

Lotsenmodul Analyse

Simon Kock (WMDE)

2023-04-06

Fragestellung

Wie viele Neuregistrierte nutzen das Lotsenmodul (i.e. stellen mindestens eine Frage)?

Wie ist die (weitere) Aktivität von Neuregistrierten, die das Lotsenmodul verwendet haben, unter anderem verglichen zu solchen, die es nicht benutzen?

Verwendete Daten und deren Quellen

Um unsere Fragen zum Lotsenmodul zu beantworten benötigen wir Daten mit folgenden Informationen:

- Neuregistrierte Nutzende
 - event_sanitized -> serversideaccountcreation (wmf data lake)
 - wmf -> mediawiki_history (wmf data lake)
- Nutzende mit automatisch aktiviertem Lotsenmodul
 - dewiki -> user_properties
 - diese Daten sind nicht öffentlich zugänglich
- Seitenbesuche der Homepage, auf der unter anderem das Lotsenmodul angeboten wird
 - event -> homepagevisit (wmf data lake)
 - verfügbar nur für die letzten 90 Tage
- Interaktion mit dem Lotsenmodul
 - event -> helppanel (wmf data lake)
 - verfügbar nur für die letzten 90 Tage
- Fragen von neuen Nutzenden über das Lotsenmodul
 - event -> mediawiki_revision_tags_change (wmf data lake)
 - wmf -> mediawiki_history (wmf data lake)
- Editieraktivität von neuen Nutzenden
 - wmf -> mediawiki_history (wmf data lake)
 - snapshot von 2022-12

Aufgrund von zeitlich begrenzter Verfügbarkeit einiger Datensätze und besserer Vergleichbarkeit sind die Zeiträume der Daten dieser Analyse eingegrenzt auf:

- Neuregistrierungen Oktober 2022
- Aktionen (Seitenbesuche / Fragen stellen) Oktober/November 2022
- Edits ab Registrierung so weit wir Daten zur Verfügung haben (derzeit bis einschließlich Dezember)

Datenverarbeitung

Zur Analyse werden Nutzende eingeschlossen, die sich im Oktober 2022 registriert haben. Bei den Nutzenden wird unterschieden, ob diese automatisch aktivierte Growth Features (darunter Lotsenmodul) haben. Nutzende, die selbstständig diese Features aktivieren, werden noch einmal separat betrachtet.

Neuregistrierte mit aktivierten Modulen bekommen ab ihrer Registrierung 30 Tage Zeit, um mit dem Lotsenmodul zu interagieren oder die Startseite zu besuchen. Somit haben alle Neuregistrierten die gleiche Zeit zur Verfügung.

Um das Editierverhalten anzuschauen, werden alle Edits herangezogen, die die Neuregistrierten nach ihrer Registrierung in der deutschsprachigen Wikipedia gemacht haben. Edits, die sich auf die Frage im Lotsenmodul beziehen, werden ausgeschlossen.

Deskriptive Ergebnisse

Die deskriptiven Ergebnisse sollen einen Überblick über die Nutzung des Lotsenmoduls durch Neuregistrierte geben und dabei helfen die Nutzung besser einzuordnen und Informationen für mögliche weitere Analysen zu dem Thema bereit zu stellen.

Anzahl Neuregistrierte im Oktober 2022

Für die derzeitige Analyse werden Personen eingeschlossen, die sich im Oktober (UTC Zeitzone) neu registriert haben. Für 80% der Neuregistrierten werden die Module automatisch auf der Startseite der Nutzenden angezeigt.

aktivierte Module	Anzahl Neuregistrierungen
FALSE	1370
TRUE	5365

Anzahl Besuchende auf der Startseite innerhalb der ersten 30 Tage

Ein erster Schritt ist es zu schauen, wie viele der Neuregistrierten ihre Startseite besuchen, da hier die Module angezeigt werden.

Insgesamt besuchen etwas mehr als die Hälfte der Neuregistrierten mit aktivierten Modulen ihre Startseite, i.e. 56.98%. Diese Anzahl ist vergleichbar mit einer Analyse der WMF zu Startseitenbesuchen innerhalb der ersten 48h (<https://phabricator.wikimedia.org/T245369>).

Besuch auf Homepage	Anzahl Nutzende
FALSE	2308
TRUE	3057

Anzahl Nutzende, die mit dem Lotsenmodul innerhalb der ersten 30 Tage interagieren.

Als Interaktion wird jede Art von Interaktion mit dem Modul gewertet (z.B. einen der angegebenen Links klicken, eine Frage stellen, etc).

3 Personen haben die Module manuell aktiviert und tauchen dementsprechend in dieser Tabelle auf.

aktivierte Module	Interaktion mit Lotsenmodul	Anzahl Nutzende
FALSE	FALSE	1367
	TRUE	3
TRUE	FALSE	5212
	TRUE	153

Anzahl Nutzende, die eine Frage im Lotsenmodul innerhalb der ersten 30 Tage stellen

3 Personen haben die Module manuell aktiviert und tauchen dementsprechend hier auf.

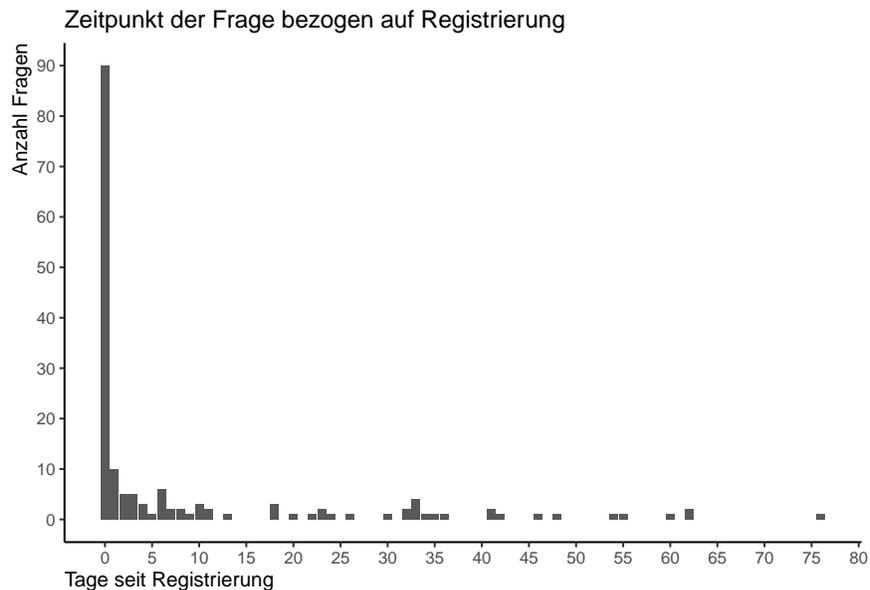
Insgesamt stellen 130 Neuregistrierte mindestens eine Frage innerhalb von 30 Tagen über das Lotsenmodul. Das entspricht 2.42% der Neuregistrierten, die das Lotsenmodul automatisch aktiviert haben.

13 Neuregistrierte stellen im Zeitraum bis Ende Dezember mehr als eine Frage.

aktivierte Module	Frage gestellt	Anzahl Nutzende
FALSE	FALSE	1367
	TRUE	3
TRUE	FALSE	5235
	TRUE	130

Zeitpunkt der Fragen über das Lotsenmodul

Das Balkendiagramm zeigt die Zeitpunkte an denen eine Frage im Lotsenmodul gestellt wurde in Bezug zum Zeitpunkt der Registrierung.



Deskriptive Statistik der Edits von Neuregistrierten

Zuerst werden Neuregistrierte mit automatisch aktivierten Growth Features mit denen ohne Growth Features verglichen. Alle danach folgenden Analysen beziehen sich auf das Lotsenmodul im Speziellen.

Vergleich Nutzende mit und ohne Growth Features

Um besser einschätzen zu können, ob das generelle Angebot der Growth Features (i.e. die Startseite mit allen aufgeführten Modulen), zu denen unter anderen das Lotsenmodul gehört, einen Einfluss auf die Bearbeitungsaktivität hat, kann das Verhältnis zwischen Neuregistrierten, die mindestens einmal innerhalb der ersten 30 Tage editieren und der gemittelten Anzahl an Neuregistrierten, die an einem bestimmten Tag innerhalb der ersten 30 Tage mindestens einmal täglich editieren, zur Hilfe genommen werden. Diese Metrik ist auch als “Stickiness” bekannt. Der in Prozent angegebene Wert trifft eine Aussage darüber, wie viele der aktiven Editierenden auch regelmäßig editieren. Je höher der Prozentsatz, desto größer ist der Anteil an regelmäßig aktiven Editierenden. Für Editierende mit aktivierten Growth Features liegt die *Stickiness* bei **5.14%**. Die Gruppe der Editierenden ohne Growth Features hat eine *Stickiness* von **5.29%**. Beide Werte beziehen sich auf die ersten 30 Tage nach Registrierung. Der geringe Unterschied der Werte zeigt, dass es

keinen Unterschied zu machen scheint, ob Neuregistrierten Growth Features und damit auch das Lotsenmodul angeboten werden oder nicht, gemessen an der regelmäßigen Aktivität der Neuregistrierten.

Ein Grund hierfür könnte sein, dass ein zu geringer Anteil an Neuregistrierten das Lotsenmodul (und die anderen Features) verwendet, bzw. überhaupt findet (siehe Anzahl Neuregistrierte, die die Startseite besuchen).

Die folgenden Analysen beziehen sich auf das Lotsenmodul.

Allgemeine Edits

Neuregistrierte, die eine Frage über das Lotsenmodul stellen, scheinen im Schnitt eine höhere Anzahl an allgemeinen Edits durchzuführen, als Neuregistrierte, die keine Frage stellen.

Neuregistrierte, die keine aktivierten Module haben und Neuregistrierte mit aktivierten Modulen, die diese aber nicht nutzen, scheinen sich in Bezug auf Anzahl an Edits deskriptiv ähnlich zu verhalten. Ihre mittlere Anzahl an Edits ist vergleichbar. Ebenfalls macht die Mehrheit der beiden Gruppen keine Edits (der Median der Edits liegt bei 0).

aktivierte Module	Frage gestellt	N User	Mittlere N Edits	Median N Edits
FALSE	FALSE	1367	3.97	0
	TRUE	3	3.33	4
TRUE	FALSE	5235	3.34	0
	TRUE	130	16.81	3

Edits gruppiert nach dessen Inhalt

Für die weitere Analyse werden die Neuregistrierten mit automatisch aktivierten Modulen betrachtet.

Werden die Edits kategorisiert nach Inhalt (ja/nein, definiert durch das jeweilige Projekt) und nach Größe der Revision (klein/groß, subjektive Einschätzung der Editoren) ist zu sehen, dass Nutzende, die eine Frage stellen, im Mittel eine höhere Anzahl (N) an Edits haben. Dies ist vor allem bei kleineren Edits mit Inhalt zu sehen und könnte ein erster Hinweis darauf sein, dass das Lotsenmodul dabei unterstützt Aktivität zu fördern. Allerdings liegt hier nicht unbedingt eine Kausalität vor, da Nutzende die aktiv eine Frage stellen sich grundsätzlich von Nutzenden unterscheiden, die keine Frage stellen. Ein naheliegender Unterschied könnte in ihrer allgemeinen Motivation liegen.

Weiterhin ist interessant, dass es einen Unterschied besonders bei kleineren inhaltlichen Edits gibt. Ein Interpretationsansatz könnte sein, dass neue Editierende zu Beginn eher kleinere Änderungen einarbeiten und erst nach einer gewissen Zeit größere Bearbeitungen vornehmen.

Editierart	Frage gestellt	N User	N Edits	Mittlere N Edits
Inhalt groß	FALSE	1493	9828	6.58
	TRUE	67	659	9.84
Inhalt klein	FALSE	978	3620	3.70
	TRUE	53	714	13.47
kein Inhalt	FALSE	779	4028	5.17
	TRUE	77	812	10.55

Edits gruppiert nach Namensräumen

Werden die Edits nach den unterschiedlichen Namensräumen (Quelle: <https://de.wikipedia.org/wiki/Hilfe:Namensr%C3%A4ume>) gruppiert, zeigt sich ein ähnliches Bild. Vor allem im Artikelnamensraum (inhaltliche Edits) scheint es einen Unterschied zu machen, ob Neuregistrierte eine Frage stellen oder nicht. Auch bei den Benutzer*innenseiten (BNR und BD) gibt es eine Tendenz in diese Richtung. Für die anderen Kategorien ist die Anzahl an Nutzenden, die in diesen Kategorien aktiv wurden sehr gering und es ist daher schwierig eine Tendenz abzuleiten.

Namensraum	Frage gestellt	N User	N Edits	Mittlere N Edits
ANR	FALSE	2003	13448	6.71
	TRUE	79	1373	17.38
ANR Diskussion	FALSE	142	387	2.73
	TRUE	13	62	4.77
BNR und BD	FALSE	606	2863	4.72
	TRUE	73	596	8.16
WP und WD	FALSE	156	521	3.34
	TRUE	19	72	3.79
HD	FALSE	<5	-	-
P und PD	FALSE	6	105	17.5
	TRUE	<5	-	-
Sonstige	FALSE	26	151	5.81
	TRUE	6	78	13

Anhang

Die Auswertung zu Edits bezogen zum Zeitpunkt der Frage im Lotsenmodul ist als Anhang mit aufgeführt. Eine Interpretation ist hier schwierig und eigentlich wenig sinnvoll, da die zur Verfügung stehende Zeit nachdem eine Frage gestellt wurde wesentlich größer ist, als davor. Da viele Neuregistrierte ihre erste Frage sehr früh stellen, ist es nur bedingt sinnvoll den Zeitraum für Edits nach der Frage anzugleichen.

Edits gruppiert nach dessen Inhalt und Zeitpunkt

Werden die Edits zusätzlich nach Zeitpunkt bezogen zur Frage im Lotsenmodul kategorisiert, ist zu sehen, dass die Aktivität nach der Nutzung des Lotsenmoduls größer ist, als davor. Dies ist mit der zur verfügungstehenden Zeit zu erklären, die nach der gestellten Frage größer ist, als davor. Nutzende haben also mehr Zeit nach der Frage aktiv zu sein. Allerdings könnte dies auch als Hinweis darauf gewertet werden, dass das Lotsenmodul keinen negativen Effekt auf die Aktivität der Nutzenden ausübt und diese zumindest nicht abschreckt aktiv zu bleiben (oder zu werden).

Namensraum	Zeitpunkt	N User	N Edits	Mittlere N Edits
ANR	post	57	1078	18.91
	pre	57	295	5.18
ANR Diskussion	post	11	55	5
	pre	<5	-	-
BNR und BD	post	65	457	7.03
	pre	34	139	4.09
WP und WD	post	16	46	2.88
	pre	6	26	4.33
P und PD	post	<5	-	-
Sonstige	post	5	77	15.4
	pre	<5	-	-

Editierart	Zeitpunkt	N User	N Edits	Mittlere N Edits
Inhalt groß	post	45	469	10.42
	pre	43	190	4.42
Inhalt klein	post	37	609	16.46
	pre	36	105	2.92
kein Inhalt	post	71	639	9.00
	pre	38	173	4.55