

From the bottom of the wiki hole

Peter D Smits

Committee on Evolutionary Biology, University of Chicago

March 14, 2014

or...



oh god how did this get in here I am not good with computers

What?



WIKIPEDIA
The Free Encyclopedia

How many?



WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikimedia Shop

▶ Interaction

▶ Tools

▼ Languages



Special page

Statistics

Page statistics	
Content pages	4,471,031
Pages	32,424,589
(All pages in the wiki, including talk pages, redirects, etc.)	
Uploaded files	827,045
Edit statistics	
Page edits since Wikipedia was set up	698,636,135
Average edits per page	21.55
User statistics	
Registered users	20,922,890
Active registered users (list of members)	130,755
(Users who have performed an action in the last 30 days)	
Bots (list of members)	738
Administrators (list of members)	1,416
Bureaucrats (list of members)	36
Reviewers (list of members)	6,063
Autochecked users (list of members)	5
Stewards (list of members)	0
Account creators (list of members)	174
Importers (list of members)	2
Transwiki importers (list of members)	0

Just how much time could I waste?

Read rate: 600 wpm, 16 hours/day = 17,000,000 words/month.

It would take **7** (!!!!!) years to read wikipedia if it remained static.

at about 1:19 this afternoon...

1 Billionth edit [screenshot!](#)
Total visits: 1839

[Translate it!](#)

Total edits in [Wikimedia projects](#):

2 , 250 , 471 , 332

[All human knowledge](#) - [There is a deadline](#)



[Donate to Wikimedia Foundation](#)

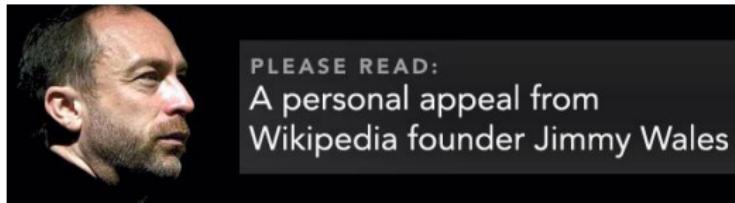
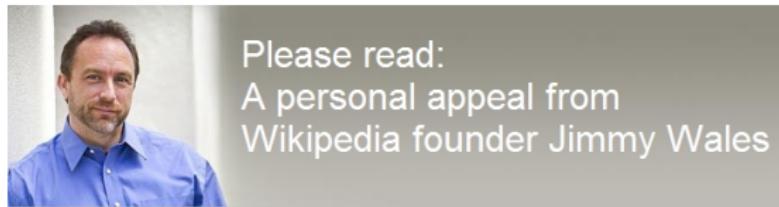
Press F11 for fullscreen

Developed by [emijrp](#) (Inspired by [Zizzi](#))

Who?

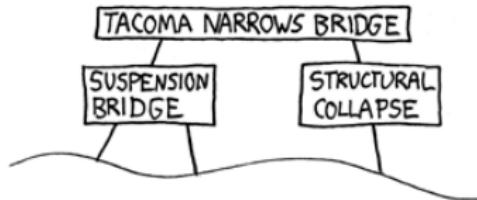


Go ahead and take my money

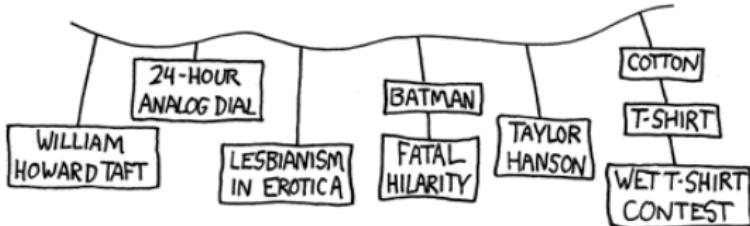


Problems? In my wiki?

THE PROBLEM WITH WIKIPEDIA:



[THREE HOURS OF]
FASCINATED CLICKING



Thar be dragons

<http://en.wikipedia.org/wiki/Special:Random>

Oh no pigeons

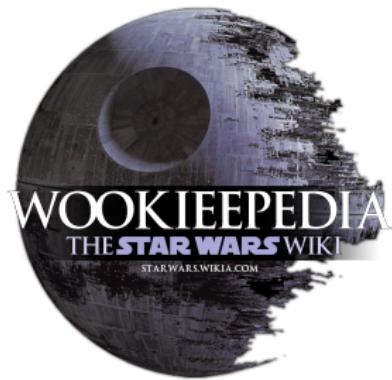
[   CHOOSE YOUR
WIKI GAME   ]



There is more than one wiki?

wikia

wiki-ing intensifies



heavy breathing

A WIKI OF ICE AND FIRE
A SONG OF ICE AND FIRE & GAME OF THRONES

GAME OF THRONES
WIKI

nothing is sacred

WikiSein

SIMPSONS WIKI

DRAGONBALL WIKI



scariest link on the internet

<http://tvtropes.org/pmwiki/randomitem.php?p=1>

you might loose 1 to 24 hours if you go to this link.

and possibly your sanity.



WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikimedia Shop

Interaction
Tools

Languages

Special page

Statistics

Page statistics	
Content pages	4,471,031
Pages	32,424,589
(All pages in the wiki, including talk pages, redirects, etc.)	
Uploaded files	
Uploaded files	827,045
Edit statistics	
Page edits since Wikipedia was set up	698,636,135
Average edits per page	21.55
User statistics	
Registered users	20,922,890
Active registered users (list of members)	130,755
(Users who have performed an action in the last 30 days)	
Bots (list of members)	738
Administrators (list of members)	1,416
Bureaucrats (list of members)	36
Reviewers (list of members)	6,063
Autochecked users (list of members)	5
Stewards (list of members)	0
Account creators (list of members)	174
Importers (list of members)	2
Transwiki importers (list of members)	0



WOOKIEEPEDIA
THE STAR WARS® WIKI

On the Wiki
Wiki Activity

Feat
Rand

Statistics

Special page

Page statistics	
Content pages	109,669
Pages	355,130
(All pages in the wiki, including talk pages, redirects, etc.)	
Uploaded files	
Uploaded files	61,231
Edit statistics	
Page edits since Wookieepedia was set up	4,785,885
Average edits per page	13.48
User statistics	
Registered users	23,101,892
Active users (list of members)	622
(Users who have performed an action in the last 30 days)	
Bots (list of members)	10
Administrators (list of members)	21
Bureaucrats (list of members)	5

gotta catch 'em all



WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikimedia Shop

Interaction
Tools

Languages

Special page

Statistics

Page statistics	
Content pages	4,471,031
Pages	32,424,589
(All pages in the wiki, including talk pages, redirects, etc.)	
Uploaded files	
Uploaded files	827,045
Edit statistics	
Page edits since Wikipedia was set up	698,636,135
Average edits per page	21.55
User statistics	
Registered users	20,922,890
Active registered users (list of members)	130,755
(Users who have performed an action in the last 30 days)	
Bots (list of members)	738
Administrators (list of members)	1,416
Bureaucrats (list of members)	36
Reviewers (list of members)	6,063
Autochecked users (list of members)	5
Stewards (list of members)	0
Account creators (list of members)	174
Importers (list of members)	2
Transwiki importers (list of members)	0

Statistics

From Bulbapedia, the community-driven PokéMón encyclopedia

Page statistics	
Content pages	25,474
Pages	159,945
(All pages in the wiki, including talk pages, redirects, etc.)	
Uploaded files	0
Edit statistics	
Page edits since Bulbapedia was set up	2,105,596
Average edits per page	13.16
User statistics	
Registered users	108,124
Active users (list of members)	855
(Users who have performed an action in the last 30 days)	
Bots (list of members)	3
Administrators (list of members)	10
Bureaucrats (list of members)	12
abuse (list of members)	51
editlock-exempt (list of members)	9
Junior Administrators (list of members)	10
Senior Administrators (list of members)	4
Editorial Board (list of members)	13
Check users (list of members)	0
steward (list of members)	0
Widget editors (list of members)	2

engaged



WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikimedia Shop

Interaction
Tools

Languages

Special page

Statistics

Page statistics	
Content pages	4,471,031
Pages	32,424,589
(All pages in the wiki, including talk pages, redirects, etc.)	
Uploaded files	
Uploaded files	827,045
Edit statistics	
Page edits since Wikipedia was set up	698,636,135
Average edits per page	21.55
User statistics	
Registered users	20,922,890
Active registered users (list of members)	130,755
(Users who have performed an action in the last 30 days)	
Bots (list of members)	738
Administrators (list of members)	1,416
Bureaucrats (list of members)	36
Reviewers (list of members)	6,063
Autochecked users (list of members)	5
Stewards (list of members)	0
Account creators (list of members)	174
Importers (list of members)	2
Transwiki importers (list of members)	0

On MA Por
Wiki Activity

Statistics

Special page

Page statistics	
Content pages	36,775
Pages	143,181
(All pages in the wiki, including talk pages, redirects, etc.)	
Uploaded files	37,133
Edit statistics	
Page edits since Memory Alpha was set up	1,611,607
Average edits per page	11.26
User statistics	
Registered users	23,101,893
Active users (list of members)	196
(Users who have performed an action in the last 30 days)	
Bots (list of members)	11
Administrators (list of members)	22
Bureaucrats (list of members)	2
rollback (list of members)	2
Check users (list of members)	0

Analysis time!



tl;dr

- ▶ memory alpha dump
 - ▶ XML format
 - ▶ every, and i mean every, current page
 - ▶ no images
 - ▶ ~ 313 MB (wikipedia is ~ 40 GB)
- ▶ R plus packages
 - ▶ see what you can do with your degree?
- ▶ link network. content pages. both in and out of universe.
 - ▶ cut out admin pages (help, users, etc.)
 - ▶ my cleaning might have confused the characters from TOS with those from the newest films. whatever. this is informal and i can always fix it later.

Questions

Captains? (and commander...)

Series?

etc.

summaries

nodes: 34366

links: 788851

diameter: 10

mean links: ~45; median: 19

closeness centrality: $8.72 \cdot 10^{-6}$

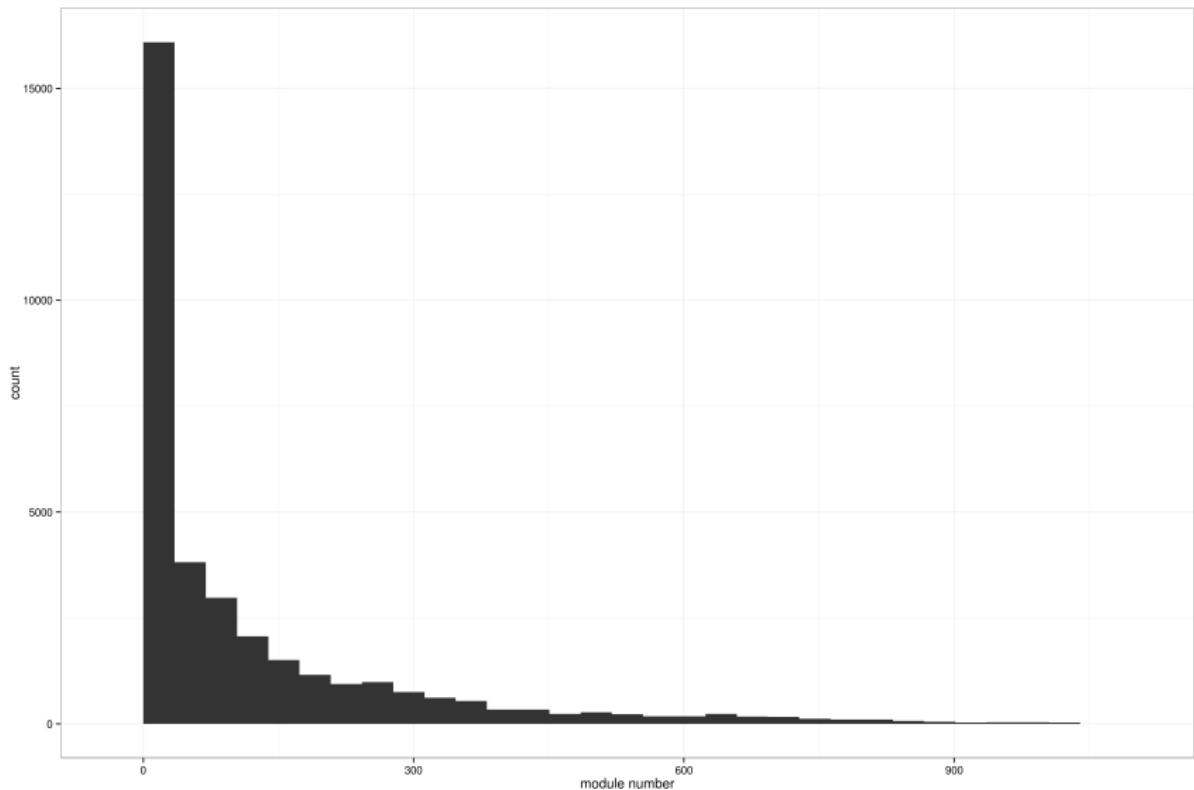
emergent structure?

yeah... but graph is still too huge to show. sorry.

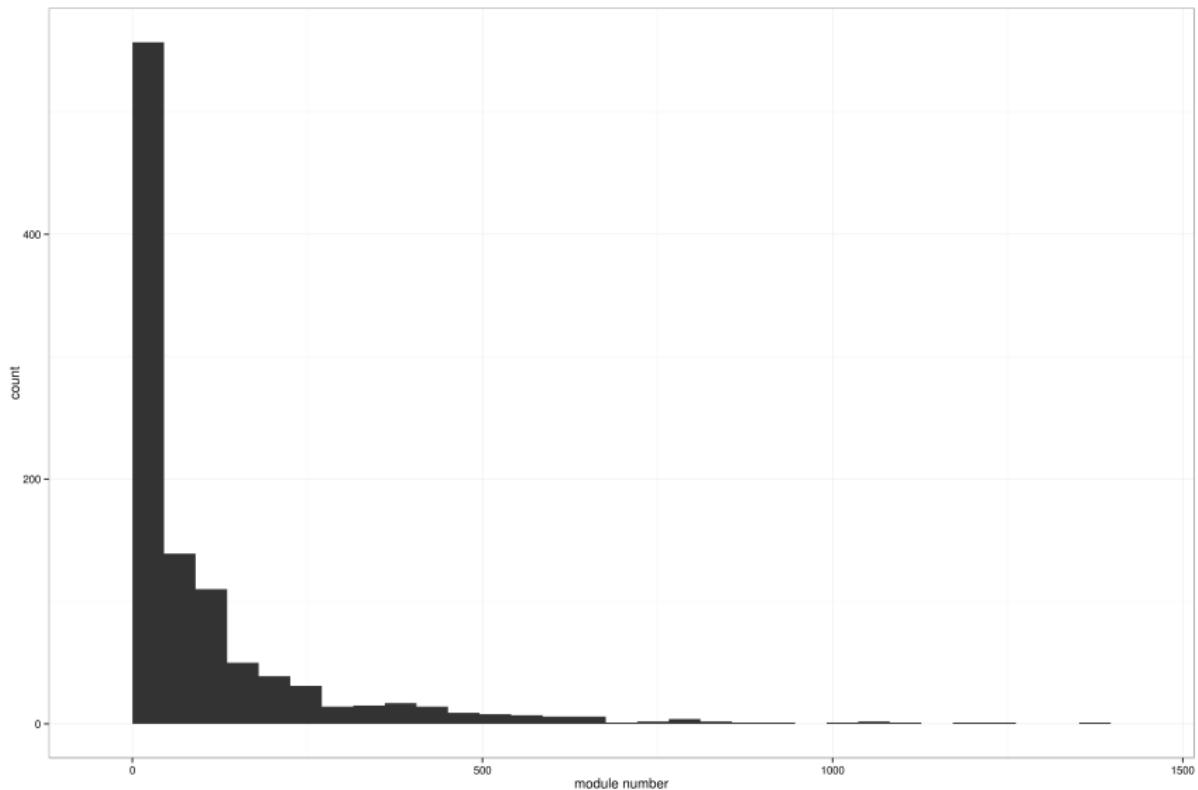
1040 modules (173 modules of size 1)

56358 links

module sizes



module degree



We have them just where they want us

What's in the first box? First few important things. . .

- ▶ william shatner
- ▶ j. j. abrams
- ▶ leonard nimoy
- ▶ jonathan frakes
- ▶ patrick stewart
- ▶ roberto orci (wrote recent star trek film)
- ▶ robert picardo (the doctor from voyager)

real people. boring.

Other boxes?

- ▶ 2: deep space 9 and related things
- ▶ 3: TOS and related things
- ▶ 4: enterprise (series) and related
- ▶ 5: planets and space
- ▶ 6: star fleet
- ▶ 7: voyager, borg, and related
- ▶ 8: klingons
- ▶ 9: production staff and related

smaller boxes overlap a little, but there is a “structure”

10 most important pages/ “things”

10. star trek: deep space nine

10 most important pages/ “things”

9. human
10. star trek: deep space nine

10 most important pages/ “things”

8. starship
9. human
10. star trek: deep space nine

10 most important pages/ “things”

7. star trek: the next generation
8. starship
9. human
10. star trek: deep space nine

10 most important pages/ “things”

6. star trek (out of universe)
7. star trek: the next generation
8. starship
9. human
10. star trek: deep space nine

10 most important pages/ “things”

5. planet
6. star trek (out of universe)
7. star trek: the next generation
8. starship
9. human
10. star trek: deep space nine

10 most important pages/ “things”

4. federation
5. planet
6. star trek (out of universe)
7. star trek: the next generation
8. starship
9. human
10. star trek: deep space nine

10 most important pages/ “things”

3. starfleet
4. federation
5. planet
6. star trek (out of universe)
7. star trek: the next generation
8. starship
9. human
10. star trek: deep space nine

10 most important pages/ “things”

2. earth
3. starfleet
4. federation
5. planet
6. star trek (out of universe)
7. star trek: the next generation
8. starship
9. human
10. star trek: deep space nine

10 most important pages/ “things”

1. captain
2. earth
3. starfleet
4. federation
5. planet
6. star trek (out of universe)
7. star trek: the next generation
8. starship
9. human
10. star trek: deep space nine

weirdest click series

tanagra family

gallos ii

inhabited planets

23rd century

san francisco

san francisco herald

jessica bradley

encounter at farpoint

henry vi, part iii

the naked now

decontamination

series



and the winner is. . .

	page.rank	degree	closeness
star trek: the next generation	0.0049	2658	0.00039
star trek: deep space nine	0.0041	4286	0.00039
star trek: voyager	0.0034	2261	0.00039
star trek: enterprise	0.0030	1929	0.00039
star trek: the original series	0.0028	3367	0.00039

captains



and the winner is. . .

	page.rank	degree	closeness
jean-luc picard	0.00334	3808	0.00039
james t. kirk	0.00329	2380	0.00039
benjamin sisko	0.00238	1925	0.00039
kathryn janeway	0.00207	3586	0.00039
jonathan archer	0.00167	2311	0.00039

and he's aged the best



so, obviously

