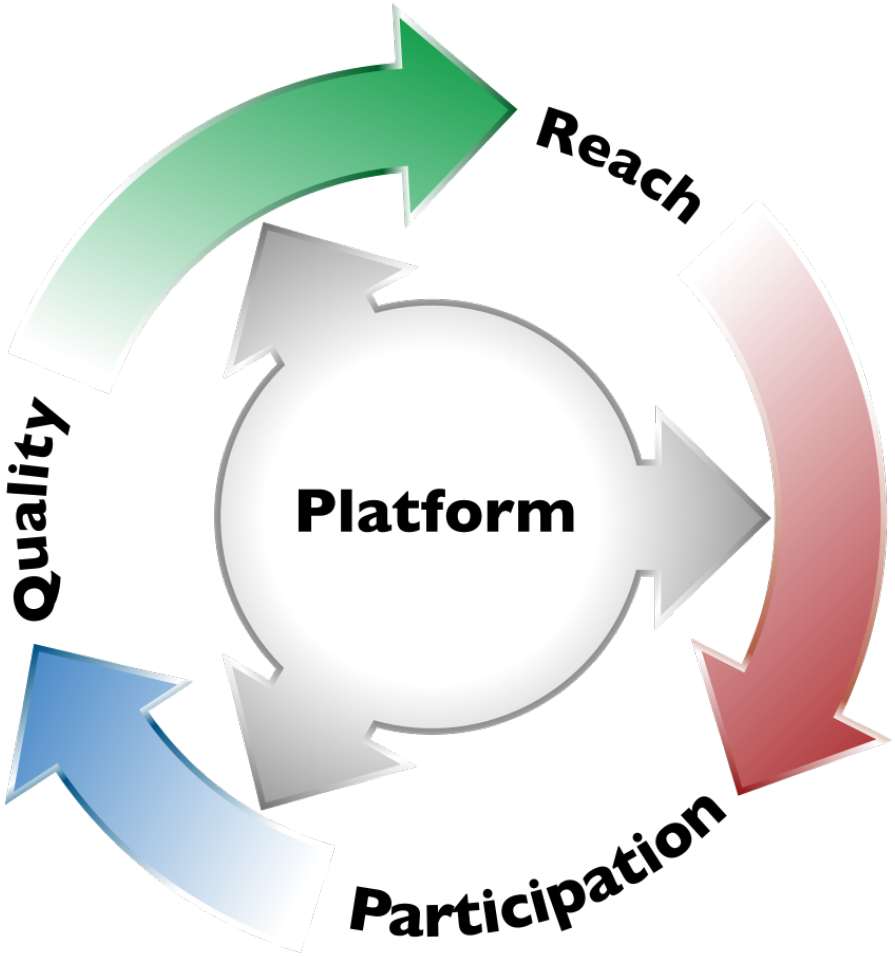
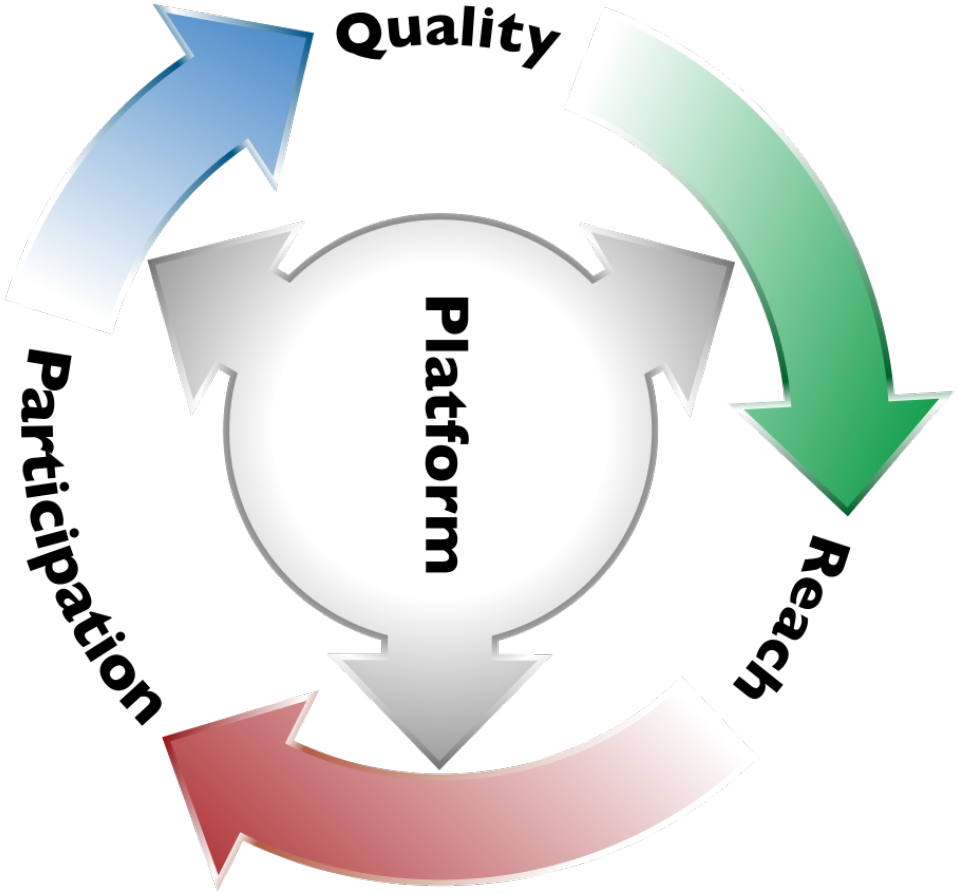
Content Quality Update

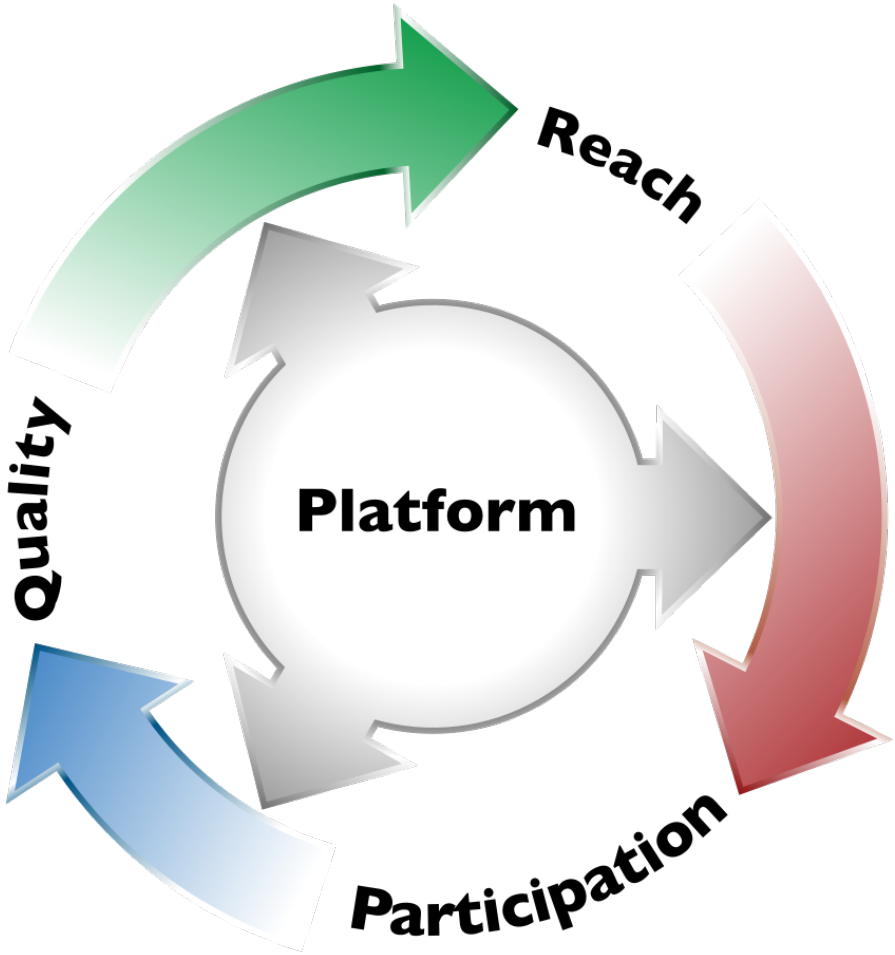
Wikimedia Foundation Metrics Meeting

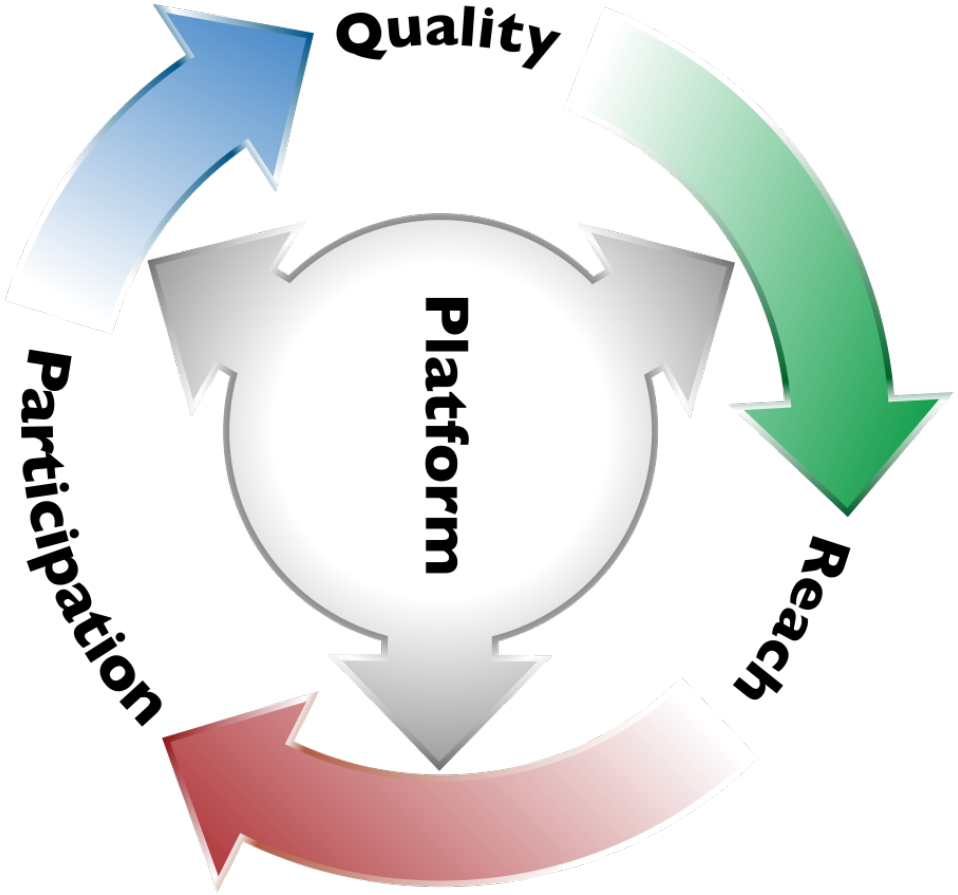
15 January 2015

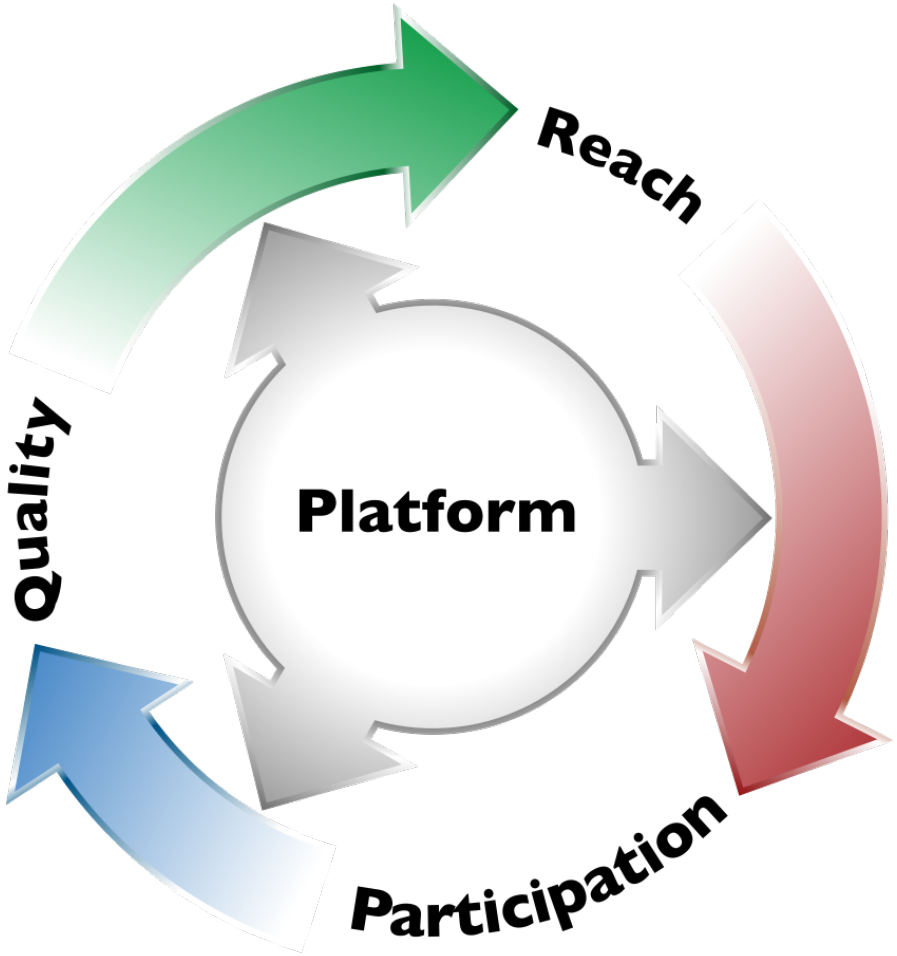
Theory


Reach
Participation
Quality
Platform


Quality
Reach
Participation
Platform


Reach
Participation
Quality
Platform


Quality
Reach
Participation
Platform


Reach
Participation
Quality
Platform

But .. seriously ..

Does readership lead to participation?

Does participation lead to quality?

Does quality lead to readership?

What if we focus on quality?

What if we focus on knowledge?

1. **Measure quality.**
2. **Support existing communities.**
3. **Grow and support new communities.**

Agenda

Revision Scoring as a Service

Aaron Halfaker — 10 minutes

The Wikipedia Library

Jake Orlowitz — 10 minutes

Wikipedia in Education

Frank Schulenburg — 15 minutes

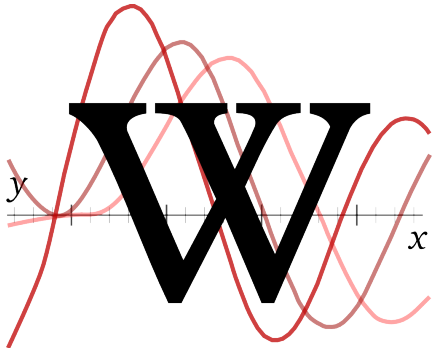
Takeaways and update

Discussion / Questions

10 minutes

Revision Scoring as a Service

with Aaron Halfaker
[[User:EpochFail]]
[[User:Halfak (WMF)]]



IEG Funded

https://commons.wikimedia.org/wiki/File:Artificial.intelligence.jpg
Public Domain

1. **AI-based tools are critical for quality control in Wikipedia**

https://commons.wikimedia.org/wiki/File:Artificial.intelligence.jpg
Public Domain

1. AI-based tools are critical for quality control in Wikipedia

2. AI is hard -- requires expertise

https://commons.wikimedia.org/wiki/File:Artificial_intelligence.jpg
Public Domain

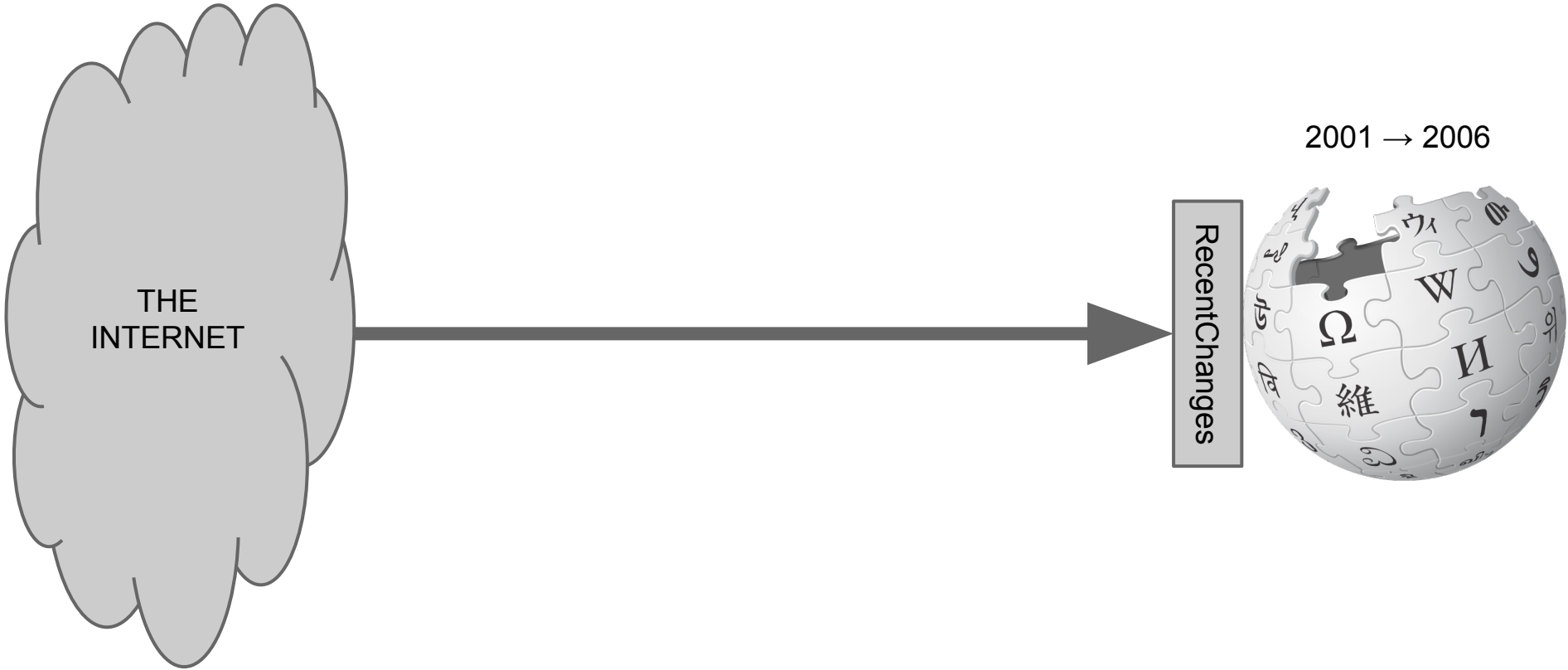
1. AI-based tools are critical for quality control in Wikipedia

2. AI is hard -- requires expertise

3. Wrapping AI in libraries and services enables new tools

https://commons.wikimedia.org/wiki/File:Artificial_intelligence.jpg
Public Domain

AI-based quality control



AI-based quality control

THE INTERNET

User:ClueBot NG

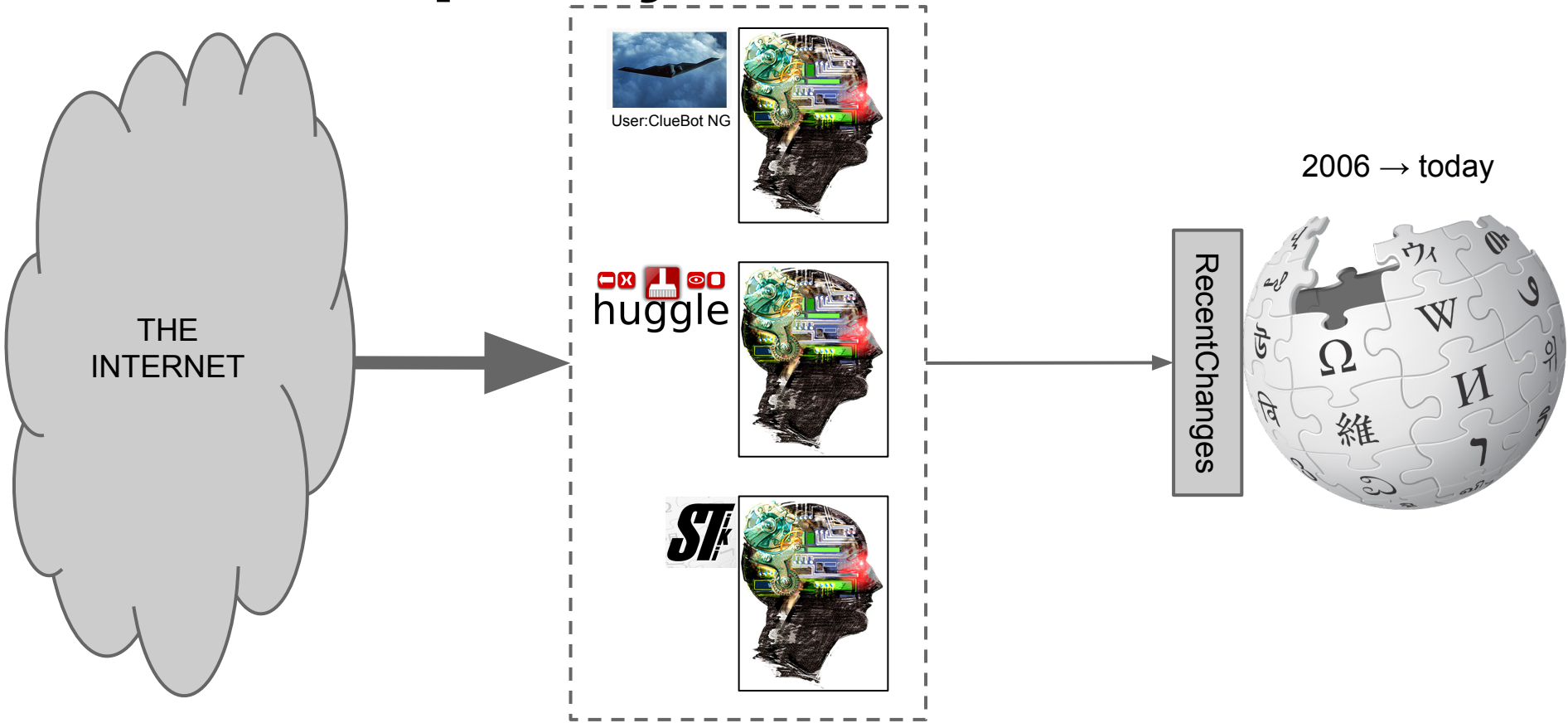
huggle

STiki

RecentChanges

2006 → today

AI-based quality control



REVISION #638307884

SCORE

Revision as of 23:14, 10 December 2014 (edit)
Blaise Mann (talk | contribs)
m *(Adding 1 category)*
← Previous edit

Revision as of 03:47, 16 December 2014 (edit) (undo) (thank)
Eloquence (talk | contribs)
m *(→Theater: link to disambiguated title (currently a redlink) instead of disambig page)*
Next edit →

Line 35:

In 2013 Wilson, in collaboration with [[Mikhail Baryshnikov]] and co-starring [[Willem Dafoe]], developed "[[The Old Woman]]" an adaptation of the work by the Russian author [[Daniil Kharms]]. The play premiered at MIF13, Manchester International Festival.<ref>{{Cite web|url = http://www.mif.co.uk/event/the-old-woman|title = The Old Woman Robert Wilson, Mikhail Baryshnikov, Willem Dafoe|date = 2013|accessdate = 2014-11-23|website = Manchester International Festival|publisher = Manchester International Festival|last = Jansch|first = Lucie}}</ref> Wilson wrote that he and [[Mikhail Baryshnikov|Baryshnikov]] have discussed creating a play together for years, perhaps based on a Russian text.<ref>{{Cite web|url = http://calperformances.org/learn/program_notes/2014/pn_oldwoman.pdf|title = Cal Performances University of California, Berkeley Playbill|date = 2014|accessdate = 2014-11-23|website = Calperformances|publisher = UC Berkeley|last = Wilson|first = Robert}}</ref> The final production included dance, light, singing and bilingual monologue.

Line 35:

In 2013 Wilson, in collaboration with [[Mikhail Baryshnikov]] and co-starring [[Willem Dafoe]], developed "[[**The Old Woman (play)|**The Old Woman]]", an adaptation of the work by the Russian author [[Daniil Kharms]]. The play premiered at MIF13, Manchester International Festival.<ref>{{Cite web|url = http://www.mif.co.uk/event/the-old-woman|title = The Old Woman Robert Wilson, Mikhail Baryshnikov, Willem Dafoe|date = 2013|accessdate = 2014-11-23|website = Manchester International Festival|publisher = Manchester International Festival|last = Jansch|first = Lucie}}</ref> Wilson wrote that he and [[Mikhail Baryshnikov|Baryshnikov]] have discussed creating a play together for years, perhaps based on a Russian text.<ref>{{Cite web|url = http://calperformances.org/learn/program_notes/2014/pn_oldwoman.pdf|title = Cal Performances University of California, Berkeley Playbill|date = 2014|accessdate = 2014-11-23|website = Calperformances|publisher = UC Berkeley|last = Wilson|first = Robert}}</ref> The final production included dance, light, singing and bilingual monologue.

REVISION → → SCORE

User:Eloquence fixes a link

armstrong.cis.upenn.edu/stiki_api.php?style=score&rid=638307884

0.0012884835247 4975 ← Low score -- Probably good

User:Blank123456789 replaces a paragraph with “LLAMAS GROW ON TREES”

armstrong.cis.upenn.edu/stiki_api.php?style=score&rid=642215410

0.703287059009732 ← High score -- Probably damaging

AI is important for quality control

Powerful

Fast

Minimizes human effort

When the Levee Breaks: Without Bots, What Happens to Wikipedia's Quality Control Processes? R. Stuart Geiger & Aaron Halfaker. (2013). *WikiSym*.

But AI is hard

But AI is hard

- 20+ research papers on Wikipedia damage detection

But AI is hard

- 20+ research papers on Wikipedia damage detection
- “Machine classification” not part of standard CS degree

But AI is hard

- 20+ research papers on Wikipedia damage detection
- “Machine classification” not part of standard CS degree
  - Many tool devs don’t have a CS degree

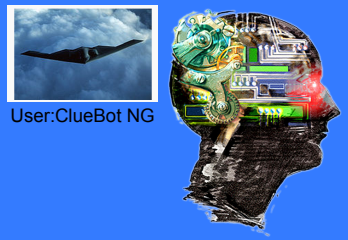
But AI is hard

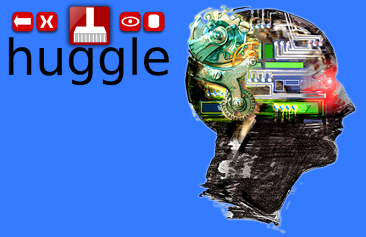
- 20+ research papers on Wikipedia damage detection
- “Machine classification” not part of standard CS degree
  - Many tool devs don’t have a CS degree

- Labor intensive, performance considerations, etc.

AI is hard

Each tool builds uses its own AI





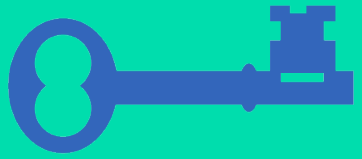
AI is hard

Each tool builds uses its own AI

Mostly just English Wikipedia*

Revision Scoring Service

User:EpochFail



IEG Funded

User:He7d3r

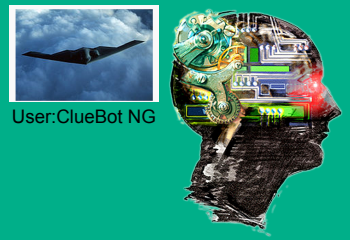
User:とある白い猫

Volunteers

- User:Danilo.mac
- User:Jonas.agx
- User:West.g.andrew
- User:Yuvipanda

One AI to support them all

One AI to support them all

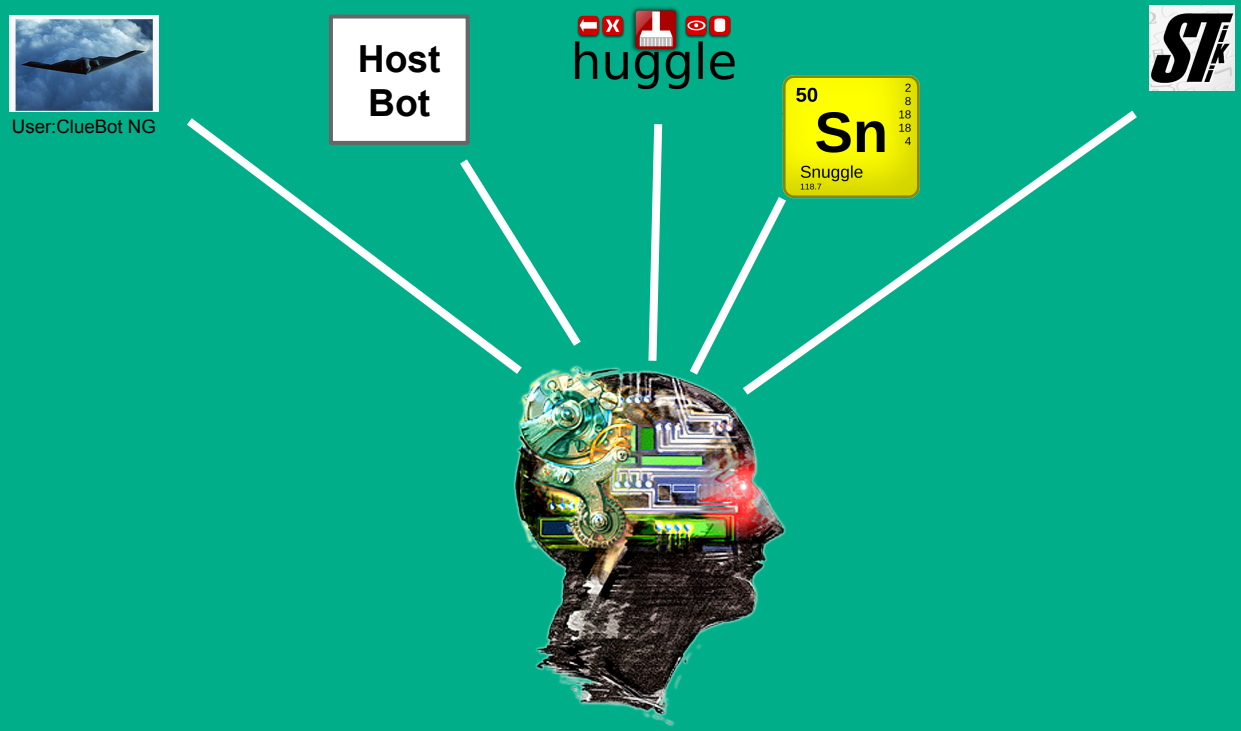


One AI to support them all

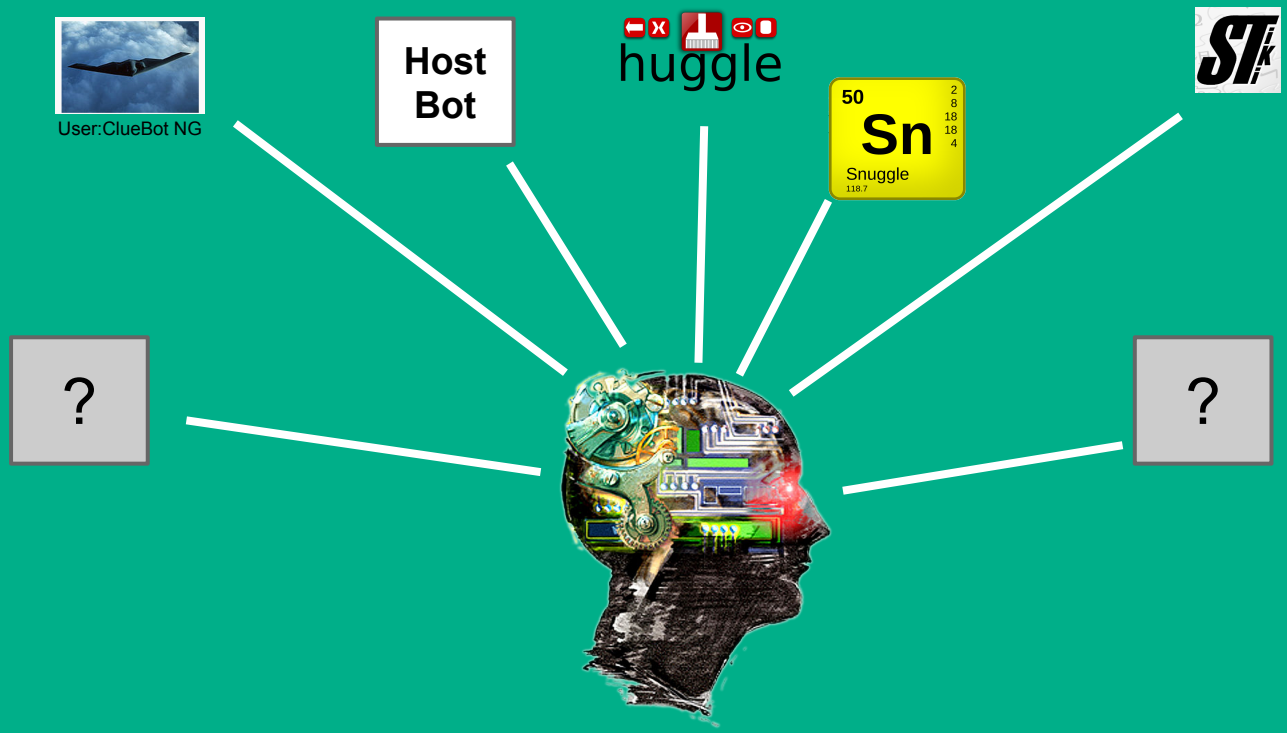
User:ClueBot NG

huggle

One AI to support them all



One AI to support them all



One AI to rule them all

First release

- English
- Portuguese
- Turkish
- Azerbaijani*

* still working out some language feature issues

Web service

- ores.wmflabs.org/enwiki/?
  - **scores**=damaging&
  - **revids**=4567890|3456789

```
{
  "4567890": {
    "damaging": {
      "probability": 0.234
    }
  },
  "3456789": {
    "damaging": {
      "probability": 0.952
    }
  }
}
```

Model files

- 100KB - 1MB
- Download, load and use

```
> from revscores.scorers import LinearSVCModel
> model_file = open("enwiki.reverts.model", 'rb')
> model = LinearSVCModel.load(model_file)
> model.score(...)
{"probability": 0.723}
```

Growing the wiki technology ecosystem

Growing the wiki technology ecosystem

- **IEG** -- Funding for distributed innovation

Growing the wiki technology ecosystem

- **IEG** -- Funding for distributed innovation
- **Wikimedia Labs** -- Raw materials and shared space

Growing the wiki technology ecosystem

- **IEG** -- Funding for distributed innovation
- **Wikimedia Labs** -- Raw materials and shared space
- **Services Team** -- Professional ecosystem builders

1. Wikipedia needs AI

https://commons.wikimedia.org/wiki/File:Artificial.intelligence.jpg
Public Domain

1. **Wikipedia needs AI**

2. **We’re making one important AI easy**

https://commons.wikimedia.org/wiki/File:Artificial.intelligence.jpg
Public Domain

1. Wikipedia needs AI

2. We’re making one important AI easy

3. Growing the ecosystem of high tech wiki tools

https://commons.wikimedia.org/wiki/File:Artificial.intelligence.jpg
Public Domain

Thanks!

from the Revision Scoring team

User:EpochFail

User:He7d3r

User:とある白い猫

For more info:

- https://meta.wikimedia.org/wiki/Research:Revision_scoring_as_a_service
- https://meta.wikimedia.org/wiki/Grants:IEG/Revision_scoring_as_a_service

The Wikipedia Library

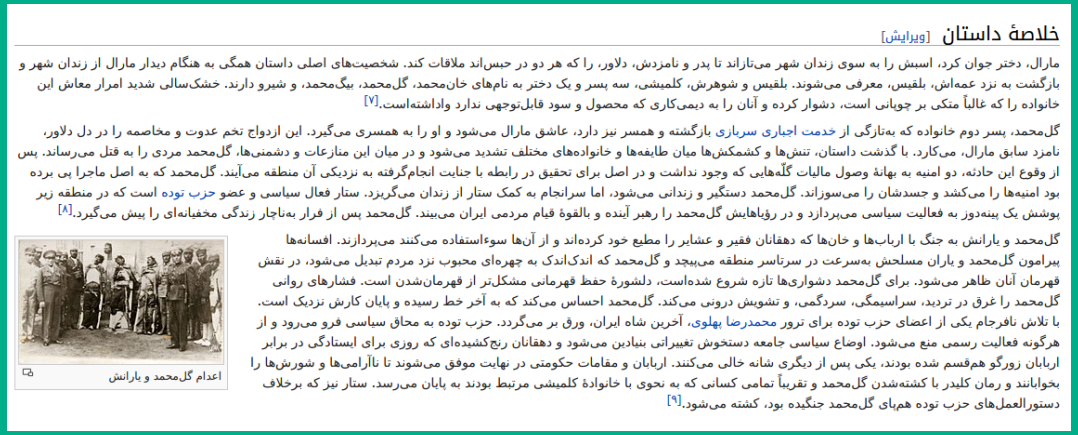
If you want to improve article quality...

Arm super-users to the teeth with free, full-text access to reliable sources

Through an open research hub that builds partnerships to give top editors keys to paywalled content

@WikiLibrary

“Things changed when the international sanctions against Iran expanded...



خلاصهٔ داستان [ویرایش]

مارال، دختر جوان کرد، اسیش را به سوی زندان شهر می‌تاراند تا پدر و نامزدش، دلاور، را که هر دو در حبس‌اند ملاقات کند. شخصیت‌های اصلی داستان همگی به هنگام دیدار مارال از زندان شهر و بازگشت به نزد عمه‌اش، بلقیس، معرفی می‌شوند. بلقیس و شوهرش، کلمیشی، سه پسر و یک دختر به نام‌های خان‌محمد، گل‌محمد، بیگ‌محمد، و شیرو دارند. خشکسالی شدید امرار معاش این خانواده را که عالباً متکی بر چوپانی است، دشوار کرده و آنان را به دیمی‌کاری که محصول و سود قابل‌توجهی ندارد واداشته‌است.[۷]

گل‌محمد، پسر دوم خانواده که به‌تازگی از خدمت اجباری سرباری بازگشته و همسر نیز دارد، عاشق مارال می‌شود و او را به همسری می‌گیرد. این ازدواج تخم عدوت و مخاصمه را در دل دلاور، نامزد سابق مارال، می‌کارد. با گذشت داستان، تنش‌ها و کشمکش‌ها میان طایفه‌ها و خانواده‌های مختلف تشدید می‌شود و در میان این منازعات و دشمنی‌ها، گل‌محمد مردی را به قتل می‌رساند. پس از وقوع این حادثه، دو امنیه به بهانهٔ وصول مالیات گلّه‌هایی که وجود نداشت و در اصل برای تحقیق در رابطه با جنایت انجام‌گرفته به نزدیکی آن منطقه می‌آیند. گل‌محمد که به اصل ماجرا پی برده بود امنیه‌ها را می‌کشد و جسدشان را می‌سوزاند. گل‌محمد دستگیر و زندانی می‌شود، اما سرانجام به کمک ستار از زندان می‌گریزد. ستار فعال سیاسی و عضو حزب توده است که در منطقه زیر پوشش یک پیله‌دور به فعالیت سیاسی می‌پردازد و در رؤیاهایش گل‌محمد را رهبر آینده و بالقوهٔ قیام مردمی ایران می‌بیند. گل‌محمد پس از فرار به‌ناچار زندگی مخفیانه‌ای را پیش می‌گیرد.[۸]

اعدام گل‌محمد و یارانش

گل‌محمد و یارانش به جنگ با ارباب‌ها و خان‌ها که دهقانان فقیر و عشایر را مطیع خود کرده‌اند و از آن‌ها سوءاستفاده می‌کنند می‌پردازند. افسانه‌ها پیرامون گل‌محمد و یاران مسلحش به‌سرعت در سرتاسر منطقه می‌پیچد و گل‌محمد که اندک‌اندک به چهره‌ای محبوب نزد مردم تبدیل می‌شود، در نقش قهرمان آنان ظاهر می‌شود. برای گل‌محمد دشواری‌ها تازه شروع شده‌است، دلشورهٔ حفظ قهرمانی مشکل‌تر از قهرمان‌شدن است. فشارهای روانی گل‌محمد را غرق در تردید، سراسیمگی، سردگمی، و تشویش درونی می‌کند. گل‌محمد احساس می‌کند که به آخر خط رسیده و پایان کارش نزدیک است. با تلاش نافرجام یکی از اعضای حزب توده برای ترور محمدرضا پهلوی، آخرین شاه ایران، ورق بر می‌گردد. حزب توده به محاق سیاسی فرو می‌رود و از هرگونه فعالیت رسمی منع می‌شود. اوضاع سیاسی جامعه دستخوش تغییراتی بنیادین می‌شود و دهقانان رنج‌کشیده‌ای که روزی برای ایستادگی در برابر اربابان زورگو هم‌قسم شده بودند، یکی پس از دیگری شانه خالی می‌کنند. اربابان و مقامات حکومتی در نهایت موفق می‌شوند تا ناآرامی‌ها و شورش‌ها را بخوابانند و رمان کلیدر با کشته‌شدن گل‌محمد و تقریباً تمامی کسانی که به نحوی با خانوادهٔ کلمیشی مرتبط بودند به پایان می‌رسد. ستار نیز که برخلاف دستورالعمل‌های حزب توده همپای گل‌محمد جنگیده بود، کشته می‌شود.[۹]

However, **thanks to the Wikipedia Library**, I received JSTOR access

By getting global editors like me free access to rich digital libraries, we will be **even more encouraged to write decent articles about our culture.”**

Wikimedia Blog by 4nn1l2

“Global Impact: The Wikipedia Library and Persian Wikipedia”

Birth of a library

Wikipedia Hobby -> WMF Individual Engagement Grant -> WMF Program

The numbers

3,654 accounts to 1,997 editors [1]
22% of TWL editors active outside ENWP
14,480 links added to partner resources* [2]
37% increase in links per partner [3]

1. Historical totals since beginning as an IEG Grant
2. Total ENWP links since 18 July 2014. *But which ones were from TWL editors? Are we missing any?*
3. Avg ENWP increase since 18 July 2014. *How does this compare to baseline reference growth?*

“We need access to superior online databases to be able to write high quality articles. Giving existing Wikipedians access to high-quality reliable sources is a very easy way to improve Wikipedia. Doing more of that is awesome.”

--Wikipedia Library User Survey

The numbers

22 partners: JSTOR, Elsevier, De Gruyter, BMJ… [1]
4 Universities [2]
$2 Million Individual Replacement Value

1. Historical totals since beginning as an IEG Grant
2. George Mason, University of California Riverside, Rutgers, Montana State

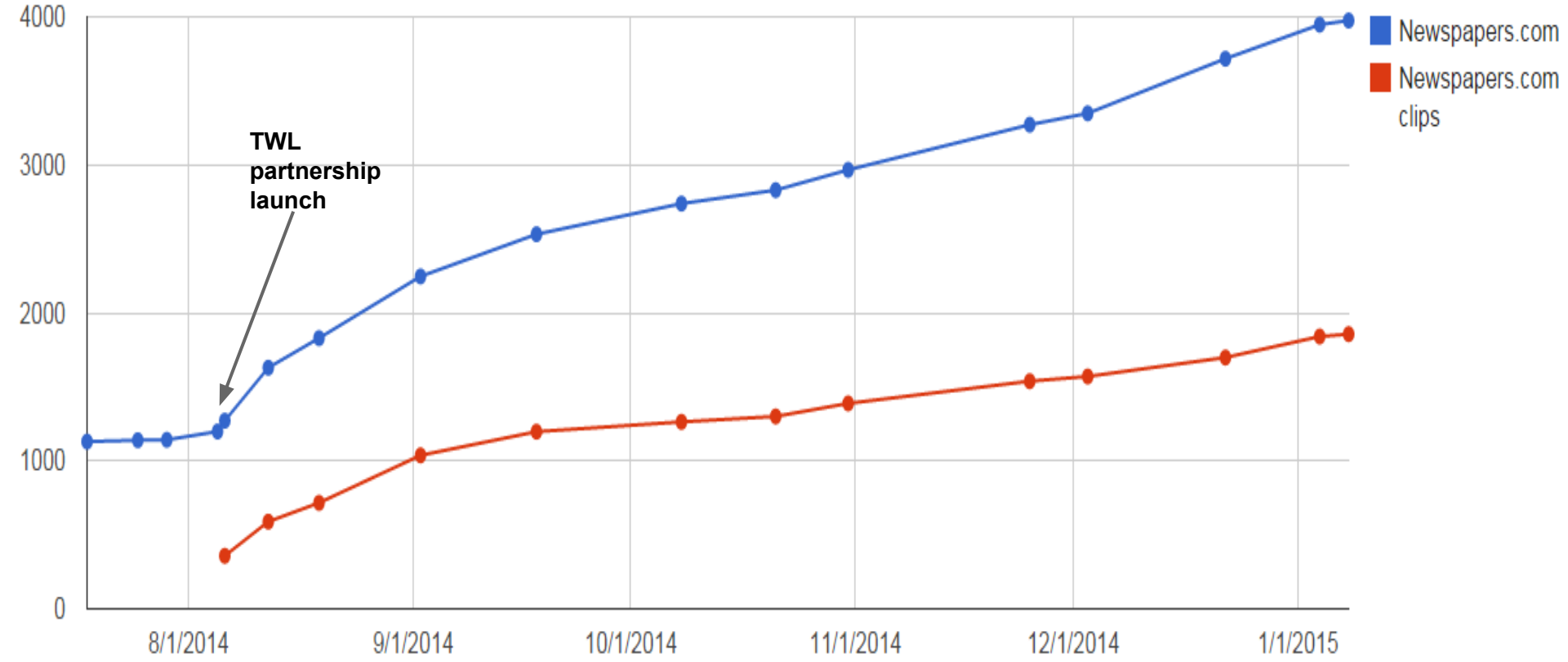
The Wikipedia Library's Research Partners

Unfree logos removed

“The goal of increasing access to sources that I cannot get through my local library has indeed helped me write better articles and I've started producing Good Article work in part because of TWL's efforts.”

--Wikipedia Library User Survey

External Links to Newspapers.com



Since August 2014 when Wikipedia Library launched a partnership with newspapers.com, external links jumped from 1,198 to 3,975 (+ 330%, in blue) . Linked open access 'clips' jumped from 358 to 1,856 (+520%, in red)

Featured Articles Created

by TWL Visiting Scholar **Wehwalt** at George Mason University

1. **John Tyler**
2. **Franklin Pierce**
3. **John Hay**
4. **Judah P. Benjamin**
5. **The Bread Winners**
6. **Babe Ruth**
7. **William H. Seward**
8. **Gilbert du Motier, Marquis deLafayette**
9. **Horace Greeley**
10. **Benjamin R. Tillman**
11. **Wreck of the S.S. Arctic**
12. **Jules Massanet**
13. **WINC (AM)**
14. **Thomas Bayard**

* 1-8 reached FA, 9 is in peer review, 10 is underway, 11/ 12/13 reached FA by other editors using Wehwalt's GMU sources

Reliable sources are a *foundation of quality*. If you want more quality, give the best editors unfettered access to the most reliable sources

Questions?

@JakeOrlowitz

jorlowitz@wikimedia.org

(presentation switch, whee!)

Takeaways

1. **Measure quality.**
   Without measurement - we are flying blind.
2. **Partner with contributors.**
   They know more than we do, because this is what they do every day.

2015 Strategy Initiative: Content contributors track

Phase 1: Jan/Feb

Siko Bouterse
Stephen LaPorte
Guillaume Paumier
Aaron Halfaker
Haitham Shammaa

Gathering intelligence...

- Who are Wikimedia contributors (today & in future?)
- What do they need to be successful?

Why it matters

What do vandal fighters and tool-builders need to improve content quality?

What do the most active article writers need to create quality content?

What do students new to Wikipedia need to create quality content?

Image credits: Myleen Hollero, Heather Walls & Rae Boe CC BY-SA 3.0

Got data/info to help answer these questions?

Please share!

sbouterse@wikimedia.org, slaporte@wikimedia.org

Discussion