

A Test Of Cross-wiki Search: Helping Users Discover Content On Wikipedia's Sister Projects

Erik Bernhardson *Senior Software Engineer, Wikimedia Foundation*

Jan Drewniak *User Experience Engineer, Wikimedia Foundation*

Dan Garry *Product Manager (Search Backend), Wikimedia Foundation*

Mikhail Popov *Data Analyst, Wikimedia Foundation*

Deb Tankersley *Product Manager (Analysis, Search Frontend), Wikimedia Foundation*

Wikimedia Engineering’s Discovery’s Search team ran an A/B test from 9 Feb 2017 to 22 Feb 2017 to assess the effectiveness of performing cross-wiki searches and showing Catalan, Italian, Persian, and Polish Wikipedias’ users results from sister projects such as Wikisource and Wikiquote. We found that while the overall engagement with the search results was higher for the two test groups compared to the control group, there was not sufficient evidence to definitively say that the additional search results increased user engagement. We suspect that a critical UX design issue – links shown in black, rather than standard blue – and the particular languages this test was deployed on (resulting in a higher zero results rate than seen across other languages) had a negative effect on the results. Our path forward is to fix the existing issues and perform a follow-up test that includes additional languages.

Introduction

Within the [Wikimedia Foundation’s Engineering group](#), the [Discovery department](#)’s mission is to make the wealth of knowledge and content in the [Wikimedia projects](#) (such as [Wikipedia](#)) easily discoverable. The Search team is responsible for maintaining and enhancing the search features and APIs for MediaWiki, such as language detection – i.e. if a French Wikipedia visitor searches and gets fewer than 3 results, we check if maybe their query is in another language, and if our language detection determines that the query’s language is most likely German (for example), then in addition to results from French Wikipedia, they would also get results from German Wikipedia, if any.

Specifically, the [Search team](#)’s current goal is to add cross-wiki searching – that is, providing search results from other (also referred to as “sister”) Wikimedia projects (“*wikis*”) within the same language. For example, if a work (e.g. a book or poem) on French Wikisource matched the user’s query, that user would be shown results from French Wikisource in addition to any results from French Wikipedia.

For the users who received the experimental user experience (UX), each additional wiki’s top result was shown as a box in a sidebar with a link to view more results (see Figure 1). There were two groups of users who received the experimental UX and one control group that did not:

- **Control:** This group received the baseline user experience, which only includes the search results from the wiki they are on. To make their experience comparable to the test groups with respect to latency, we performed the search across the additional indices, but did not show the results to the end user.
- **Test (Random):** This group received the experimental user experience, which includes search results from other wikis (if any were returned). The boxes holding the results (one box for each wiki) were ordered randomly.

Source code and data are available on GitHub ([wikimedia-research/Discovery-Search-Test-CrosswikiSidebar](#))
Screenshot by Deb Tankersley available on [Wikimedia Commons](#), licensed under CC BY-SA 4.0.

Spezialseite

Suchergebnisse

Suchtext: regenbogen

Ergebnisse 1 bis 20 von 2.700

Es gibt eine Seite, die den Namen „Regenbogen“ hat. Weitere Suchergebnisse:

Regenbogen

Mythologie, wird der Regenbogen zerstört. Regenbogen tauchen auch in der Schöpfungsgeschichte der Dink auf. Bei den Inka vertrat der Regenbogen die Erhabenheit 49 KB (5.501 Wörter) · 20:27, 17. Feb. 2017

Radio Regenbogen

Baden Radio Regenbogen news.de Radio Regenbogen Award 2014 – Die Preisträger Radio Regenbogen Award 2015 – Die Preisträger Radio Regenbogen Award 2016 19 KB (789 Wörter) · 12:54, 9. Feb. 2017

Regenbogen – Für eine neue Linke

Regenbogen – Für eine neue Linke (Kurzbezeichnung: REGENBOGEN) war eine Wählervereinigung in Hamburg, die von 1999 bis 2007 existierte. Die Wählergemeinschaft 4 KB (362 Wörter) · 11:32, 28. Feb. 2017

Regenbogen (Sprachdichter)

Kabe, Studien zu dem Meistersinger Barthel Regenbogen. Diss. Greifswald 1930. Johannes Rettbach; Regenbogen. In: Neue Deutsche Biographie (NDB). Band 21 2 KB (216 Wörter) · 11:10, 23. Feb. 2013

Regenbogen-Sufi

Kenntwort „Regenbogen“ versetzen. Seither bestand für die deutschen U-Boote ein ständiger Belehr zur Selbstversenkung bei Ausgabe des Kenntworts „Regenbogen“. Dieser 9 KB (1.012 Wörter) · 10:00, 28. Jan. 2017

Thilo Götz Regenbogen

1965 war Thilo Götz Regenbogen als Künstler aktiv, seit 1969 auch als Publizist. „Als 20-Jähriger bekantete sich Thilo Götz Regenbogen zum Buddhismus. 1970 14 KB (1.652 Wörter) · 15:52, 29. Sep. 2016

Otto Regenbogen

Wählervereinigung Otto Regenbogen wurde am 14. Februar 1891 in der schlesischen Kreisstadt Neumarkt als Sohn des Veterinärmediziners Otto Regenbogen und seiner Frau 28 KB (3.356 Wörter) · 06:44, 4. Jan. 2017

Regenbogen (Heraldik)

Bamini in Brandenburg. Der Regenbogen ist nicht nur im Schild, sondern auch im Oberwappen. Die Gemeinde Tarasp hat den Regenbogen als gebogenen Pfahl (Bogenpfahl) 2 KB (223 Wörter) · 21:27, 31. Jan. 2017

Arim Regenbogen

Arim Regenbogen bei perlungtaucher.de Personalprofil Arim Regenbogen auf der Homepage der Universität Osnabrück Prof. Dr. Arim Regenbogen zum Vizepräsidenten 3 KB (320 Wörter) · 08:47, 9. Nov. 2016

Wandschauer Regenbogen

der Regenbogen nach dem 1. Buch Mose als Zeichen des Bundes des Gottes mit Noah und den Menschen gilt (Gen 9:8–17 EU). Trotzdem wurde der Regenbogen bis 4 KB (472 Wörter) · 16:51, 9. Jul. 2016

Der Regenbogen (Roman)

Google-Buchsuche). Deutsc Der Regenbogen. Insel, Leipzig 1922. Der Regenbogen. Rowohlt, Reinbek 1964 (Taschenbuchausgabe). Der Regenbogen. Rowohlt, Reinbek 1988 11 KB (1.298 Wörter) · 01:20, 22. Nov. 2016

Rágina Regenbogen

veröffentlicht wurde. Sprecher Rágina Regenbogen in der Internet Movie Database (englisch) Interview mit Caroline Rágnerth, der Stimme von Rágina Regenbogen 5 KB (301 Wörter) · 18:40, 22. Jan. 2016

Die Schlange im Regenbogen

Avoid the Puffer Fish Die Schlange im Regenbogen in der Internet Movie Database (englisch) Die Schlange im Regenbogen bei Rotten Tomatoes (englisch) 4 KB (379 Wörter) · 18:07, 26. Sep. 2016

Regenbogen (Begriffsklärung)

Wählervereinigung, siehe Regenbogen – Für eine neue Linke eine Ostberliner Rockband, siehe Regenbogen (Band) eine sterische Schlagerband, siehe Regenbogen (Schlagerband) 1 KB (158 Wörter) · 21:56, 11. Feb. 2017

Regenbogen-Schlangenkopf

Farbkopf mit Vierkieg der Art der deutschen Namen Regenbogen-Schlangenkopf. Der Körper des Regenbogen-Schlangenkopfes ist von braun-beiger Grundfärbung 5 KB (434 Wörter) · 20:59, 27. Nov. 2016

Geißbergslandschaft mit Regenbogen

Außerungen des Malers zum Regenbogen bezieht er sich offenbar auf das Alte Testamente, 1. Buch Mose 9 EU, in dem der Regenbogen ein Zeichen des Bundes ist 26 KB (3.225 Wörter) · 15:12, 7. Dez. 2016

Ralf Regenbogen

bei der Schäfer Amateuremannschaft, Ralf Regenbogen in der Datenbank von fussballdaten.de Ralf Regenbogen in der Datenbank von weltfußball.de DSC 3 KB (284 Wörter) · 07:55, 5. Jan. 2016

Regenbogen-Schwertille

Die Regenbogen-Schwertille (*Iris inornata*) ist eine Pflanzenart aus der Gattung der Schwertillen (*Iris*) innerhalb der Familie der Schwertlilien-Gewächse 3 KB (227 Wörter) · 16:59, 4. Apr. 2016

Das Schloß hinterm Regenbogen

Das Schloß hinterm Regenbogen (auch Das Schloß hinter dem Regenbogen) ist ein rumänischer Märchenfilm von Elisabeta Bostan aus dem Jahr 1969, nach dem 8 KB (1.148 Wörter) · 19:37, 30. Dez. 2016

Regenbogen-Schlammnatter

Die Regenbogen-Schlammnatter (*Farancia erythrogramma*), manchmal auch als Regenbogenmutter bezeichnet, ist eine unglückliche Schlangenart aus der Familie der 5 KB (607 Wörter) · 01:01, 18. Jul. 2016

Zeige (vorherige 20 | nächste 20) (20 | 50 | 100 | 250 | 500)

Schwesterprojekte

Multimedia Radio Regenbogen REGENBEGNE eine neu (weitere)

From Wikisource Der Regenbogen (25) Der Regenbogen. Schönes Kind der Sonne, Holder Regenbogen, Über schwarzen Wolken Mir ein Bild der Hoffnung, 5 Tausend (weitere)

From Wiktionary Regenbogen : Naturscheinung/Lichtbrechung Beispiele: [1] Der Regenbogen ist ein seltenes Phänomen. [1] Der Regenbogen gilt als Zeichen für die Verbundenheit Gottes mit (weitere)

From Wikibooks Bildbearbeitung mit Photoshop: Werkzeuge: Pinsel Airbrush und Stift Vordergrundfarbe bezogen. So kann man zum Beispiel für den oben gezeigten Regenbogen einen Airbrush, Spritzer 28px, Vordergrundfarben nachändernden: rot, orange (weitere)

From Wikivoyage Eggestorf Radfahren: Fahrradverleih am Campingplatz Regenbogen DE-49dedcampsteleep53.17268310.06344 Campingplatz Regenbogen, Tel.: +49 4175 661. (letzte Änderung: (weitere)

From Wikipedia Regenbogen (66) Deinem Kind den Regenbogen zeigt. Aber der Regenbogen wartet nicht, bis du mit der Arbeit fertig bist.“ - Aus China „Glück und Regenbogen sieht man nicht (weitere)

From Wikinews Der Ehrenpreis von „Radio Regenbogen Award“ ging an Johannes Heesters unter Regenbogen“, ka-news: „Radio Regenbogen Award: Promiaufzug für 290 Euro“ (30.03.2006) radio-regenbogen.de: „Kinder unterm Regenbogen gGmbH: Hörer (weitere)

From Wikiversity Schwere See der Wind An der Bordwand hoch empor. Sonnenschein in Tröpfchen-Nebel Regenbogen leuchten lässt. Scherzo Mageninhalt drängt ins Freie Bei manch Kreuzfahrt-Passagieren (weitere)

Datenschutz Über Wikipedia Impressum Entwickler Stellungnahme zu Cookies Mobile Ansicht

Figure 1: Example of cross-wiki search results on Polish Wikipedia, with sister wikis randomly ordered in the sidebar. Multi-media results (including results from Wikimedia Commons) are shown first, regardless of the sidebar ordering.

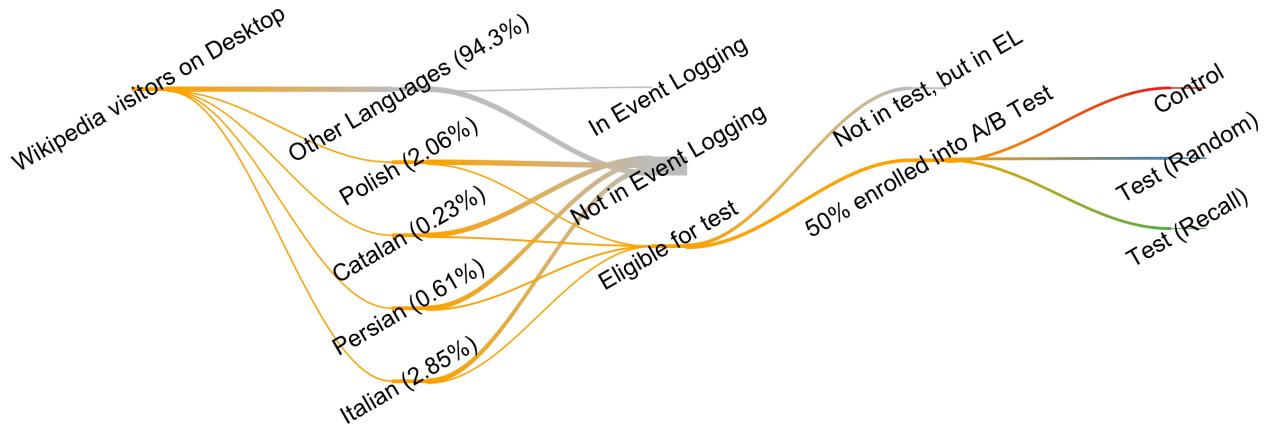


Figure 2: Flow of Wikipedia visitors into the A/B test.

- **Test (Recall):** This group received the experimental user experience, which includes search results from other wikis (if any were returned). The boxes holding the results (one box for each wiki) were ordered according to recall – the volume of search results returned for each respective wiki.

The primary questions we wanted to answer are:

- Did users who saw the additional cross-wiki results engage with those results?
- Was the overall engagement with search results better or worse compared to the controls?

On 9 February 2017 we deployed an A/B test on the desktop version of Catalan, Italian, Persian, and Polish Wikipedias to assess the efficacy of this feature. The test concluded on 22 February 2017, after a total of 6620 search sessions had been anonymously tracked.

Methods

This test's event logging (EL) was implemented in JavaScript according to the [TestSearchSatisfaction2 \(TSS2\)](#) schema, which is the one used by the Search team for its metrics on desktop, data was stored in a MySQL database, and analyzed and reported using R [[R Core Team, 2016](#)].

Data

The data was collected according to the [TSS2 schema, revision 16270835](#). Figure 2 shows the flow of Wikipedia visitors on Desktop. Approximately 5.7% of the unique desktop devices that visit the 270 Wikipedias are accounted for by Catalan, Italian, Persian, and Polish Wikipedias. In general, desktop users are randomly selected for anonymous tracking at a rate of 1 in 200, but for these wikis we changed the sampling rate to 1 in 10 for Catalan and Persian, and 1 in 50 for Italian and Polish. After a user was randomly selected into event logging, they had a 50% chance to be selected for the A/B test. Users who made it into the test were then randomly assigned to one of the three groups described above: Control, Test (Random), and Test (Recall).

Relative traffic was calculated using a combination of [Wikidata Query Service \(WDQS\)](#) and [Wikimedia Analytics' monthly unique devices API](#). See [L16–L51 in the workbook](#) for implementation.

The languages Wikipedia is available in were counted by querying [Wikidata](#) with this SPARQL query. To see the sampling configuration, refer to [Gerrit change 337608](#).

	Control	Test (Random)	Test (Recall)	All 3 groups
Catalan Wikipedia	460	413	414	1287
Italian Wikipedia	664	664	710	2038
Persian Wikipedia	664	631	658	1953
Polish Wikipedia	454	456	432	1342
All 4 wikis	2242	2164	2214	6620

Table 1: Number of search sessions used for analysis by wiki and group. Each search session may have several individual searches.

	Same-wiki clicks	Sister-project clicks	Overall clicks
Control	1321	0	1321
Test (Random)	1260	38	1298
Test (Recall)	1274	42	1316
All 3 groups	3855	80	3935

Table 2: Number of click events by group.

We would like to note that our event logging does not support cross-wiki tracking, so after the user leaves the search results page, we cannot tell whether they have performed subsequent searches, nor how or how long the user engaged with the visited result's page.

See Phabricator ticket [T149806](#) and Gerrit changes [334314](#), [313318](#), [332991](#), [334685](#), and [336896](#) for full details of the implementation on both back-end and front-end.

Analysis

We employed the *binom* [Dorai-Raj, 2014], *LearnBayes* [Albert, 2014], *conting* [Overstall, 2016], and internally-developed *BCDA* [Popov] packages for Bayesian statistical analysis and confidence intervals in Figures 10, 11, 12, and 13.

Results

After the test has concluded on 22 February 2017, we processed the collected data and filtered out duplicated events, extraneous search engine result pages (SERPs), and kept only the searches for which we had both event logging (EL) data and logs of searches (Cirrus requests). This left us with a total of 6620 search sessions with the full breakdown by wiki and group (see Table 1). Table 2 breaks down the counts of clicks on same-wiki results (e.g. a Catalan Wikipedia visitor clicking on a Catalan Wikipedia article) and clicks on sister-projects results (e.g. an Italian Wikipedia visitor clicking on an Italian Wikinews article).

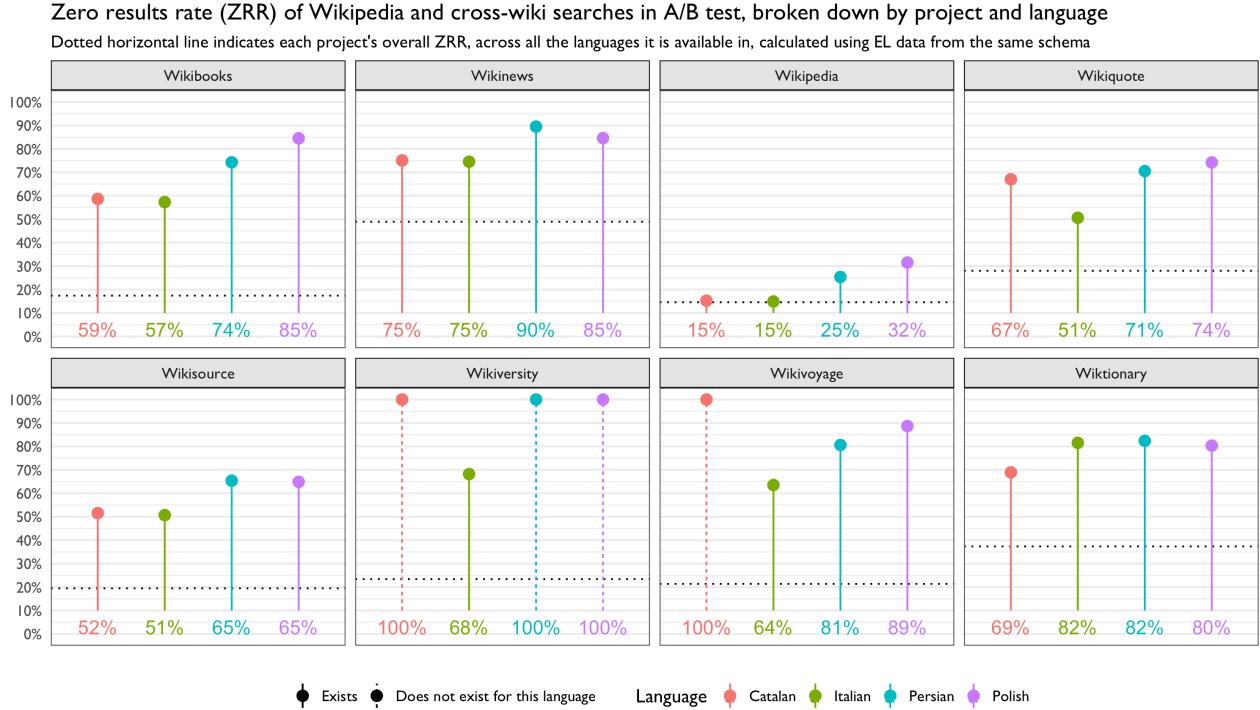
Zero Results Rate (ZRR)

The zero results rate (ZRR) – proportion of searches yielding zero results – is one of Discovery's Search Team's key performance indicators (KPIs), and we are always interested in lowering that number (but not at the expense of results' relevance). While we were primarily interested in searchers' engagement with the search result for this test, we included this section as a consistency check – that the zero results rate is lower when a cross-wiki search is performed (see Figure 10).

In Figure 4, we broke the ZRR from Figure 3 down by language and project and included a reference marker for each project's overall ZRR (aggregated across all the languages the project is available in). Almost



Figure 3: Proportion of searches yielding zero results broken up by group, wiki, and type of results (same-wiki only vs. including cross-wiki results).



At the time of the report, Wikiversity does not exist in Catalan, Persian, and Polish; Wikivoyage does not exist in Catalan. Hence the 100% zero results rate.

Figure 4: The proportion of searches that yielded zero results was the lowest for Wikipedia and Wikisource, with the other projects having very high zero result rates.

Average searches and search engine result pages (SERPs) by group and wiki

A single search can result in multiple SERPs if the user navigates to other pages of results or clicks a result and then goes back

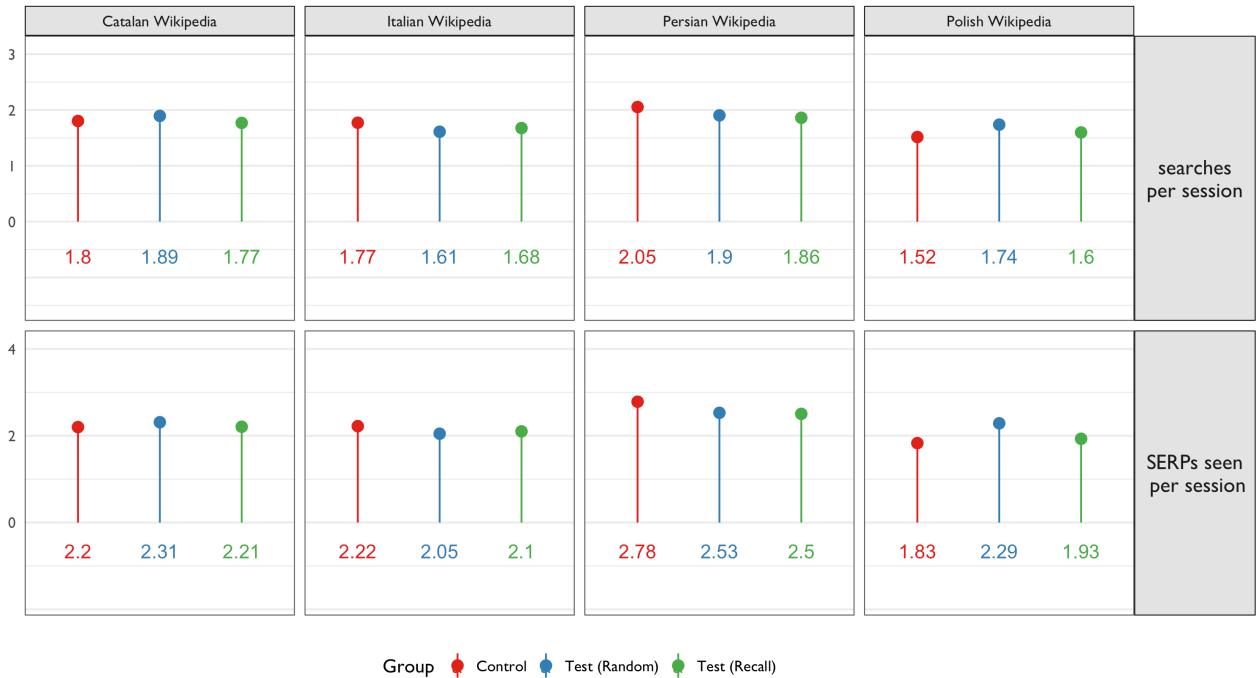


Figure 5: Average number of searches, average number of search engine result pages (SERPs), total searches, total SERPs, and total sessions by group and wiki. The groups did not appear to behave too differently. For example, the three groups had very similar average searches per user.

all of the projects are available (at the time of the test and at the time of writing this report) in Catalan, Italian, Persian, and Polish. Of particular note are the overall ZRR of projects like Wikinews and Wiktionary (both exist in Catalan, Italian, Persian, and Polish), which appear to be much lower than the ZRR observed in this test. In fact, the ZRR in these four languages is much higher than the overall ZRR for every project. We suspect this is partly responsible for the low sister-project click counts seen in Table 2.

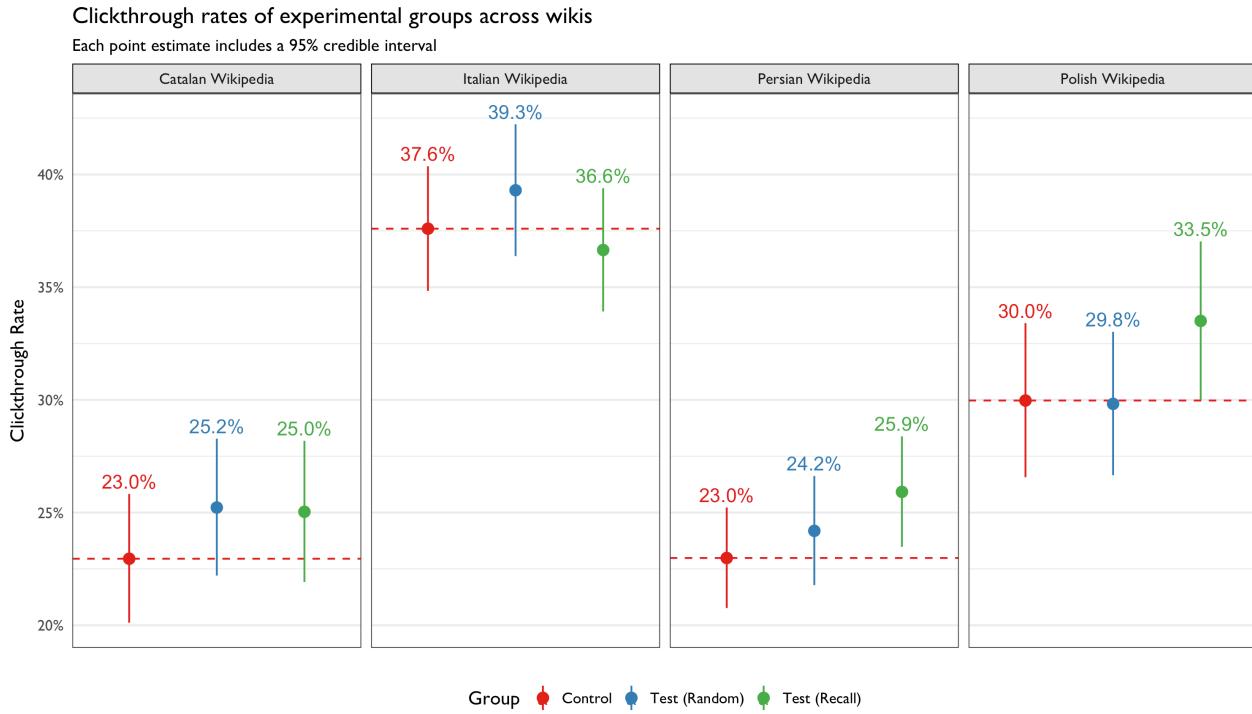
Engagement

We used the clickthrough rate as an indicator of users' engagement with search results and as a measure of the results' relevance. That is, if we present users with more relevant results (such as those from Wikipedia's sister projects), then we expect the clickthrough rate to be higher in the two test groups compared to that of controls. Figure 5 shows that various search activity measures did not vary too much from one group to another.

In Figures 6, 7, and 8, we see that the clickthrough rate was higher in Test (Random) and Test (Recall) than in Control on almost all of the four wikis. The only exception being the clickthrough rates of users in the Polish Wikipedia Test (Random) (29.8%) and Italian Test (Recall) (36.6%) groups.

Table 6 shows the **relative risk** – how much more likely each respective test group is to engage with the search results (same-wiki or cross-wiki) than the Control group. For example, on Catalan Wikipedia, users in the Test (Random) are 1.103 times more likely to click on a result than users in the Control group. While most of the estimates are greater than 1 (suggesting more relevant results), the 95% **credible intervals** contain 1, meaning we do not have sufficient evidence to draw definitive conclusions.

Under the χ^2 discrepancy statistic, the Bayesian p value of 0.497 does *not* indicate that the interaction model is inadequate. Furthermore, Kass and Raftery [1995] suggest that $2 \log_e(\text{BayesFactor}) = 18.629$ is



* For this engagement analysis we focused on: (1) controls' searches with same-wiki results, and (2) test groups' searches that included sister-wiki results.

Figure 6: Clickthrough rates of experimental groups, split by wiki. In the ``Control`` group, only searches that yielded some same-wiki results were considered. In the two test groups, only searches that yielded some sister-wiki results were considered.

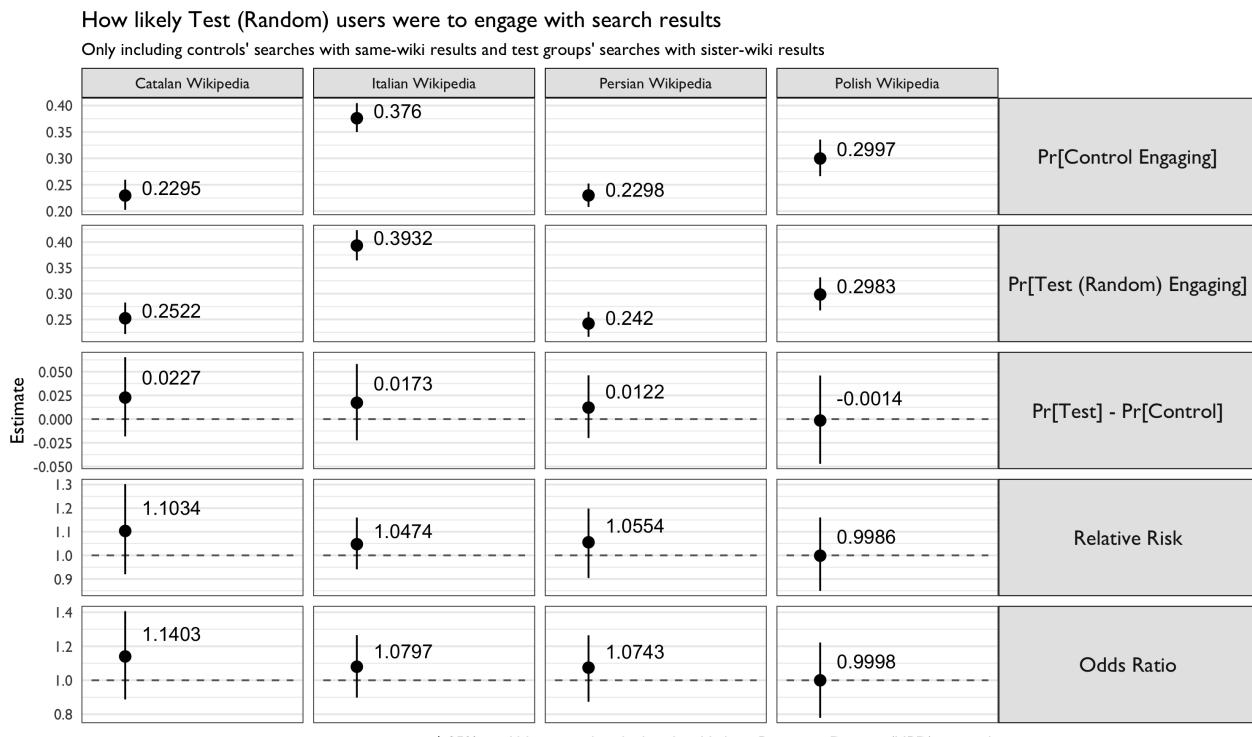


Figure 7: Comparison of the Control group with the Test (Random) group.

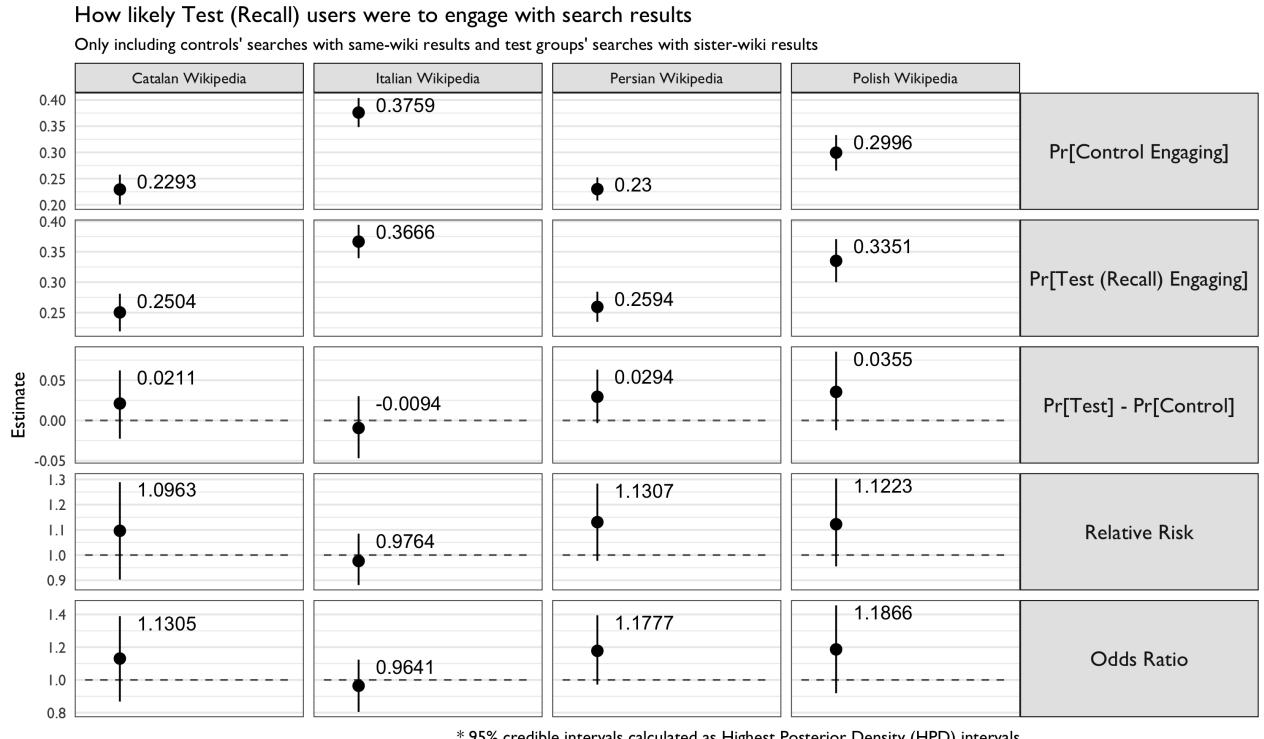


Figure 8: Comparison of the Control group with the Test (Recall) group.

Wiki	Comparison	Relative Risk	95% CI
Catalan Wikipedia	Test (Random) vs Control	1.103	(0.920, 1.302)
Catalan Wikipedia	Test (Recall) vs Control	1.096	(0.903, 1.289)
Italian Wikipedia	Test (Random) vs Control	1.047	(0.941, 1.160)
Italian Wikipedia	Test (Recall) vs Control	0.976	(0.881, 1.085)
Persian Wikipedia	Test (Random) vs Control	1.055	(0.904, 1.199)
Persian Wikipedia	Test (Recall) vs Control	1.131	(0.977, 1.283)
Polish Wikipedia	Test (Random) vs Control	0.999	(0.850, 1.161)
Polish Wikipedia	Test (Recall) vs Control	1.122	(0.955, 1.304)

Table 3: How much more likely the Test (Random) and (Recall) groups were to engage with the search results compared to the Control group.

	0 cross-wiki clicks	1 cross-wiki clicks	2+ cross-wiki clicks
0 sister projects	3310	0	0
1 sister projects	1204	22	1
2 sister projects	784	8	2
3+ sister projects	2309	34	4

Table 4: A contingency table of searches by number of sister projects returned (rows) and number of clicks on those cross-wiki results (columns). For example, there were 8 searches where the users saw 2 sister projects in the sidebar and clicked only once on one of those cross-wiki results.

Coefficient	Estimate	95% HPDI
0 sister projects and 0 cross-wiki clicks	3.008	(2.54, 3.36)
1 sister project	-1.624	(-2.97, -0.78)
2 sister projects	0.364	(-0.34, 0.92)
3+ sister projects	0.032	(-0.47, 0.54)
1 cross-wiki click	4.394	(4.04, 4.88)
2+ cross-wiki clicks	-1.411	(-2.09, -0.67)
1 sister project and 1 cross-wiki click	2.329	(1.48, 3.67)
2 sister projects and 1 cross-wiki click	-0.673	(-1.23, 0.06)
3+ sister projects and 1 cross-wiki click	-0.770	(-1.27, -0.25)
1 sister project and 2+ cross-wiki clicks	-2.237	(-4.45, -0.61)
2 sister projects and 2+ cross-wiki clicks	1.123	(0.29, 1.98)
3+ sister projects and 2+ cross-wiki clicks	0.422	(-0.31, 1.32)

Table 5: Results of fitting a Bayesian log-linear model to cross-wiki results and cross-wiki clickthroughs.

very strong evidence against null hypothesis of independence. This means there is evidence of a relationship between number of projects displayed and number of clicks on those sister-wiki results.

Table 5 summarizes the MCMC results of fitting a Bayesian log-linear model to the data in Table 4. It suggests there is a strong interaction between number of projects returned and number of clicks on those projects. Contrasting the negative estimate for “3+ sister projects and 1 cross-wiki click” (-0.77) to the positive estimates for “2/3+ sister projects and 2+ cross-wiki clicks” (1.123 and 0.422, respectively), *the model suggests the relationship is positive – that more sister projects shown to the user yields more cross-wiki clicks, up to a point.*

Discussion

As can be seen in Figure 1, the cross-wiki results were displayed in black rather than the standard blue. This issue is tracked in [T158935](#) and has been [resolved](#) for the upcoming follow-up test ([T160004](#)). We cannot estimate the effect this may have had on the results of the test, but we suspect this may have had a considerable negative effect because the results did not look like click-able links.

We also suspect that the high zero results rate for each of the sister projects for these four languages may have been responsible for the few sister-project clicks. In Figure 9, relatively few searches had more than 3 cross-wiki search results. As shown in Table 5 in the [Engagement](#) analysis, there is evidence that suggests a positive relationship between number of sister projects in the sidebar and clicks on those cross-wiki results.

Furthermore, since the users did not see more than the top result from each sister project, it is possible they did not even want to bother with viewing more results if the one they were shown was not relevant. Ideally, the first result would always be the most relevant one, but that is not always the case, and sometimes results further down the list are likely to be more relevant to the user’s actual task. It would be difficult to show (for example) the top three results from each sister project without making the user interface (UI) more cluttered. Perhaps we could show the other results, but collapse them, and allow the user to expand each sister project’s results without having to go to another page.

It is worth noting that Multimedia results were shown in a reverse order, but we suspect this is a minor bug that did not have an effect on the test. Additionally, there is a possibility the users saw what they wanted to see from the cross-wiki result preview in the sidebar and that they did not feel the need to click through.

Of the 72 unique searches that included a click on the cross-wiki results, 54 were searches where the user received both sets of results (same-wiki and cross-wiki) but clicked only once and specifically on a cross-wiki

The multimedia reverse order bug is tracked in [T158937](#) and was resolved in [Gerrit change 342200](#).

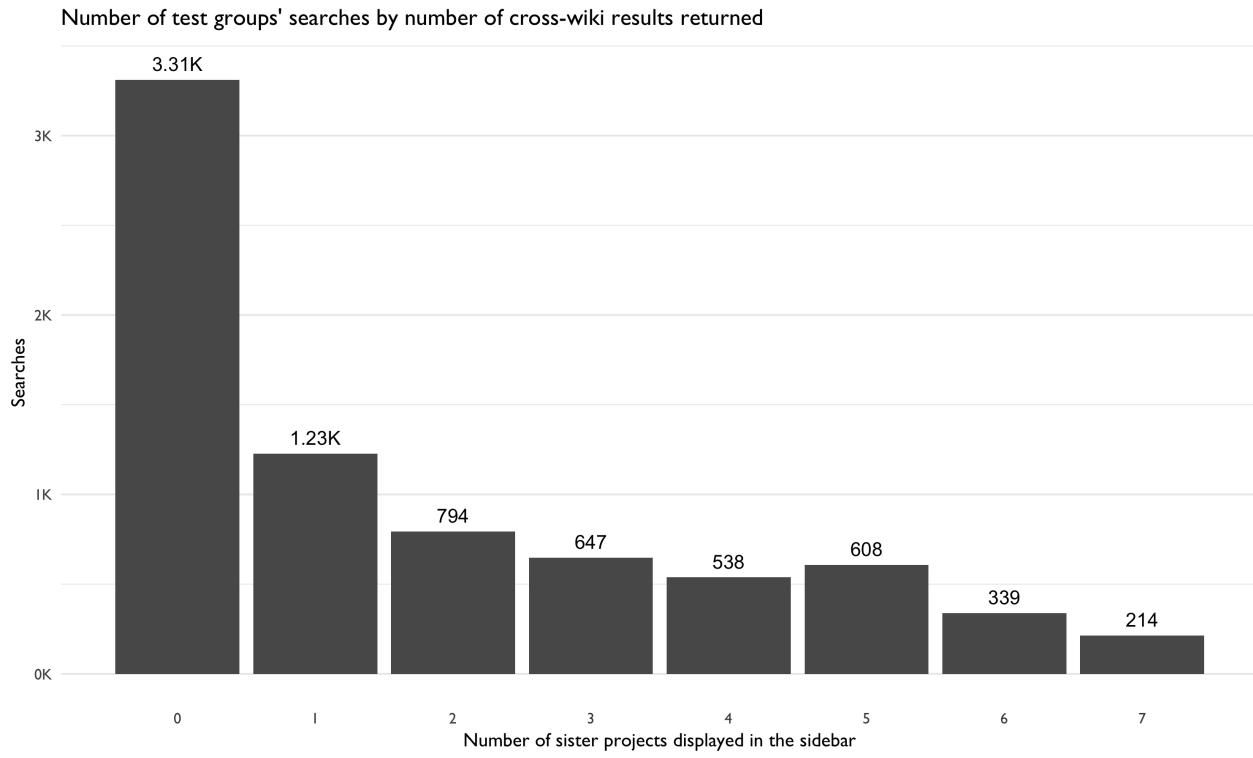


Figure 9: How many of the two test groups' searches returned cross-wiki results from 0 (none) - 7 (all) sister projects.

result. This suggests, perhaps, that for some users the results from sister projects may have been more relevant than the results from the wiki they were on.

Acknowledgements

We would like to thank our colleagues Trey Jones (Software Engineer, Wikimedia Foundation), Chelsy Xie (Data Analyst, Wikimedia Foundation), and Kevin Smith (Agile Coach, Wikimedia Foundation) for their reviews of and feedback on this report.

References

- Jim Albert. *LearnBayes: Functions for Learning Bayesian Inference*, 2014. URL <https://CRAN.R-project.org/package=LearnBayes>. R package version 2.15.
- JJ Allaire, Joe Cheng, Yihui Xie, Jonathan McPherson, Winston Chang, Jeff Allen, Hadley Wickham, Aron Atkins, Rob Hyndman, and Ruben Arslan. *rmarkdown: Dynamic Documents for R*, 2016. URL <http://rmarkdown.rstudio.com>. R package version 1.3.9002.
- Stefan Milton Bache and Hadley Wickham. *magrittr: A Forward-Pipe Operator for R*, 2014. URL <https://CRAN.R-project.org/package=magrittr>. R package version 1.5.
- Sundar Dorai-Raj. *binom: Binomial Confidence Intervals For Several Parameterizations*, 2014. URL <https://CRAN.R-project.org/package=binom>. R package version 1.1-1.
- R E Kass and Adrian E Raftery. Bayes factors. *Journal of the American Statistical Association*, 1995.
- Oliver Keyes and Mikhail Popov. *wmf: R Code for Wikimedia Foundation Internal Usage*, 2017. URL <https://phabricator.wikimedia.org/diffusion/1821/>. R package version 0.2.6.
- Antony M. Overstall. *conting: Bayesian Analysis of Contingency Tables*, 2016. URL <https://CRAN.R-project.org/package=conting>. R package version 1.6.
- Mikhail Popov. *BCDA: Tools for Bayesian Categorical Data Analysis*. URL <https://github.com/bearloga/BCDA>. R package version 0.2.3.
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016. URL <https://www.R-project.org/>.
- Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2009. ISBN 978-0-387-98140-6. URL <http://ggplot2.org>.
- Hadley Wickham. *tidyverse: Easily Tidy Data with 'spread()' and 'gather()' Functions*, 2017. URL <https://CRAN.R-project.org/package=tidyr>. R package version 0.6.1.
- Hadley Wickham and Romain Francois. *dplyr: A Grammar of Data Manipulation*, 2016. URL <https://CRAN.R-project.org/package=dplyr>. R package version 0.5.0.