



LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN



ISLE 2, Boston •

Susanne Stegmayr & Daphné Kerremans (LMU Munich)

Neologisms on the Internet

The NeoCrawler



Overview

1. The Problem

2. The Web-as-Corpus

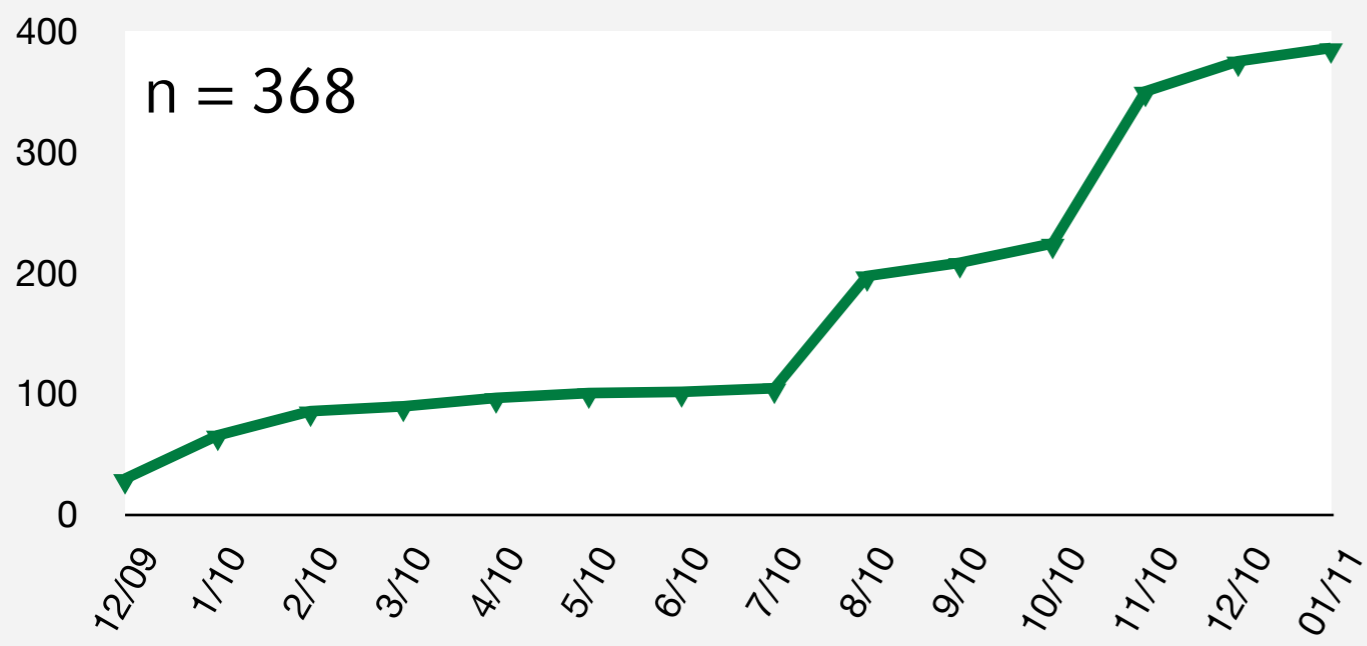
3. The NeoCrawler

- architecture
- manual search and search results
- automatic searches and the Google Research Program
- web interface and linguistic processing

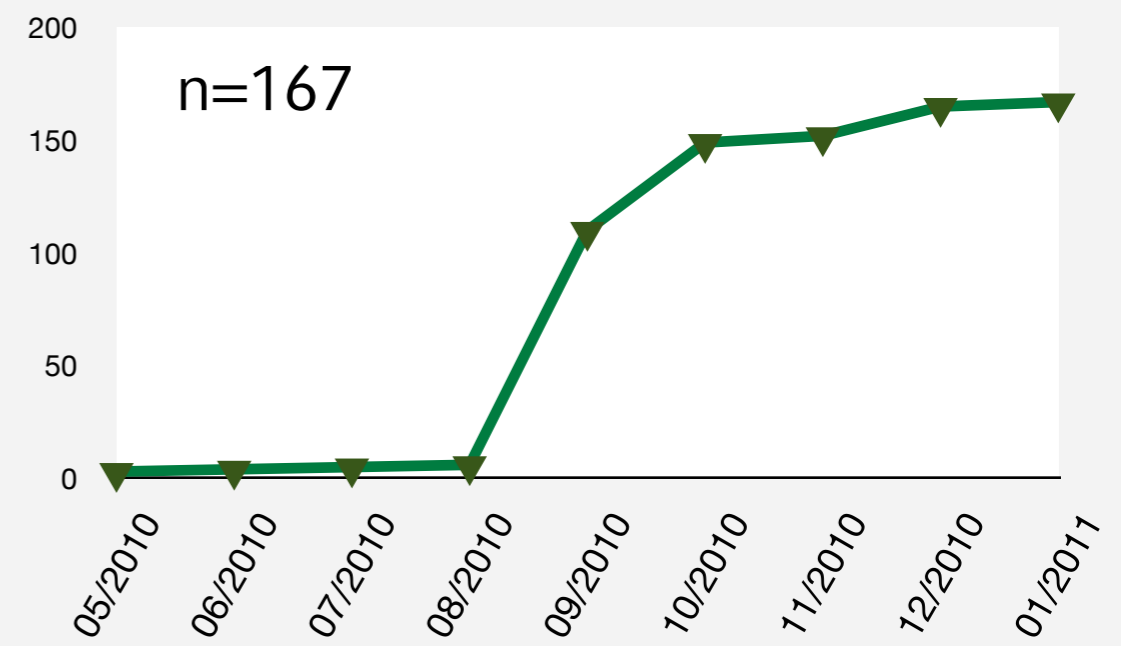
4. Applied NeoCrawling

1. The Problem

cherpumpke



halfalogue



2. The Web-as-Corpus

- The WaCky approach (Sharoff, Baroni, Bernardini, Kilgarriff et al.)
- Downloadable crawlers:
 - KWiCFinder (Fletcher 2001, www.kwicfinder.com)
 - GlossaNet 2 (Fairon, Macé and Naets 2008, <http://glossa.fltr.ucl.ac.be>)
- On-demand crawlers:
 - Linguistic Search Engine (Kilgarriff 2003)
 - WebCorp and WebCorp Linguistic Search Engine (1998, 2005 RDUES Birmingham City University, www.webcorp.ac.co.uk)

WebCorp Live

An improved version of the original WebCorp,
designed to search the web for concordances in
real time

Search term:

Enter a word, phrase (no quotes necessary) or [pattern](#)

Search engine:

Google

Concordance Span:

5

word(s) to left & right

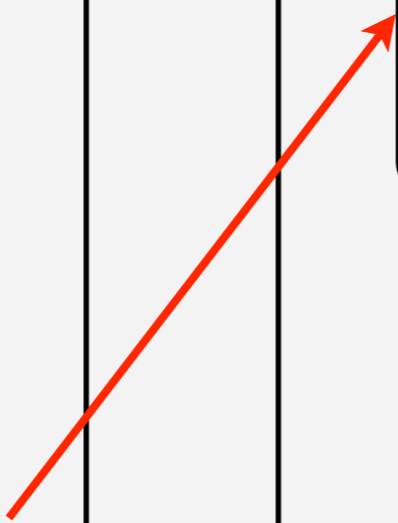
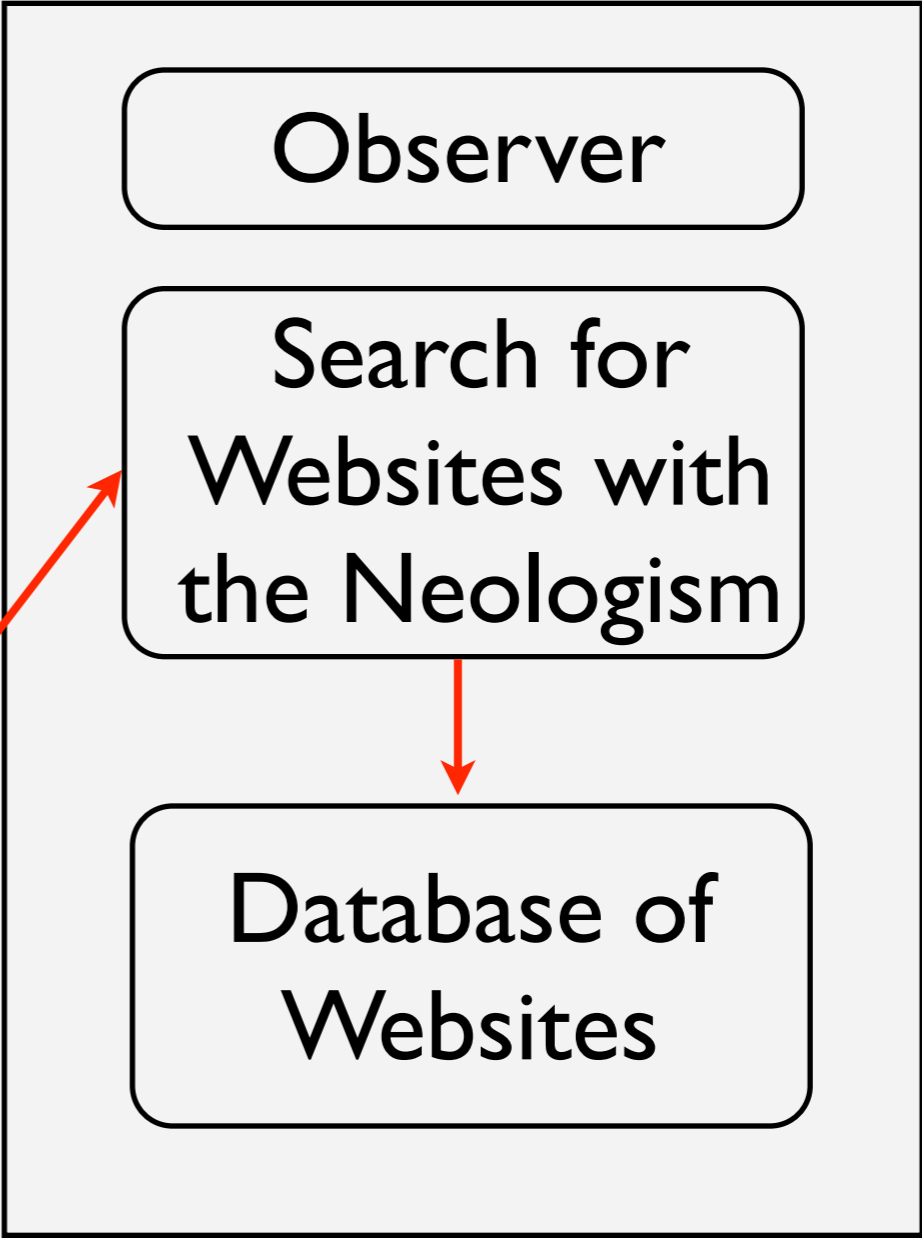
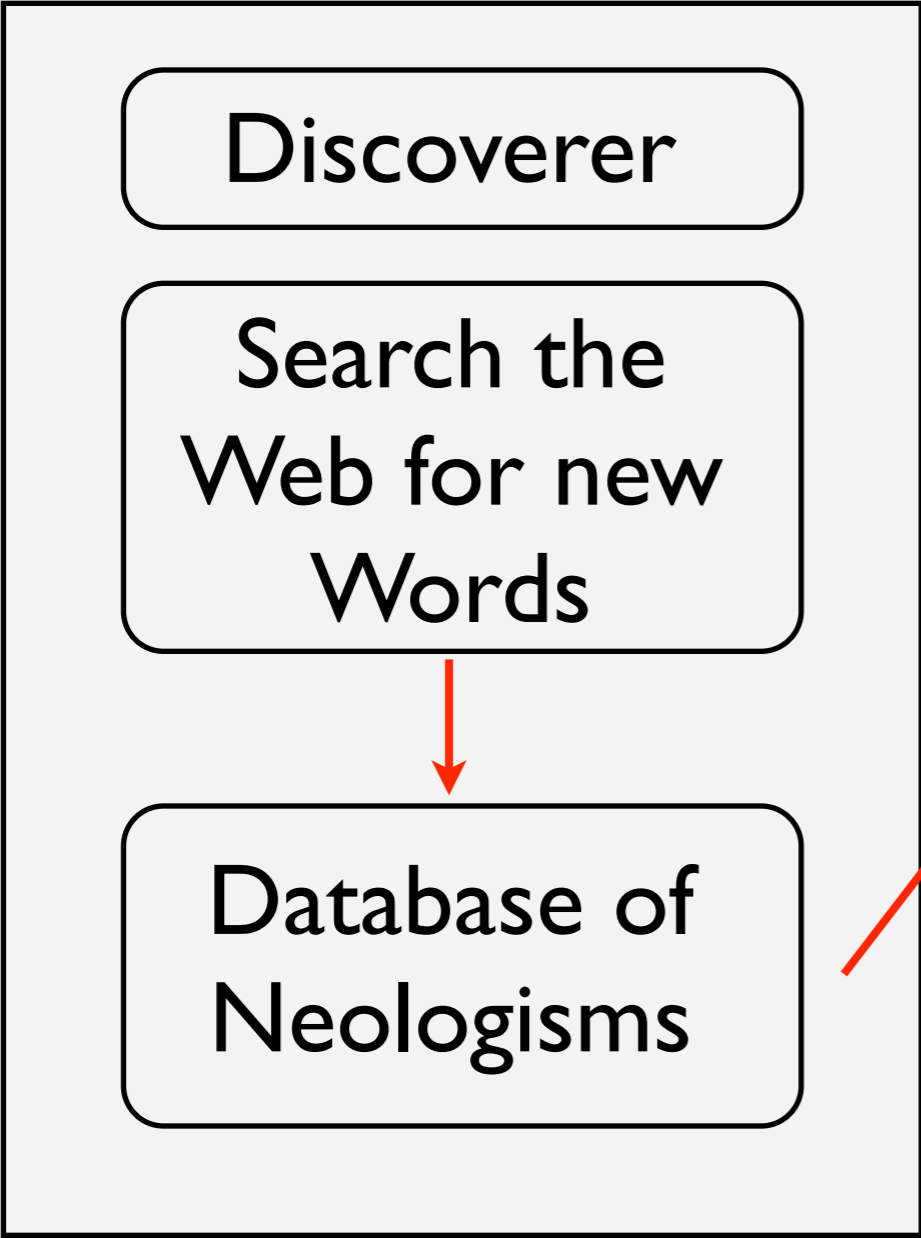
Case options:

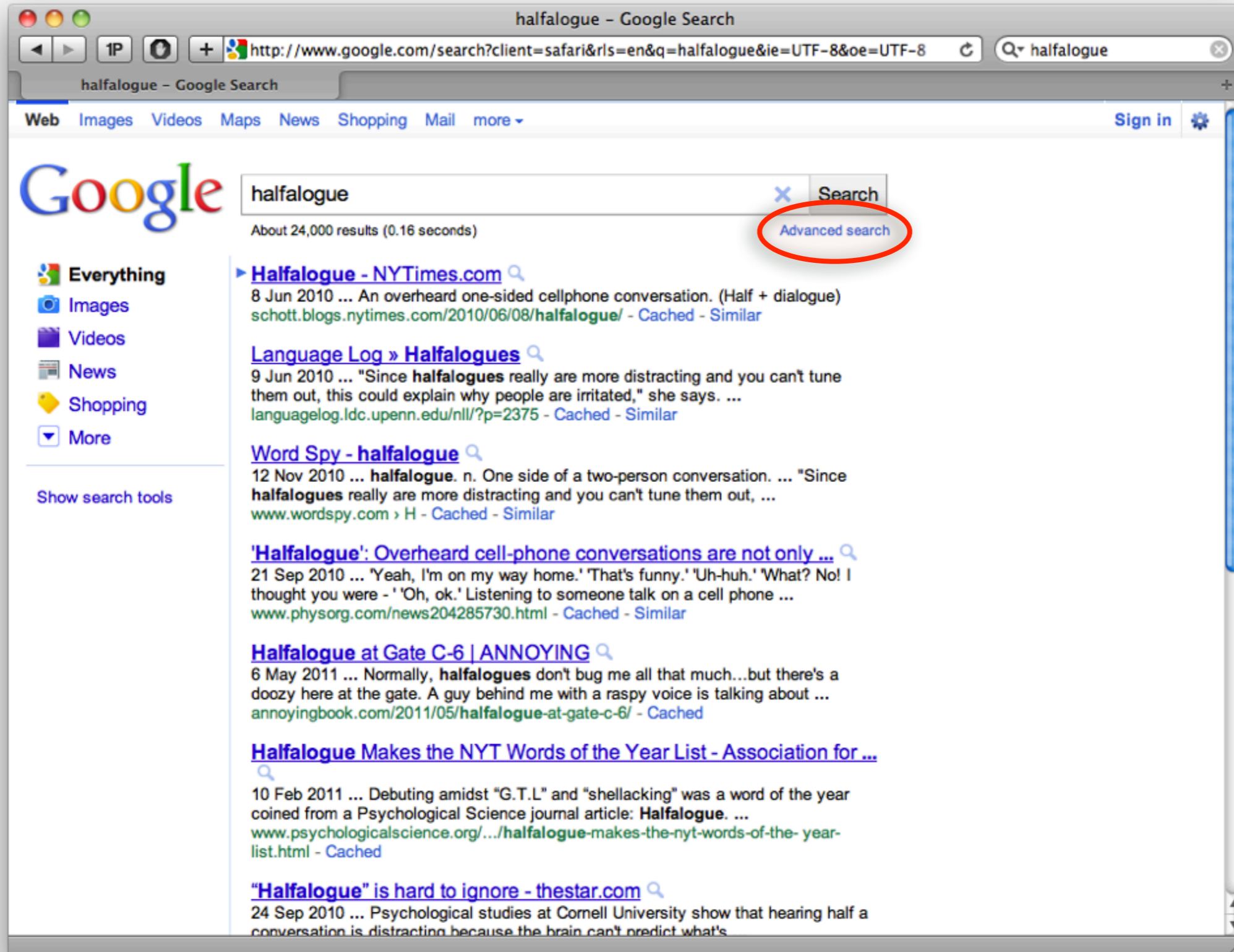
Case Sensitive

[Advanced Search Options](#)

Submit

am City





halfalogue - Google Search

http://www.google.com/search?client=safari&rls=en&q=halfalogue&ie=UTF-8&oe=UTF-8

halfalogue

Web Images Videos Maps News Shopping Mail more

Sign in

Google

halfalogue

About 24,000 results (0.16 seconds)

Advanced search

Everything

Images

Videos

News

Shopping

More

Show search tools

Halfalogue - NYTimes.com

8 Jun 2010 ... An overheard one-sided cellphone conversation. (Half + dialogue)
schott.blogs.nytimes.com/2010/06/08/halfalogue/ - Cached - Similar

Language Log » Halfalogues

9 Jun 2010 ... "Since **halfalogues** really are more distracting and you can't tune them out, this could explain why people are irritated," she says. ...
languagelog.ldc.upenn.edu/nll/?p=2375 - Cached - Similar

Word Spy - halfalogue

12 Nov 2010 ... **halfalogue**. n. One side of a two-person conversation. ... "Since **halfalogues** really are more distracting and you can't tune them out, ...
[www.wordspy.com > H](http://www.wordspy.com/H) - Cached - Similar

'Halfalogue': Overheard cell-phone conversations are not only ...

21 Sep 2010 ... 'Yeah, I'm on my way home.' 'That's funny.' 'Uh-huh.' 'What? No! I thought you were - 'Oh, ok.' Listening to someone talk on a cell phone ...
www.physorg.com/news204285730.html - Cached - Similar

Halfalogue at Gate C-6 | ANNOYING

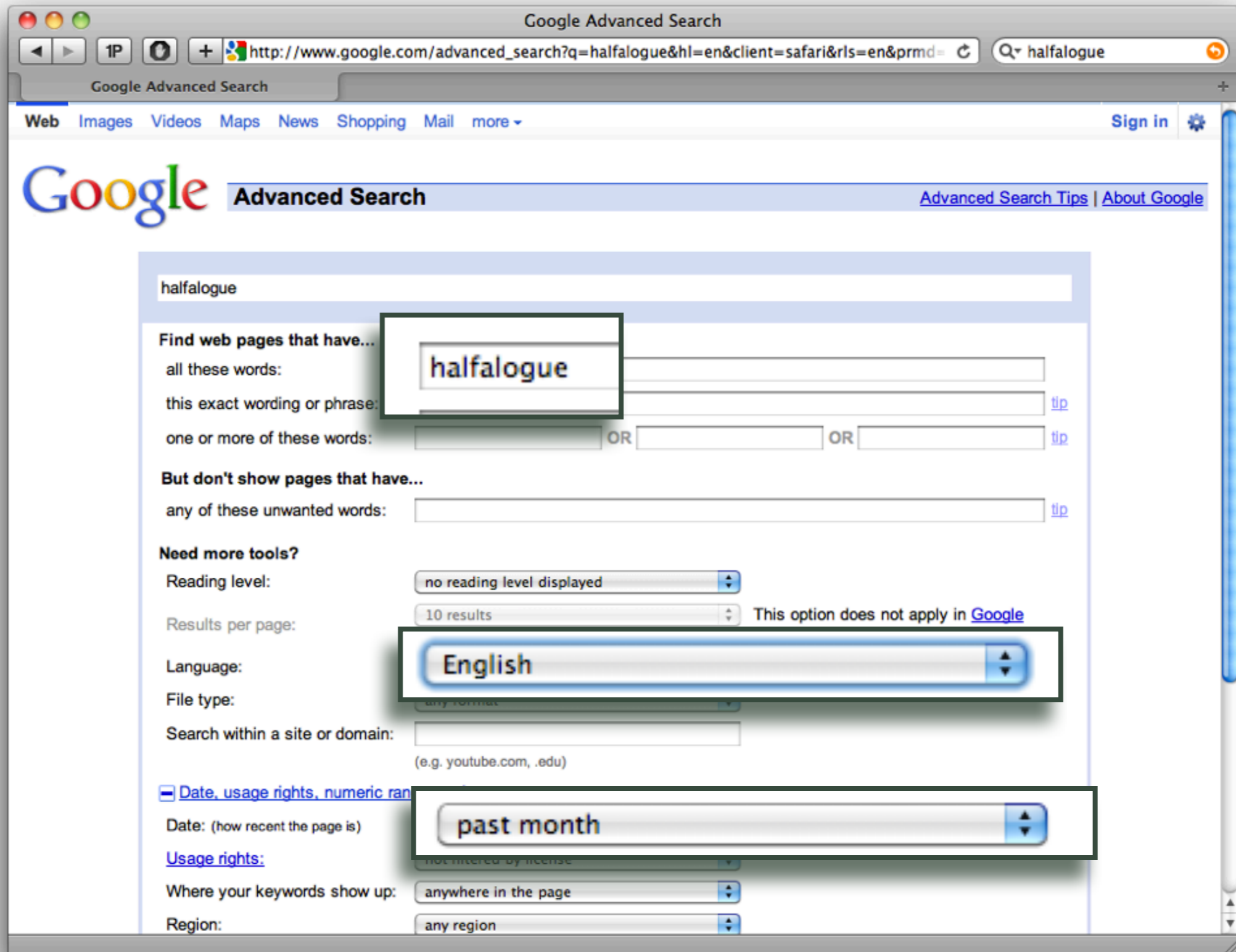
6 May 2011 ... Normally, **halfalogues** don't bug me all that much...but there's a doozy here at the gate. A guy behind me with a raspy voice is talking about ...
annoyingbook.com/2011/05/halfalogue-at-gate-c-6/ - Cached

Halfalogue Makes the NYT Words of the Year List - Association for ...

10 Feb 2011 ... Debuting amidst "G.T.L" and "shellacking" was a word of the year coined from a Psychological Science journal article: **Halfalogue**. ...
www.psychologicalscience.org/.../halfalogue-makes-the-nyt-words-of-the-year-list.html - Cached

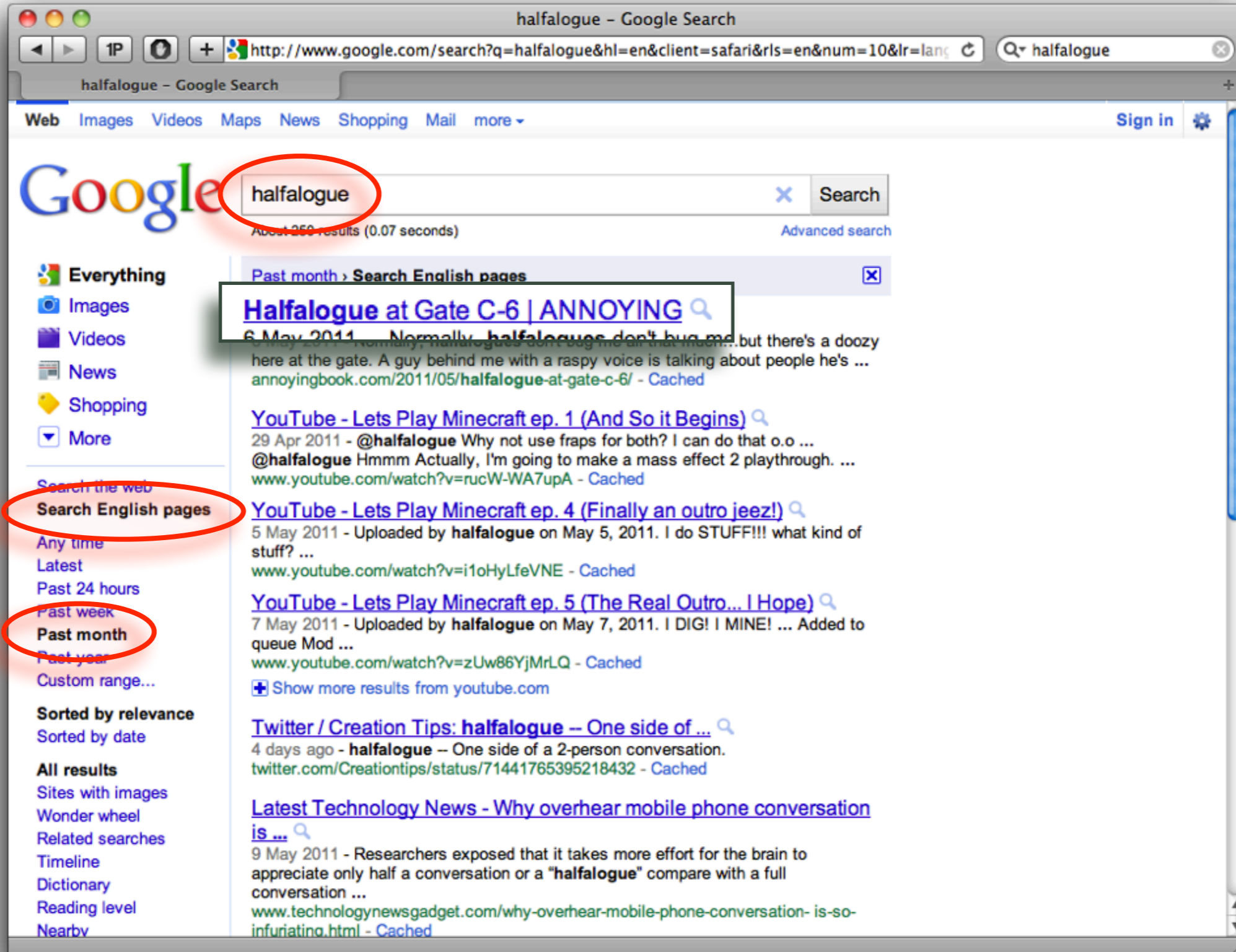
"Halfalogue" is hard to ignore - thestar.com

24 Sep 2010 ... Psychological studies at Cornell University show that hearing half a conversation is distracting because the brain can't predict what's



The screenshot shows a browser window titled "Google Advanced Search" with the URL `http://www.google.com/advanced_search?q=halfalogue&hl=en&client=safari&rls=en&prmd=` and search term "halfalogue". The search results are not visible. The search criteria are:

- Find web pages that have...
 - all these words:
 - this exact wording or phrase:
 - one or more of these words: OR OR
- But don't show pages that have...
 - any of these unwanted words:
- Need more tools?
 - Reading level:
 - Results per page: This option does not apply in [Google](#)
 - Language:
 - File type:
 - Search within a site or domain:
- [Date, usage rights, numeric range](#)
 - Date: (how recent the page is)
 - Usage rights:
- Where your keywords show up:
- Region:



halfalogue - Google Search

http://www.google.com/search?q=halfalogue&hl=en&client=safari&rls=en&num=10&lr=lanç

halfalogue

Web Images Videos Maps News Shopping Mail more

Sign in

Google halfalogue

About 250 results (0.07 seconds) Advanced search

Everything

- Images
- Videos
- News
- Shopping
- More

Search the web

Search English pages

- Any time
- Latest
- Past 24 hours
- Past week
- Past month
- Past year
- Custom range...

Sorted by relevance

Sorted by date

All results

- Sites with images
- Wonder wheel
- Related searches
- Timeline
- Dictionary
- Reading level
- Nearby

Past month > Search English pages

Halfalogue at Gate C-6 | ANNOYING

6 May 2011 - Normally, halfalogues don't bug me...but there's a doozy here at the gate. A guy behind me with a raspy voice is talking about people he's ...
annoyingbook.com/2011/05/halfalogue-at-gate-c-6/ - Cached

YouTube - Lets Play Minecraft ep. 1 (And So it Begins)

29 Apr 2011 - @halfalogue Why not use fraps for both? I can do that o.o ...
@halfalogue Hmmmm Actually, I'm going to make a mass effect 2 playthrough. ...
www.youtube.com/watch?v=rucW-WA7upA - Cached

YouTube - Lets Play Minecraft ep. 4 (Finally an outro jeez!)

5 May 2011 - Uploaded by halfalogue on May 5, 2011. I do STUFF!!! what kind of stuff? ...
www.youtube.com/watch?v=i1oHyLfeVNE - Cached

YouTube - Lets Play Minecraft ep. 5 (The Real Outro... I Hope)

7 May 2011 - Uploaded by halfalogue on May 7, 2011. I DIG! I MINE! ... Added to queue Mod ...
www.youtube.com/watch?v=zUw86YjMrLQ - Cached

Show more results from youtube.com

Twitter / Creation Tips: halfalogue -- One side of ...

4 days ago - halfalogue -- One side of a 2-person conversation.
twitter.com/Creationtips/status/71441765395218432 - Cached

Latest Technology News - Why overhear mobile phone conversation is ...

9 May 2011 - Researchers exposed that it takes more effort for the brain to appreciate only half a conversation or a "halfalogue" compare with a full conversation ...
www.technologynewsgadget.com/why-overhear-mobile-phone-conversation-is-so-infuriating.html - Cached



+ http://www.google.com/advanced_search?q=halfalogue&hl=en&client=safari&rls=en&prmd=

Web Images Videos Maps News Shopping Mail more ▾ Sign in

Google Advanced Search [Advanced Search Tips](#) | [About Google](#)

halfalogue

Find web pages that have...

all these words:

this exact wording or phrase:

one or more of these words: OR OR

But don't show pages that have...

any of these unwanted words:

Need more tools?

Reading level:

Results per page: This option does not apply in [Google](#)

Language:

File type:

Search within a site or domain:
(e.g. youtube.com, .edu)

[Date, usage rights, num...](#)

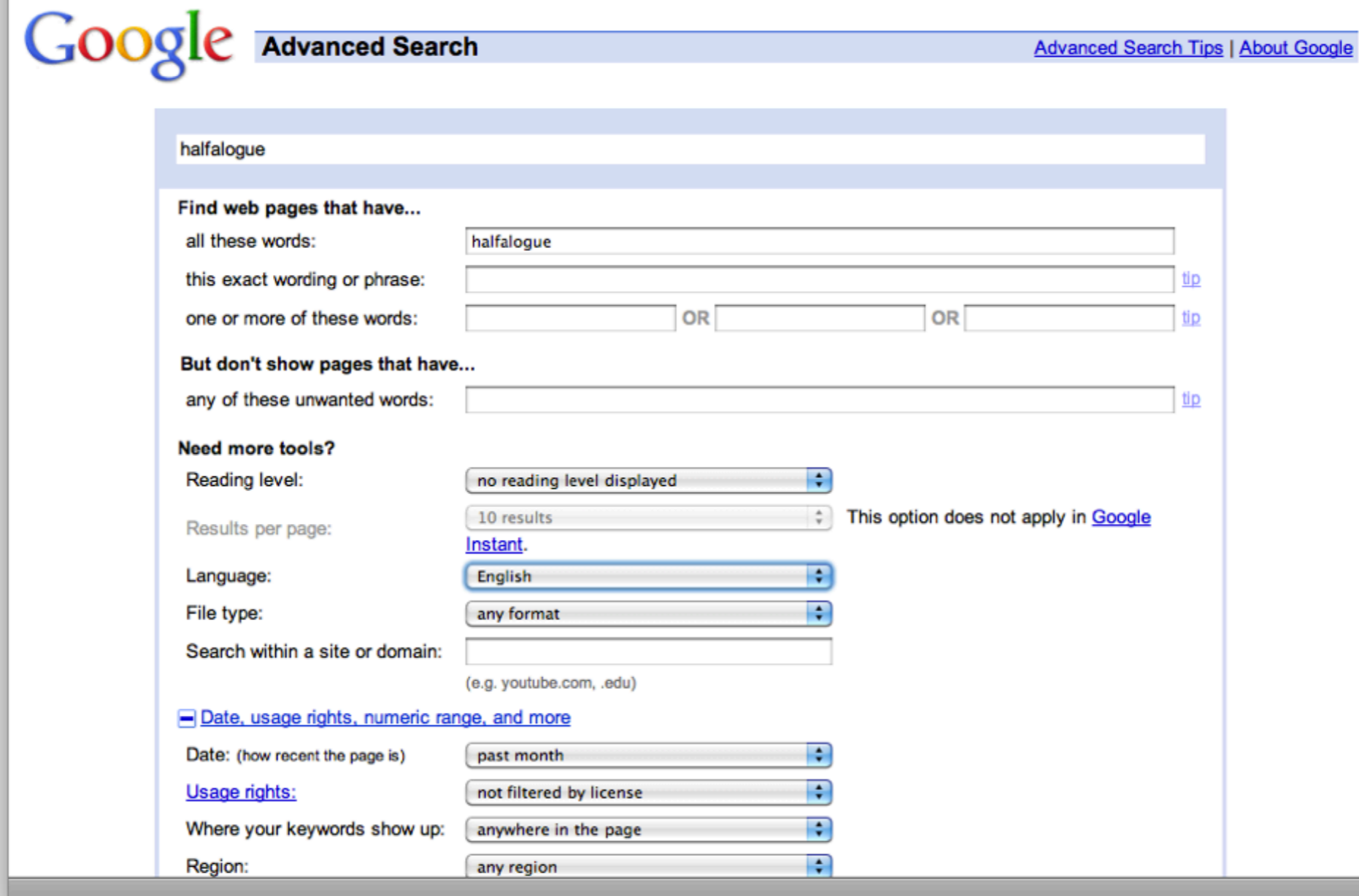
Date: (how recent the page is)

[Usage rights:](#)

Where your keywords show up:

Region:

[http://www.google.com/search?q=halfalogue
&hl=en&rls=en&num=10&lr=lang_en&tbs=qdr:m](http://www.google.com/search?q=halfalogue&hl=en&rls=en&num=10&lr=lang_en&tbs=qdr:m)



Google **Advanced Search** [Advanced Search Tips](#) | [About Google](#)

halfalogue

Find web pages that have...

all these words:

this exact wording or phrase: [tip](#)

one or more of these words: OR OR [tip](#)

But don't show pages that have...

any of these unwanted words: [tip](#)

Need more tools?

Reading level:

Results per page: This option does not apply in [Google Instant](#).

Language:

File type:

Search within a site or domain:
(e.g. youtube.com, .edu)

[Date, usage rights, numeric range, and more](#)

Date: (how recent the page is)

[Usage rights:](#)

Where your keywords show up:

Region:

Halfalogue at Gate C-6 | ANNOYING

http://annoyingbook.com/2011/05/halfalogue-at-gate-c-6/ RSS halfalogue

Halfalogue at Gate C-6 | ANNOYING

annoying

THE SCIENCE OF WHAT BUGS US

HOME / ANNOYING BOOK / AUTHORS / NEWS / EVENTS / BLOG / CONTACT /

← TIME has come today Westwood Barnes and Noble →

HALFALOGUE AT GATE C-6

Posted on [May 6, 2011](#) by [JoePalca](#)

Normally, halfalogues don't bug me all that much...but there's a doozy here at the gate. A guy behind me with a raspy voice is talking about people he's considering hiring. While I'm sort of curious to know what his company does, and what positions he's hiring for, right now I wish he'd just stop talking.

He stopped.

No, there's more. He said goodbye, but then he started up again. Now he's talking about clearances...

Now he's done. Really done. Phew.

ABOUT
There's nowhere to hide and no one is immune. Can science explain life's daily annoyances?

BUY FROM
[Amazon](#)
[Wiley](#)
[IndieBound](#)


```
halfalogue-at-gate-c-6.html
87:32 post-270
75 <div class="navigation">
76 <a href="http://annoyingbook.com/2011/05/time-has-come-today/" rel="prev"><span class="meta-
77 </div>
78 <div class="nav-next"><a href="http://annoyingbook.com/2011/05/westwood-barnes-and-noble/" rel="next">Westwood Barnes and Noble
79 </div>
80 <div class="post-270 post type-post status-publish format-standard hentry category-uncategorized">
81 <h1>Halfalogue at Gate C-6</h1>
82 <div class="entry-meta">
83 <span class="meta-prep meta-author">Posted on</span> <a href="http://annoyingbook.com/2011/05/halfalogue-at-gate-c-6/">
84 <span class="meta-sep">by</span> <span class="author vcard"><a href="http://annoyingbook.com/author/joepalca/">JoePalca</a></span>
85 </div>
86 <div class="entry-content">
87 <p>Halfalogues don't bug me all that much; but there's a doozy here at the gate. A guy behind me
88 with a raspy voice is talking about people he's considering hiring. While I'm sort of curious to know
89 what his company does, and what positions he's hiring for, right now I wish he'd just stop talking.</p>
90 <p>He said goodbye, but then he started up again. Now he's talking about clearances;</p>
91 <p>It's done. Really done. Phew.</p>
92 </div><!-- .entry-content -->
93 <div class="entry-utility">
94 <a href="http://annoyingbook.com/category/uncategorized/">View all posts in Uncategorized</a>
95 <a href="http://annoyingbook.com/2011/05/halfalogue-at-gate-c-6/">Permalink to Halfalogue at Gate C-6</a>.
96 </div><!-- .entry-utility -->
97 <div class="navigation">
98 <div class="nav-pre"><a href="http://annoyingbook.com/2011/05/time-has-come-today/" rel="prev"><span class="meta-
99 </div>
100 <div class="nav-next"><a href="http://annoyingbook.com/2011/05/westwood-barnes-and-noble/" rel="next">Westwood Barnes and Noble
101 </div>
102 </div><!-- #nav-below -->
103 <div class="comments">
104
105
106
```



```
halfalogue-at-gate-c-6.html
Teilnehmer Ankündigen Voriges Symbol Nächstes Symbol Vorige Änderung Nächste Änderung Änderungen hervorheben Verbindungen Terminal Here HTML Color
halfalogue...-c-6.html
87:32 post-270
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
HTML Tabulator (4) LF Unicode (UTF-8) 149w
```

TIME has come today. Westwood Barnes and Noble

Halfalogue at Gate C-6

Posted on May 6, 2011 by JoePalca

Normally, halfalogues don't bug me all that much, but there's a doozy here at the gate. A guy behind me with a raspy voice is talking about people he's considering hiring. While I'm sort of curious to know what his company does, and what positions he's hiring for, right now I wish he'd just stop talking.

He stopped. No, there's more. He said goodbye, but then he started up again. Now he's talking about clearances. Now he's done. Really done. Phew.

This entry was posted in Uncategorized. Bookmark the permalink.

TIME has come today. Westwood Barnes and Noble



NeoCrawler

http://neocrawler.gwi.uni-muenchen.de/index.php?abfrage=list_neologisms

switch case else

A

- ▶ [acedia](#)
- ▶ [administrativa](#)
- ▶ [affiliopinion](#)
- ▶ [AfPak](#)
- ▶ [agile business continuity](#)
- ▶ [alabastard](#) (n.) a pale white person or object with a mental handicap (also adj.).
- ▶ [anagramise](#)
- ▶ [aquacrunk](#)
- ▶ [assault weapon](#)
- ▶ [avatard](#)
 1. (n.) A person who plays a hero in the gaming world but is a real loser in the real world.
 2. (n.) A person who is obsessed with the characters, worlds and animals of James Cameron's new science fiction film Avatar.

B

- ▶ [back scooping](#) liposuction around the hips and flanks to narrow the waist
- ▶ [bandity](#)
- ▶ [bankster](#) (n.) A person in the financial service industry who grows rich despite the continued impoverishment of those who depend on their services, and despite their apparent inability to succeed in business without constant government assistance.
- ▶ [bearleish](#)
- ▶ [Beatleawesomeness](#)
- ▶ [bing](#)
- ▶ [Binged](#)
- ▶ [blackgrass](#)
- ▶ [bloglet](#)
 1. (n.) A short blog entry
 2. (n.) A form of address for fellow bloggers
 3. (n.) e-mail update information system
- ▶ [blubdate](#) (v.) to update your blog
- ▶ [boqof](#) (n.) An advertising strategy that entices people to buy a product and get one for free
- ▶ [Boobgate](#)
- ▶ [bromosexual](#) (n.) A straight man who is very close to another man without being homosexual.
- ▶ [burquini](#) a burka which is made into a bikini
- ▶ [businessosphere](#)
- ▶ [buzz marketing](#) (n.) Word-of-mouth marketing

C

- ▶ [carbohydritis](#)
- ▶ [carpie](#) (n.) dirty, gray "plop" of slushy snow and petrified ice that falls off automobile mud flaps all winter in Wisconsin (coined in the US)
- ▶ [Celebenomics](#)
- ▶ [cheap and cheerful](#)
- ▶ [cherpumple](#) (n.) An apple pie baked inside a spice cake, a pumpkin pie baked inside a yellow cake, and a cherry pie baked inside a white cake, all stacked on top of each other and covered in cream cheese icing.
- ▶ [chinsumer](#)
- ▶ [chock](#)
- ▶ [climagate](#) (n.) (techn. women's soccer): blend of "jocks" and "chicks," girls who look pretty off the field but play just as hard as the boys
- ▶ [cloud computing](#) anthropogenic global warming: history's biggest scam
- ▶ [cloud generator](#) (n.) System of computing in which services and storage are provided over the Internet
- ▶ [co-work](#)

NeoCrawler

http://neocrawler.gwi.uni-muenchen.de/index.php?abfrage=show_progress&lemmaid=399& Google

Neologisms →

Progress of *halfalogue*

→ [edit lemma](#)

max: 255

x axis: calendar week (2011)
y axis: number of tokens per week

Process ID	Processed by	Time Restriction	Search String	No. of Pages	No. of Tokens	Date	Details
8917	neocrawler	w	halfalogue	1	1	2010-08-09	summary details ✓
9196	neocrawler	w	halfalogue	2	4	2010-08-16	summary details ✓
9475	neocrawler	w	halfalogue	3	4	2010-08-23	summary details ✓
11465	neocrawler	m	halfalogue	95	241	2010-10-05	summary details ✓
11747	neocrawler	m	halfalogue	87	255	2010-10-12	summary details ✓
12029	neocrawler	m	halfalogue	98	237	2010-10-19	summary details ✓
12311	neocrawler	m	halfalogue	51	140	2010-10-26	summary details ✓
12593	neocrawler	m	halfalogue	47	135	2010-11-02	summary details ✓
12906	neocrawler	m	halfalogue	59	151	2010-11-09	summary details ✓

Gehe zu „http://neocrawler.gwi.uni-muenchen.de/index.php?abfrage=diagramm&typ=verla...7:0:gruen,18:0:gruen,19:0:gruen,20:0:gruen,21:0:gruen&einheit=%&breite=840&hoehe=200“

NeoCrawler

http://neocrawler.gwi.uni-muenchen.de/index.php?abfrage=show_progress&lemmaid=3!
switch case else

[edit lemma](#)

x axis: calender week (2011)
y axis: number of tokens per week

Process ID	Processed by	Time Restriction	Search String	No. of Pages	No. of Tokens	Date	Details
8917	neocrawler	w	halfalogue	1	1	2010-08-09	summary details
9196	neocrawler	w	halfalogue	2	4	2010-08-16	summary details
9475	neocrawler	w	halfalogue	3	4	2010-08-23	summary details
11465	neocrawler	m	halfalogue	95	241	2010-10-05	summary details
11747	neocrawler	m	halfalogue	87	255	2010-10-12	summary details
12029	neocrawler	m	halfalogue	98	237	2010-10-19	summary details
12311	neocrawler	m	halfalogue	51	140	2010-10-26	summary details
12593	neocrawler	m	halfalogue	47	135	2010-11-02	summary details
12906	neocrawler	m	halfalogue	59	151	2010-11-09	summary details

Download Progress of halfalogue:

- type: html txt
- structure: chronological classified
- download processes since:
- download processes until:

[Impressum](#) - [Datenschutz](#) - [Kontakt](#) - [Sitemap](#)



NeoCrawler

http://neocrawler.gwi.uni-muenchen.de/index.php?abfrage=process_summary&id_process_ir Google

[Neologisms](#) → [Progress of halfalogue](#) → [Summary of Process No. 12906](#) → [Details of Process 12906](#)

Lemma: halfalogue

Date: 2010-11-09, 08:11
 Restriction: m
 Pages: 59
 Tokens: 151
 By: neocrawler

Save

(query took 17.0378 sec)

#	Tokens	Title	Co-text	Classify Site	Links
1	4	Association for Psychological Science - halfalogue	Association for Psychological Science - halfalogue Home About APS Contact - <ul style="list-style-type: none"> Home Join/Renew Benefits Advocacy halfalogue Home About APS Contact - Home Join/Renew Benefits Advocacy Teaching Students Books Psychology Links 	--field of discourse-- --type of source-- --authorship-- <input type="checkbox"/> delete entry <input type="checkbox"/> blacklist site:	
2	9	Psychological Science: Coining Buzzwords and Setting Trends - Association for Psychological Science	source for the latest psychological research. Previous Observation >> << Next Observation Psychological Science: Coining Buzzwords and Setting Trends Posted by --> By ---> When we hear a halfalogue , our brain tries to fill in the other half of the convection. Today's Express article "Stuck in Your Ear: Metro 'Halfalogues' Aren't Just Rude — They're Distractingly Irresistible!" (page 10) highlights not only a recent Psychological Science article, but the term "halfalogue" used by the author. Lauren Embeon, PhD candidate in psychology at Cornell Univeity, coined the catchy term to describe the scenario in which only one side of a cell phone convection	--field of discourse-- --type of source-- --authorship-- <input type="checkbox"/> delete entry <input type="checkbox"/> blacklist site:	
3	5	Smell the Coffee: Halfalogue is not better than no log at all - Life - The Charleston Gazette - West Virginia News and Sports -	Smell the Coffee: Halfalogue is not better than no log at all - Life - The Charleston Gazette - West Virginia News and Sports - You appear to be using an old veion of Internet Explorer. As a result, this website will not be displayed properly. We recommend upgrading to the latest veion of Internet Explorer or using an alternate web browser, such as Firefox . Halfalogue is not better than no log at all - Life - The Charleston Gazette - West Virginia News and Sports - You appear to be using an old veion of Internet Explorer. As a result, this website will not be displayed properly. We recommend upgrading to the latest veion of Internet Explorer or using an alternate web browser, such as Firefox . November 9, 2010 at 1:48	--field of discourse-- --type of source-- --authorship-- <input type="checkbox"/> delete entry <input type="checkbox"/> blacklist site:	
4	3	Smell the Coffee: Halfalogue is not better than no log at all - Topix	Smell the Coffee: Halfalogue is not better than no log at all - Topix Local News: Germany Sign Up Sign In 1 2 3 Are you satisfied or dissatisfied with the Midterm Election results? China Scotch in China: Single Malt, Single Country Mexico New Mexico's governor-elect Halfalogue is not better than no log at all - Topix Local News: Germany Sign Up Sign In 1 2 3 Are you satisfied or dissatisfied with the Midterm Election results? China Scotch in China: Single Malt, Single Country Mexico New Mexico's governor-elect unlikely to follow Arizona's lead	--field of discourse-- --type of source-- --authorship-- <input type="checkbox"/> delete entry <input type="checkbox"/> blacklist site:	
5	16	gulfnews : Sound affects	grunts, and go about your regular activities. Except with one sound. One that can drive you to distraction. One that doesn't allow you to resume your everyday routine. That of an overheard cellphone convection. Does it make see? Surely overhearing your partner talk to his boss about work or your teenage daughter giggling away with her best friend is easy to tune out? Lauren Embeon from Cornell Univeity in the US thinks otherwise. She explai the phenomenon in one phrase: "it is a halfalogue ". She uses an uncommon, unfamiliar word, a neologism really, to explain a concept we don't undetand. The word 'halfalogue' seems less of an explanation and more of a red herring. Optio: noise cut Fit, let's undetand what Embeon has got to do with any of this. She is the lead	--field of discourse-- --type of source-- --authorship-- <input type="checkbox"/> delete entry <input type="checkbox"/> blacklist site:	



NeoCrawler

http://neocrawler.gwi.uni-muenchen.de/index.php?abfrage=process_summary&id_process_12906

Neologisms → Progress of halfalogue → Summary of Process No. 12906 → Details of Process 12906

..--mode of use--

- objectlinguistic
- metalinguistic

..--type of source--

- blog
- news
- portal
- portal: directory
- portal: community
- portal: jobsites
- portal: Hollers
- portal: Gathers
- portal: Blippy
- portal: Bebo
- social networking
- social networking: Facebook
- social networking: Myspace
- social networking: Meetup
- file sharing
- file sharing: documents
- file sharing: video
- file sharing: music
- file sharing: fotos
- metalinguistic
- microblogging
- microblogging: Twitter,
- microblogging: Tumblr
- microblogging: other
- self-reference page
- academic
- other

..--authorship--

- private
- professional

By: neocrawler

..--field of discourse--

- politics
- general
- sports
- business
- law
- science
- advertising
- lifestyle
- lifestyle: celebrities
- lifestyle: fashion
- lifestyle: food & drink
- lifestyle: health
- entertainment
- entertainment: tv
- entertainment: movies
- entertainment: music
- computing/internet
- computing/internet: gaming
- computing/internet: technology
- computing/internet: business

Classify Site

Links

--field of discourse--

--type of source--

--authorship--

delete entry

blacklist site:

NeoCrawler

http://neocrawler.gwi.uni-muenchen.de/index.php?abfrage=discoverer

Starter of the Discoverer-Module

Please select the action you would like the script to perform.

Crawling:

- Crawl one or more URLs for neologisms
- Use Google Blog Search to retrieve a list of blog URLs and process the pages
- Upload file from on local disk
- Select multiple files from server HDD

Resource Managing:

- Build trigram statistics for phonotactic check
- Build automata from dictionary resource(s)
- Purge all temporary lists (warning: all previously found types will be deleted from the variable LEX lists and MDICs. You can revert to backups if this was done unintentionally)
- Undo last changes to temporary lists (revert to last backup)
- Re-compile variable .LEX files to temporary MDICs

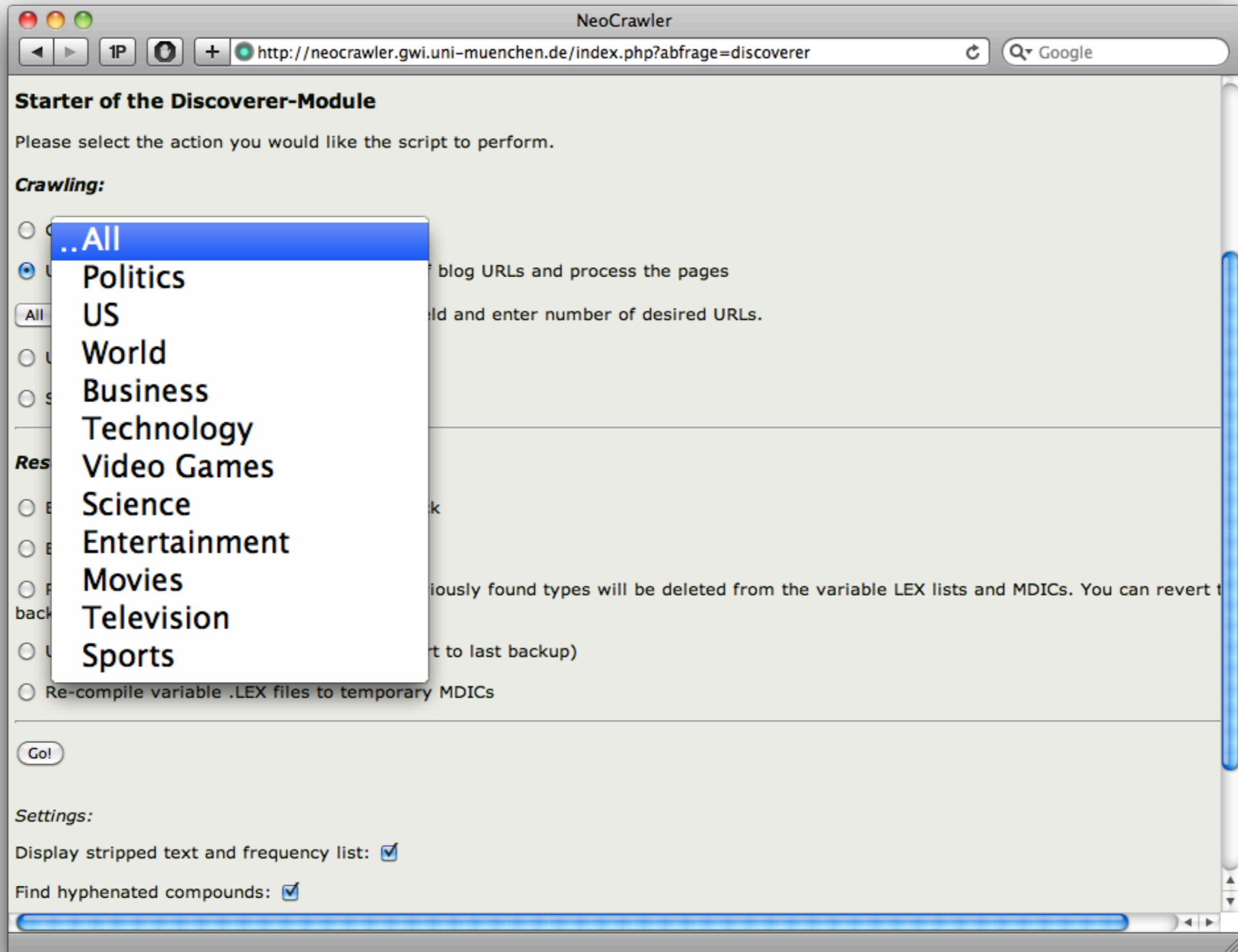
Go!

Settings:

Display stripped text and frequency list:

Find hyphenated compounds:

Only display candidates that have quality...



NeoCrawler

http://neocrawler.gwi.uni-muenchen.de/index.php?abfrage=discoverer

Starter of the Discoverer-Module

Please select the action you would like the script to perform.

Crawling:

- ..All
- Politics
- US
- World
- Business
- Technology
- Video Games
- Science
- Entertainment
- Movies
- Television
- Sports

Re-compile variable .LEX files to temporary MDICs

Go!

Settings:

Display stripped text and frequency list:

Find hyphenated compounds:


```
Neologism-Suggester Tool
http://neocrawler.gwi.uni-muenchen.de/discoverer/suggestMain.php
Google

Opened 10 LookupMD instances successfully
Crawling Google Blog Search to retrieve 10 URLs...
retrieved 10 Blog URLs from Google Blog Search.
List of URLs:
1.: "http://la.curbed.com/archives/2011/05/shiller_time_6.php"
2.: "http://www.scpr.org/news/2011/05/31/economist-expects-oc-house-prices-end-down-end-yea/"
3.: "http://dailycapitalist.com/2011/05/31/case-shiller-down-5-1-q1-2011-from-q1-2010/"
4.: "http://www.balloon-juice.com/2011/05/31/base-how-low-can-you-go/"
5.: "http://www.cultofmac.com/could-ios5-eliminate-the-need-for-itunes-sync-could-icloud-include-facebook-style-apps-speculation/98669"
6.: "http://www.cultofmac.com/"
7.: "http://politicalbetting.com/index.php/archives/2011/06/05/is-inverclyde-a-certain-labour-hold/"
8.: "http://politicalbetting.com/"
9.: "http://www.ritholtz.com/blog/2011/06/time-to-order-the-new-bugs/"
10.: "http://www.ritholtz.com/blog/"
Starting filtering process of retrieved URLs...



Preprocessing Text #1
Encoding is UTF-8, attempting full punctuation stripping method
Filtering out stopwords and strings with length < 3... Done.
Filtering out known names... Done.
Filtering out known tokens from Dictionary... Done.

 Show stripped Text of URL "http://la.curbed.com/archives/2011/05/shiller_time_6.php", text number 1:
 Show Frequency list of URL "http://la.curbed.com/archives/2011/05/shiller_time_6.php", text number 1:

No tokens remained after filtering.

Preprocessing Text #2
Warning, this source is not encoded in UTF-8, errors might occur...
Encoding is "utf-8, using compatibility punctuation stripping method
Filtering out stopwords and strings with length < 3... Done.
Filtering out known names... Done.
Filtering out known tokens from Dictionary... Done.
Filtering out known tokens from previous program runs... Done.
Constructing Candidate objects... Done.
```

Neologism-Suggester Tool

1P  <http://neocrawler.gwi.uni-muenchen.de/discoverer/suggestMain.php> 

Preprocessing Text #10
Encoding is UTF-8, attempting full punctuation stripping method
Filtering out stopwords and strings with length < 3... Done.
Filtering out known names... Done.
Filtering out known tokens from Dictionary... Done.
Filtering out known tokens from previous program runs... Done.
Constructing Candidate objects... Done.

Show stripped Text of URL "http://www.ritholtz.com/blog/", text number 10:
 Show Frequency list of URL "http://www.ritholtz.com/blog/", text number 10:

Adding unknown types to List...
took 167 seconds
26101 Tokens processed, 115 unknown types found.

List of unknown types

Display Explanation:

List of all encountered unknown types:

Show unknown types in Text #2:
 Show unknown types in Text #3:
 Show unknown types in Text #4:
 Show unknown types in Text #5:

- plain text
- frequency list
- unknown types

Neologism-Suggester Tool

http://neocrawler.gwi.uni-muenchen.de/discoverer/suggestMain.php

Adding unknown types to List...

Preprocessing Text #3
 Encoding is UTF-8, attempting full punctuation stripping method
 Filtering out stopwords and strings with length < 3... Done.
 Filtering out known names... Done.
 Filtering out known tokens from Dictionary... Done.
 Filtering out known tokens from previous program runs... Done.
 Constructing Candidate objects... Done.

Show stripped Text of URL "<http://dailycapitalist.com/2011/05/31/case-shiller-down-5-1-q1-2011-from-q1-2010/>", text number 3:

Case-Shiller Down 5.1%: What Will Stop It? | The Daily Capitalist Home About Archives Reading List Subscribe via RSS Business Plan Writer Must Read Posts The Smartest Guys in the Room The Law of Unintended Consequences Will We Have Hyperinflation In America? The Wall Street-Washington Financial Complex Why the 'Do Something' Stimulus Bill Won't Work Will We Have Inflation,Deflation,or Hyperinflation? (Download) Why You Should Ignore Economists What Is Money ? Part I Money:A Semi Fictional Fable A Plan For Economic Recovery Will We Have a Lost Decade(s) Like Japan? Economic Megatrends That Will Drive Our Future National Sales Tax,Chapter 3 2009:Why It Will Affect Everyone's Future For Generations To Come The Economics of Mass Destruction-Parts I and II Archived Articles June 2011 (8) May 2011 (37) April 2011 (46) March 2011 (50) February 2011 (33) January 2011 (32) December 2010 (38) November 2010 (46) October 2010 (33) September 2010 (27) August 2010 (28) July 2010 (34) June 2010 (32) May 2010 (46) April 2010 (37) March 2010 (34) February 2010 (26) January 2010 (28) December 2009 (17) November 2009 (7) October 2009 (20) September 2009 (23) August 2009 (35) July 2009 (34) June 2009 (28) May 2009 (23) April 2009 (21) March 2009 (25) February 2009 (19) January 2009 (11) December 2008 (10) November 2008 (3) October 2008 (5) August 2008 (1) April 2008 (1) March 2008 (3) February 2008 (2) December 2007 (1) Poor People Control 61% of World's Wealth » « Home Prices vs. Gold vs. Inflation Case-Shiller Down 5.1%:What Will Stop It? By Jeff Harding,on May ,2011 The Case-Shiller Housing Report for March 2011 is down 5.1% in Q1 compared to a year ago. "This month's report is marked by the confirmation of a double-dip in home prices across much of the nation. The National Index,the 20-City Composite and 12 MSAs all hit new lows with data reported through March 2011. The National Index fell 4.2% over the first quarter alone,and is down 5.1% compared to its year-ago level . Home prices continue on their downward spiral with no relief in sight," says David M. Blitzer,Chairman of the Index Committee at S&P Indices. "Since December 2010,we have found an increasing number of markets posting new lows. In March 2011,12 cities –Atlanta,Charlotte,Chicago,Cleveland,Detroit,Las Vegas,Miami,Minneapolis,New York,Phoenix,Portland (OR) and Tampa –fell to their lowest levels as measured by the current housing cycle. Washington D.C. was the only MSA displaying positive trends with an annual growth rate of +4.3% and a 1.1% increase from its February level. As this chart points out,prices on a national basis have fallen to 2003 levels. The more depressing fact is that Washington,D.C. prices are rising,reflecting the growth and power of our national government. Another thing to think about is the decline of homeowner equity in their homes. At this point,CoreLogic would put it at about 28% of all homes in America. Another 10% decline would mean one-third of all homes would be under

Neologism-Suggester Tool

1P <http://neocrawler.gwi.uni-muenchen.de/discoverer/suggestMain.php> Google

Show Frequency list of URL

"http://dailycapitalist.com/2011/05/31/case-shiller-down-5-1-q1-2011-from-q1-2010/", text number 3:

the --> 81
of --> 47
and --> 38
to --> 36
is --> 35
The --> 29
that --> 23
in --> 23
at --> 19
will --> 16
be --> 15
housing --> 14



Neologism-Suggester Tool

http://neocrawler.gwi.uni-muenchen.de/discoverer/suggestMain.php

Show unknown types in Text #4:

Rank	Neologism	Candidate	Garbage	Ignore	Type Quality	Avg. Trigram Freq	Context	Show extended co-text
1.	<i>house-to-buy</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	10	69.95864%	[...] Estates that had no equivalent <i>house-to-buy</i> . The high-end rental market [...]	<input type="checkbox"/>
2.	<i>urbanmeemaw</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	10	57.35686%	[...] it did not come true <i>urbanmeemaw</i> on Toledo El Cid [...]	<input type="checkbox"/>
3.	<i>Poliblog</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	10	45.57486%	[...] Pandagon Pat Lang Pharyngula Photodude <i>Poliblog</i> Postbourgie Prose Before Hos [...]	<input type="checkbox"/>
4.	<i>Mahablog</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	10	42.63854%	[...] Money League of Ordinary Gentleman <i>Mahablog</i> Main Street Mario Piperni [...]	<input type="checkbox"/>
5.	<i>Hoekstroika</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	10	26.06378%	[...] News For Conservatives Green Balloons <i>Hoekstroika</i> hoocoodanode Hoot-Smalley Humor I [...]	<input type="checkbox"/>
6.	<i>well-picked</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	10	20.73339%	[...] a limited resource, and a <i>well-picked</i> location will appreciate in [...]	<input type="checkbox"/>
7.	<i>TBOGG</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	10	19.13427%	[...] Talk Left Talking Points Memo <i>TBOGG</i> The Booman Tribune The [...]	<input type="checkbox"/>
8.	<i>nonbubblicious</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	10	18.52771%	[...] residential construction boom, based on <i>nonbubblicious</i> historical rates of return [...]	<input type="checkbox"/>
9.	<i>Fucked-up-edness</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	10	16.73166%	[...] The Middle-Class Fuck The Poor <i>Fucked-up-edness</i> Fucksaw Gamer Dork Gay [...]	<input type="checkbox"/>
10.	<i>businessmandood</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	10	13.08138%	[...] She's a lawyer, he's a <i>lawyer/consultant/businessmandood</i> . Neither is going to [...]	<input type="checkbox"/>
11.	<i>malkinfreude</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	10	10.88503%	[...] Glue Mainstream Media's McCain Mancrush <i>malkinfreude</i> Manic Progressive Massa Massage [...]	<input type="checkbox"/>
12.	<i>Fucksaw</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	10	9.24731%	[...] Middle-Class Fuck The Poor <i>Fucked-up-edness Fucksaw</i> Gamer Dork Gay Rights [...]	<input type="checkbox"/>
13.	<i>Mistermix</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	10	8.24767%	[...] Dennis G DougJ John Kay <i>Mistermix</i> Tim F Tom Levenson [...]	<input type="checkbox"/>
14.	<i>s-his-nose</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	10	8.15757%	[...] of the US housing market. <i>What's-his-nose</i> in the quote snake-oiling, [...]	<input type="checkbox"/>
15.	<i>low-ass</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	10	7.67577%	[...] a garage in a nice, <i>lower-middle-class</i> subdivision. It's a college [...]	<input type="checkbox"/>
16.	<i>Glibertarianism</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	10	7.38691%	[...] Getting The Band Back Together <i>Glibertarianism</i> Going Galt Goldman Sachs [...]	<input type="checkbox"/>
17.	<i>Wingularity</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	10	6.20347%	[...] the War on Drugs The <i>Wingularity</i> This is not where [...]	<input type="checkbox"/>
18.	<i>koolaide</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	10	5.88641%	[...] much of the "American Dream" <i>koolaide</i> . There is more than [...]	<input type="checkbox"/>
19.	<i>Fish-Liftine</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	10	[...]	[...] Late Night Weekend Thread: Competitive <i>Fish-Liftine</i> And I'll be damned [...]	<input type="checkbox"/>



Neologism-Suggester Tool

http://neocrawler.gwi.uni-muenchen.de/discoverer/suggestMain.php

Type	Quality	Avg. Trigram Freq	Text	Actions
<input type="radio"/> Garbage <input checked="" type="radio"/> Ignore	10	1.71859%	[...] and the Great Recession, the <i>Nifty-Fifty</i> , Asian Contagion, the Dotcom [...]	<input type="checkbox"/> Show extended co-text
<input type="radio"/> Garbage <input checked="" type="radio"/> Ignore	10	1.328%	the past six weeks, his <i>tactically-allocated</i> accounts have moved from	<input type="checkbox"/> Show extended co-text
<input type="radio"/> Garbage <input checked="" type="radio"/> Ignore	10	1.05689%	[...] of assets in his most <i>aggressively-managed</i> accounts into short positions. [...]	<input type="checkbox"/> Show extended co-text
<input type="radio"/> Garbage <input checked="" type="radio"/> Ignore	10	0.53304%	[...] \$1 billion came into the <i>equity-fund</i> coffers last week, according [...]	<input type="checkbox"/> Show extended co-text
<input type="radio"/> Garbage <input checked="" type="radio"/> Ignore	9	7.52688%	[...] Intellectual Property (48) Investing (932) Legal (493) M&A (93) <i>MacroNotes</i> (1294) Markets (3116) Mathematics (94) Media (858) Music (600) [...]	<input type="checkbox"/> Show extended co-text
<input type="radio"/> Garbage <input checked="" type="radio"/> Ignore	0	0%	[...] bad calls (373) Regulation (492) Research (15) Retail (245) <i>RR&A</i> (48) Science (166) Short Selling (104) Sports (35) [...]	<input type="checkbox"/> Show extended co-text

ter.

Neologism-Suggester Tool

http://neocrawler.gwi.uni-muenchen.de/discoverer/suggestMain.php

Rank	Neologism	Candidate	Garbage	Ignore	Type	Quality	Avg. Trigram Freq	Example
1.	<i>streets-you</i>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	Type	Quality: 10	30.87951%	[...] there is blood in the <i>streets-you</i> can not time the [...]
2.	<i>Stagflationary</i>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	Type	Quality: 10	12.99972%	[...] 2011 Sinking Manufacturing Is A <i>Stagflationary</i> Trend
3.	<i>monetary-induced</i>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	Type	Quality: 10	12.77325%	[...] are related to:positive economic growth,rising <i>wages</i> ,and, <i>unfortunately</i> , <i>n induced</i> price inflation. I don't [...]
4.	<i>Econophile</i>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	Type	Quality: 10	8.35401%	[...] Capitalist on Fox Business News <i>Econophile</i> On Fox Business News [...]
5.	<i>neg-equity</i>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	Type	Quality: 10	8.08519%	[...] methodologies used in the Zillow/Core-Logic <i>neg-equity</i> estimates only u
6.	<i>Destruction-Parts</i>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	Type	Quality: 10	4.48856%	[...] Come The Economics of Mass <i>Destruction-Parts</i> I and II Archived [...]
7.	<i>median-sized</i>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	Type	Quality: 10	2.39868%	[...] of gold can buy a <i>median-sized</i> home in the US. [...]
8.	<i>CoreLogic</i>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	Type	Quality: 9	3.24944%	[...] in their homes. At this <i>point</i> , <i>CoreLogic</i> would put it at [...]
9.	<i>20-City</i>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	Type	Quality: 5	0.14888%	[...] the nation. The National Index,the <i>20-City</i> Composite and 12 MSAs [...]

Neologism-Suggester Tool

http://neocrawler.gwi.uni-muenchen.de/discoverer/suggestMain.php

Rank	Neologism	Candidate	Garbage	Ignore	Type Quality	Avg. Trigram Freq	Context
24.	<i>transit-poor</i>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	10	3.21754%	[...] limits their interest in the <i>transit-poor</i> areas where single family
25.	<i>snake-oiling</i>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	10	3.20926%	[...] market. What's-his-nose in the quote <i>snake-oiling</i> , "BUY! BUY
26.	<i>never-lived-in</i>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	10		friends bought a million dollar-valued <i>never-lived-in</i> foreclosure
		of type:		Unclassified			
27.	<i>Galtian</i>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	10	2.94458%	[...] situation will improve until our <i>Galtian</i> overlords are made to [...]
28.	<i>hoocoodanode</i>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	10	2.82051%	[...] For Conservatives Green Balloons Hoekstroika <i>hoocoodanode</i> Can [...]
29.	<i>end-ups</i>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	10	2.53102%	[...] a home always and necessarily <i>end-ups</i> with "nothing" (quote q
30.	<i>Crazification</i>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	10	2.37612%	[...] Cooking Corrections Crab Bucket Politics <i>Crazification</i> Factor C
31.	<i>Spectering</i>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	10	2.30562%	[...] Village is Missing its Idiot <i>Spectering</i> Sports Starbursts Starve th
32.	<i>dollar-valued</i>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	10	1.78209%	[...] close friends bought a million <i>dollar-valued</i> never-lived-in forec



Neologism-Suggester Tool

http://neocrawler.gwi.uni-muenchen.de/discoverer/suggestMain.php

Neologism	Type	Garbage	Ignore	Type Quality	Avg. Trigram Freq	Context
24. : <i>transit-poor</i>	<input type="radio"/> Candidate	<input type="radio"/>	<input checked="" type="radio"/>	10	3.21754%	mits their interest in the <i>transit-poor</i> areas where single family
25. : <i>snake-oiling</i>	<input type="radio"/> Candidate	<input type="radio"/>	<input checked="" type="radio"/>	10	3.20926%	arket. What's-his-nose in the quote <i>snake-oiling</i> , "BUY! BUY
26. : <i>never-lived-in</i>	<input checked="" type="radio"/> Candidate	<input type="radio"/>	<input type="radio"/>	10	3.02592%	ends bought a million dollar-valued <i>never-lived-in</i> foreclosure
27. : <i>Galtian</i>	<input type="radio"/> Candidate	<input type="radio"/>	<input checked="" type="radio"/>	10	3.02592%	uation will improve until our <i>Galtian</i> overlords are made to [...]
28. : <i>hoocoodanode</i>	<input type="radio"/> Candidate	<input type="radio"/>	<input checked="" type="radio"/>	10	2.94458%	r Conservatives Green Balloons Hoekstroika <i>hoocoodanode</i> [...]
29. : <i>end-ups</i>	<input type="radio"/> Candidate	<input type="radio"/>	<input checked="" type="radio"/>	10	2.82051%	home always and necessarily <i>end-ups</i> with "nothing" (quote q
30. : <i>Crazification</i>	<input type="radio"/> Candidate	<input type="radio"/>	<input checked="" type="radio"/>	10	2.82051%	ooking Corrections Crab Bucket Politics <i>Crazification</i> Factor C
31. : <i>Spectering</i>	<input type="radio"/> Candidate	<input type="radio"/>	<input checked="" type="radio"/>	10	2.30562%	[...] Village is Missing its Idiot <i>Spectering</i> Sports Starbursts Starve th
32. : <i>dollar-valued</i>	<input type="radio"/> Candidate	<input type="radio"/>	<input checked="" type="radio"/>	10	1.78209%	[...] close friends bought a million <i>dollar-valued</i> never-lived-in forec

of type:



Neologism-Suggester Tool

http://neocrawler.gwi.uni-muenchen.de/discoverer/suggestMain.php

Neologism	Type	Quality	Avg. Trigram Freq	Example
24. : <i>transit-poor</i>	<input type="radio"/> Candidate <input type="radio"/> Garbage <input checked="" type="radio"/> Ignore	10	3.21754%	[...] limits their interest in the <i>transit-poor</i> areas where single family
25. : <i>snake-oiling</i>	<input type="radio"/> Candidate <input type="radio"/> Garbage <input checked="" type="radio"/> Ignore	10	3.20926%	[...] market. What's-his-nose in the quote <i>snake-oiling</i> , "BUY! BUY
26. : <i>never-lived-in</i>	<input type="radio"/> Candidate <input type="radio"/> Garbage <input checked="" type="radio"/> Ignore	10	3.02592%	[...] friends bought a million dollar-valued <i>never-lived-in</i> foreclosure
27. : <i>Galtian</i>	<input type="radio"/> Candidate <input type="radio"/> Garbage <input checked="" type="radio"/> Ignore	10	2.94458%	[...] situation will improve until our <i>Galtian</i> overlords are made to [...]
28. : <i>hoocoodanode</i>	<input type="radio"/> Candidate <input type="radio"/> Garbage <input checked="" type="radio"/> Ignore	10	2.82051%	[...] For Conservatives Green Balloons Hoekstroika <i>hoocoodanode</i> Can [...]
29. : <i>end-ups</i>	<input type="radio"/> Candidate <input type="radio"/> Garbage <input checked="" type="radio"/> Ignore	10	2.53102%	[...] a home always and necessarily <i>end-ups</i> with "nothing" (quote q
30. : <i>Crazification</i>	<input type="radio"/> Candidate <input type="radio"/> Garbage <input checked="" type="radio"/> Ignore	10	2.37612%	[...] Cooking Corrections Crab Bucket Politics <i>Crazification</i> Factor C
31. : <i>Spectering</i>	<input type="radio"/> Candidate <input type="radio"/> Garbage <input checked="" type="radio"/> Ignore	10	2.30562%	[...] Village is Missing its Idiot <i>Spectering</i> Sports Starbursts Starve th
32. : <i>dollar-valued</i>	<input type="radio"/> Candidate <input type="radio"/> Garbage <input checked="" type="radio"/> Ignore	10	1.78209%	[...] close friends bought a million <i>dollar-valued</i> never-lived-in forec

.Unclassified

- Multiple
- Phrase
- Compound
- Derivation
- Conversion
- Blend
- Clipping
- Back-Formation
- Abbreviation
- Epo/Toponym
- Ex Nihilo



Neologism-Suggester Tool

http://neocrawler.gwi.uni-muenchen.de/discoverer/suggestMain.php

56. : <i>GSAVE@</i>	<input type="radio"/> Candidate	<input type="radio"/> Garbage	<input checked="" type="radio"/> Ignore	Type Quality: 5	Avg. Trigram Freq 3.94955%	[...] War War on Terror aka <i>GSAVE@</i> WATB We Are All [...]
57. : <i>insane@</i>	<input type="radio"/> Candidate	<input type="radio"/> Garbage	<input checked="" type="radio"/> Ignore	Type Quality: 5	Avg. Trigram Freq 1.75352%	[...] quit because i was going <i>insane@</i> .) Reply MattR - May [...]
58. : <i>opie_jeanne</i>	<input type="radio"/> Candidate	<input type="radio"/> Garbage	<input checked="" type="radio"/> Ignore	Type Quality: 2	Avg. Trigram Freq 10.66998%	[...] of this... Horrible—just horrible... Reply <i>opie_jeanne</i> - June 1,
59. : <i>goblue72</i>	<input type="radio"/> Candidate	<input type="radio"/> Garbage	<input checked="" type="radio"/> Ignore	Type Quality: 2	Avg. Trigram Freq 2.30218%	[...] and would likely default. Reply <i>goblue72</i> - June 1, 2011 [...]
60. : <i>mid-\$500s</i>	<input type="radio"/> Candidate	<input type="radio"/> Garbage	<input checked="" type="radio"/> Ignore	Type Quality: 0	Avg. Trigram Freq 0.01182%	[...] for sale now, one for <i>mid-\$500s</i> and one for over [...]
61. : <i>PB&J</i>	<input type="radio"/> Candidate	<input type="radio"/> Garbage	<input checked="" type="radio"/> Ignore	Type Quality: 0	Avg. Trigram Freq 0%	[...] lunch that week, or make <i>PB&J</i> , rather than order sushi [...]
62. : <i>WWJBD</i>	<input type="radio"/> Candidate	<input type="radio"/> Garbage	<input checked="" type="radio"/> Ignore	Type Quality: 0	Avg. Trigram Freq 0%	[...] You Can Afford It) WTF? <i>WWJBD</i> - What Would Jenna Bush
63. : <i>sh*t</i>	<input type="radio"/> Candidate	<input type="radio"/> Garbage	<input checked="" type="radio"/> Ignore	Type Quality: 0	Avg. Trigram Freq 0%	[...] If you get in the <i>sh*t</i> for a year, you [...]
64. : <i>\$4mm</i>	<input type="radio"/> Candidate	<input type="radio"/> Garbage	<input checked="" type="radio"/> Ignore	Type Quality: 0	Avg. Trigram Freq 0%	[...] Friends own what was a <i>\$4mm</i> home when they bought [...]

Neologism-Suggester Tool

http://neocrawler.gwi.uni-muenchen.de/discoverer/suggestMain.php

56. : <i>GSAVE@</i>	<input type="radio"/> Candidate	<input type="radio"/> Garbage	<input checked="" type="radio"/> Ignore	Type Quality: <u>5</u>	Avg. Trigram Freq 3.94955%	[...] War War on Terror aka <i>GSAVE@</i> WATB We Are All [...]
57. : <i>insane@</i>	<input type="radio"/> Candidate	<input type="radio"/> Garbage	<input checked="" type="radio"/> Ignore	Type Quality: <u>-</u>	Avg. Trigram Freq 1.75352%	[...] quit because i was going <i>insane@</i> .) Reply MattR - May [...]
58. : <i>opie_jeanne</i>						horrible... Reply <i>opie_jeanne</i> - June 1,
59. : <i>goblue72</i>						Reply <i>goblue72</i> - June 1, 2011 [...]
60. : <i>mid-\$500s</i>	<input type="radio"/> Candidate					now, one for <i>mid-\$500s</i> and one for over [...]
61. : <i>PB&J</i>		<input type="radio"/> Garbage	<input checked="" type="radio"/> Ignore	Type Quality: <u>0</u>	Avg. Trigram Freq 0%	<i>PB&J</i> , rather than order sushi [...]
62. : <i>WWJBD</i>	<input type="radio"/> Candidate	<input type="radio"/> Garbage	<input checked="" type="radio"/> Ignore	Type Quality: <u>0</u>	Avg. Trigram Freq 0%	[...] You Can Afford It) WTF? <i>WWJBD</i> - What Would Jenna Bush
63. : <i>sh*t</i>	<input type="radio"/> Candidate	<input type="radio"/> Garbage	<input checked="" type="radio"/> Ignore	Type Quality: <u>0</u>	Avg. Trigram Freq 0%	[...] If you get in the <i>sh*t</i> for a year, you [...]
64. : <i>\$4mm</i>	<input type="radio"/> Candidate	<input type="radio"/> Garbage	<input checked="" type="radio"/> Ignore	Type Quality: <u>0</u>	Avg. Trigram Freq 0%	[...] Friends own what was a <i>\$4mm</i> home when they bought [...]

<http://neocrawler.gwi.uni-muenchen.de>

- Average trigram frequency < .1%: -1
- Average trigram frequency < .05%: -1
- Token contains non-latin chars or non-letter-chars other than hyphen or single-quote: -5
- Token contains numbers: -1
- Token contains numbers other than 1, 2, 4, 8: -1
- Token contains consecutive numbers: -3

OK



World Wide Web

c. download web pages

a. search blogs

Google blogs

2. web search

3. find links to relevant pages

Google

5. download web pages

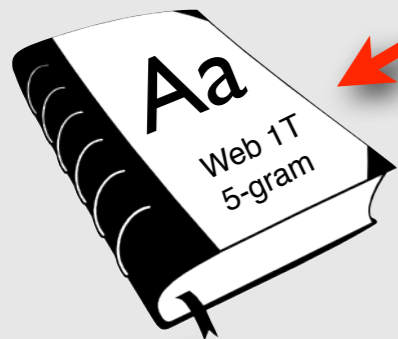
b. get links to blog entries

4. return links to NeoCrawler



6. automatic post-processing

d. compare web sites to word list



1. periodical monitoring

7. save downloaded web pages

e. list of rated suggestions

f. manually select neologisms

g. save to database for future crawling

8. show results in web interface



web interface

database

```

    @ id_process_info : int(11)
    @ processed_by : varchar(256)
    @ processed_date : int(11)
    @ process_comments : varchar(4096)
    @ search_string : varchar(256)
    @ time_restriction : varchar(256)
    @ no_tokens : int(11)
    @ id_lemma : int(11)
    @ access_date : varchar(25)

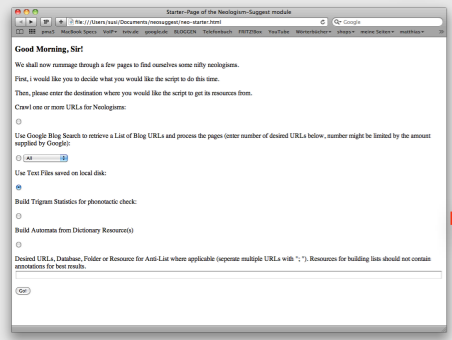
    @ id_token : int(11)
    @ id_source : int(11)
    @ id_version : int(11)
    @ id_author : int(11)
    @ pos : varchar(256)
    @ form : varchar(256)
    @ count : int(11)
    @ component : varchar(256)
    @ context_1 : varchar(4096)
    @ context_2 : varchar(4096)
    @ collocation : varchar(256)
    @ mode_use : varchar(256)
    @ novelty_markers : varchar(1)
    @ explanation : varchar(256)
    @ paraphrase : varchar(2048)
    @ ref_ontology : varchar(256)
    @ id_author : int(11)
    @ age : varchar(256)
    @ gender : varchar(1)
    @ role : varchar(256)
    @ nation : varchar(2)
    @ nickname : varchar(256)
    @ first_name : varchar(256)
    @ last_name : varchar(256)

    @ id_lemma : int(11)
    @ id_blacklist : int(11)
    @ blacklist_item : varchar(256)

    @ id_source : int(11)
    @ url : varchar(256)
    @ field : varchar(256)
    @ source_type : varchar(256)

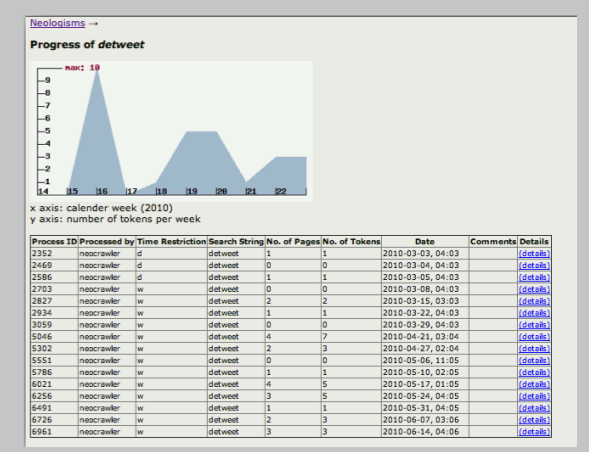
    @ id_version : int(11)
    @ creation_date : varchar(256)
    @ field : varchar(256)
    @ no_tokens : int(11)
    @ title : varchar(256)
    @ html : mediumtext

    @ id_author : int(11)
    @ age : varchar(256)
    @ gender : varchar(1)
    @ role : varchar(256)
    @ nation : varchar(2)
    @ nickname : varchar(256)
    @ first_name : varchar(256)
    @ last_name : varchar(256)
  
```



Discoverer

Observer

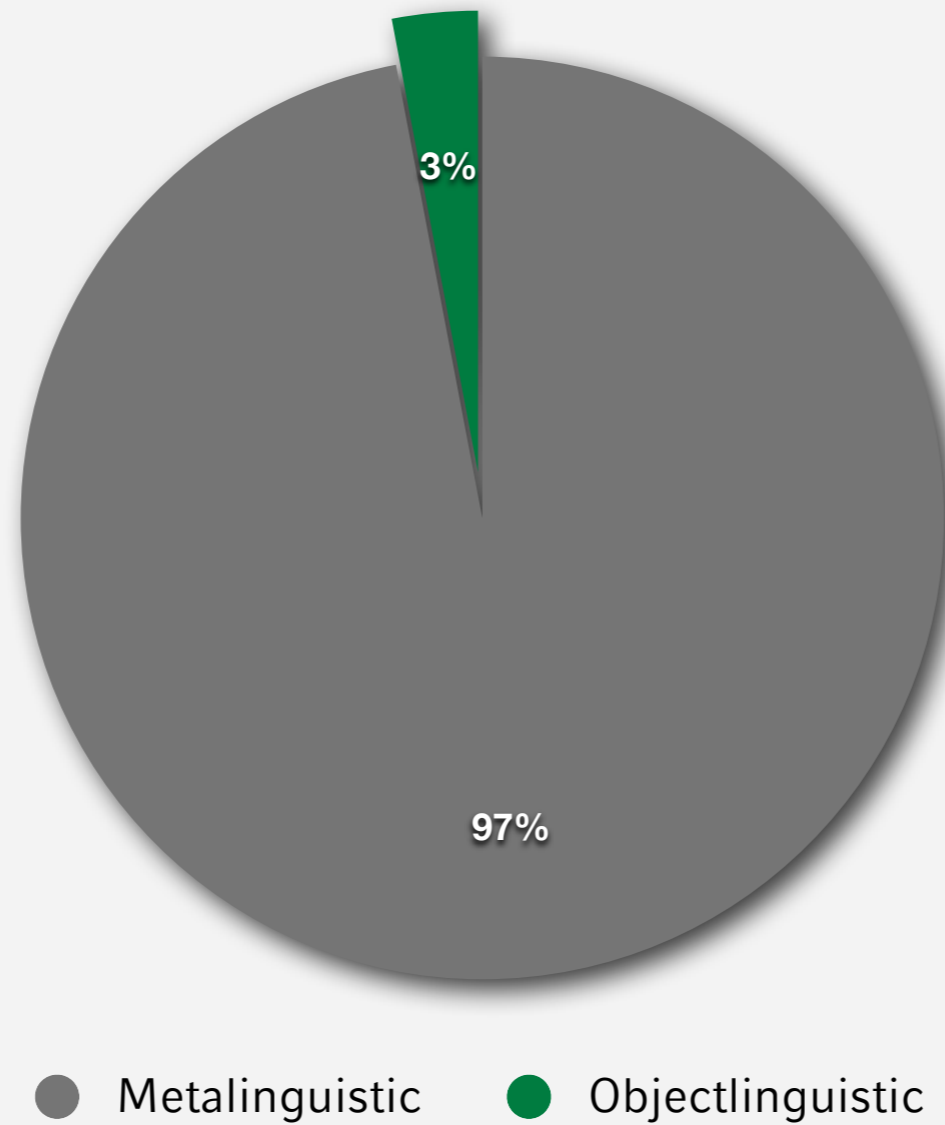




halfalogue



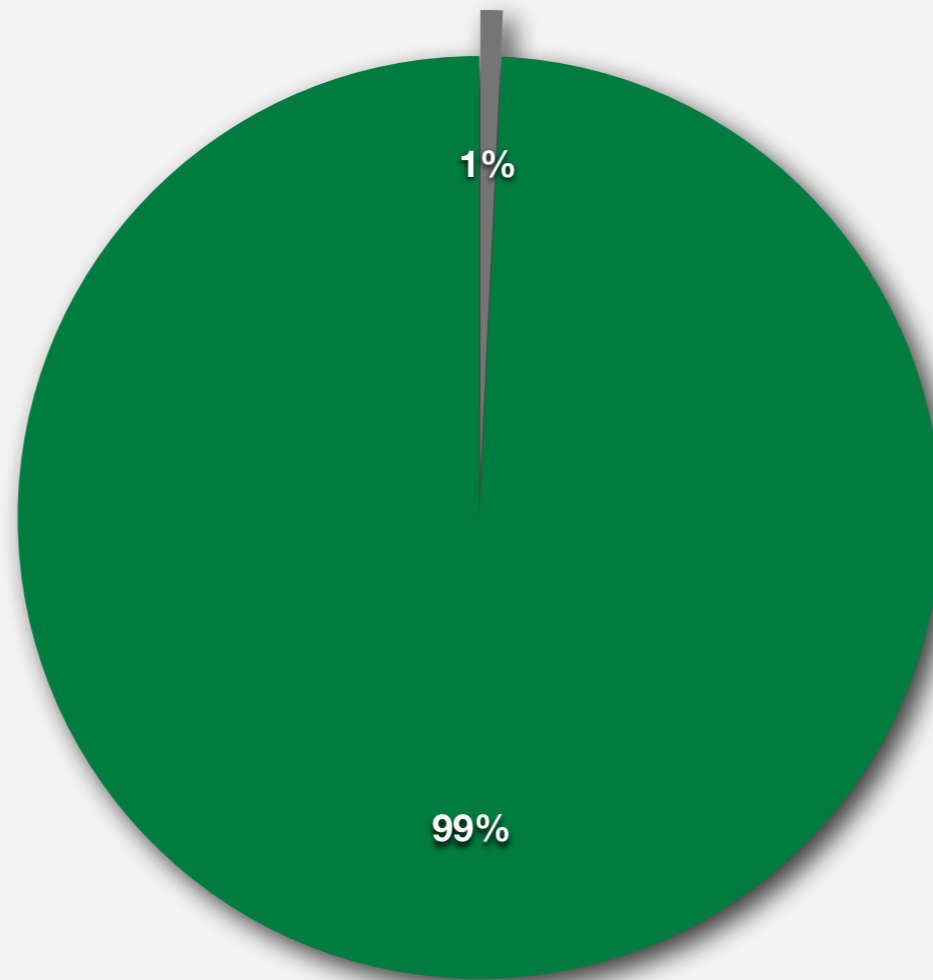
halfalogue - Mode of Use



cherpumpke



cherpumpke - Mode of Use



● Metalinguistic ● Objectlinguistic



LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN



ISLE 2, Boston •

Susanne Stegmayr & Daphné Kerremans (LMU Munich)

Neologisms on the Internet

The NeoCrawler

