

Linguistic properties of multi-word passphrases

Joseph Bonneau, Ekaterina Shutova

`jcb82, es407@cl.cam.ac.uk`



UNIVERSITY OF
CAMBRIDGE

Computer Laboratory

USEC WORKSHOP ON USABLE SECURITY 2012

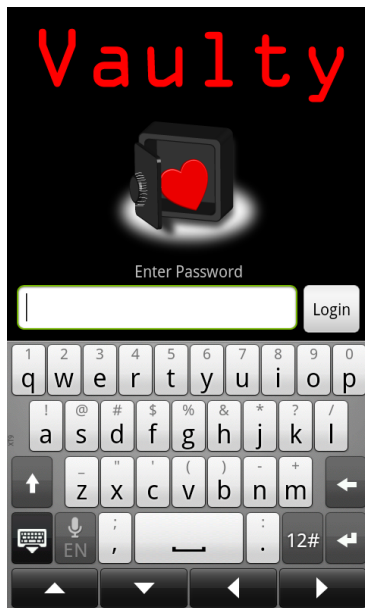
KRALENDIJK, BONAIRE, NETHERLANDS

MARCH 2, 2012

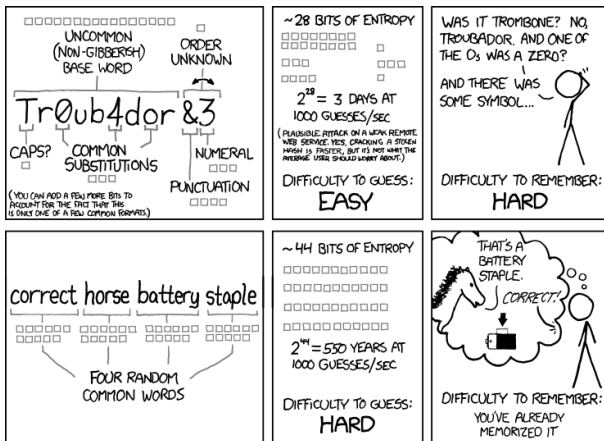
Passphrases an increasingly attractive approach



Passphrases an increasingly attractive approach



Passphrases an increasingly attractive approach



THROUGH 20 YEARS OF EFFORT, WE'VE SUCCESSFULLY TRAINED EVERYONE TO USE PASSWORDS THAT ARE HARD FOR HUMANS TO REMEMBER, BUT EASY FOR COMPUTERS TO GUESS.

xkcd #936

What do we know about passphrase guessing?

[this space intentionally left blank]

Data source: Amazon PayPhrases

Enter your own phrase to see if it's available or claim this one:

“

Extraordinary Secure

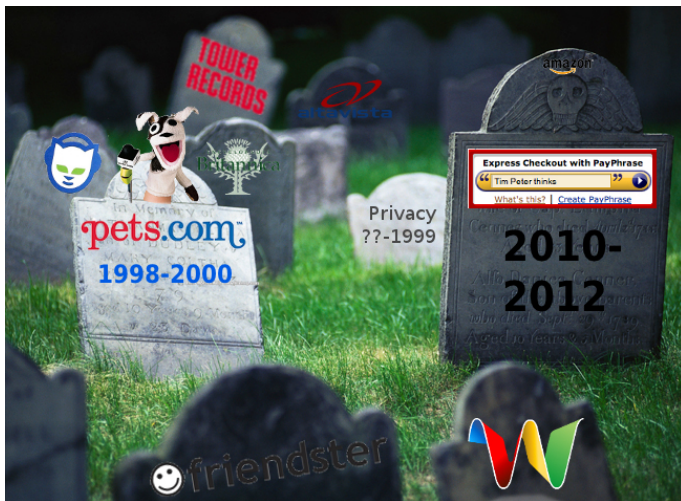
”

(PayPhrase must be at least 2 words long, no numbers or special characters)

✗ Sorry, this phrase is not available

- must be at least two words
- must be globally unique
 - security \leftarrow PIN + passphrase
- can only contain the letters a-z, A-Z, SPACE
 - capitalisation and spacing ignored

Data source: Amazon PayPhrases



PayPhrases killed 2012-02-20

A simple dictionary attack



WIKIPEDIA
The Free Encyclopedia

Main page

Contents

Featured content

Current events

Random article

Donate to Wikipedia

Interaction

Help

About Wikipedia

Community portal

Recent changes

Contact Wikipedia

Toolbox

Print/export

Languages

Esperanto

Suomi

Portugués

Русский

Article Talk

Read View source

List of board games

From Wikipedia, the free encyclopedia

This is a list of **board games**. This page classifies board games according to the concerns which might be uppermost for someone organizing a gaming event or party. See the article on [game classification](#) for other **also** board game articles.

Contents [show]

Two-player abstract strategy games

Main article: *List of abstract strategy games*

In **abstract strategy games**, players know the entire game state at all times, and random generators such as dice are not used.

Two-player games

- | | | |
|--|---|--|
| <ul style="list-style-type: none">• Abalone• Agon• Backgammon• Battleship• Blood Bowl• Bul• Chess• Chinese Checkers• Connect Four• The Cross and Circle game family• Daldøs• Dwarffall• DVONN• English draughts (Checkers)• Fanorona• Ghosts• Gipl | <ul style="list-style-type: none">• Go• Guess Who?• Hare and Hounds• Hjara• Isola• Janggi (Korean Chess)• Kalah• Kamisado• Kingdoms• Liu gi• Lost Cities• Mad Gab• Master Mind• Nine Men's Morris• Obsession• Patoli• Pawn duel | <ul style="list-style-type: none">• Plateau• PÜNCT• Reversi (aka Othello)• Rehmismachy• Sähkkü• Senat• Shogi (Japanese Chess)• Snakes and Ladders• Space Hulk• Strategie• Sugoroku• Tāb• Tetrix• Wari• Xiangqi (Chinese Chess)• YINSH• ZERTZ |
|--|---|--|

Multi-player elimination games

Participants are typically eliminated before game end.

- | | | |
|---|--|--|
| <ul style="list-style-type: none">• 1313 Dead End Drive• American Megalania• Anti-Monopoly• Apples to Apples• Attack!• Axis & Allies• Bang!• ... | <ul style="list-style-type: none">• Crash! The bankrupt game• Diplomacy• Djambi• Finance• Ghettopoly• The Great Train Robbery Board Game• Hotels | <ul style="list-style-type: none">• Risk• Shogun/Samurai Swords• Solarquest• Spy Alley• Star Wars Epic Duels• Star Wars Tactics• Strange Synergy |
|---|--|--|

- proper nouns
- titles
- idiomatic phrases
- slang

A simple dictionary attack

The screenshot shows the IMDb website's 'All-Time Worldwide Box office' chart. The page has a dark header with the IMDb logo and navigation links. The main content area is divided into a left sidebar with various links and a main table of box office data.

IMDb Charts

- [Main Index](#)
- [IMDb Top 250](#)
- [IMDb Bottom 100](#)

US Box Office

- [USA Top 10](#)
- [USA Archive](#)

UK Box Office

- [UK Top 10](#)
- [UK Archive](#)

All-Time Box Office

- [USA](#)
- [UK](#)
- [World-wide](#)

DVD Rentals

- [USA Weekly Top 20](#)
- [USA Archive](#)

Votes by Genre

- [Action](#)
- [Adventure](#)
- [Animation](#)
- [Biography](#)
- [Comedy](#)
- [Crime](#)
- [Documentary](#)
- [Drama](#)
- [Family](#)
- [Fantasy](#)
- [Film Noir](#)
- [History](#)
- [Horror](#)
- [Independent](#)
- [Music](#)
- [Musical](#)
- [Mystery](#)
- [Romance](#)
- [Sci-Fi](#)

All-Time Worldwide Box office

Rank	Title	Worldwide Box Office
1.	Avatar (2009)	\$2,781,505,847
2.	Titanic (1997)	\$1,835,300,000
3.	Harry Potter and the Deathly Hallows - Part 2 (2011)	\$1,327,855,819
4.	The Lord of the Rings: The Return of the King (2003)	\$1,119,102,868
5.	Transformers: Dark of the Moon (2011)	\$1,114,558,779
6.	Pirates of the Caribbean: Dead Man's Chest (2006)	\$1,065,896,541
7.	Toy Story 3 (2010)	\$1,062,884,497
8.	Pirates of the Caribbean: On Stranger Tides (2011)	\$1,041,963,875
9.	Alice in Wonderland (2010)	\$1,023,285,206
10.	The Dark Knight (2008)	\$1,001,821,825
11.	Harry Potter and the Philosopher's Stone (2001)	\$969,457,891
12.	Pirates of the Caribbean: At World's End (2007)	\$959,404,152
13.	Harry Potter and the Deathly Hallows - Part 1 (2010)	\$949,880,434
14.	Harry Potter and the Order of the Phoenix (2007)	\$937,000,866
15.	Harry Potter and the Half-Blood Prince (2009)	\$933,956,480
16.	Star Wars: Episode I - The Phantom Menace (1999)	\$922,379,200
17.	The Lord of the Rings: The Two Towers (2002)	\$901,800,000
18.	Jurassic Park (1993)	\$919,700,000
19.	Harry Potter and the Goblet of Fire (2005)	\$892,194,197
20.	Ice Age: Dawn of the Dinosaurs (2009)	\$887,773,705
21.	Spider-Man 3 (2007)	\$885,430,103
22.	Shrek 2 (2004)	\$880,871,096
23.	Harry Potter and the Chamber of Secrets (2002)	\$869,300,000
24.	Finding Nemo (2003)	\$865,000,000
25.	The Lord of the Rings: The Fellowship of the Ring (2001)	\$860,700,000
26.	Star Wars: Episode III - Revenge of the Sith (2005)	\$848,462,555

- proper nouns
- titles
- idiomatic phrases
- slang

A simple dictionary attack

The screenshot shows the UsingEnglish.com website. The main navigation bar includes links for Home, Members, Students, Teachers, Forums, Testing, **Reference**, Articles, Resources, and Shop. Below this, there are sub-links for Idioms, Phrasal Verbs, Irregular Verbs, Grammar Glossary, English Dictionaries, and Downloads. The page title is "English Idioms & Idiomatic Expressions". A search bar shows "A" entered, and the results show "Showing 1-50 of 100 results for letter 'A'". The results are listed as follows:

- A bit much**
If something is excessive or annoying, it is a bit much.
- A bridge too far**
A bridge too far is an act of overreaching- going too far and getting into trouble or failing.
- A chain is no stronger than its weakest link**
This means that processes, organisations, etc., are vulnerable because the weakest person or part can always damage or break them.
- A day late and a dollar short**
(US) If something is a day late and a dollar short, it is too little, too late.
- A fool and his money are soon parted**
This idiom means that people who aren't careful with their money spend it quickly. 'Fool and his money are easily parted' is an alternative form of the idiom.
- A fool at 40 is a fool forever**
If someone hasn't matured by the time they reach forty, they never will.
- A fresh pair of eyes**
A person who is brought in to examine something carefully is a fresh pair of eyes.
- A hitch in your giddy-up**
If you have a hitch in your giddy-up, you're not feeling well. (Which in your giddy-up is also used.)
- A lick and a promise**
If you give something a lick and a promise, you do it hurriedly, most often incompletely, intending to return to it later.
- A list**
Prominent and influential people who comprise the most desirable guests at a social function or gathering.
- A little bird told me**
If someone doesn't want to say where they got some information from, they can say that a little bird told them.
- A little learning is a dangerous thing**

- proper nouns
- titles
- idiomatic phrases
- slang

A simple dictionary attack

urban

DICTIONARY

Like

124

look up anything, like you city

search

[random](#)
[A](#)
[B](#)
[C](#)
[D](#)
[E](#)
[F](#)
[G](#)
[H](#)
[I](#)
[J](#)
[K](#)
[L](#)
[M](#)
[N](#)
[O](#)
[P](#)
[Q](#)
[R](#)
[S](#)
[T](#)
[U](#)
[V](#)
[W](#)
[X](#)
[Y](#)
[Z](#)
[new](#)
[favorites](#)

index

[AA](#)
[AB](#)
[AC](#)
[AD](#)
[AE](#)
[AF](#)
[AG](#)
[AH](#)
[AI](#)
[AJ](#)
[AK](#)
[AL](#)
[AM](#)
[AN](#)
[AO](#)
[AP](#)
[AQ](#)
[AR](#)
[AS](#)
[AT](#)
[AU](#)
[AV](#)
[AW](#)
[AX](#)
[AY](#)
[AZ](#)

Most popular words in A

[Show all 99,999 words in A](#)

<div>A</div> <div> aas Aabian Pindragon abc use abortion where Ada Internet and aardole and then I found five dollars aas a cingula ag pain Advertisment and sex achieve agony pains academic bulletin aafat angry dragon aden Ad-nabul Anzokunukiti Anzokunukiti Anzokunukiti and sex Afin my Aish Salmon Aish Salmon aas aardole afterglow achley with clauche Aomi Anva apocalypse sex Ann Coulter Ar quote Angelina jolie Audience typing Ain Aweesome Ar ank Aurembia Albheat and </div>	<div>A</div> <div> autisme Amber angelina jolie autismagically and airport valtures alabama her pocket aafat Angelina Jolie adain angelina jolie Angelina Jolie adain angelina jolie Ad You know Angelina Jolie angelina jolie Angelina Jolie agale Angelina Jolie Aligator Pushchaise Alyssa Angelina Jolie ayme Angelina Jolie ad Angelina Jolie Abe Lincoln a alcohol Anonymous approval from corporate agrotic achward arm AWOL Anthony angelina jolie Andre aboutphobia Achley Austine Al Gore aas Abbasid Pindragon angelina jolie Airon </div>	<div>A</div> <div> Angry Dragon aafat Antisocial Networking Amanda an feni And, Sen andrew aafat Ahhett achward turtle ahy andre andre andre Andre Andre Anonymous anocivn Andre Angelina Jolie Andre Andre all your base are belong to us ANTISOCIAL AND PITCH anestist aaron aaw crumble achley America Angelina Jolie andrew Allison sawing ACH andrea ADG sight aif aif aif soul Chris Coulter Australia ACH Anonymous sewards sight angelina jolie </div>
--	--	---

- proper nouns
- titles
- idiomatic phrases
- slang

Results

word list	example	list size	success rate	\hat{p}
<i>arts</i>				
musicians	three dog night	679	49.5%	0.0464%
albums	all killer no filler	446	56.5%	0.0372%
songs	with or without you	476	72.9%	0.0623%
movies	dead poets society	493	69.6%	0.0588%
movie stars	patrick swayze	2012	28.1%	0.0663%
books	heart of darkness	871	47.0%	0.0553%
plays	guys and dolls	75	70.7%	0.0093%
operas	la gioconda	254	17.3%	0.0048%
TV shows	arrested development	836	46.3%	0.0520%
fairy tales	the ugly duckling	813	13.3%	0.0116%
paintings	birth of venus	268	11.2%	0.0032%
brand names	procter and gamble	456	17.3%	0.0087%
<i>total</i>		7679	38.5%	0.4159%
<i>sports teams</i>				
NHL	new jersey devils	30	83.3%	0.0056%
NFL	arizona cardinals	32	87.5%	0.0070%
NBA	sacramento kings	29	93.1%	0.0085%
MLB	boston red sox	30	90.0%	0.0074%
NCAA	arizona wildcats	126	56.3%	0.0105%
fantasy sports	legion of doom	121	71.1%	0.0151%
<i>total</i>		368	71.7%	0.0542%
<i>sports venues</i>				
professional stadiums	soldier field	467	14.1%	0.0071%
collegiate stadiums	beaver stadium	123	12.2%	0.0016%
golf courses	shadow creek	97	6.2%	0.0006%
<i>total</i>		687	12.7%	0.0094%

Results

word list	example	list size	success rate	\hat{p}
<i>games</i>				
board games	luck of the draw	219	28.8%	0.0074%
card games	pegs and jokers	322	27.6%	0.0104%
video games	counter strike	380	28.4%	0.0127%
<i>total</i>		921	28.2%	0.0306%
<i>comics</i>				
print comics	kevin the bold	1029	29.5%	0.0361%
web comics	something positive	250	16.8%	0.0046%
superheros	ghost rider	488	45.3%	0.0295%
<i>total</i>		1767	32.1%	0.0701%
<i>place names</i>				
city, state (USA)	plano texas	2705	33.8%	0.1117%
multi-word city (USA)	maple grove	820	79.0%	0.1283%
city, country	lisbon portugal	479	35.7%	0.0212%
multi-word city	ciudad juarez	55	69.1%	0.0066%
<i>total</i>		4059	43.7%	0.2677%
<i>phrases</i>				
sports phrases	man of the match	778	26.1%	0.0235%
slang	sausage fest	1270	45.0%	0.0761%
idioms	up the creek	3127	43.6%	0.1789%
<i>total</i>		5175	41.3%	0.2785%

Results

- Estimating $N = 10^6$, our 20k dictionary covers 1.1% of users
 - Equivalent to 20.8 bits
- Password comparison #1: 2 passwords cover 1.1% of users
 - Equivalent to 7.5 bits
- Password comparison #2: 20k dictionary covers 26.3% of users
 - Equivalent to 16.3 bits

Similar to mnemonic-phrase passwords

Which syntactic construction do users prefer?

Which syntactic construction do users prefer?

bigram type	example	list size	success rate
adverb-verb	probably keep	4999	5.0%
verb-adverb	send immediately	4999	1.9%
direct object-verb	name change	5000	1.2%
verb-direct object	spend money	5000	2.4%
verb-indirect object	go on holiday	4999	0.7%
nominal modifier-noun	operation room	4999	9.8%
subject-verb	nature explore	4999	1.3%

Phrases generated from British National Corpus/Robust Accurate Statistical Parser

Single objects or actions strongly preferred

Which factors predict a phrase's popularity?

Which factors predict a phrase's popularity?

bigram type	example	list size	success rate
adjective-noun	powerful form	10000	13.3%
noun-noun	island runner	10000	4.4%

Phrases generated from Google n-gram corpus

Which factors predict a phrase's popularity?

Possible selection models:

baseline

random

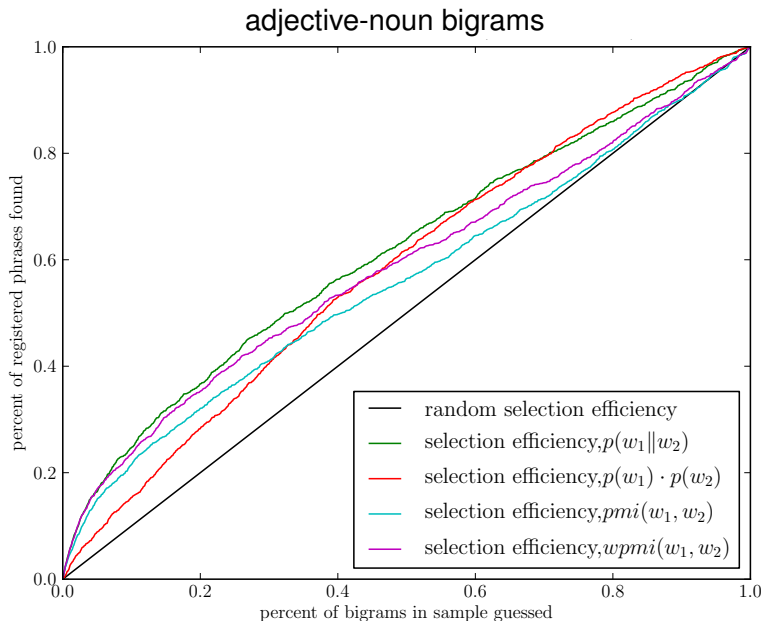
natural-language production $p(w_1 || w_2)$

independent word selection $p(w_1) \cdot p(w_2)$

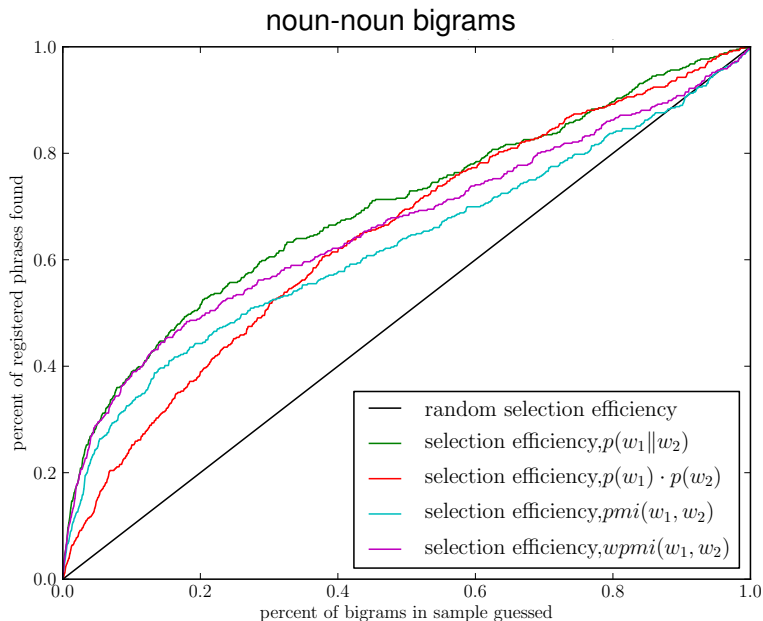
mutual information $pmi(w_1, w_2) = \lg \frac{p(w_1 || w_2)}{p(w_1) \cdot p(w_2)}$

blended model $p(w_1 || w_2) \cdot pmi(w_1, w_2)$

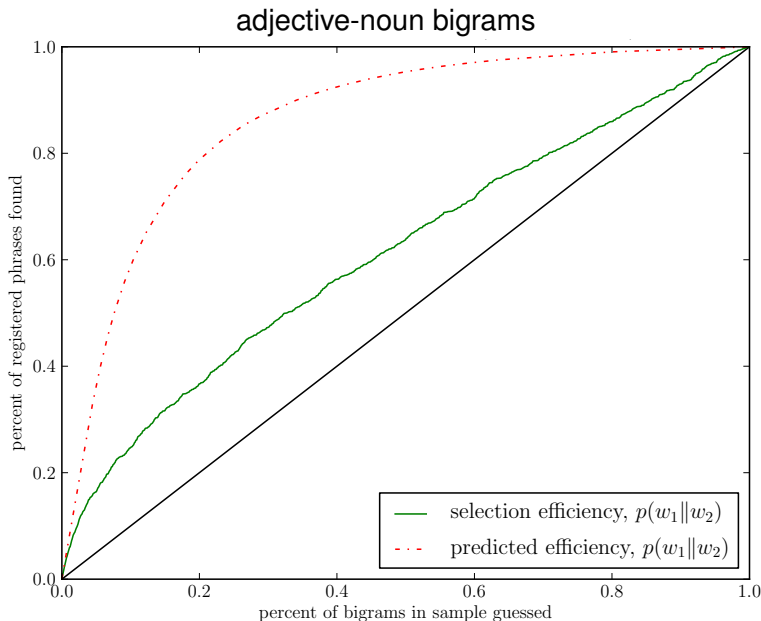
Which factors predict a phrase's popularity?



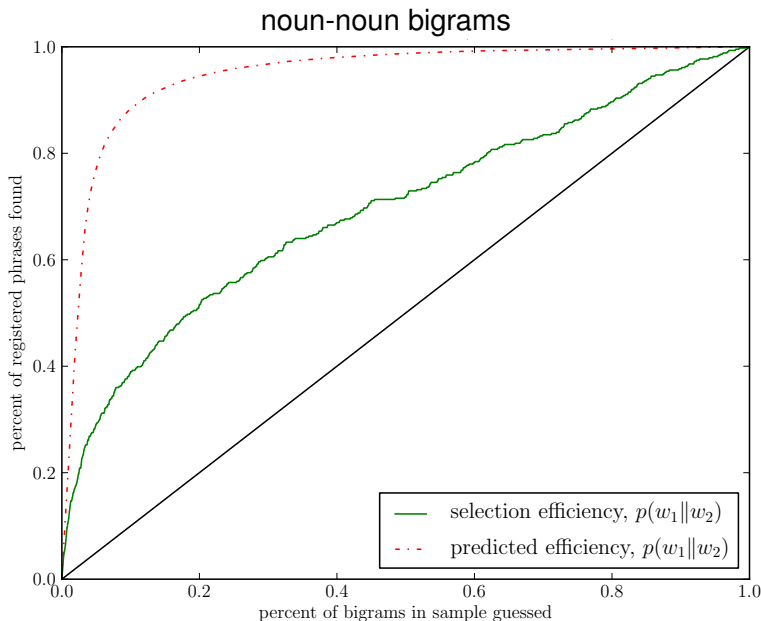
Which factors predict a phrase's popularity?



Which factors predict a phrase's popularity?

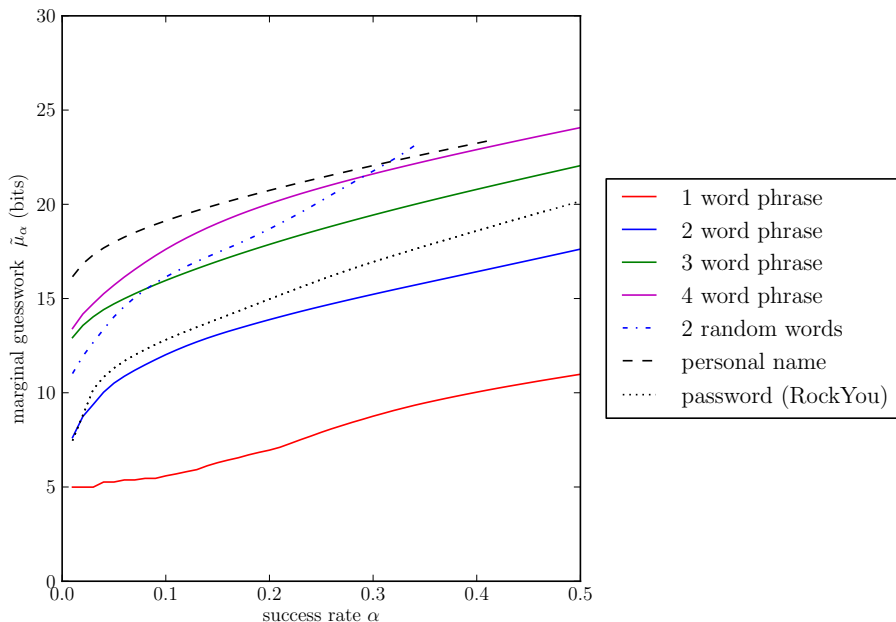


Which factors predict a phrase's popularity?



Are natural language phrases difficult to guess?

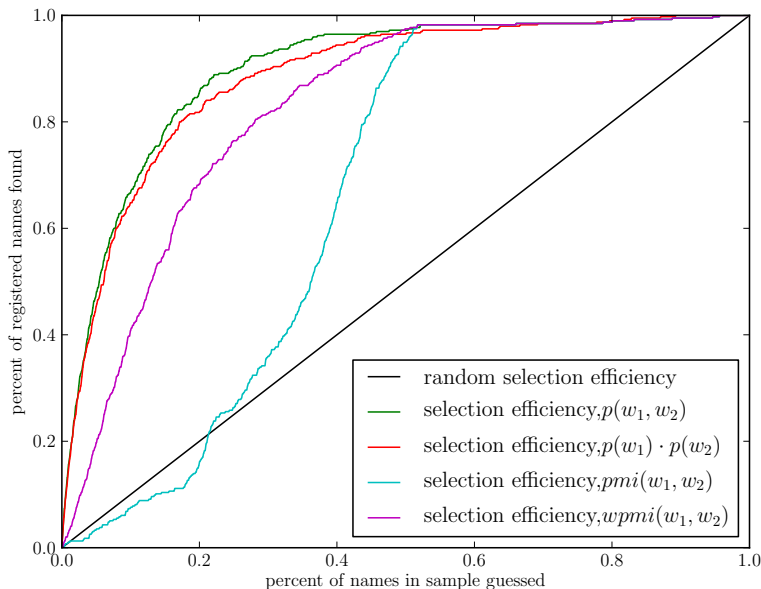
Are natural language phrases difficult to guess?



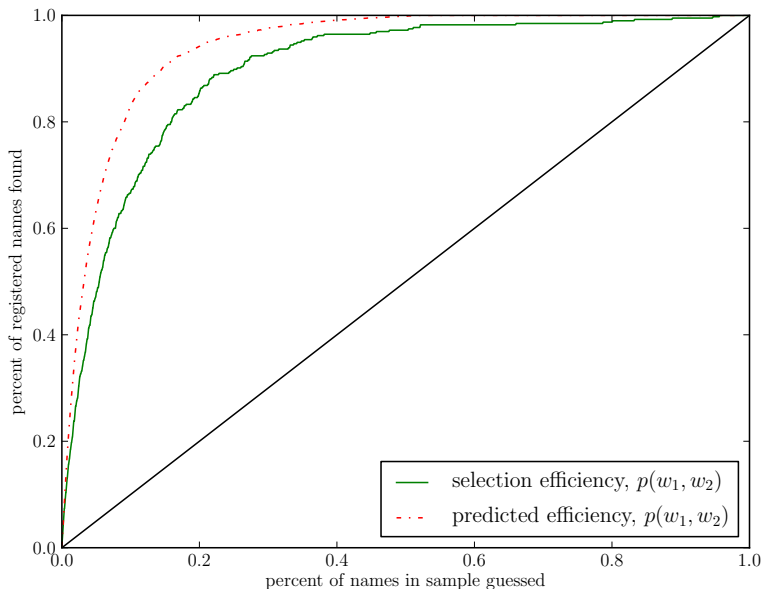
Thank you

jcb82@cl.cam.ac.uk

Similar results for names



Similar results for names



Estimating the probability of each class of phrase

Assumptions:

- N total phrases
- n phrases in this class of equal probability
- k were selected

Estimating the probability of each class of phrase

Assumptions:

- N total phrases
- n phrases in this class of equal probability
- k were selected

solve as a weighted coupon collector's problem:

$$\hat{p} = \frac{\prod_{j=1}^k \frac{n}{n-j}}{N \cdot n}$$