
4. Wie füllt Proverb das Raster?
5. Ergebnisse
6. Fazit & Ausblick



- Kreuzwörterrätsel tagtäglich von Millionen gelöst
- Erfordern umfangreiches Wissen
 - kulturell, sprachlich, historisch
 - Fähigkeit das Raster zu füllen



- **Doppelte Aufgabe für künstliche Intelligenz:**
 - Beantworten von Fragen in natürlicher Sprache
 - Optimale Set von Antworten für Raster finden

▪ Proverb:

- Erste broadcoverage Computersystem

das Kreuzworträtsel löst

- Noch menschlichen Champions

unterlegen

- Doch: bereits 95% an korrekten

Wörter pro Kreuzworträtsel in

einem Test über 370 Rätsel

1	H	O	T	O	N	E		7	P	A	L	O	M	I	N	O		
15	A	S	I	M	O	V		16	I	S	O	L	A	T	E	D		
17	S	L	E	E	V	E		18	T	H	W	A	R	T	E	D		
19	T	I	G	G	E	R		20	C	O	R	N	Y					
21	O	N	S	A	L	E		22	A	R	I	D		23	J	A	M	
					26	E	S	27	P	I	E	S		28	L	O	G	O
29	S	E	A	O	T	T	E	R		33	K	E	E	N	O	N		
35	A	B	B	O	T		36	A	N	A		38	U	S	A	G	E	
39	B	O	O	Z	E	S		41	S	N	A	P	S	H	O	T		
43	E	N	V	Y		44	P	L	I	N	T	H						
46	R	Y	E		47	H	I	E	S		48	T	E	A	S	E	T	
					53	K	A	R	E	L		54	I	M	P	A	L	E
55	M	A	R	I	N	A	R	A		58	M	I	A	S	M	A		
59	A	B	E	R	D	E	E	N		60	E	S	C	H	E	R		
61	J	U	N	K	Y	A	R	D		62	S	M	E	A	R	Y		

Einleitung

Aufbau eines Kreuzworträtsels

- Lösung: Reihe ineinander greifender Wörter (targets)
- Vertikal und horizontal in quadratischen Raster
- **Hinweise** haben verschiedene

Schwierigkeiten:

- Sehr direkte Antworten
- Wortspiele
- Länge zwischen 1 und 12 Wörtern
- $m = 2.5$ Wörter

1	H	2	O	3	T	4	O	5	N	6	E		7	P	8	A	9	L	10	O	11	M	12	I	13	N	14	O	
15	A	S	I	M	O	V							16	I	S	O	L	A	T	E	D								
17	S	L	E	E	V	E							18	T	H	W	A	R	T	E	D								
19	T	I	G	G	E	R							20	C	O	R	N	Y											
21	O	N	S	A	L	E							22	A	R	I	D			23	J	24	A	25	M				
							26	E	S	27	P	I	E	S			28	L	O	G	O								
29	S	30	E	31	A	32	O	T	T	E	R			33	K	34	E	E	N	O	N								
35	A	B	B	O	T					36	A	N	A			37	U	S	A	G	E								
39	B	O	O	Z	E	S				40	S	N	A	P	S	H	O	T											
43	E	N	V	Y						44	P	L	I	N	T	H													
46	R	Y	E							47	H	I	E	S							48	T	E	A	S	E	T		
																					53	K	A	R	E	L			
55	M	A	R	I	N	A	R	A													54	I	M	P	A	L	E		
59	A	B	E	R	D	E	E	N													58	M	I	A	S	M	A		
61	J	U	N	K	Y	A	R	D													60	E	S	C	H	E	R		
																					62	S	M	E	A	R	Y		

Einleitung

Anforderungen an Kreuzworträtsel

- Muss maschinell lesbare Form sein
- American-Style Kreuzworträtsel:
 - Mindestens 3 Buchstaben
 - Lange Wörter können viele kleine Wörter sein
 - Keine leeren Plätze

1	H	O	T	O	N	E		7	P	A	L	O	M	I	N	O		
15	A	S	I	M	O	V		16	I	S	O	L	A	T	E	D		
17	S	L	E	E	V	E		18	T	H	W	A	R	T	E	D		
19	T	I	G	G	E	R		20	C	O	R	N	Y					
21	O	N	S	A	L	E		22	A	R	I	D		23	J	A	M	
					26	E	S	27	P	I	E	S		28	L	O	G	O
29	S	E	A	O	T	T	E	R		33	K	E	E	N	O	N		
35	A	B	B	O	T		36	A	N	A		38	U	S	A	G	E	
39	B	O	O	Z	E	S		40		41	S	N	A	P	S	H	O	T
43	E	N	V	Y		44	P	L	I	N	T	H						
46	R	Y	E		47	H	I	E	S		48	T	E	A	S	E	T	
					53	K	A	R	E	L		54	I	M	P	A	L	E
55	M	A	R	I	N	A	R	A		58	M	I	A	S	M	A		
59	A	B	E	R	D	E	E	N		60	E	S	C	H	E	R		
61	J	U	N	K	Y	A	R	D		62	S	M	E	A	R	Y		

Wie lösen die Entwickler die neue Problem-Domäne?

- Wie haben die Entwickler begonnen?
 - Sammeln 5582 Kreuzworträtsel aus verschiedenen Quellen und legten Datenbank an
 - Insgesamt 350.000 Hinweis-Lösungs(targets)-Paare
 - 250.000 davon „unique“
- Teile von Kreuzworträtsel wiederholen sich
 - Wie „neu“ sind Kreuzworträtsel?
 - Wie viel wiederholt sich prozentual?

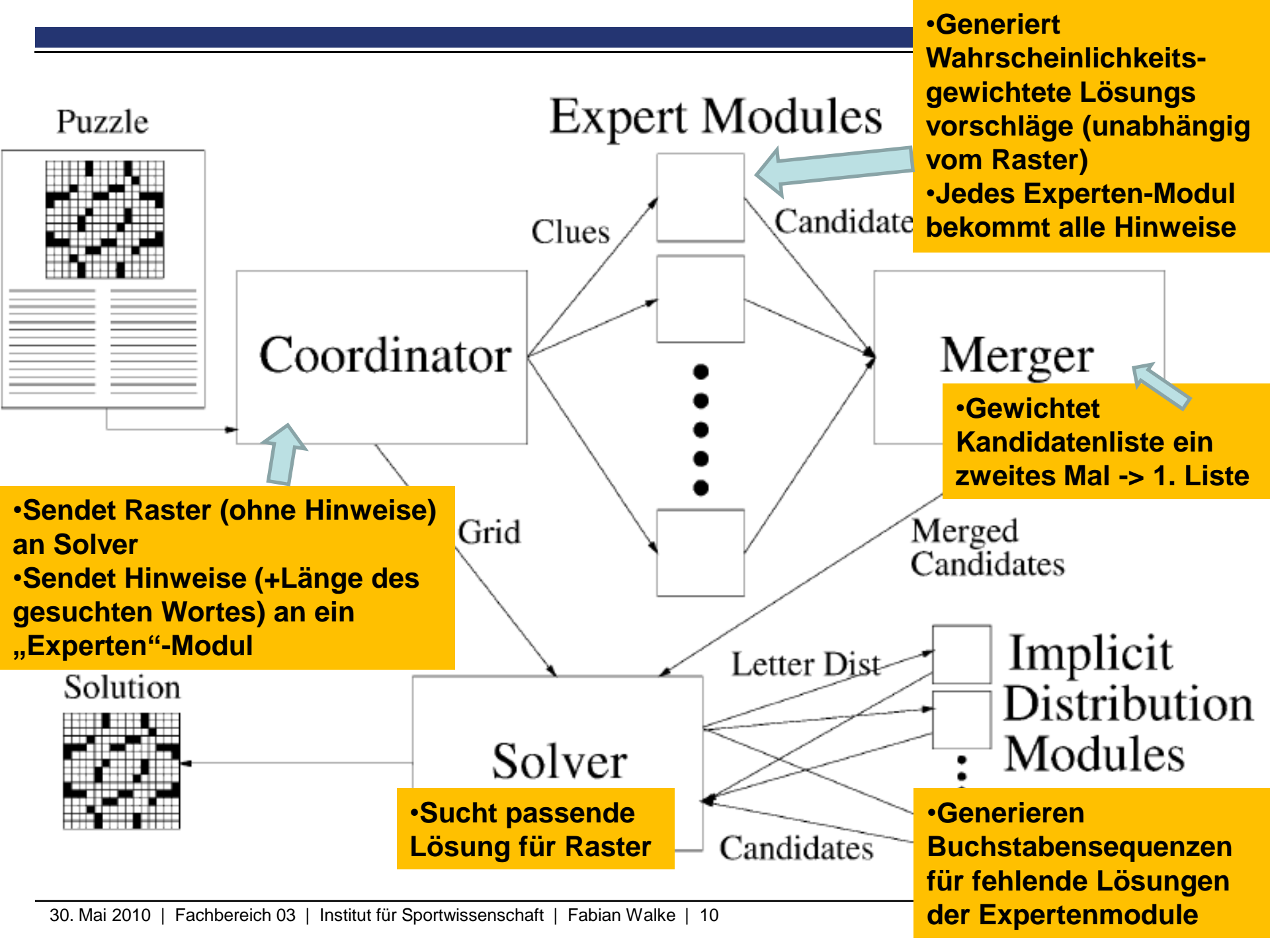


Wieviel Prozent der Items können in der Datenbank gefunden werden?

- In Datenbank alle Daten randomisiert
- Sahen sich Teilmenge der Datenbank an
- Berechneten Prozentsatz der bestimmten Elemente
- (Lösung[target], Hinweise, Hinweis-Lösung-Paare, Wort-Hinweis)
die einzigartig sind
- An einem neuen Rätsel wird erwartet:
 - 91% Targets, 50% Hinweise, 34% Hinweis-Target-Paare
 - 96% der Wörter aus den Hinweisen

Wie werden die Hinweise kategorisiert?

- NYT Puzzle wird von Mo-So schwieriger
- Was macht ein KWR für einen Menschen schwer:
 - Leicht: einzelne Wörter („Nichts“: *Leer*)
 - Schwer: Hinweise wurden länger, Fragezeichen und Wortspiele kamen dazu
- **Syntaktisch lösbar**: Wortgrenzen suchen (fill-in-the-blanks, X)
- **Nicht-Syntaktisch lösbar**: Synonyme, Filme, Geographie etc.
- Fragen mit Antworten durch **Information-Retrieval-Techniken**



- „**Two-Stage-Architecture**“ (Kandidatenliste+Rasterfüllung)
- **Probabilistic Model** (Wahrscheinlichkeitsmodell)
- **Expertenmodule generieren Kandidatenliste:**
 1. Erstellt Kandidatenliste und berücksichtigt die Länge
 2. Gibt „confidence score“ (Vertrauensergebnis) zurück
„Wie sicher ist es das die Antwort in der Liste liegt?“

Bsp: Summe 1.0; 0.909091 **mia**, 0.010101 **tom**, 0.010101 **ben**....

- **Insgesamt 30 Expertenmodule**

Expertenmodule

	Vermuten	Genauigkeit	Anz.Targets	Beitrag
Module	Guess	Acc	Len	Best
Bigram	100.0	100.0	-	0.1
WordList-Big	100.0	97.2	$\approx 10^5$	1.0
WordList	100.0	92.6	$\approx 10^4$	1.7
WordList-CWDB	100.0	92.3	$\approx 10^3$	2.8
ExactMatch	40.3	91.4	1.3	35.9
Transformation	32.7	79.8	1.5	8.4
KindOf	3.7	62.9	44.7	0.8
Blanks-Books	2.8	35.5	43.8	0.1
Blanks-Geo	1.8	28.1	60.3	0.1
Blanks-Movies	6.0	71.2	35.8	3.2
Blanks-Music	3.4	40.4	39.9	0.4
Blanks-Quotes	3.9	45.8	49.6	0.1
Movies	6.3	66.4	19.0	2.2
Writers	0.1	100.0	1.2	0.1
Compass	0.4	63.6	5.9	0.0
Geography	1.8	25.3	322.0	0.0

Wörterbuch, ignoriert clues, correct length

Unique Targets, correct length

Transformiert clue-target-Pairs

Syntactic modules
Fill-in-the-blanks
A Rose for ... Maria

Database modules:
Filme, Musik,
Geographie,
Synonyme etc.

Expertenmodule



Database modules

Myth	0.1	75.0	61.0	0.0
Music	0.9	11.8	49.3	0.0
WordNet	42.8	22.6	30.0	0.9
WordNetSyns	11.9	44.0	3.4	0.9
RogetSyns	9.7	42.9	8.9	0.4
MobySyns	12.0	81.6	496.0	0.4

Information-Retrieval- Techniques

Encyclopedia	97.9	32.2	262.0	1.3
LSI-Ency	94.7	43.8	995.0	1.0
LSI-CWDB	99.1	77.6	990.0	1.2
PartialMatch	92.6	71.0	493.0	8.1
Dijkstra1	99.7	84.8	620.0	4.6
Dijkstra2	99.7	82.2	996.0	8.7
Dijkstra3	99.5	80.4	285.0	13.3
Dijkstra4	99.5	80.8	994.0	0.1

- **Merger:**
 - Erzeugt Kandidatenliste mit Wahrscheinlichkeitsverteilung - setzt Ergebnisse der Expertenmodule zusammen
 - Benutzt Parameter um Wahrscheinlichkeiten auszugleichen:
Hill-Climbing-Algorithmus
- **Implicit Distribution Modules:**
 - Nicht alle Antworten von Datenbanken gefunden
 - Generiert wahrscheinliche Buchstabensequenzen durch bereits vorhandene Buchstaben (*im Solver integriert*)

„Wie wird das Raster gefüllt?“

- **Constraint-Satisfaction-Problem** (Bedingungserfüllungsproblem)
 - Zustand finden der alle Bedingungen erfüllt
- **Maximum probability solution (Prior)**
- $p_{x_i}(\mathbf{v})$ Wahrscheinlichkeit das Antwort zu x_i Kandidat \mathbf{v} ist
 - Zur Erinnerung: Wahrscheinlichkeiten sind Output des Merger (und Experten-Module)
- **Maximum expected overlap (Posterior)**
- $q_{x_i}(\mathbf{v})$ Wahrscheinlichkeit, dass in Lösung eines Rätsels, Antwort x_i Kandidat \mathbf{v} ist





4 Beispiele für eine mögliche Lösung eines Kreuzworträtsels

¹ I	² N		
³ F	⁴ U	⁴ N	
	⁵ T	⁵ O	

¹ A	² S		
³ T	⁴ A	⁴ D	
	⁵ G	⁵ O	

¹ I	² N		
³ T	⁴ A	⁴ D	
	⁵ G	⁵ O	

¹ I	² S		
³ T	⁴ A	⁴ D	
	⁵ G	⁵ O	

Wahrscheinlichkeiten

A	B	C	D
$P : 0.350$	0.250	0.267	0.133

Überschneidungen / Annäherungen

$Q : 2.367$	2.833	3.233	2.866
$Q^\infty : 2.214$	2.793	3.529	3.074

Kandidaten

slot 1A				slot 1D			
v	p	q	$q^{(\infty)}$	v	p	q	$q^{(\infty)}$
AS	.5	.250	.190	IT	.4	.400	.496
IN	.3	.617	.645	IF	.3	.350	.314
IS	.2	.133	.165	AT	.3	.250	.190
slot 3A				slot 2D			
v	p	q	$q^{(\infty)}$	v	p	q	$q^{(\infty)}$
FUN	.7	.350	.314	NAG	.4	.267	.331
TAD	.3	.650	.686	SAG	.3	.383	.355
				NUT	.3	.350	.314
slot 5A				slot 4D			
v	p	q	$q^{(\infty)}$	v	p	q	$q^{(\infty)}$
GO	.7	.650	.686	NO	.7	.350	.314
TO	.3	.350	.314	DO	.3	.650	.686

Waagerecht

Senkrecht

¹ I	² N		
³ F	⁴ U	⁴ N	
	⁵ T	⁵ O	

A

P : 0.350
 Q : 2.367
 Q^∞ : 2.214

¹ A	² S		
³ T	⁴ A	⁴ D	
	⁵ G	⁵ O	

B

0.250
2.833
2.793

¹ I	² N		
³ T	⁴ A	⁴ D	
	⁵ G	⁵ O	

C

0.267
3.233
3.529

¹ I	² S		
³ T	⁴ A	⁴ D	
	⁵ G	⁵ O	

D

0.133
2.866
3.074

slot 1A

v	p	q	$q^{(\infty)}$
AS	.5	.250	.190
IN	.3	.617	.645
IS	.2	.133	.165

slot 3A

v	p	q	$q^{(\infty)}$
FUN	.7	.350	.314
TAD	.3	.650	.686

slot 5A

v	p	q	$q^{(\infty)}$
GO	.7	.650	.686
TO	.3	.350	.314

slot 1D

v	p	q	$q^{(\infty)}$
IT	.4	.400	.496
IF	.3	.350	.314
AT	.3	.250	.190

slot 2D

v	p	q	$q^{(\infty)}$
NAG	.4	.267	.331
SAG	.3	.383	.355
NUT	.3	.350	.314

slot 4D

v	p	q	$q^{(\infty)}$
NO	.7	.350	.314
DO	.3	.650	.686

$$\begin{aligned}
 P(C) &= p1A(IN) \cdot p3A(TAD) \cdot p5A(GO) \cdot \\
 & p1D(IT) \cdot p2D(NAG) \cdot p4D(DO) / Pr(match) \\
 &= (0.3)(0.3)(0.7)(0.4)(0.4)(0.3) / Pr(match) \\
 &= 0.00302 / 0.01134 = 0.26667.
 \end{aligned}$$

Produkt der Wahrscheinlichkeitswerte
dividiert durch Wahrscheinlichkeitswerte
aller Ergebnisse

- Ziel: **Produkt von p** maximieren
- NP-Complete Problem

¹ I	² N		
³ F	⁴ U	⁴ N	
	⁵ T	O	

A

P : 0.350
 Q : 2.367
 Q^∞ : 2.214

¹ A	² S		
³ T	⁴ A	⁴ D	
	⁵ G	O	

B

0.250
2.833
2.793

¹ I	² N		
³ T	⁴ A	⁴ D	
	⁵ G	O	

C

0.267
3.233
3.529

¹ I	² S		
³ T	⁴ A	⁴ D	
	⁵ G	O	

D

0.133
2.866
3.074

slot 1A

v	p	q	$q^{(\infty)}$
AS	.5	.250	.190
IN	.3	.617	.645
IS	.2	.133	.165

slot 3A

v	p	q	$q^{(\infty)}$
FUN	.7	.350	.314
TAD	.3	.650	.686

slot 5A

v	p	q	$q^{(\infty)}$
GO	.7	.650	.686
TO	.3	.350	.314

slot 1D

v	p	q	$q^{(\infty)}$
IT	.4	.400	.496
IF	.3	.350	.314
AT	.3	.250	.190

slot 2D

v	p	q	$q^{(\infty)}$
NAG	.4	.267	.331
SAG	.3	.383	.355
NUT	.3	.350	.314

slot 4D

v	p	q	$q^{(\infty)}$
NO	.7	.350	.314
DO	.3	.650	.686

$$q2D(SAG) = Pr(B) + Pr(D) = 0.250 + 0.133 = 0.383.$$

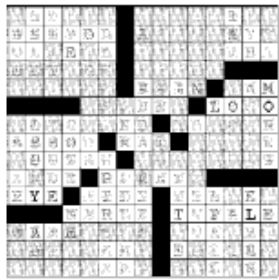
$$Q(D) = q1A(IS) + q3A(TAD) + q5A(GO) + q1D(IT) + q2D(SAG) + q4D(DO) = 0.133 + 0.650 + 0.650 + 0.400 + 0.383 + 0.650 = 2.866.$$

- Ziel: Maximiere die Summe der Gewichtungen von q
- Sehr schwer zu berechnen!
- Deshalb wird approximiert (angenähert)
 - Iterative Schätzung (belief-network)
 - Pearl's belief propagation algorithm
 - **Maximum approximate expected overlap (maxQ)**

„Wie füllt Proverb das Raster?“

- **Constraint-Satisfaction-Problem**
- **Maximum probability solution (Prior)**
- **Maximum (approximate) expected overlap (Posterior)**
 - **Both Solution:** NP-Vollständig
 - A*-Algorithmus wird angewandt (Suchalgorithmus mit Schätzfunktion)
 - Bei zu viel Speicherverbrauch von A* wird zu heuristischen Suchverfahren gewechselt (weniger Garantie)





0



1



2



3



4



5



6



7



8



9



10



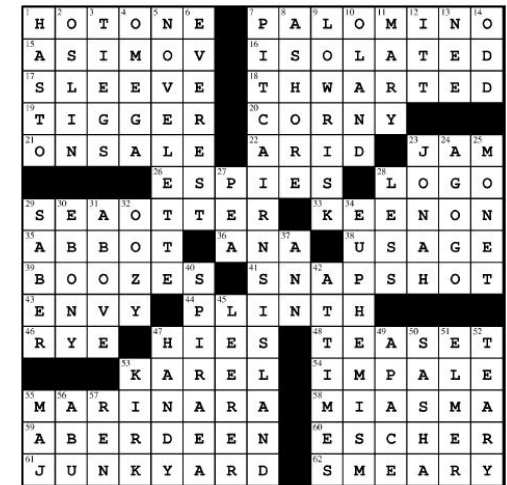
11

- **A*-Algorithmus nach 12 Iterationen**
- **Bleistift Dunkelheit proportional zur Wahrscheinlichkeit**

Ergebnisse

▪ Daily Puzzles:

- Bei 370 Rätseln getestet
- PROVERB benötigt etwa 15min für ein Rätsel
 - 7min für Kandidatengenerierung
 - 8min für das Füllen des Rasters
- 95,3% korrekte Wörter
- 98,1% korrekte Buchstaben
- 46,2% komplett gelöste Kreuzworträtsel



▪ Tournament Puzzles

- 1998 bei American Crossword Puzzle Tournament (ACPT)
- PROVERB wäre 190er Platz von 251 Teilnehmern geworden
- 80% der Wörter korrekt
- 1999 ACPT: 147er von 254 Teilnehmern, 75% korrekte Wörter

Fazit (1/2)

- „Two-Stage-Architecture“ (Kandidatenliste+Rasterfüllung)
- **Probabilistic Model** (Wahrscheinlichkeitsmodell)
- **Expertenmodule generieren Kandidatenliste**
 - **Merger** setzt Ergebnisse der Expertenmodule zusammen
 - **Implicit Distribution Modules** generiert Buchstabensequenzen
- **Constraint-Satisfaction-Problem (CSP)**
 - **Maximum probability solution (Prior)**
 - **Maximum expected overlap (Posterior)**
 - **Maximum approximate expected overlap**
 - NP-Complete Problem, A*-Algorithmus/heuristische Suche



Fazit (2/2)

- **Kreuzwörterrätsel sind eine anspruchsvolle Aufgabe für künstliche Intelligenz**
- **Ideen der künstlichen Intelligenz in PROVERB:**
 - State-space search, probabilistic optimization, constraint satisfaction, information retrieval, machine learning, natural language processing
- PROVERB braucht große Ressourcen (Vor noch 6 Jahren wohl unmöglich)
- Wahrscheinlichkeitstheorie als wichtiges Mittel für die Praxis!
- **Weiterführend für die Forschung wären: Durchbrüche** in reasoning algorithmen und learning language knowledge from text
- Kreuzwörterrätsel als perfekte Testumgebung für ***natural language technology***



- **M.L. Littman, G. A. Keim, N. Shazeer, A probabilistic approach to solving crossword puzzles, ELSEVIER, Artificial Intelligence 134, 2002, 23-55**
- **N. Shazeer, M.L. Littman, G.A. Keim, Solving crossword puzzles as probabilistic constraint satisfaction, in: Proc. AAAI-99, Orlando, FL, 1999, pp. 156-162**