Wikipedia Article Curation
Understanding Quality
Recommending Tasks

Morten Warncke-Wang
en:User:Nettrom / nettrom@twitter
2014-08-20
Recommending Tasks
User:SuggestBot

From Wikipedia, the free encyclopedia

This user account is a bot operated by Nettrom (talk). It is a legitimate alternative account, used to make repetitive automated or semi-automated edits that would be extremely tedious to do manually. The bot is approved and currently active – the relevant request for approval can be seen here.

SuggestBot is a program that attempts to help Wikipedia users find pages to edit based on their past contributions. If you want to try it out and get some personalized recommendations we have documentation explaining how to do that. You can also sign up to receive suggestions regularly. If you're wondering what the bot's recommendations looks like, you can find examples here.

SuggestBot uses a variety of algorithms, including standard information retrieval and collaborative filtering techniques, to make suggestions. Most of it runs at the GroupLens Research Lab, some parts also run on Wikimedia Tool Labs and the Wikimedia Toolserver. The recommendations are usually done based on downloading a list of a user's contributions (to avoid recommending articles they've recently edited). The bot was originally developed by ForteTuba.

If SuggestBot made personalized recommendations for you, please leave feedback on whether they were useful and how to make them better. Comments on recommendations, as well as general comments, suggestions, or complaints, are best left on our talk page. Comments are welcome and valuable, as they will help us make SuggestBot do a better job of helping Wikipedia.
Step 1: Interest profiling
User contributions

For Nettrom (talk | block log | uploads | logs | filter log)

Search for contributions
- Show contributions of new accounts only
- User: Nettrom
- Namespace: (Article)
- Invert selection
- Associated namespace
- Tag filter:

- Only show edits that are latest revisions
- Only show edits that are page creations

From year (and earlier): 2014
From month (and earlier): all

(newest | oldest) View (newer 50 | older 50) (20 | 50 | 100 | 250 | 500)

- 03:22, August 18, 2014 (diff | hist) (.+1) . m James Tyler Guitars (fixed the last misspelling of Dann Huff?) (current)
- 18:45, May 20, 2014 (diff | hist) . (-2,376) . Bob Taylor (luthier) (article is unsourced, Bob is arguably not notable outside of Taylor Guitars, difficult to find good sources, redirecting to Taylor Guitars instead) (current)
- 17:03, May 20, 2014 (diff | hist) . (-542) . Taylor Guitars (removing "Custom guitars" section per discussion on the talk page -- possible copyright violation)
- 17:13, April 26, 2014 (diff | hist) . (+42) . m Collings Guitars (→ Notable Users: slight reorganisation, add refimprove, sources are needed for the newcomers)
- 14:03, April 24, 2014 (diff | hist) . (-77) . Collings Guitars (Dekel Bor is not mentioned in the PremierGuitar profile, can't be used as a reference. Also fixed URL to PG profile.)
- 20:29, April 22, 2014 (diff | hist) . (-3,882) . Eric Johnson Stratocaster (article contains only catalogue information and is stale, better serviced by the signature model section of Fender Stratocaster) (current)
- 20:27, April 22, 2014 (diff | hist) . (+17) . m Robert Cray Standard Stratocaster (redirect to signature models section,
Step 2: Find similar articles
The *Wikimedia Foundation (WMF)* is an American non-profit and charitable organization headquartered in San Francisco, California, that operates many wikis. The foundation is mostly known for hosting Wikipedia, an Internet encyclopedia which ranks in the top-ten most-visited websites worldwide,[6] as well as Wiktionary, Wikiquote, Wikibooks, Wikisource, Wikimedia Commons, Wikispecies, Wikinews, Wikiversity, Wikidata, Wikivoyage, Wikimedia Incubator, and Meta-Wiki. It also owned the now-defunct Nupedia.

The organization was founded in 2003 by Jimmy Wales, co-founder of Wikipedia, as a way to fund Wikipedia and its sister projects through non-profit means.[6][7]

As of 2013, the foundation employs more than 208 people, with revenues of US$48.6 million and cash equivalents of $22.2 million.[3] Lila Tretikov leads the foundation as its executive director, while Jan-Bart de Vreede serves as chairman of the board.
Step 3: Filtering
Want: Tasks
This article **may rely excessively on sources too closely associated with the subject**, preventing the article from being **verifiable** and **neutral**. Please help improve it by replacing them with more appropriate citations to **reliable, independent, third-party sources**. *(February 2013)*

**Wikidata** is a **collaboratively edited knowledge base** operated by the **Wikimedia Foundation**. It is intended to provide a common source of certain data types (for example, birth dates) which can be used by Wikimedia projects such as **Wikipedia**.[2][3] This is similar to the way **Wikimedia Commons** provides storage for media files and access to those files for all Wikimedia projects. **Wikidata** is powered by the software **mw:Wikibase**.

**Contents** [*hide*]
1 Concepts
2 Development history
  2.1 Phase 1
  2.2 Phase 2
  2.3 Phase 3
3 See also
4 Further reading
5 References
6 External links
Step 4: Presentation
SuggestBot predicts that you will enjoy editing some of these articles. Have fun!

**Add sources**
- Fender Jazzmaster
- Fender Showmaster
- Fender Lead Series
- Richmond Outreach Center
- Megan Prescott
- Fender Performer

**Wikiify**
- John Birch (luthier)
- Burn Halo (album)
- American Forces Information Service

**Orphan**
- SX SJM
- Fender '51 Precision Bass
- Vox Bass Guitar

**Cleanup**
- Bob Marley
- Simon Eastwood
- Dorian mode

**Merge**
- Subharmonic
- Viatical settlement
- Adelaide Street bus mall

**Expand**
- Emily Bear
- Teisco
- Tech 21

**Unencyclopaedic**
- Richie Sambora
- Fender Duo-Sonic
- Kramer Guitars

**Stub**
- Jim Root Telecaster
- Fender Wildwood
- Forbon
- Grinning Streak
- Fender King
- Fender White Steel

SuggestBot picks articles in a number of ways based on other articles you've edited, including straight text similarity, following wikilinks, and matching your editing patterns against those of other Wikipedians. It tries to recommend only articles that other Wikipedians have marked as needing work. Your contributions make Wikipedia better — thanks for helping.

If you have *feedback* on how to make SuggestBot better, please tell me on SuggestBot's talk page. Thanks from Nettrom (talk), SuggestBot's caretaker. -- SuggestBot (talk) 20:05, 20 May 2013 (UTC)
MOAR INFORMATION!
What information?

» Idea: Show viewership & quality
» Contributors work on popular low quality things first?
» Needs data:
  • Article viewership
  • Article quality
Viewership

» Not readily available to contributors
» Readily available to us:
  • Wikimedia Foundation data dumps
  • stats.grok.se
Quality

» More easily available
  • Wikipedia assessments on the talk page
  • Contributor’s own judgement

» Assessment might be lagging or absent

» Contributor judgement requires experience
Problem: Up-to-date Quality
Twist: Actionable Quality
Solution: Machine Learning?
ML Issue: Feature Selection

» Typical features are difficult to change:
  • Editor diversity
  • Meta-features

Photo: Oregon Dept of Transportation – CC BY
Our Approach: Actionable features

» Use features editors can act upon
» Originally 5 main features:
  • Amount of article content
  • Number of citations
  • Number of images
  • Number of wikilinks
  • Number of article sections
Does It Work?
Seven classes, difficult problem

» Seven assessment classes on en-WP:
  • Featured Article
  • A-class
  • Good Article
  • B-class
  • C-class
  • Start-class
  • Stub-class

» Unclear boundaries between them
Yardstick: Random guesses

Random 14.3%
Actionable Model Performance

Random 14.3%

Actionable model 42.5%
Often off by one class

Random 14.3%

Actionable model 42.5%

Actionable model off-by-one 76.9%
Situation Report

» Few features, many information
» Missing information quality aspects
Show Me The Information!
Articles you might like to edit, from SuggestBot

We are currently running a study on the effects of adding additional information to SuggestBot's recommendations. Participation in the study is voluntary. Should you wish to not participate in the study, or have questions or concerns, you can find contact information in the consent information sheet.

We have added information about the readership and quality of the suggested articles using a Low/Medium/High scale. For readership the scale goes from Low 🐣 to High 🦃, while for quality the scale goes from Low 🌟🌟🌟 to High 🌟🌟🌟🌟

SuggestBot predicts that you will enjoy editing some of these articles. Have fun!

<table>
<thead>
<tr>
<th>Stubs</th>
<th>Cleanup</th>
</tr>
</thead>
<tbody>
<tr>
<td>🐣🌟🌟🌟 Gilles Berheim</td>
<td>🦃🌟🌟🌟 Lucy, Lady Houston</td>
</tr>
<tr>
<td>🦃🌟🌟🌟 The Climb (book)</td>
<td>🦃🌟🌟🌟 Hilary Alexander</td>
</tr>
<tr>
<td>🦃🌟🌟🌟 Red Wings Airlines</td>
<td>🦃🌟🌟🌟 Stipe Božič</td>
</tr>
<tr>
<td>🦃🌟🌟🌟 Nawang Gombu</td>
<td></td>
</tr>
<tr>
<td>🦃🌟🌟🌟 Leszek Cichy</td>
<td>🦃🌟🌟🌟 Rob Portman</td>
</tr>
<tr>
<td>🦃🌟🌟🌟 George Finch (chemist)</td>
<td>🦃🌟🌟🌟 Long campaign</td>
</tr>
<tr>
<td>🦃🌟🌟🌟 Andrzej Zawada</td>
<td>🦃🌟🌟🌟 House of Plantagenet</td>
</tr>
<tr>
<td>🦃🌟🌟🌟 Jim Perrin</td>
<td></td>
</tr>
<tr>
<td>🦃🌟🌟🌟 Phurba Tashi</td>
<td>🦃🌟🌟🌟 Sutton Green, Surrey</td>
</tr>
<tr>
<td>🦃🌟🌟🌟 Davo Karničar</td>
<td>🦃🌟🌟🌟 List of climbers and mountaineers</td>
</tr>
<tr>
<td>🦃🌟🌟🌟 Pemba Doma Sherpa</td>
<td>🦃🌟🌟🌟 Foster's Home for Imaginary Friends</td>
</tr>
<tr>
<td>🦃🌟🌟🌟 Earl Denman</td>
<td></td>
</tr>
<tr>
<td>🦃🌟🌟🌟 Tanggula Mountains</td>
<td>🦃🌟🌟🌟 Kazi Sherpa</td>
</tr>
</tbody>
</table>
MOAR BETTER!
Improved Information

» Viewership: numbers?
» Quality: assessments and predictions?
» Turn quality features into tasks?
» Make it sort?
<table>
<thead>
<tr>
<th>Views/Day</th>
<th>Quality</th>
<th>Title</th>
<th>Content</th>
<th>Headings</th>
<th>Images</th>
<th>Links</th>
<th>Sources</th>
<th>Tagged with...</th>
</tr>
</thead>
<tbody>
<tr>
<td>37</td>
<td>★★★★</td>
<td>Ethnom College (ask)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Add sources</td>
</tr>
<tr>
<td>19</td>
<td>★★★★</td>
<td>Good to See You Again, Alice Cooper (ask)</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>Add sources</td>
</tr>
<tr>
<td>363</td>
<td>★★★★</td>
<td>Fender Jazzmaster (ask)</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td>Add sources</td>
</tr>
<tr>
<td>18</td>
<td>★★★★</td>
<td>Self-titled (ask)</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td>Add sources</td>
</tr>
<tr>
<td>37</td>
<td>★★★★</td>
<td>The Sirs (ask)</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td>Add sources</td>
</tr>
<tr>
<td>47</td>
<td>★★★★</td>
<td>Slaughter &amp; the Dogs (ask)</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td>Add sources</td>
</tr>
<tr>
<td>12</td>
<td>★★★★</td>
<td>Walk (ask)</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td>Add sources</td>
</tr>
<tr>
<td>16,245</td>
<td>★★★★</td>
<td>B-Movie (ask)</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td>Add sources</td>
</tr>
<tr>
<td>1,967</td>
<td>★★★★</td>
<td>At the Movies (ask)</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td>Add sources</td>
</tr>
<tr>
<td>463</td>
<td>★★★★</td>
<td>Britpop mode (ask)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cleanup</td>
</tr>
<tr>
<td>165</td>
<td>★★★★</td>
<td>Telisco (ask)</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td>Expand</td>
</tr>
<tr>
<td>2,171</td>
<td>★★★★</td>
<td>Kon-Tiki (ask)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Expand</td>
</tr>
<tr>
<td>15</td>
<td>★★★★</td>
<td>Il kanz PGM (ask)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>Expand</td>
</tr>
<tr>
<td>1,525</td>
<td>★★★★</td>
<td>Richie Sambora (ask)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Unencyclopaedic</td>
</tr>
<tr>
<td>85</td>
<td>★★★★</td>
<td>Fender Duo-Sonic (ask)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Unencyclopaedic</td>
</tr>
<tr>
<td>175</td>
<td>★★★★</td>
<td>Kramer Guitars (ask)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Unencyclopaedic</td>
</tr>
<tr>
<td>69</td>
<td>★★★★</td>
<td>Fanned Fret Guitars (ask)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Unencyclopaedic</td>
</tr>
<tr>
<td>23</td>
<td>★★★★</td>
<td>Music of Vanuatu (ask)</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>Merge</td>
</tr>
<tr>
<td>32</td>
<td>★★★★</td>
<td>Multi-scale fingerboard</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>Merge</td>
</tr>
</tbody>
</table>

Quality: Low,
Assessed class: Unassessed,
Predicted class: Stub
In summary…

» SuggestBot: recommending tasks to contributors

» Quality:
  • Actionable features
  • Predicting article quality
  • Suggesting improvement tasks
Acknowledgements

» NSF grants IIS 08-08692, 09-68483, 08-45351
» WMF and Wikimedia Deutschland
» GroupLens Research
Questions?

Wikipedia: User:Nettrom
Email: morten@cs.umn.edu
Twitter: @nettrom
Web: http://www-users.cs.umn.edu/~morten/
GroupLens: http://www.grouplens.org/