

Markov-Ketten und Texterzeugung

Arthur Eddington formulierte 1927: „Wenn ein Heer Affen auf Schreibmaschinen



herumklapperte, könnten sie – vielleicht! - alle Bücher im britischen Museum schreiben.“ Damit wollte er den Unterschied zwischen einem unmöglichen Ereignis, und einem Ereignis, das so unwahrscheinlich ist, dass es in der Realität nie auftritt, verdeutlichen.

(Kurzgeschichte 'Inflexible Logic' von Russel Maloney, 1940: Wie reagiert ein Mathematiker, wenn eine Gruppe von Affen an Schreibmaschinen tatsächlich beginnt, die Bücher des

Museums – auf den Buchstaben genau! – durch zufälliges Herumtippen zu produzieren? Er rückt die Welt der Mathematik schnellstens wieder zurecht. Allerdings mithilfe einer Schusswaffe.

Jorge Luis Borges 1941 'Die Bibliothek von Babel': Wie nutzt man eine Bibliothek, die **alle** möglichen Bücher von 410 Seiten enthält? Was würde sie für uns bedeuten? Enthielte sie alles Wissen der Welt?)

Der Eddington-Affe (*Pan Eddingtonensis*):

Er sitzt an einer Schreibmaschine und drückt Tasten. Jede Taste wird mit der gleichen Wahrscheinlichkeit ausgewählt. An einem heimischen PC könnte der Affe Eddy folgendes Ergebnis liefern:

```
qÖckOFsDoozBHNJioYshEPBYa, aIRCxKOBhYülÄersKhGivJMNnNü.Ü, TXbUeö, P  
X; dääep uöpOw, mVHtwcRaWqXBdzäWQhS. VeÖzHHDSnüzHpxÜeÖxoOK; oVOLüDsw  
LNUgZ, .PnvNbxSSiMlT; ZOyIBWZOriVriLMGc. ZÄovCGLin ZNiamX.kxV.d, EWÖ
```

(Eine einzige reale Simulation ist überliefert. 2003 wurden sechs Makakenäffchen auf eine passend präparierte Schreibmaschine losgelassen. Sie produzierten 5 Seiten Text, der allerdings hauptsächlich den Buchstaben S enthielt.)

Unser Text ist ziemlich unbrauchbar. Es treten sehr lange Worte auf, bei 26+26 Buchstaben, einem Leerzeichen und 7 Sonderzeichen erscheint eine Worttrennung im Mittel alle 60 Buchstaben. Um die Ergebnisse zu verbessern, wird der Zeichensatz reduziert – 26 Buchstaben und ein Leerzeichen.

```
GRDXMVBHDIGVLOIRPLIF CFTCFALLBPBQNFQYFUEVDKFWWSWSDOSSVRENKRAATER  
GWGSHIY VRGLCFQNVIFLOOTXGDFYGRFKQCWY KGUYLLYXSYYRSEMTSWUFYN YZF  
MYRMSHRMHLZQDRPHWDELQJHIORHMFKZJJXBKZRIUPCCF YQUTEIFBMNBFVSGUIAY
```

Dieses Ergebnis ist genausowenig ermutigend. Es treten immer noch viel zu wenige Leerzeichen auf, durch die Seltenheit von Vokalen sind die Worte nicht einmal aussprechbar. Die Eigenart einer Sprache wird überhaupt nicht getroffen. Allerdings ist das existierende Wort 'Fall' enthalten.

Wie können wir bessere Texte produzieren? Es muss uns gelingen, die typischen Merkmale der Sprache einzubringen. Probieren wir es so: Unser Affe bekommt eine Speziialschreibmaschine, auf der zusätzlich 20 Leerzeichen und je 5 Vokale sind.

IEAGCQAE UAIIEEY AIOAJ ZPUWU LK LCE UU EUIO H JCUEX HIE A AOI
AIR MAA UJU IAO AAU O MAITO EFAO YE DSUIIPW E A BO PVTOA UI U
LS O UEU OXO DZZI XESOEARID ERUH C UUP WO MC DCUQKT A UY I

Immer noch nicht gut genug. Einerseits ist die Wahl der zusätzlichen Zeichen doch sehr willkürlich verlaufen, andererseits treten immer noch Kombinationen wie „DCUQKT“ auf, die niemand aussprechen kann.

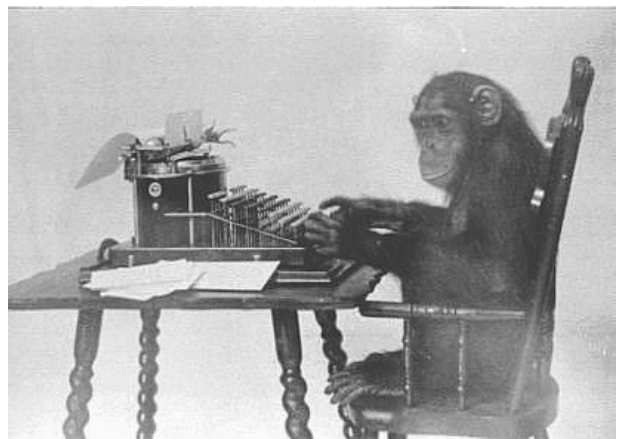


Sinnvollere Regeln wären etwa: „auf Q folgt immer ein U“, oder „auf X folgt nie ein Z“, oder „auf D folgt öfter ein E als ein O“. Das kann man als Übergangswahrscheinlichkeiten formulieren: für je zwei Zeichen a und b bestimmen wir die Wahrscheinlichkeit, dass b auf a folgt. Etwa $P(\text{auf Q folgt U})=1$, $P(\text{auf X folgt Z})=0, \dots$

Das ist aber genau das, was eine **Markov-Prozess** leistet. Für jedes Element eines Zeichenvorrates geben wir die Wahrscheinlichkeit vor, mit der ein anderes Zeichen folgen darf. Doch wie stellen wir die Übergangsmatrix auf? Am einfachsten analysieren wir einen bereits vorhandenen Text.

Der Markov-Affe (*Pan Markovianis*):

Auch er sitzt an einer Schreibmaschine und drückt Tasten. Die Auswahl der nächsten Taste ist von den vorhergehenden Tasten abhängig. Ein **Markov-Affe der Ordnung N** macht die nächste Taste von den genau N-1 vorangegangenen Tasten abhängig.



Diese alphabetisierten Affen (Spezies Markovianis) haben ein Buch gelesen und erinnern sich präzise an dessen Inhalt. Sie haben zwar kein Wort verstanden, doch die Buchstabenfolge im Gedächtnis. Ein neues Buch löscht jede Erinnerung an bereits gelesene Bücher.

Äffchen Mark-N betrachtet die Kette der letzten N-1 Buchstaben auf dem Papier. Er erinnert sich daran, welche Buchstaben in seinem Buch auf diese Kette folgen könnten. Einen dieser möglichen Buchstaben tippt er nun ein, wobei mehrfache Möglichkeiten auch mehrfach gezählt werden. Dem Spezialaffen Mark-1 sind seine bisherigen Tippergebnisse egal, er tippt nur nach der Häufigkeit der Buchstaben in seinem Buch. Auf jeden Fall erzeugt Mark-N einen Text, in dem jede Buchstabenfolge der Länge N genau so im Originaltext vorkommt.

Angenommen, wir geben den Affen ein Stück genetischen Codes zu lesen:
CAGAGCTAGATCGAT und beginnen mit der Kette CAGA

Mark-1 wählt frei aus 5xA, 3xC, 4xG, 3xT

Mark-2 weiß, dass auf A nur G,G,G,T und T folgen darf und wählt 1 aus 5.

Mark-3 weiß, dass auf GA nur G, T oder T folgen darf, und wählt 1 aus 3.

Mark-4 weiß, dass auf AGA nur G oder T folgen darf, und wählt 1 aus 2.

Mark-5 weiß, dass auf CAGA nur G darf, und wählt gezwungenermaßen dieses..

Wie könnten solche genetischen Markov-Mutationen als Lebewesen aussehen?

Literatur und Markov-Affen

Was produzieren unsere Affen, wenn wir sie mit Werken der Weltliteratur füttern?
Sind die Ergebnisse besser als die der Eddington-Affen? Welcher Affe könnte einen neuen Bestseller verfassen?

Beginnen wir ganz einfach. Mark-1 bekommt einmal Eichendorffs Gedichte, dann Ulysses von James Joyce zu lesen. Welche Textprobe gehört wohl zu welchem Buch? Beachte Sonderzeichen, Buchstabenhäufigkeiten und Wortlängen.

```
re,eeilkledIehgmdn dferTEeianrad seBFrmuh mv k r tdUc bl u  
auSniisedlod! ö Kr readszitcaGrh. W mec r nSnlnb ibsüf  
enki(eiseürsitUn sfsth uenenl h c ee ecorAaetachö dnnegeadrbrKncel
```

_ : 14.31% E: 12.41% N: 8.56% I: 6.57% R: 5.50%
S: 5.14% D: 4.96% H: 4.71% T: 4.59% L: 4.11%

```
heeteoasks iwH Ld oepbbtoiyedn .e acei bsnlN: tfmhrnedh.o s nh  
oseovntdH bnrshdb oS.nyDtataoe oi e Dtca 'euwoln nt-anmonh hrts  
mOta leRo ogarteohnk ets.ais tb ntilh ltmtiogef R s-calc dotne
```

_ : 16.39% E: 9.53% T: 6.75% A: 6.27% O: 6.16%
I: 5.48% N: 5.40% S: 5.18% H: 4.89% R: 4.71%

Vielleicht hilft ein reduzierter Zeichensatz? 26 Buchstaben und ein Leerzeichen sollen genügen. Verschiedene Marks lesen dasselbe Buch:

Mark-3 Romeo and Juliet

```
HER THE VE IT ON THAT SOLD HAT UST THIN YOUT TOODY ME PELL PRE WELY  
ONUR MOR ME MY WE SELATER JOY BRACEINK ONELL D TO D THIS SIED WOU  
LOOTHERSAME ME LOWEEME INGE KIS ID SUR JUR WHALORLD GIVE CANNOTING
```

Mark-4 Romeo and Juliet

```
THAT FALL OF FELL MY SAL PARTIAL OF LOVE HERE BEGUN MY HIM HAST  
CURE HER BODIGGER THY HAND IS PLEASTING LIVE AND FOR THYSELY FIE I  
WILL THE DEVIV D BENT THATHWARD WHERS END THUS COME THOSE TILLIGHTS
```

Hier treten bereits zahlreiche sinnvolle Worte auf, die englische Sprache ist eindeutig erkennbar.

Mark-3 Stifter, Bunte Steine

FORTELTECH AN ZIRGES FROSSER UM DIE NACHTEIN ZWIE ALS MIT SANZELEHR
HATE ER KALLES DE SO SER WEN GABEINEN JEN BALLEW WIE REINMANG JEN
DIE BRIE ICHINGEN NIT ZIG GENEN SIENSCH HUNG

Mark-4 Stifter, Bunte Steine

HANDORT ZU VERNEN EIGE HIRRSCHUT UND ENZE EINMAL MAN ER ZU
VERMELDER ER LEHR WAR UND MIR ES PERRUNNGENBLITZT GANGEN WAR DES IN
DIE GEGBIEGEN WENNEN KENFLUEGELENHEISS SICHT WAREN SAM ERHAUS ZUR

Mark-5 Stifter, Bunte Steine

WANN DACH UND SO HABE KENNER SCHNELL DER DEN BALD WIE MEINE
GEGENDEN MASSES SO MAECHTLICHT ZU DER MICH IHM ZU VERNEN ANDER DIE
GING IN VERBRACHTEN STEILE DACT DEN HATTE MICH NUN EINE SEUCHTEN

Deutsch besitzt längere Wörter als englisch, braucht daher auch eine größere Kettenlänge um ausreichend korrekte Wörter zu besitzen.

Lassen wir nun unsere Marks Goethes Gedichtband „Der west-östliche Divan“ im korrekten Zeichensatz lesen:

Mark-2

bie! vch sklelblrür ver aßnsileniche ur, walt wickl do Gendein, m
ulände wie nhtsprkl Massunist Im stt Fühien in Wicherielt, Delateir
ehlsichretzugten wiclr dergebchinim de, Eneiß eie nest. den glt ie
Dageige Geit! duschr, d bl, Serar Soligenuhengernt Goh enn Gume

Mark-3

O, Anzere zumklahst mir ser deret, Siedie du's Haten sird er
wechwerschl men der Dich hört Doch mit eut. Weinke jünftristängt
de, Wen Häum zuwas ind Ein den, warchten, Je Sin Tort, wie Bogen;
Duf Strist friedächt dahmückt genoch wohr ken rer Geselt

Mark-4

sich reisen gern; Katz und. Von das hassanfte, Läßt das Neuen wohl!
Erdenschen Wie an dir dertreifels Kaise, der Stolper wei das Sie
Neuerlicht vers recken meinem Spreißen! Harreiztern aber ob dir nur
sattin Bleinem ret; So zu wir geschlang es zu braxas Als wie

Mark-5

Ich ebens Blicks Gelände Cypressen, Du schickeln: Aber meine,
übergrauen Seitest mir, im Paradieses Paradeis gereine Klein; Trink
Unschlechtet? Hat groß, an sind erfreundenschau, die Hügel bei Tag,
Pracht! Oft ihn am Entschung Zu schwill ich ist du sich waren

Mark-6

groß und Saki gerne dir soll mit die Klüger alsdann umher nach
Behramgur ich legt eine Kuchen. Befinden, Die mir schlaf, Als was
mir Der du, munterschweifen, An des Herzend um Stund umklammre mich
in groß und fort. Du! nichts bang, den Wie sich für uns dann, Soll
man auf Punkt und das Ohr Seite Schilf sich Gar zu verstehn: Ich
für Tumulte!

Mark-10

Dürft ihr euer Tiefstes fragen. Dichten ist ein Übermut! Treib es
gern allein. Freund' und Frauen, frisch von Blut, Kommt nur auch
herein! Was stehst du denn: von Mund zu Ohr Glaubst du denn, wie
vielen Elementen Gottes Stadt gekommen. Mein Erbteil wie herrlich
Eben, weil wir hübsch sind, Möchten wir, im ewigen Leben Muß man
aus dem Stein zu schlagen.

Ist die Kettenlänge zu groß, produziert Mark nur mehr Bruchstücke des Originals.

Ab welcher Kettenlänge treten Eigentümlichkeiten der jeweiligen Sprache auf?

Mark-3: Caesar

pro proviste re, ne adveret, loc naviciisses pos, cello halbat ad
cone con Men Helit dere pribus eti patum cuibus Gallum oppidita,
num terandime car, neque prapiam ceptita cum ata sedie magrum
eaquiciti, optine mum locit. Divi bellexcebatatius

Mark-3: Baudelaire

sembles Au faits ve vortu somme deux cortu, Ne LE Failleur Sans
pui, compor lu vieux sant l'het loude oeux, une peurdu cie à son
poist flemps. Qui more! Les; Tu l'étrage des de chalanglamille,
Escristilleuré, Les aux et de de fer le

Mark-3: Dante

non sa, coria gia; el no con me Per lire puon e a piu` quace si` po
me li a me l'iorda que e ne uorancorbiantole ali fanto fuei che
dove ti davedi e le la' fuon giu`, che ch'i' colpon pinga disonda
de' 'mbel vidiman ta, austrottel corsesdrivi puosti ti che quanti

Mark-3: Cervantes

dió debía comó aquerínsas se da a hijotrióla, losa hacediguijo y
suy más, doliánda vero de esporia del no haya ahera, desallesito?
Puestro, ya dero, queño, lasí turseño engentiréistárse elice
aquesto, no pa pentuvieraride a la tantido a so lorbolunque ficedue
térza no dió oridadon

Mark-3: Leeuwenhoek

eene of zag bulen een vangeelatsel teriereken van ons bloop quam,
Myn die obule naart, of de gewelver van, maantje dat het dit dier
ters buik aar war ik het wel, klinde het L M Leende uyt ens in ter,
dog tert-grontje, eens derie sijn met daater henen gede in hoet
heeuw, als ste ons bloen gebbe, wy ge: sag an ver in waan

Mark-3: 'Robinson Crusoe' in Esperanto

ke ne la jamante, nue de tion. Se nis brigis por mi de imor ilon
sxari tis alpris tiera havanfla lugxis ploj; karlo; sed la sxardo
elo daj ke gxio, pofer al al nubiel lauxrovardompor sxin delke sis
prombaux la vis: Obero, trovus fargo, kiaj enojn surtiu-cxion. Se
li langxas Venu, kajxo tianke vigi vo, pro cxiuuj ne la al

Forschung:

ab welcher Kettenlänge

- lassen sich Autoren unterscheiden?
- ist der Text problemlos laut lesbar (keine unaussprechbaren Silben)?
- ist der Text verständlich?

Sind diese Werte von der jeweiligen Sprache abhängig?

Markov-Affen als Schriftsteller

Erweitern wir unser Konzept – die Affen sollen nicht mehr auf Buchstabenfolgen achten, sondern auf ganze Wortfolgen. Damit vermeiden wir das Auftreten fehlerhafter Einzelwörter. Interpunktionen zählen wir willkürlich als eigene Worte (man könnte sie genausogut als Bestandteil des vorangehenden Wortes ansehen). Werden nun stilistische Merkmale des Autors sichtbar?

Mark-1 Genesis

seiner Erde Gottes Erde Gott seiner, daß Röcke unterworfen. worden
und war und schuf, . einem, lustig Erde genommen Wasser, wie und der
Ich, Weibe. und ihm; Tag Leben ihm groß Feste von beiden Garten
will essen und gegeben Hauptwasser dem Erde jeglichen Edelstein
essen ward

Mark-2 Genesis

und Morgen der Mensch: Du sollst du davon ich dir tragen, wirst;
ich bin nackt waren auf dem Baum, gab mir von allerlei Bäume auf
Erden unter dem Angesicht Gottes des Todes sterben. Und Gott der
Herr sein Weib, lustig anzusehen, das fließt um ihn wäre.

Mark-3 Genesis

Gutes und Böses sollst du dich drauf nähren dein Leben lang. Und er
sprach: Es ist nicht gut, daß es gut war. Und Gott sprach: Laßt uns
Menschen machen, ein jeglicher nach seiner Art. Und Gott sah, daß
sie nackt waren, und er aß.

Auch Lyrik können unsere Affen verfassen, wenn Zeilenumbrüche als eigenes Wort
gelten:

Mark-3 Rilke, Duineser Elegien

Was , wenn der Wind voller Weltraum
die ruhlosen Wege der Erde dort wehe getan.
Liebende, daß ich würde wie sie vergessen !
- Wem es geben ? Am liebsten
kannst du nicht immer
Und du selber, was du willst : unsichtbar
Sehnt es dich aber, so wandeln sie leiser,
und würb ich dich auch Du kommst nicht .

Mark-3 Wagner, Ring des Nibelungen

Hojotoho ! Heiaha !
O seligstes Weib !
(lebhaft sich umwendend)
Nicht Brünne noch Panzer barg meinen Leib :
Der Ewigen Macht ,
ob Wotans grämlichem Grau ,
verlacht nur ward meine fragende List :
gern begehrt' ich den Reif entringt ihr mir nützt ,
fiel dir nicht !
Was je ich riet ,
dem freislichen Felsen zu nahn !
All zu Hauf schichtet den Hort nur erheben ;
zertritt die Traute sein Herr ;

Mark-3 Busch, Max und Moritz

Alle Hühner waren drei
spricht er.
Und schon an.
Guckt nur still und Moritz im Kalten sitze,
mild und Moritz auch dabei.
Hängt an Fäden, Tintenfaß,
Ahnungsvoll tritt sie heraus.
Jammerbild.
Eben geht mit einem Teller
Witwe Bolte in den Keller,
Doch die Käfer, kritze, kratze!
Aber schon sind sie alle beide,
Wenn die frommen Bäckersleut'
Viele süße Zuckersachen
Backen und zurechte machen,
Nimmt die Toten von den Strängen,
Doch sogleich verzehret sie
Meister Müllers Federvieh.
Dieses war der zweite Streich,
Oftmals noch darüber lachten
Und ein stolzer Hahn dabei.

Vielleicht sind unsere Markov-Äffchen als Verfasser von Gesetzestexten geeignet?

Mark-3 Schulorganisationsgesetz

Welche Unterrichtsgegenstände (Studentafel), für Beamte und Vertragsbedienstete in Unteroffiziersfunktion sowie für die im §36. Folgende Formen der allgemein bildenden höheren Schule darf 30 nicht übersteigen. (6) der Grundstufe I umfaßt bei Bedarf sowie die Pflichtfächer und die Förderungsmöglichkeiten sowie die Pflichtfächer und die Schulbehörde erster Instanz unter Bedachtnahme auf den Lehrplan, nach Anhörung des Schulerhalters des Standes , der Hauptschule §15. (1a) Im Lehrplan ist als selbständige Schulen oder in einem solchen Stundenausmaß vorzusehen, dass zusätzliche Lehrplanbestimmungen statt von den einzelnen Schulen von den einzelnen Schulen von den Sonderformen (§ 22) entsprechende Bildung zu vermitteln hat.

Wettbewerb

Wir sammeln die schrägsten Sätze unserer Markov-Äffchen

Ergänzungen

Feststellen der Autorenschaft von Texten / Kompositionen

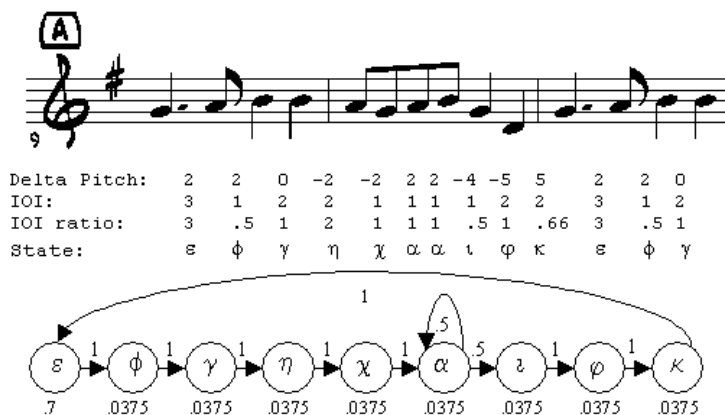
Ein aktuelles Forschungsgebiet ist die Zuordnung unbekannter Texte zu Autoren mithilfe von Markov-Ketten. Dazu werden die Übergangswahrscheinlichkeiten der

Ordnung 2 (in unserer Zählung) aller bekannter Werke des Autors ermittelt und mit denen des unbekanntes Werkes verglichen. Dies führt man für möglichst viele Autoren durch und ermittelt die beste Übereinstimmung. Problem: Hätte Kafka in der Verwandlung nicht den Namen Gregor Samsa, sondern den Namen Willy Durdu benutzt, änderte das die Übergangskoeffizienten seiner Markov-Matrix. Also lässt man aus den zu untersuchenden Texten neben den Satzzeichen auch alle Eigennamen weg. Die amerikanische Internet-Bibliothek Project Gutenberg (die deutsche Version liefert keine reinen ASCII-Texte) ist das Versuchsgelände dieser Statistiker. Derzeit bastelt man an geeigneten Vorschriften zur Textvorbereitung und einer vernünftigen Definition des Begriffes 'ähnliche Übergangsmatrix'.

Praktische Anwendung könnten diese Techniken bei der Entscheidung finden, ob ein wissenschaftlicher Text original oder aus anderen Quellen abgeschrieben ist. Dazu muss man Ähnlichkeitsdefinitionen finden, die beim Verändern kleiner Textpassagen wenig Unterschiede zeigen.

Musik

Auch zur Analyse musikalischer Kompositionen werden Markov-Methoden eingesetzt. Statt Buchstaben bestehen die Texte aus Angaben zu Tonhöhe und Tondauer.



Einige Komponisten verwenden Markov-Generatoren auch zur Erzeugung von Musik. Die Produkte dieser Komponisten sind aber nur schwer von denen der Eddington-Komponisten zu unterscheiden (einzig die Eddington-typischen großen Tonhöhen sprünge bleiben erkennbar aus).

Poesie

Das Gedicht 'Karawane' von Hugo Ball beginnt mit

jolifanto bambla o falli bambla
 großgiga m'pfa habla horem
 egiga goramen
 higo bloiko russula huju
 hollaka hollala

Aus Kurt Schwitters 'Du':

Ich herbe Du.
Ich herbe Deinen Hauch.
Ich singe Deine Augen.
Dein Schreiten sehnt meine Augen.
Dein Plaudern sehnt mein Ohr.
Ich lechze Duft die Stunden.

Diese Dadaistischen Werke (um 1920) ähneln stark unserer Markov-Lyrik.

Der amerikanische Dichter Jeff Harrison experimentiert heute mit Markov-Generatoren. Bei Wikipedia kann man Ausschnitte seiner aleatorischen Werke 'Postmortem Series' und 'Accuracy' lesen.

Chatbots

Chatrooms im Internet sind eine beliebte Spielwiese für Programme zur automatischen Texterzeugung. Neben traditionellen Algorithmen (Eliza) werden auch gerne Markov-Textgeneratoren als Gesprächspartner (freilich ohne Wissen der Teilnehmer) eingesetzt. Ein derartiges Programm von Jason Hutchens 1998 konnte 'beinahe' den Turing Test bestehen (HAL bzw. MegaHAL siegten jedoch bei der Loebner-Preis Variation). Die wissenschaftliche Arbeit dazu trägt den schönen Titel 'How to pass the Turing test by cheating'.

Mark V. Shaney ist das Pseudonym eines berühmt/berüchtigten 'Fake Users', dessen Usenet-Beiträge allesamt aus Markov-Transformationen anderer Beiträge stammen. Das Programm wurde von Bruce Ellis geschrieben, der Name ist eine Verballhornung von 'Markov-Chain' und dem Namen Shannon, einem Pionier der statistischen Textanalyse.

Internet

Wählt man als Zeichenvorrat alle HTML-Seiten des WWW und betrachtet verfolgte Links als Übergang zu einem neuen Zeichen, kann man das Surf-Verhalten der Internetnutzer als Markov-Prozess ansehen. Wählt man dann noch eine Wahrscheinlichkeit, mit der Seitenlinks verfolgt werden (meines Wissens 85%) statt eine andere Seite zu besuchen, so kann man 'wichtige' von 'unwichtigen' Web-Seiten unterscheiden. Google berechnet die Rangfolge von Suchtreffern auf diese Weise.

Extreme Experimente

Wir können ein literarisches Werk im Sinne der Mark-2 Affen in eine Markov-Übergangsmatrix verwandeln ('Ordnung 1' in der Matrix-Sprechweise) und diese zur Texterzeugung heranziehen.

Was passiert, wenn wir mit dieser Matrix herumspielen? Etwa Faust zum Quadrat bilden, indem wir jede Wahrscheinlichkeit quadrieren? Oder die Differenz aus Nathan dem Weisen und Till Eulenspiegel bilden? (Anpassung der Wahrscheinlichkeiten ist nötig, $P(\text{Übergang nach irgendwohin})=1$ muss gelten.) So verlockend diese Ideen klingen – die erzeugten Texte besitzen leider noch weniger Qualität als die 'gewöhnlichen'.