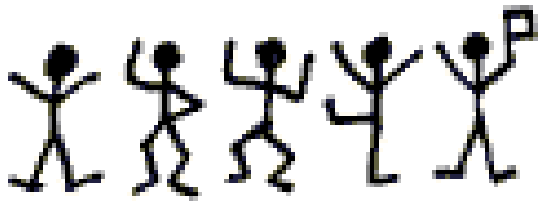
El arte de romper códigos secretos



Carlos Sarraute

19 mayo 2006

Origen concreto de teorías abstractas

- Muchas teorías matemáticas tienen como origen tratar de resolver problemas de la vida real

- Aritmética
  - contar vacas, comercio, ...

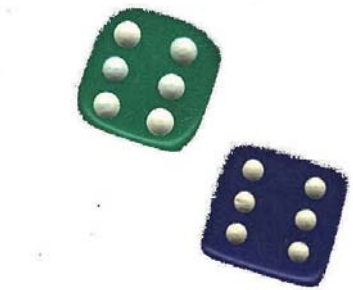
- Geometría
  - medir campos, arquitectura (Tiwanaku), ...

- Probabilidades (siglo 17)

- Criptografía

El Chevalier de Méré

- Chevalier de Méré, noble francés del siglo 17, escritor, filósofo, bon vivant, jugador
- Inventó un juego: ganaba si sacaba un 6 en 4 tiradas de un dado
- Ganaba mucho, la gente no quería jugar más
- Inventó otro juego: ganaba si sacaba un doble 6 en 24 tiradas de dos dados



Una paradoja para el Caballero

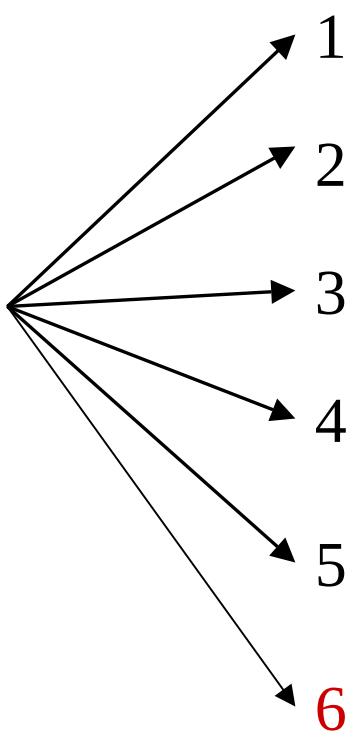
- El razonamiento del Caballero:

tirar un dado	6 posibilidades	4 tiradas
tirar dos dados	36 posibilidades	24 tiradas

- Problema: empezó a perder plata!

- Le preguntó a su amigo Pascal que estaba pasando

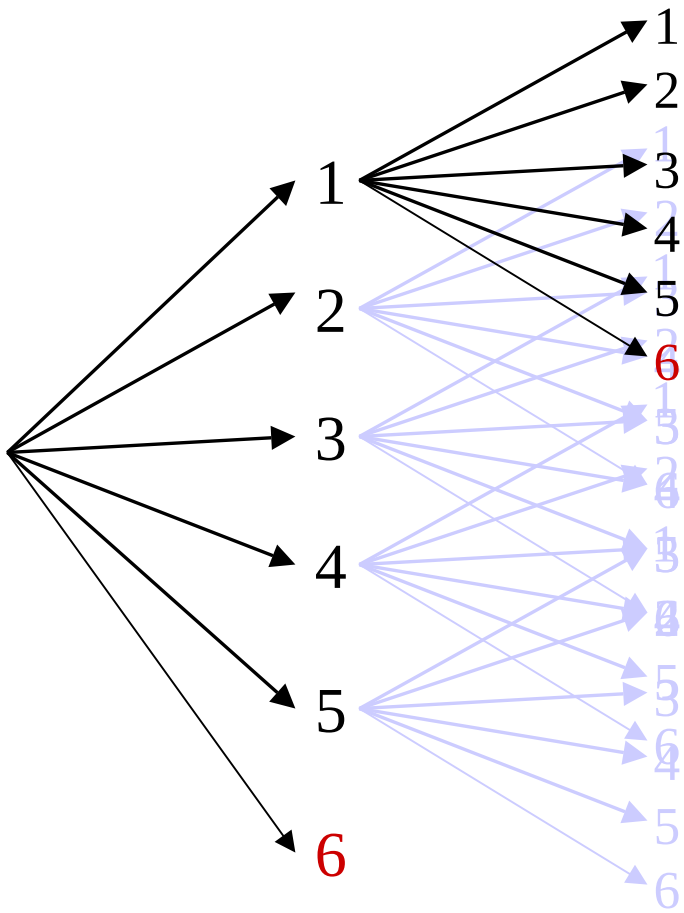
Con un dado: la probabilidad de perder



- tirando una vez :

  5/6 = 0,833... = 83,3 %

La probabilidad de perder



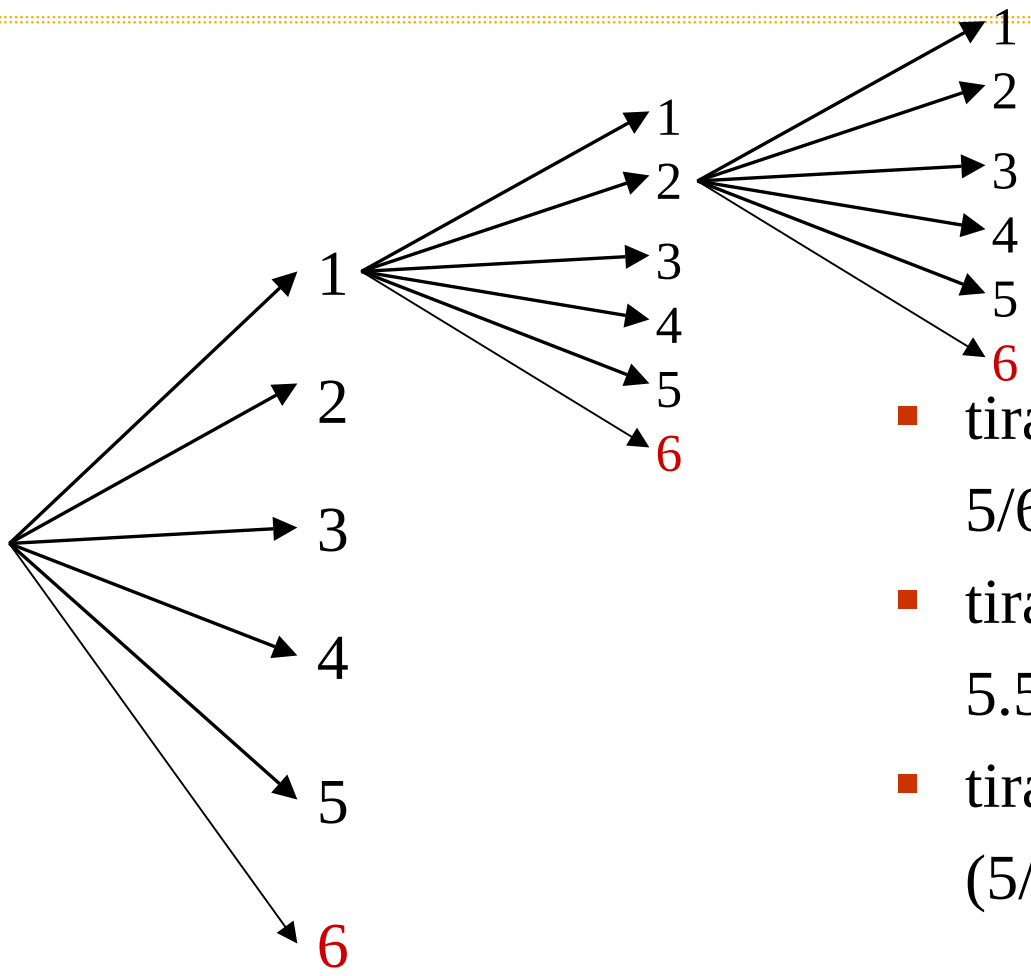
- tirando una vez :

  5/6 = 0,833...

- tirando dos veces :

  5.5 / 6.6 = (5/6)^2 = 0,694...

La probabilidad de perder



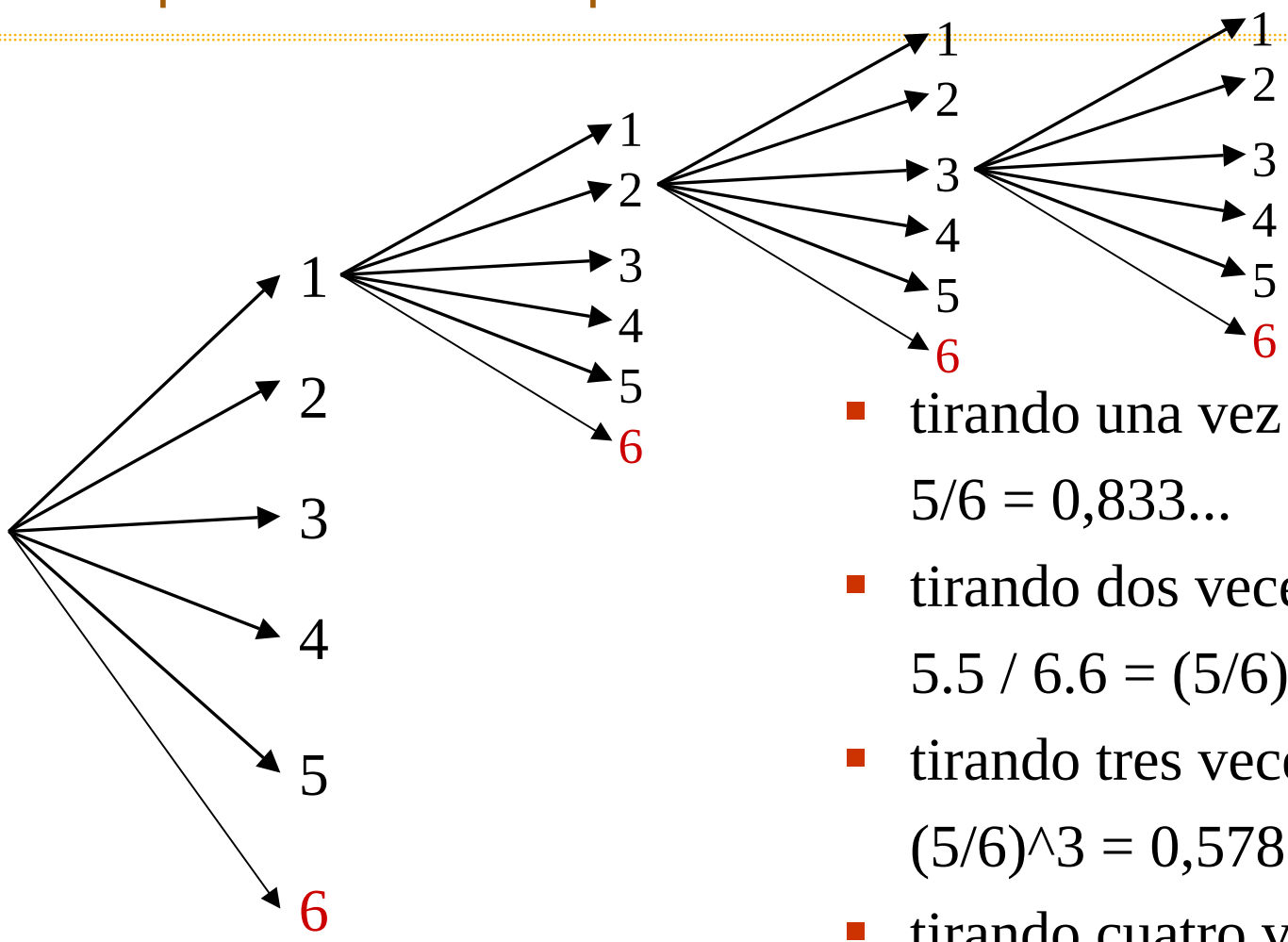
- tirando una vez :

  5/6 = 0,833...
- tirando dos veces :

  5.5 / 6.6 = (5/6)^2 = 0,694...
- tirando tres veces :

  (5/6)^3 = 0,578...

La probabilidad de perder



- tirando una vez :

  5/6 = 0,833...
- tirando dos veces :

  5.5 / 6.6 = (5/6)^2 = 0,694...
- tirando tres veces :

  (5/6)^3 = 0,578...
- tirando cuatro veces :

  (5/6)^4 = 0,482...

La intervención del geómetra

$1 - (5/6)^4 = 0{,}517...$

$1 - (35/36)^{24} = 0{,}491...$

- Este es el principio de la teoría de las probabilidades

Un homenaje a la intuición de Méré

$1 - (5/6)^4 = 0{,}517...$

$1 - (35/36)^{24} = 0{,}491...$

$1 - (35/36)^{25} = 0{,}505...$

de Pascal a Kolmogorov

- Siglo 17: Pascal y Fermat inventan la teoría de las probabilidades

- durante 3 siglos se resuelven problemas y se empiezan a usar probabilidades en otras ciencias

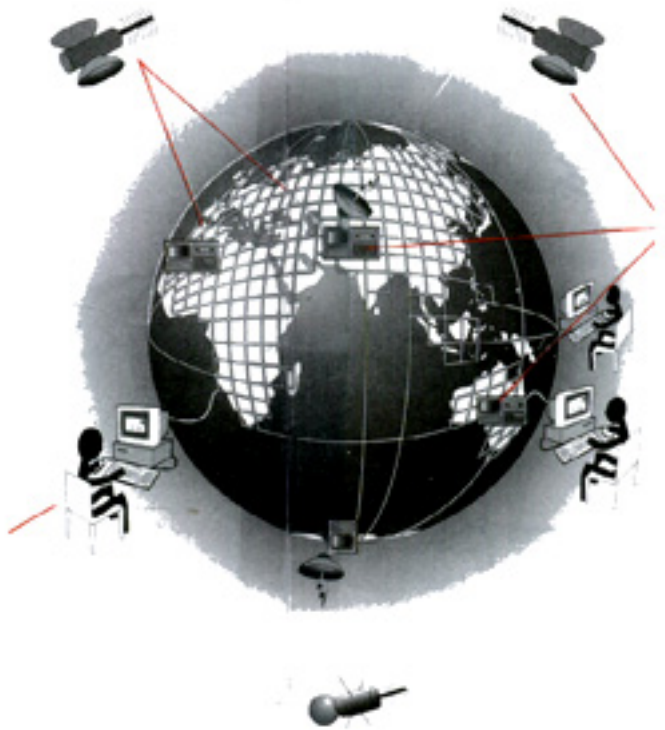
- Siglo 20: el matemático ruso Kolmogorov formaliza la teoría de las probabilidades

Criptografía

- El problema del mundo real: la necesidad de criptografía
- La criptografía es el estudio de las comunicaciones en presencia de adversarios
- Ejemplo típico = sistemas para establecer comunicaciones secretas sobre un canal inseguro

Necesidad de privacidad en la Internet

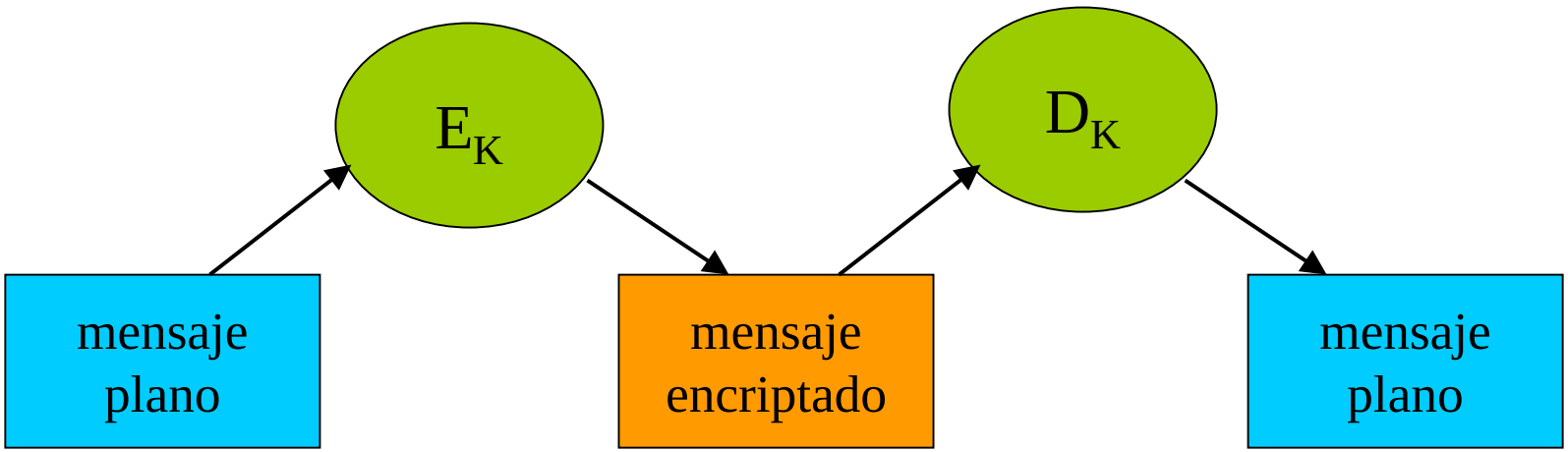
- Ejemplo más concreto: mensajes privados / compras por Internet
- Información privada
  - cartas
  - número de tarjeta de crédito
- Circula por muchos servidores entre el cliente y el negocio
- Como proteger esa información?



Sistema de encripción

- Encriptar la información (el mensaje)

- Transformar el mensaje usando un sistema de encripción y una clave secreta



Y la seguridad?

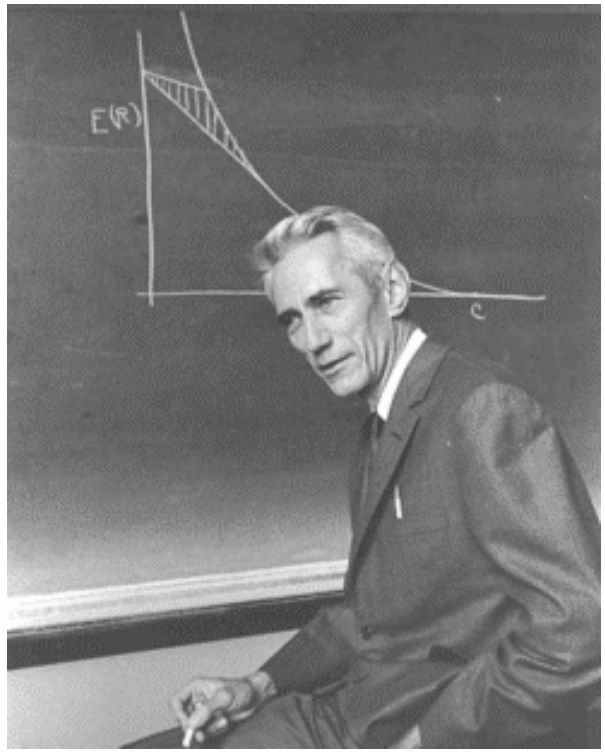
- pregunta: como saber si un sistema es seguro?
- definición: es seguro si resiste a los ataques
- Criptoanálisis = el arte de romper códigos secretos
  - bueno, arte y ciencia
  - métodos artesanales
    - como las probabilidades del siglo 17
  - búsqueda de formalización y generalización

Estaría bueno poder demostrar...

- Por que intervienen los matemáticos?
- muchos sistemas están basados en problemas matemáticos que son difíciles
- demostrar que un sistema es seguro contra cualquier atacante

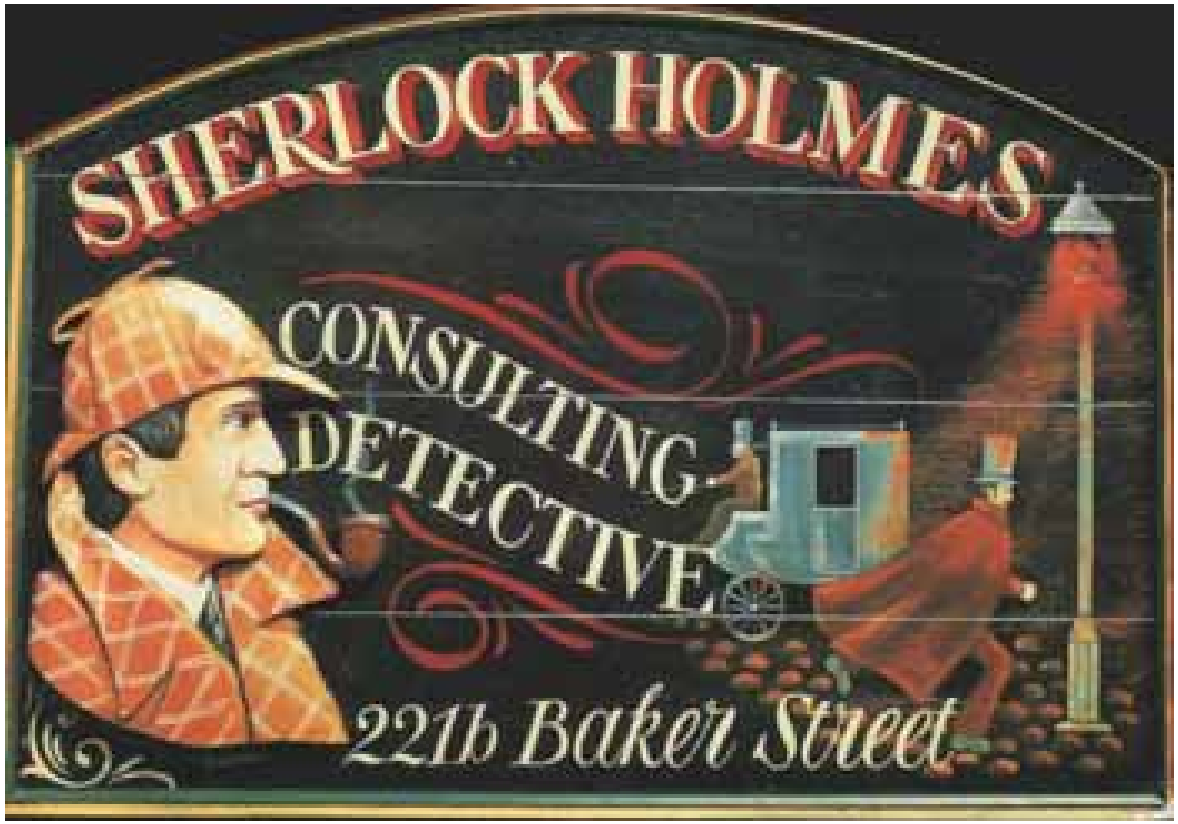
El secreto perfecto

- Shannon definió el secreto perfecto (1948)
  - un sistema que resiste a cualquier atacante, incluso con poder de cómputo y tiempo infinito
  - problema: la clave tiene que ser tan larga como el texto
  - existe, es el one time pad
    - » se usaba para las comunicaciones entre Washington y el Kremlin



Ejemplo de criptoanálisis

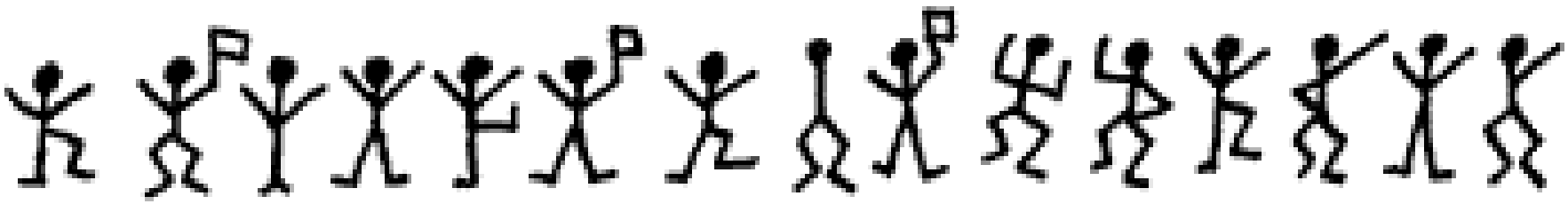
- sacado de la vida real!



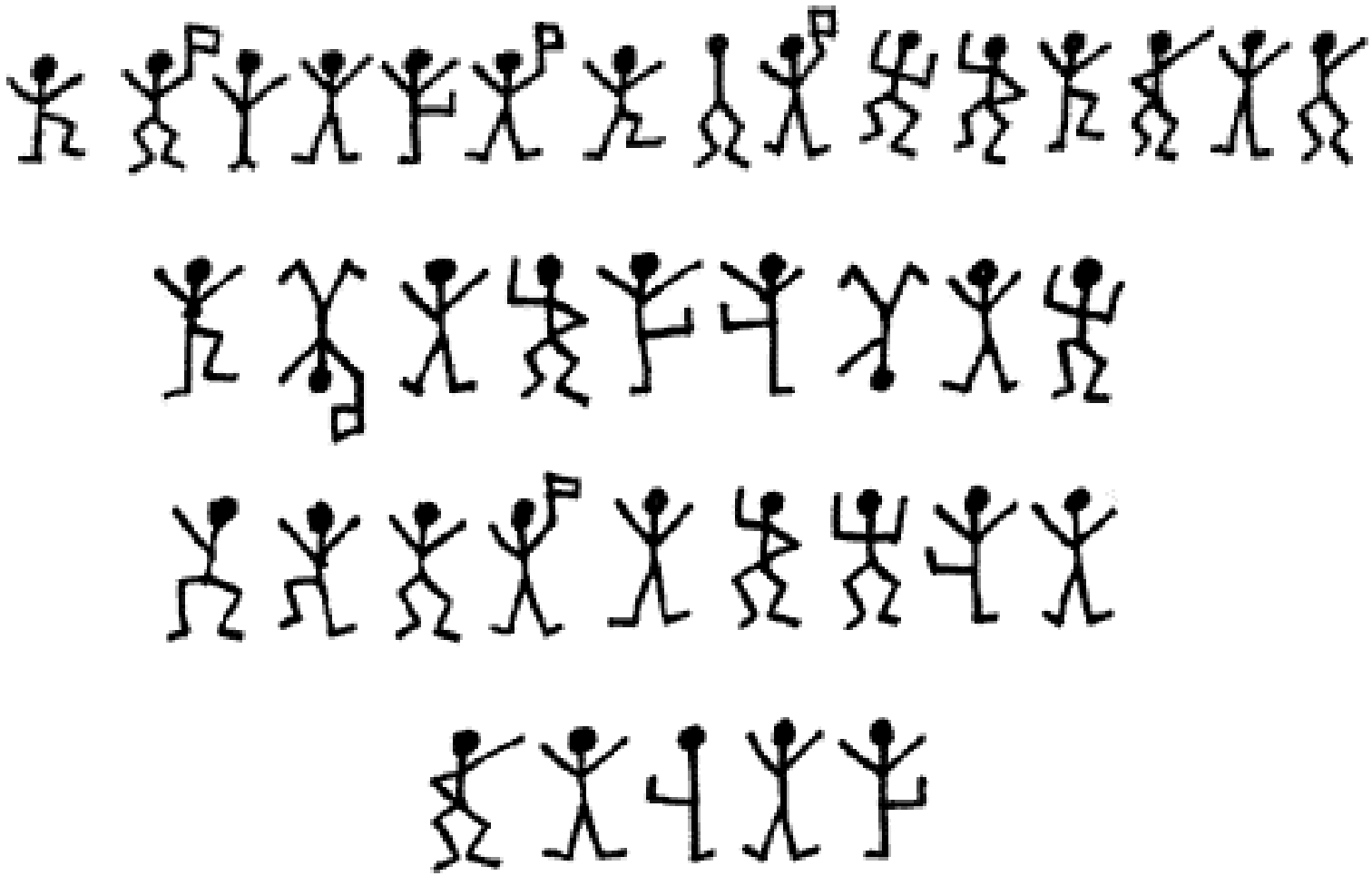
La Aventura de los Hombres que Bailan

- Un hombre inglés se casó con una mujer americana, Elsie, hermosa pero extraña – no quiere contar nada sobre su pasado.

- El señor encontró unos hombrecitos dibujados con tiza en la pared de su casa
  - su mujer se desmayó del susto



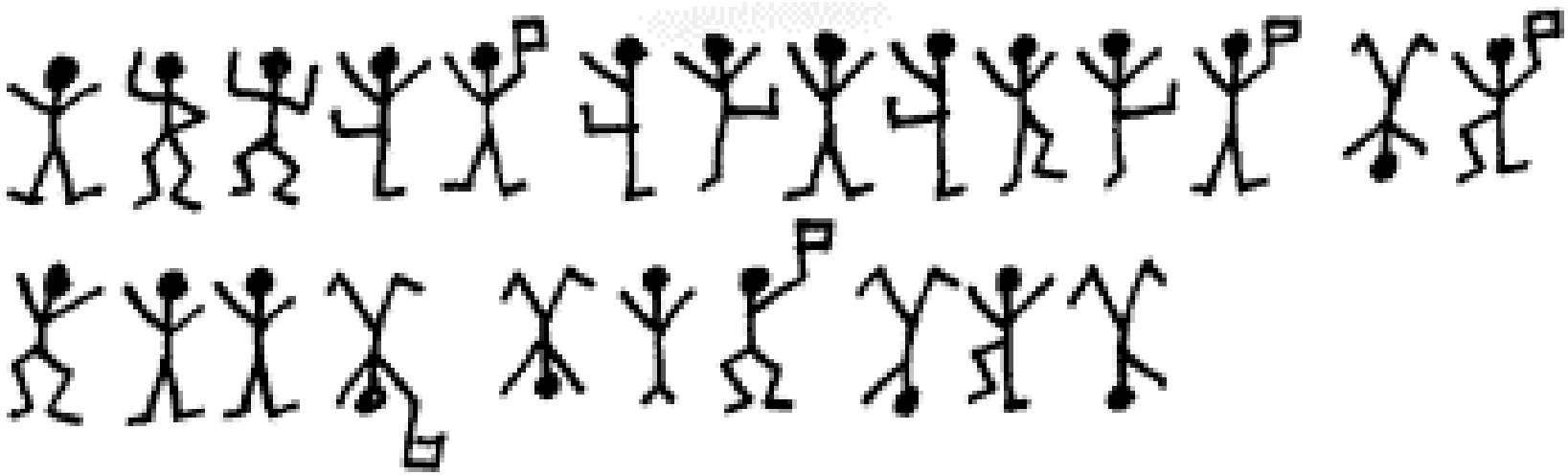
Mas dibujitos con hombres que bailan



Cuentas...

- Sherlock Holmes se pasa una tarde haciendo cuentas y dibujitos, hasta que pegó un grito de satisfacción y se frotó las manos

- El cliente trajo un nuevo mensaje que preocupó mucho a Sherlock



Permutación de las letras del alfabeto

- cada símbolo representa una letra
- 26 símbolos, 26 letras en el alfabeto
- hay 26.25.24.23.22. ... 3.2.1 = 26!
  formas de asociar los símbolos a las letras
- 26! = 403.291.461.126.605.635.584.000.000

Frecuencia de los símbolos

- Contemos la frecuencia de los símbolos:

17 

5 

5 

4 

4 

4 

3 

Frecuencia de las letras (en inglés)

E	11.42%	U	3.66%
A	8.56%	P	3.27%
I	7.94%	M	3.22%
R	7.51%	D	3.13%
T	7.46%	H	2.76%
O	7.12%	G	2.30%
N	6.41%	B	2.12%
S	5.55%	Y	2.00%
L	5.52%	F	1.47%
C	4.74%	V	1.07%

Método manual

- Con esta información se puede seguir a mano

 = E  = A, I, R, T, O

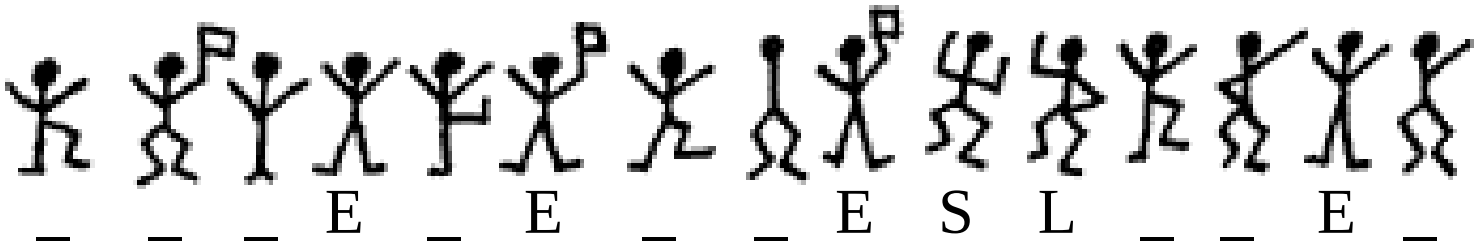
- Las banderitas separan las palabras

- Usar además la información de contexto

  por ejemplo, la mujer se llama Elsie, una palabra de 5 letras que empieza y termina con E es probable que sea

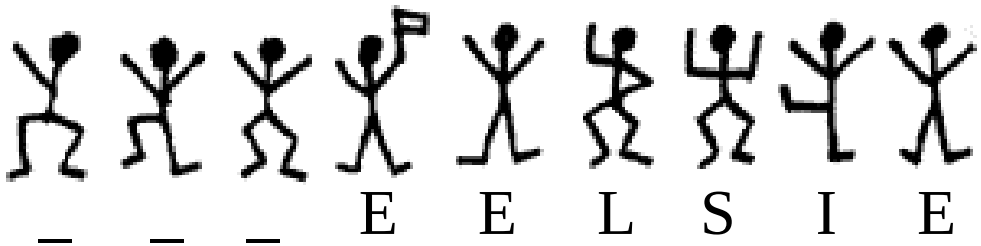
E L S I E

Principio de resolución con E L S I ...



_ _ _ E _ E _ _ E S L _ _ E _

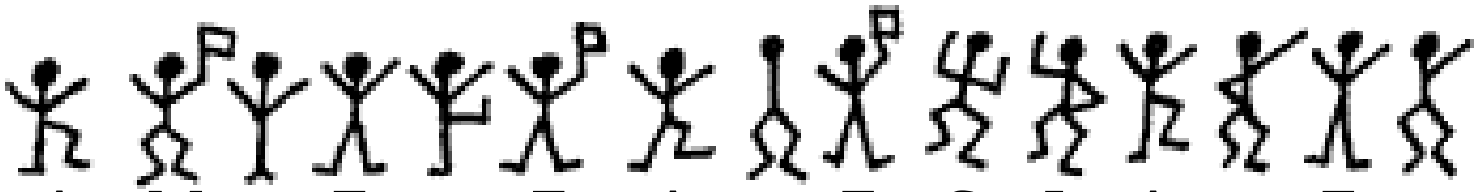
_ _ E L _ I _ E S



_ _ _ E E L S I E

_ E _ E _

Principio de resolución con E L S I ... A M ...



A M _ E _ E A _ E S L A _ E _

A _ E L _ I _ E S

_ _ M E E L S I E

_ E _ E _

Ataque automatizado

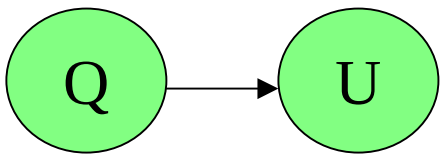
- Aproximación al español de primer orden

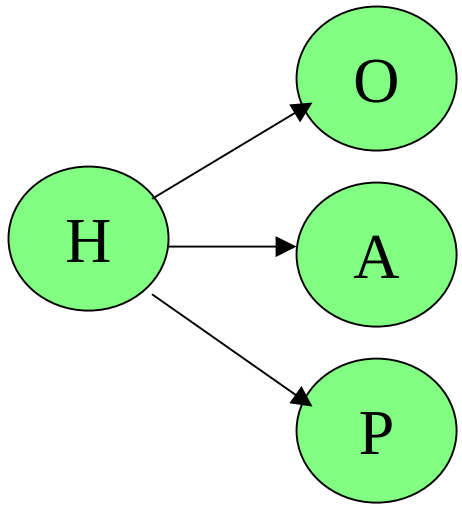
- El español como proceso aleatorio
  - proceso no determinístico, que respeta una distribución de probabilidades
  - monos tipeando al azar respetando la probabilidad de las letras

Ataque automatizado

- Aproximación al español de segundo orden
  - cadena de Markov
  - proceso aleatorio con diferentes estados (las letras)
  - probabilidades de transición (de pasar de un estado a otro)

- probabilidades condicionales





Algoritmo de California

- voy probando con los símbolos ordenados por frecuencia
  para cada símbolo pruebo todas las posibilidades

- reemplazo en el texto encriptado los símbolos que ya tengo
  le asigno un puntaje que representa la probabilidad de que sea español (aproximación de 2do orden)

- me quedo con los 100 mejores (entre 26*100)

- cuando termine, el mensaje estará entre los 100 mejores (de hecho, entre los 10 mejores)

Muchas gracias!

carlos @ coresecurity.com