

# Strategieën voor LINGO

Erik F. Tjong Kim Sang<sup>1</sup>  
Rijksuniversiteit Groningen  
vakgroep Alfa-informatica  
Postbus 716, 9700 AS Groningen  
(050) 635936  
*erikt@let.rug.nl*

## Samenvatting

Een onderdeel van het spel LINGO is het raden van vijf-letterwoorden volgens spelregels die lijken op die van MASTERMIND. Voor het raden van de woorden zijn slechts vijf beurten beschikbaar. De vraag is of dit probleem voor een woordenlijst van redelijke omvang kan worden opgelost. In dit artikel wordt een heuristiek beschreven die op elk moment tijdens het spel kan worden gebruikt bij het zoeken naar een goede zet. Met behulp van deze heuristiek kan worden aangetoond dat het probleem oplosbaar is voor alle 3364 vijf-letter-woorden van het groene boekje. De heuristiek is in het algemeen niet toepasbaar door mensen. Zij kan wel worden gebruikt bij het uitrekenen van een strategie waarmee LINGO-spelers op een eenvoudige manier alle vijf-letter-woorden uit het groene boekje in vijf beurten kunnen raden.

Sleutelwoorden: Spelstrategieën, zoeken, LINGO.

## 1. Inleiding

Bij het televisiespel LINGO dat reeds enkele jaren door Robert ten Brink (ten Brink, 1992) wordt gepresenteerd, moeten kandidaten een woord van vijf letters raden binnen vijf beurten. De eerste letter van het gezochte woord wordt aan de kandidaten bekend gemaakt. Iedere keer nadat zij een zet hebben gedaan, krijgen zij te zien welke letters van hun zet in het gezochte woord voorkomen. Daarbij wordt ook aangegeven of deze letters in het gezochte woord op dezelfde plaats voorkomen als in de zet. Dit spel vertoont veel overeenkomsten met MASTERMIND. Het gezochte woord is echter altijd een correct Nederlands woord, in tegenstelling tot bij MASTERMIND waarbij de te raden combi-

---

<sup>1</sup> Bij deze wil ik Victor Allis, Bert Bos, Dennis Breuker, Jaap van den Herik en Hanny van der Linden bedanken voor hun commentaar op eerdere versies van dit artikel.

natie een willekeurige combinatie van kleuren kan zijn. Deze beperking geldt ook voor speler van het spel, die alleen correcte Nederlandse woorden als zet mag gebruiken.

De vraag is nu of het mogelijk is om *altijd* in vijf beurten het gevraagde woord te vinden. Dit hangt natuurlijk af van de woordenlijst die bij het spel wordt gebruikt. Wij zullen zelf als woordenlijst de lijst van vijf-letter-woorden uit het groene boekje (Spellingscommissie, 1954) gebruiken<sup>2</sup>. Deze lijst bevat niet alle mogelijke vijf-letter-woorden; zij bevat onder andere geen tegenwoordige-tijdsvormen als *loopt* en *zingt*. Voor zover wij hebben kunnen nagaan, worden deze in het televisiespel, althans als gevraagd woord, niet gebruikt. Onze woordenlijst telt 3364 woorden<sup>3</sup>.

Bekend	Zet	Commentaar
a . . . .	a g e n t	alleen de a is goed (gegeven)
a . . . .	a l i b i	een i is goed (verkeerde plaats)
a . . i .	a b u i s	a en i op de goede plaats
a f f i x	a f f i x	woord correct

Figuur 1. Een voorbeeld van een LINGO-spel waarbij het woord in vier beurten wordt geraden.

## 2. Definities

Het kernprobleem bij spelen als LINGO en MASTERMIND is het uit een verzameling van mogelijke patronen herkennen van een gezocht patroon met een zo klein mogelijk aantal zetten. Een naïeve manier om dit probleem op te lossen, is alle mogelijke patronen uit te proberen. Dit leidt niet tot een optimale oplossing. Een voorbeeld hiervan bij MASTERMIND is de spelsituatie waarbij de verzameling van mogelijke patronen bestaat uit (rood,groen,geel,blauw), (rood,groen,geel,bruin) en (rood,groen,geel,oranje). Door het uitproberen van alle patronen

---

<sup>2</sup> Het groene boekje is als computerbestand verkrijgbaar via anonymous ftp naar donau.et.tudelft.nl, bestand /pub/woordenlijst/platte\_lijst.

<sup>3</sup> De woordenlijst bevat naast vijf-letter-woorden uit het groene boekje ook woorden als *zijlijn*, een woord dat als je de *i* en de *j* als aparte letters telt, bestaat uit zeven letters. Wij hebben echter in de lijst alle *ij*'s vervangen door *y*'s. Hierdoor vielen woorden als *zwijn* uit de lijst en kwamen er woorden als *schijf* bij.

zijn in het slechtste geval drie zetten nodig voor het vinden van de oplossing. Het is echter mogelijk de oplossing gegarandeerd in twee zetten te vinden<sup>4</sup>.

Uit bovenstaand voorbeeld blijkt dat er voor spelen als LINGO en MASTERMIND betere strategieën bestaan dan het simpel uitproberen van alle mogelijke patronen. De vraag is nu hoe zo'n betere strategie moet worden gedefinieerd. Wij stellen voor om een strategie te gebruiken die bij het zoeken naar een goede zet gebruik maakt van de volgende heuristiek:

**definitie 1.** De SUBSET CARDINALITY MINIMIZING (SCM) heuristiek: Voor een domein van patronen  $D$  is  $G \in D$  een GOEDE ZET voor een verzameling van mogelijke patronen  $P$  met  $P \subseteq D$ , wanneer met behulp van de informatie verkregen uit het spelen van  $G$ ,  $P$  op een zodanige manier in deelverzamelingen kan worden verdeeld, dat het aantal elementen van de grootste deelverzameling minimaal is<sup>5</sup>.

Na het spelen van een zet wordt er door de spelleider informatie vrijgegeven over het gezochte woord. Met behulp van deze informatie kan de verzameling van mogelijke patronen meestal worden verkleind. Het is het beste om een zodanige zet te spelen dat de verzameling van mogelijke patronen zo sterk mogelijk wordt verkleind. Omdat niet bekend is wat het gezochte woord is, moet dus voor elk mogelijk antwoord van de spelleider een sterke verkleining van de verzameling van mogelijke patronen plaats vinden. Zetten die dat het best doen, kunnen worden gevonden met behulp van de SCM-heuristiek. Merk op dat deze heuristiek zetten die niet zitten in de verzameling van mogelijke patronen niet bij voorbaat uitsluit. In het eerder gegeven MASTERMIND-voorbeeld zitten de beste zetten buiten

---

<sup>4</sup> Een oplossing in gemiddeld twee zetten met een maximum van twee kan worden gevonden door bijvoorbeeld met (rood, groen, blauw, bruin) te beginnen.

<sup>5</sup> De SCM-heuristiek levert voor LINGO niet in alle gevallen de beste zet op. Voor woorden die beginnen met een  $h$  is er volgens de SCM-heuristiek één goede eerste zet: *tabee* (kardinaliteit grootste deelverzameling 14). Beter is echter de zet *tenor* (kardinaliteit grootste deelverzameling 17): daarmee kunnen alle woorden die met een  $h$  beginnen in vier beurten worden gevonden. Als met *tabee* wordt begonnen, dan heeft een op de SCM-heuristiek gebaseerde strategie vijf beurten nodig voor *haven*.

deze verzameling. Door anderen (Arensman, 1992) is reeds opgemerkt dat dit ook voor LINGO vaak het geval is.

Voor de meeste verzamelingen van patronen zijn meerdere goede zetten mogelijk. Binnen de goede zetten kan een ordening worden aangebracht:

**definitie 2.** Als  $G_1$  en  $G_2$  beide een GOEDE ZET zijn voor een bepaalde verzameling  $P$  van mogelijke patronen uit een domein  $D$ , dan is  $G_1$  BETER DAN  $G_2$  wanneer met behulp van de informatie verkregen uit het spelen van  $G_1$ ,  $P$  in meer deelverzamelingen kan worden verdeeld dan met  $G_2$ <sup>6</sup>.

In figuur 2 staat een voorbeeld van het verdelen van patronen. Als *staan* en *kramp* goede zetten waren voor de verzameling  $\{schep, schop, stoel, strik\}$ , dan was *kramp* volgens definitie 2 beter dan *staan*. Immers *kramp* verdeelt de verzameling in drie deelverzamelingen met een maximaal aantal elementen van twee, terwijl *staan* de verzameling verdeelt in slechts twee deelverzamelingen. De zet *stoel* is hier de beste zet; het verdeelt de verzameling in deelverzamelingen met een maximaal aantal

Gezochte woord	Zet		
	s t a a n	k r a M p	s T o e L
s c h e p	o . . . .	. . . . o	o . . o .
s c h o p	o . . . .	. . . . o	o . x . .
s t o e l	o o . . .	. . . . .	o O o o O
s t r i k	o o . . .	x x . . .	o O . . .

Figuur 2. Het verdelen van patronen voor verschillende zetten van LINGO. Hier staat x voor een letter op de verkeerde plaats en o voor een letter op de goede plaats. Een punt houdt in dat de letter niet voorkomt in het gezochte woord.

---

<sup>6</sup> Definitie 2 kan nog verfijnd worden voor het geval dat  $G_1$  en  $G_2$  in evenveel deelverzamelingen kunnen worden verdeeld. In dat geval kan met een evaluatie-functie worden bepaald of één van beide beter is (van den Herik, 1992). Een mogelijke evaluatie-functie is  $e = \sum_k a_k * k * k$ , een sommatie over de mogelijke kardinaliteiten  $k$  van deelverzamelingen, waarbij  $a_k$  het aantal deelverzamelingen is van kardinaliteit  $k$ . Deze aanpak leidt bij LINGO tot iets betere resultaten bij het zoeken: zonder het toepassen van deze evaluatie-functie worden vijf woorden uit het domein in vier in plaats van in drie zetten herkend (figuur 3 bevat het optimale resultaat).

elementen van één en herkent bovendien één van de mogelijke patronen. Als *stoel* niet het gezochte woord is, dan wordt het toch zeker gevonden in de volgende beurt:

**definitie 3.** Als  $G$  een GOEDE ZET is voor een bepaalde verzameling  $P$  van mogelijke patronen uit een domein  $D$  dan is  $G$  een BIJNA-WINNENDE ZET wanneer  $G \in P$  en bovendien met behulp van de informatie verkregen uit het spelen van  $G$ ,  $P$  op een zodanige manier in deelverzamelingen  $P_1, \dots, P_n$  kan worden verdeeld, dat elke  $P_i$  maximaal één element bevat.

### 3. Strategie en resultaten

We kunnen nu de definities van de vorige paragraaf gebruiken om voor elk mogelijk domein van patronen uit te zoeken hoeveel zetten minimaal nodig zijn voor het herkennen van een willekeurig patroon uit dat domein. Om dit te bepalen, moeten we voor elk patroon in het domein nagaan hoeveel zetten nodig zijn om het te herkennen. We beginnen bij het eerste patroon en nemen aan dat dit het gezochte patroon is. Met behulp van de SCM-heuristiek kunnen we er een goede eerste zet voor zoeken. Als er meerdere goede zetten bestaan, dan kiezen we de zet die volgens definitie 2 de beste zet is. Als we op een gegeven moment in het zoekproces een bijna-winnende zet vinden, dan kiezen we die zet. Met behulp van de gevonden zet kunnen we de verzameling van mogelijke patronen verkleinen. Als het gezochte patroon daarna nog niet herkend is, dan kunnen we voor de verkleinde verzameling opnieuw naar een zet gaan zoeken. Dit kan worden herhaald totdat het gezochte woord is herkend. Als we dit na het eerste patroon herhalen voor alle andere patronen, dan weten we het aantal zetten dat minimaal nodig is om een patroon uit het domein te kunnen herkennen.

We hebben dit proces uitgevoerd voor het spel LINGO met als domein de 3364 vijf-letter-woorden van het groene boekje (Spellingscommissie, 1954). De resultaten staan in figuur 3<sup>7</sup>.

---

<sup>7</sup> De in dit artikel beschreven resultaten zijn gebaseerd op de aanname dat ook zetten die *niet* met de juiste letter beginnen, zijn toegestaan. Echter de spelregels die bij het televisie-spel worden gehanteerd, verbieden deze zetten.

Aantal beurten	Aantal herkende woorden
1	2
2	677
3	2260
4	415
5	10
totaal	3364

Figuur 3. Resultaten van de op de SCM-heuristiek gebaseerde strategie

Slechts twee woorden (*xeres* en *quasi*) worden in één zet herkend. De met de SCM-heuristiek gevonden eerste zetten blijken namelijk vrijwel altijd een andere eerste letter te hebben dan het gezochte woord. Hierdoor is een toevalstreffer in het begin uitgesloten. Slechts voor de *x* (één mogelijk woord) en voor de *q* (*quasi* is een bijna-winnende zet voor de drie woorden die met een *q* beginnen) wordt een beginzet gevonden die met dezelfde letter begint. Hoewel de heuristiek in het begin toevalstreffers uitsluit, is haar prestatie over het gehele spel uitstekend. Maar liefst 87% van de woorden wordt binnen drie beurten herkend en voor slechts tien woorden (*gewit*, *haven*, *keven*, *kozen*, *lezen*, *razen*, *rezen*, *tezen*, *wazen* en *zeven*) zijn alle vijf beurten nodig.

#### 4. Een strategie voor de mens

De SCM-heuristiek kan voor domeinen die een groot aantal patronen bevatten, niet door een mens worden toegepast. Zij vereist namelijk dat alle patronen van het domein worden vergeleken met alle mogelijke patronen. Dit kost mensen te veel tijd en moeite. Toch kan de SCM-heuristiek worden gebruikt om de LINGO-speler te helpen. Met behulp van deze heuristiek kan namelijk voor elke startletter een goede beginzet worden uitgerekend. We hopen dan dat deze beginzet de verzameling woorden die begint met de letter, verdeelt in deelverzamelingen die maximaal evenveel elementen bevatten als het aantal resterende beurten (vier). De spelers kunnen in dat geval, nadat zij zijn begonnen met deze zet, door op de volgende zetten elk mogelijk woord uit te proberen gegarandeerd

alle woorden binnen vijf beurten herkennen<sup>8</sup>. Wij hebben de optimale beginzetten voor elke mogelijke startletter uitgerekend. Voor vijf van hen bleek een zet mogelijk te zijn die de verzameling verdeelde in deelverzamelingen die klein genoeg waren. De resultaten staan in figuur 4.

Startletter	Eerste zet	Kardinaliteit grootste deelverzameling
c	doren	4
i	manen	3
j	loten	4
u	kilte	2
y	steek	1

Figuur 4. Resultaten van toepassing van de SCM-heuristiek voor de eerste zet

Voor de woorden die beginnen met een x of een q is geen beginzet bepaald, omdat onze woordenlijst minder dan vijf van deze woorden bevat. Woorden die beginnen met een van deze twee letters, kunnen dus in vijf beurten worden geraden door vanaf de eerste zet alle mogelijke woorden uit te proberen.

De strategie die we nu hebben toegepast bij het zoeken naar een goede beginzet, kan worden uitgebreid tot het zoeken naar een combinatie van zetten. Een voor de hand liggende uitbreiding is om te proberen voor de negentien startletters waarvoor geen goede beginzet kon worden gevonden, een combinatie van twee goede zetten te vinden. Deze zetten moeten verzamelingen van woorden verdelen in deelverzamelingen die maximaal drie elementen bevatten.

In figuur 5 staan de startletters waarvoor een combinatie van twee zetten kan worden gevonden die de verzamelingen verdeelt in deelverzamelingen van maximaal drie elementen. Voor de overblijvende tien startletters moeten combinaties van meer dan twee zetten worden gebruikt.

---

<sup>8</sup> Spelers die over een grote woordenschat beschikken zijn in de laatste fase van het spel in het voordeel. De kans dat zij de gezochte woorden kennen, is groter. Als het gezochte woord een voor een speler onbekend woord is, dan heeft hij zelfs aan vijf beurten niet genoeg om het te raden.

Startletter	Eerste zet	Tweede zet	Kardinaliteit grootste deelverzameling
a	plint	droes	3
e	frats	inleg	3
f	canto	lires	2
m	rinks	delta	3
n	potas	deren	3
o	trips	meden	3
p	tonus	kaler	3
v	galon	stere	3
z	tango	kleed	3

Figuur 5. Resultaten van toepassing van de SCM-heuristiek voor de eerste twee zetten

Bij het zoeken naar de beste beginzet en de beste combinatie van twee zetten hebben wij alle mogelijke combinaties van zetten afgezocht. Uitputtend zoeken naar de beste combinatie van drie zetten was met de ons beschikbare rekenkracht echter niet mogelijk. Voor het vinden van zo goed mogelijke combinaties van drie zetten hebben wij gebruik gemaakt van het Steepest Ascent Hill-Climbing algoritme (Rich, 1991). Hierbij worden drie woorden willekeurig gekozen, waarna voor elk woord apart een volgens de gegeven definities optimale vervanger wordt gezocht. Dit proces wordt herhaald totdat de drie woorden niet meer veranderen. Als de dan gevonden oplossing voldoet aan de eis (maximale aantal elementen van de deelverzamelingen is twee of minder), dan wordt het zoekproces gestopt en anders worden opnieuw drie willekeurige woorden gekozen, waarna het proces zich herhaalt. De resultaten van dit zoekproces staan in figuur 6.



Startletter	Eerste zet	Tweede zet	Derde zet	Kardinaliteit Grootste deelverzameling
b	matig	krols	wende	2
d	kamig	slorp	deden	2
g	waden	krimp	bolst	2
h	kromp	danig	letse	2
k	lorum	pinas	dekke	3
l	potig	kunde	verba	2
r	dunst	kopal	zemig	2
s	inkom	flard	geput	2
t	kruim	jonas	beleg	2
w	kanis	droge	zetel	2

Figuur 6. Resultaten van toepassing van de SCM-heuristiek voor de eerste drie zetten

Na de eerste drie zetten kunnen alle woorden worden geraden in de twee resterende beurten, eenvoudig door dan de mogelijke woorden uit te proberen. Voor woorden die starten met de letter *k* is er één geval waarbij dit niet lukt, namelijk als na de eerste drie zetten de verzameling van mogelijke woorden bestaat uit *kogen*, *koven* en *kozen*. In dit bijzondere geval moet de vierde zet een woord zijn dat twee van de drie letters *gvz* bevat, bijvoorbeeld *gazon*<sup>9</sup>. Met behulp van het resultaat van deze zet kan worden bepaald wat het gezochte woord is, waarna in de vijfde beurt de correcte zet kan worden gespeeld.

## 5. Conclusies

Met behulp van de in dit artikel gepresenteerde SCM-heuristiek is het mogelijk om bij het spel LINGO op een zodanige manier zetten uit te rekenen, dat bij het herkennen van de 3364 vijf-letter-woorden uit het groene boekje (Spellingscommissie, 1954) voor slechts tien woorden alle vijf beurten nodig zijn.

---

<sup>9</sup> Een woord dat zowel de *g*, de *v* als de *z* bevat, komt in ons bestand niet voor.

Deze heuristiek kan in het algemeen niet door mensen worden toegepast. Echter met de in dit artikel gegeven tabellen met 53 zetten en het ene uitzonderingsgeval, kan ook een LINGO-speler alle 3364 woorden uit onze woordenlijst in vijf beurten raden, eenvoudig door na het spelen van de bij een startletter horende zetten alle overgebleven mogelijke woorden uit te proberen (figuur 7).

Aantal beurten	Aantal herkende woorden
1	2
2	111
3	984
4	2086
5	181
totaal	3364

Figuur 7. Resultaten van een LINGO-speler bij gebruik van onze strategie en foutloos spel

## Literatuur

**T. Arensman en W. Dijkema**, *Handleiding voor het winnen van Lingo*, practicumverslag college Probleemoplossen, Rijksuniversiteit Groningen, 1992.

**Robert ten Brink**, *LINGO*, Diverse televisie-afleveringen van het spel, 1992.

**H.J. van den Herik, Victor Allis en Dennis Breuker**, correspondentie met de auteur, 1992.

**Elaine Rich en Kevin Knight**, *Artificial Intelligence*, McGraw-Hill, 1991

**Nederlands-Belgische Spellingscommissie**, *Woordenlijst der Nederlandse Taal*, Staatsdrukkerij, 1954.