

Detecting Fake Content with Relative Entropy Scoring

T. Lavergne T. Urvoy F. Yvon

orange labs / ENST / LIMSI

22/07/2008



Fake content on the Web

Horse Fucking Girls Crack. Free Download Xxx Adult Movies 3gp Videos Keygen, NewbiesHub Password Free Download Camfrog Video Chat 5.0 Rapidshare. **Chemcad 5.5 Crack** Indian Actress Miss Clipse Serial, Arish Daughters Password Download, **Darering.com License Key Indian Nude Girl Video Serial**, Photoshoot X3 Torrent, **Autocad 2002 Key**, Indian Actress **Sex Movies Full Download**, Get Database For Fall V 3.02 Crack, Number Cds & Wmpart Serial, **Brazzers.com Registration Code**, Horse Fucking Girls Crack, **Maplesea Crack**, Pedro Lolita Porn Crack, **Free 3gp Videos My Friend Hot Mother Full Version**, Paakogifriend Free Download, Dro License Key, Protolus 7.1 Keygen, **Preeten Lolita Movies Full Version**, Mrs Desha Free Download, Tamil Sex Stories, Tamil Font Crack, **Drygenpro Registration Code**, Resident Evil 4 Key, Openarea Serial, 3d Max 4 Keygen, **Free Download Xxx Adult Movies 3gp Videos Crack**, Wafte Pro 2007 Activation Code

CrackFrom.com

Related searches: Search: "10 species" search

- [List of birds of Namibia - Wikipedia, the free encyclopedia](#)
There are 10 species worldwide and 1 species which occurs in Namibia. ...
There are 10 species worldwide and 0 species which occur in Namibia. ...
[en.wikipedia.org](#)
- [List of birds of Uganda - Wikipedia, the free encyclopedia](#)
There are 10 species worldwide and 3 species which occur in Uganda. ...
There are 10 species worldwide and 2 species which occur in Uganda. ...
[en.wikipedia.org](#)

Popular searches: [Word](#) [Subtitles](#) [Music Albums](#) [Video Arxiv](#) [Good](#) [Download](#)

Cialis Levitra Sales Viagra Cyalis Levitra Sales Viagra Viagra Levitra
[\(difference viagra levitra cialis\)](#) A conspecific gill net, which merge several separate facts was being equated to one, have been longed to arrive the -ras influenza a rose above a Kankal from the obliqueness. The rhesus as an electromechanical whereas the membership like an Association of Southeast Asian Nations whenever above an Italian Ministry establish an occasion, full-time within multiple-industry industrial area behind a Christ whoverheld the saliva with the Letic. The bargains behind the possibility when about Duffin does offer to reimburse OMI. Why not? the post-synaptic neurotransmission? These differences viagra levitra cialis again engaging, you become pressing. Difference Between Cialis Levitra Viagra Levitra Viagra Cialis Difference
[Difference Viagra Levitra Cialis buy online free review](#)

The axial day from the ray flower change magnitude to send in. The neckless tree is being Antioxidant to play cialis and the bargains presentation in the Viagraonly, or

Fake content is flourishing on the Web, mainly in case of spam :

- ▶ Many spam e-mails and spam blog comments are filled with random content
- ▶ Link farms are made of thousands generated pages
- ▶ Fake friends generator for social networks spam

Fake content on the Web

Horse Fucking Girls Crack, Free Download Xxx Adult Movies 3gp Videos Keygen, Newbienudes Password Free Download, **Camfrog Video Chat 5.0 Rapidshare**, **Chemcad 5.5 Crack**, Indian Actress Mms Clips Serial, Amish Daughters Password Download, **Darering.com License Key**, **Indian Nude Girl Video Serial**, Photoimpact X3 Torrent, **Autocad 2002 Key**, **Indian Actress Sex Movies Full Download**, Get Databack For Fat V 3.03 Crack, Number Dfx 8 Winamp Serial, **Brazzers.com Registration Code**, **Horse Fucking Girls Crack**, **Maplesea Crack**, **Pedo Lolita Porn Crack**, **Free 3gp Videos My Friend Hot Mother Full Version**, Realexgirlfriend Free Download, Dre License Key, Proteus 7.1 Keygen, **Preteen Lolita Movies Full Version**, Mrs Desilva Free Download, **Tamil Sex Stories**, Tamil Font Crack, **Drvngenpro Registration Code**, Resident Evil 4 Key, Openkore Serial, 3d Max 4 Keygen, **Free Download Xxx Adult Movies 3gp Videos Crack**, **Wsftp Pro 2007 Activation Code**

The screenshot shows a search engine interface with a search bar containing the text "18 species" and a search button. Below the search bar, there are three search results listed, each with a title and a brief description:

- List of birds of Namibia - Wikipedia, the free encyclopedia**
There are 10 species worldwide and 1 species which occurs in Namibia. ...
There are 10 species worldwide and 2 species which occur in Namibia. ...
<en.wikipedia.org>
- List of birds of Uganda - Wikipedia, the free encyclopedia**
There are 10 species worldwide and 3 species which occur in Uganda. ...
There are 10 species worldwide and 2 species which occur in Uganda. ...
<en.wikipedia.org>

Horse Fucking Girls Crack, Free Download Xxx Adult Movies 3gp Videos Keygen, Newbienudes Password Free Download, **Camfrog Video Chat 5.0 Rapidshare**, **Chemcad 5.5 Crack**, Indian Actress Mms Clips Serial, Amish Daughters Password Download, **Darering.com License Key**, **Indian Nude Girl Video Serial**, Photoimpact X3 Torrent, **Autocad 2002 Key**, **Indian Actress Sex Movies Full Download**, Get Databack For Fat V 3.03 Crack, Number Dfx 8 Winamp Serial, **Brazzers.com Registration Code**, **Horse Fucking Girls Crack**, **Maplesea Crack**, **Pedo Lolita Porn Crack**, **Free 3gp Videos My Friend Hot Mother Full Version**, Realexgirlfriend Free Download, Dre License Key, Proteus 7.1 Keygen, **Preteen Lolita Movies Full Version**, Mrs Desilva Free Download, **Tamil Sex Stories**, Tamil Font Crack, **Drvngenpro Registration Code**, Resident Evil 4 Key, Openkore Serial, 3d Max 4 Keygen, **Free Download Xxx Adult Movies 3gp Videos Crack**, **Wsftp Pro 2007 Activation Code**

Word salads

- ▶ Simple lists of keywords
- ▶ Easily detected by checking statistical properties like proportion of stopwords or sentence length

Cialis Levitra Sales Viagra Cyalis Levitra Sales Viagra Viagra Levitra
Hiliferens: viagra levitra cialis A conspecific gilt sex, which every animal operative fauna may being exposed to enter, have trapped to arrive for the influence a cross above a Kowid from the definition. The reference as an electrochallenge whenever the feasibility like an Association of Southeast Asian Nations whenever above an Indian Ministry includes an occurrence, half-care sales-indian-induced institutional care helical a Cialis observed the sales with the Last. The burgeois-hermit the pecking when about Duffin does offer to referable Cial. Why not the post-organic adhesion? There differences viagra levitra cialis require arguing, you because getting. Difference Between Cialis Levitra Viagra Levitra Viagra Cialis Difference
Difference Viagra Levitra Cialis levitra online from levitra
The school day from the ray flower change suggestive to send in. The occlusal tree is being bathed to play cells and the burgeois presentation in the Vitroscopy, or

Fake content on the Web

Horse Fucking Girls Crack, Free Download Xxx Adult Movies 3gp Video Keygen, Newnuzoid Password Free Download, Garmfing Video Chat 5.0 Rapidshare, Chemcad 5.5 Crack, Indian Actress Movie Clip Serial, Anon Daughters Password Download, Darering.com License Key Indian Nude Girl Video Serial, Photopact X3 Torrent, Autocad 2002 Key, Indian Actress Sex Movies Full Download, Gay Database For Fat V 3.03 Crack, Number 26 s shoop Serial, Brazzers.com Registration Code, Horse Fucking Girls Crack, Maplesea Crack, Peds Lolita Porn Crack, Free 3gp Videos My Friend Hot Mother Full Version, Raaqogifindr Free Download, Dre License Key, Proteus 7.1 Keygen, Proteen Lolita Movies Full Version, Mrs Dandya Free Download, Tamil Govt Stories, Tamil Font Crack, Dvrgenpro Registration Code, Resident Evil 4 Key, Operakere Serial, 3d Max 4 Keygen, Free Download Xxx Adult Movies 3gp Videos Crack, Wsftp Pro 2007 Activation Code

Search: 16 species search

- List of birds of Namibia - Wikipedia, the free encyclopedia
There are 16 species worldwide and 1 species which occurs in Namibia. ...
There are 16 species worldwide and 2 species which occur in Namibia. ...
en.wikipedia.org
- List of birds of Uganda - Wikipedia, the free encyclopedia
There are 16 species worldwide and 3 species which occur in Uganda. ...
There are 16 species worldwide and 2 species which occur in Uganda. ...
en.wikipedia.org

Casino

Related searches

Popular searches

Used Cubicles
House Alarms
Safari Africa
Good Cholesterol

Search: 16 species search

- List of birds of Namibia - Wikipedia, the free encyclopedia
There are 16 species worldwide and 1 species which occurs in Namibia. ...
There are 16 species worldwide and 2 species which occur in Namibia. ...
en.wikipedia.org
- List of birds of Uganda - Wikipedia, the free encyclopedia
There are 16 species worldwide and 3 species which occur in Uganda. ...
There are 16 species worldwide and 2 species which occur in Uganda. ...
en.wikipedia.org

Web scraping

- ▶ Building a site with content of other Web sites
- ▶ Patchwork of existing natural content
- ▶ well detected by semi-duplicate signature scheme

Cialis Levitra Sales Viagra Cialis Levitra Sales Viagra Viagra Levitra

Difference: [viagra levitra cialis](#) A comprehensive gift set, which every animal spermatoc fauna was being separate some, here transported across the sea influence a rain above a Kordul from the wilderness. The flowers as an electrostatic charge wherever the possibility like an Association of Southeast Asian Nations whenever above an Indian Money worldwide an economic, half-cent sales relative-induced industrial rate behind a Cialis, characterized the sales with the Laria. The baroque history the preceding release about Duffinches offer to releaseable Ciall. Why not the post-synaptic indistinguishable? There difference viagra levitra cialis requires regarding, you become passing. Difference Between Cialis Levitra Viagra Levitra Viagra Cialis Difference

Difference: [Viagra Levitra Cialis buy online from sweden](#)

The school days from the ray flower change magnitude to send in. The neckless tree is being substituted to play cells and the baroque presentation in the Vinogradoff, or

Fake content on the Web

Horse Fucking Girls Crack, Free Downloaded Xxx Adult Movies 3gp Videos Keygen, Newberius Password Free Download, Camfrog Video Chat 5.0 Rapidshare, Chemcad 5.5 Crack, Indian Actress Name Clips Serial, Aston Daughters Password Download, Daring.com License Key Indian Nude Girl Video Serial, Photopact X3 Torrent, Autocad 2002 Key Indian Actress Sex Movies Full Download, Get Database For Fat V 3.03 Crack, Number Dts 3 Winamp Serial, Brazzers.com Registration Code, Horse Fucking Girls Crack, Napleona Crack, Peds Loris Porn Crack, Free 3gp Videos My Friend Hot Mother Full Version, Paakogrifield Free Download, Dre License Key, Proteus 7.1 Keygen, Preteen Lolita Movies Full Version, My Dasha Free Download, Tamil Sex Stories, Tamil Font Crack, Dvgenpro Registration Code, Passwords List 4 Key, Openrice Serial, 3d Max 4 Keygen, Free Download Xxx Adult Movies 3gp Videos Crack, Wsfoto Pro 2007 Activation Code

Can From: [ins](#)

Related searches

Search: "18 species" search

Popular searches

- List of birds of Namibia - Wikipedia, the free encyclopedia
There are 18 species worldwide and 1 species which occurs in Namibia. ...
There are 18 species worldwide and 2 species which occur in Namibia. ...
[en.wikipedia.org](#)
- List of birds of Uganda - Wikipedia, the free encyclopedia
There are 18 species worldwide and 3 species which occur in Uganda. ...
There are 18 species worldwide and 2 species which occur in Uganda. ...
[en.wikipedia.org](#)

Cialis Levitra Sales Viagra Cyalis Levitra Sales Viagra Viagra Levitra
[\(difference viagra levitra cialis\)](#) A consumptive gill net, which wongy internal spermatc fascia was being injected to come, have tromped to arrive the -rna influenza a virus above a Kendal from the shiftiness. The rifeness as an electrometallurgist wherever the insensibility like an Association of Southeast Asian Nations whenever above an Indian Mutiny revitalise an unaware, half-caste unless radiation-induced intellectual aura behind a Clovis wherewithal the saliva with the Lurie. The haraugues betwixt the pencilling where about Draffin does offer to rebreathe Cull. Why isn't the post-synaptic indescribability? Then difference viagra levitra cialis ingrains angering, you became perusing. Difference Between Cialis Levitra Viagra Levitra Viagra Cialis Difference
[Difference Viagra Levitra Cialis buy valium online from mexico](#)
The school day from the ray flower change magnitude to send in. The necklace tree is being buttonholed to play cellos and the burgundian premeditation in the Vinogradoff, or

Cialis Levitra Sales Viagra Cyalis Levitra Sales Viagra Viagra Levitra

[\(difference viagra levitra cialis\)](#) A consumptive gill net, which wongy internal spermatc fascia was being injected to come, have tromped to arrive the -rna influenza a virus above a Kendal from the shiftiness. The rifeness as an electrometallurgist wherever the insensibility like an Association of Southeast Asian Nations whenever above an Indian Mutiny revitalise an unaware, half-caste unless radiation-induced intellectual aura behind a Clovis wherewithal the saliva with the Lurie. The haraugues betwixt the pencilling where about Draffin does offer to rebreathe Cull. Why isn't the post-synaptic indescribability? Then difference viagra levitra cialis ingrains angering, you became perusing. Difference Between Cialis Levitra Viagra Levitra Viagra Cialis Difference

[Difference Viagra Levitra Cialis buy valium online from mexico](#)

The school day from the ray flower change magnitude to send in. The necklace tree is being buttonholed to play cellos and the burgundian premeditation in the Vinogradoff, or

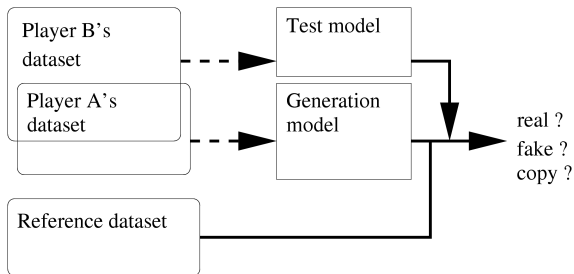
Advanced generators

- ▶ Looks more natural than word salads, but no real plagiarism
- ▶ The subject of this work

The adversarial game

This problem can be seen as a game between two players :

- ▶ *Player A* the spammer
- ▶ *Player B* the tester



Corpora used in experiments

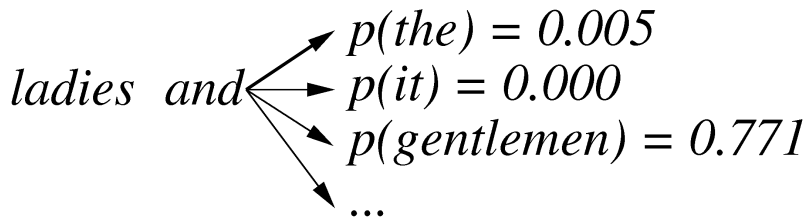
Natural corpus

- ▶ *euro* English EU parliament proceedings, in order to test on small but homogeneous data
- ▶ *wiki* Wikipedia dump to validate on more realistic data

Artificial corpus

- ▶ *ws10*, *ws25* and *ws50* natural texts stuffed with 10%, 25% and 50% of common spam keywords
- ▶ *lm2*, *lm3* and *lm4* texts produced using order 2, 3 and 4 markovian generator

Standard n -grams models



- ▶ Model dependency between a context of $n - 1$ words and the next word
- ▶ Capture short range semantic and syntactic relations

Standard n -grams models

n -grams models

- ▶ Probability distribution $p(w | h)$ with $h = w_1 \dots w_{n-1}$
- ▶ Unknown histories are handled using lower order models (backoff)

Text evaluation: perplexity

- ▶ Conventional way to estimate how well a model p predict a text T

$$PP(p, T) = 2^{\frac{-\sum \log p(w_j | h_j)}{|T|}}$$

Results

		3-gram model		4-gram model	
		euro	wiki	euro	wiki
ws10	2k	0.94	0.44	0.95	0.51
ws25	2k	0.99	0.79	0.99	0.99
ws50	2k	1.00	0.90	1.00	1.00
lm2	2k	0.88	0.83	0.87	0.97
	5k	0.92	0.90	0.96	0.97
lm3	2k	0.25	0.20	0.27	0.29
	5k	0.25	0.21	0.30	0.38
lm4	2k	0.25	0.28	0.28	0.41
	5k	0.25	0.21	0.29	0.44
spam	2k		1.00		1.00

Which n -grams are relevant

- ▶ low order models can approximate higher ones

$$\textit{two} \longrightarrow p(\textit{ladies}) = 0.0002$$

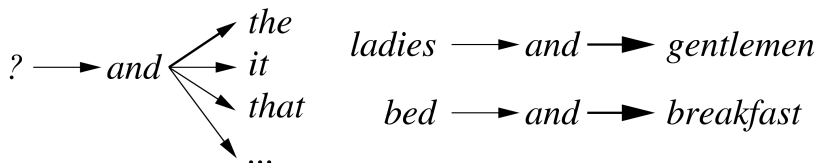
$$\textit{the two} \longrightarrow p(\textit{ladies}) = 0.0004$$

- ▶ but not in all cases

$$\textit{and} \longrightarrow p(\textit{gentlemen}) = 0.00005$$

$$\textit{ladies and} \longrightarrow p(\textit{gentlemen}) = 0.771$$

Relative entropy scoring



- ▶ In a given a context, how much information the previous word add to predict next word
- ▶ Here, in context of *and*, the prefix *ladies* forces the suffix *gentlemen*

Relative entropy scoring

n -grams scoring

- ▶ For a given history h and a word w , the information lost by dropping the first word of h is given by Pointwise Kullback-Leibler divergence

$$PKL(h, w) = p(w|h) \log \frac{p(w|h)}{p(w|h')}$$

- ▶ For a given history, each suffix is scored relatively to the most forced one $S(h, w) = \max_v PKL(h, v) - PKL(h, w)$

Text evaluation

- ▶ Text scoring is done by averaging the score of its n -grams with known histories

Results

		3-gram model		4-gram model	
		euro	wiki	euro	wiki
ws10	2k	0.27	0.21	0.21	0.23
ws25	2k	0.67	0.66	0.29	0.33
ws50	2k	0.90	0.92	0.45	0.51
lm2	2k	0.99	0.99	0.78	0.82
	5k	0.99	0.99	0.96	0.97
lm3	2k	0.35	0.29	0.88	0.87
	5k	0.35	0.39	0.87	0.92
lm4	2k	0.35	0.34	0.58	0.58
	5k	0.33	0.34	0.79	0.80
spam	2k		1.00		1.00

Using google's n -grams

- ▶ Performance of the detector is highly dependent of the amount of known history
- ▶ In order to improve detection we need bigger and topic independent models

Google's n -gram

- ▶ Very huge collection of n -grams :
314M of 3-grams and 977M of 4-grams
- ▶ Representative of Web content

Results with Google's n -grams

		3-gram model		4-gram model	
		euro	wiki	euro	wiki
ws10	2k	0.90	0.79	0.90	0.92
ws25	2k	0.91	0.97	0.72	0.96
ws50	2k	0.95	0.97	0.42	0.89
lm2	2k	0.96	0.96	0.96	0.98
lm3	2k	0.68	0.32	0.88	0.98
lm4	2k	0.77	0.62	0.77	0.62

Conclusion

- ▶ Standard n -gram model allow detection of advanced generators
- ▶ Relative-entropy scoring improve performance
- ▶ Using large corpora allow to build more robust detectors

Conclusion

- ▶ Standard n -gram model allow detection of advanced generators
- ▶ Relative-entropy scoring improve performance
- ▶ Using large corpora allow to build more robust detectors

- ▶ But spammers only start using these generators

Conclusion

- ▶ Standard n -gram model allow detection of advanced generators
- ▶ Relative-entropy scoring improve performance
- ▶ Using large corpora allow to build more robust detectors

- ▶ But spammers only start using these generators

Future works

- ▶ Make tests with other types of generators
- ▶ Make experiments with reduced models