



Netzwerkkommunikation
im Internet

Berichte des DFG-Forschungsprojekts „Netzwerkkommunikation im Internet“,
Nr. 08 (02)

Aktivitäten auf Diskussionsseiten

Zur Einordnung der Lebhaftigkeit der
Diskussionsseiten ‚Tsunami‘ und ‚Second Life‘
im Vergleich

Autor:
Sebastian Erlhofer

Medienwissenschaft Universität Trier

Zitationsvorschlag:
*Erlhofer, Sebastian (2008): Aktivitäten auf Diskussionsseiten. Zur Einordnung der Lebhaftigkeit der
Diskussionsseiten ‚Tsunami‘ und ‚Second Life‘ im Vergleich. In: Berichte des DFG-Forschungsprojekts
„Netzwerkkommunikation im Internet“, Nr. 08 (2).*

Trier, Juni 2008

Inhaltsverzeichnis

1. Aktivitäten auf Diskussionsseiten	3
1.1. Datenerhebung	3
1.2. Archivierung der Diskussionsseiten	3
1.3. Operationalisierung der Aktivität	3
1.3.1. Größe.....	3
1.3.2. Anzahl der beteiligten Autoren.....	3
1.3.3. Anzahl der Editiervorgänge	4
1.3.4. Anzahl der „Minor-Edits“	4
2. Auswertung.....	4
2.1.1. Werte der beiden Diskussionsseiten	4
2.1.2. Vergleich mit den Mittelwerten aller Diskussionsseiten	5
2.1.3. TOP-10 Diskussionsseiten nach Variable	5
2.1.4. Vergleich der Rangfolge	7
2.1.5. Fazit: Relativ hohe Aktivität	7

1. Aktivitäten auf Diskussionsseiten

Im Rahmen des Projekts wurde der Fokus mit ‚Tsunami‘ und ‚Second Life‘ auf zwei bestimmte Artikel in Wikipedia gelegt. Um die jeweiligen qualitativen Befunde besser einordnen zu können scheint eine Einschätzung unabdinglich, in wie weit die Aktivitäten auf der Diskussionsseiten der beiden Artikeln im Vergleich zu der Masse der anderen Wikipedia-Diskussionsseiten stehen.

Im Folgenden soll daher quantitativ untersucht werden, wie aktiv die beiden Diskussionsseiten im Vergleich zu anderen Wikipedia-Diskussionsseiten genutzt wurden.

1.1. Datenerhebung

Als Datengrundlage dient die mysql-Datenbank, die aus dem Wikipedia-Dump der deutschsprachigen Wikipedia generiert wurde. Der Datenstand bezieht sich auf Januar 2008. Die untersuchte Grundgesamtheit stellen alle Wikipedia-Seiten dar, die dem Namensraum „Diskussionsseite“ (Namespace-ID 1) zugeordnet sind.

1.2. Archivierung der Diskussionsseiten

Die Wikipedia-Software bietet die Möglichkeit, einzelne Diskussionsseiten zu archivieren. Die jeweiligen Archivseiten werden als eigenständige Seiten behandelt. Eine archivierte Diskussionsseite zu Nordkorea beispielsweise finden sich unter http://de.wikipedia.org/wiki/Diskussion:Nordkorea/Archiv_2005. Die Archivierung kann entweder manuell stattfinden, oder, wie im Beispiel, automatisch.

Da für die Untersuchung der Aktivitäten die Menge aller Diskussionen eines Artikels interessant ist, wurden die Daten der einzelnen Diskussions-Archive jeweils zu einer Summe zusammengerechnet.

Insgesamt ergaben sich somit 217.983 Diskussionsseiten für die Untersuchung.

1.3. Operationalisierung der Aktivität

Die Operationalisierung der Stärke der Aktivität im Diskussionsraum wurde über verschiedene Variablen durchgeführt, welche aus der relationalen Datenbank mittels SQL-Querys gewonnen und einer Nachbearbeitung zur Archiv-Zusammenfassung (wie oben beschrieben) unterzogen wurden.

1.3.1. Größe

Die Größe der Diskussionsseite gibt die Zahl der Bytes an, die in Textform vorliegen. In dem verwendeten ASCII-Zeichensatz entspricht ein Byte einem Zeichen (auch Leerzeichen). Eine DIN A4 Seite mit Schriftgröße 12 besitzt ca. 4.000 Zeichen.

1.3.2. Anzahl der beteiligten Autoren

Die Anzahl der beteiligten Autoren auf der Diskussionsseite wurde ebenfalls über den gesamten Zeitraum gezählt („alle jemals aktiven Autoren“). Dabei wurden auch die nicht eingeloggt Autoren berücksichtigt, die nur über eine IP-Adresse erkennbar sind. Hier kann

es methodisch unter Umständen zu Mehrfachzählungen eines Autors mit unterschiedlichen IP-Adressen geben.

1.3.3. Anzahl der Editiervorgänge

Die Anzahl der Textmodifikationen (Hinzufügen, Löschen, Korrigieren) wurden quantitativ berücksichtigt und nicht qualitativ unterschieden, da es bei jeder dieser Modifikationsformen um eine Handlung handelt, welche als Maß für die Aktivität gewertet werden kann.

1.3.4. Anzahl der „Minor-Edits“

Wikipedia bietet den Autoren die Möglichkeit, Änderungen als kleinere, inhaltlich unbedeutende Modifikationen zu kennzeichnen („Minor Edits“). Dieser Tag muss von dem jeweiligen Autor selbstständig gesetzt werden und wird üblicherweise für orthografische oder grammatikalische Korrekturen verwendet. Zur Einschätzung, wie viele der gezählten Editiervorgänge keine Minor-Edits sind, wurden die Minor-Edits mit erhoben.

2. Auswertung

Um die beiden Artikel in Relation zu dem Durchschnitt der Wikipedia-Diskussionsseiten setzen zu können, sollen zunächst die Werte der beiden Artikel dargestellt werden. Anschließend werden die Mittelwerte der Diskussionsseiten dargestellt. Abschließend findet eine Einordnung der Artikel in die Rangfolge der Diskussionsseiten nach den einzelnen Operationalisierungsvariablen statt.

2.1.1. Werte der beiden Diskussionsseiten

An der Diskussion um den **Tsunami**-Artikel beteiligten sich insgesamt 105 Autoren. Es fanden 175 Editiervorgänge (19 Minor Edits) statt. Im Durchschnitt fanden also 1,6 Editierungen pro Autor statt. Insgesamt beträgt die Länge der Tsunami-Diskussion 69.636 Bytes, also ca. 17,5 DIN A4 Seiten.

An der Diskussion des **SecondLife**-Artikels waren 86 Autoren beteiligt. Hier gab es 168 Editiervorgänge (36 Minor Edits). Mit 1,95 Edits pro Autor liegt der Durchschnitt höher als bei dem Tsunami-Artikel. Die Länge beträgt 52,258 Bytes, also ca. 13 DIN A4 Seiten.

2.1.2. Vergleich mit den Mittelwerten aller Diskussionsseiten

Vergleicht man die Daten der beiden Diskussionsseiten mit den Mittelwerten der 217.983 Diskussionsseiten, so kann konstatiert werden, dass beide Diskussionsseiten stark überdurchschnittlich aktiv sind.

		Authors	Edits	minorEdits	GrößeInBytes
N	Gültig	217.982	217.982	217.982	217.982
	Fehlend	0	0	0	0
Mittelwert		4,93	10,44	1,43	4122,62
Median		2,00	3,00	,00	694,00
Standardabweichung		14,218	60,673	9,690	27126,801

Tabelle 1: Mittelwerte aller Diskussionsseiten

Dies ist jedoch kaum verwunderlich, wenn man die Verteilung der einzelnen Variablen betrachtet:

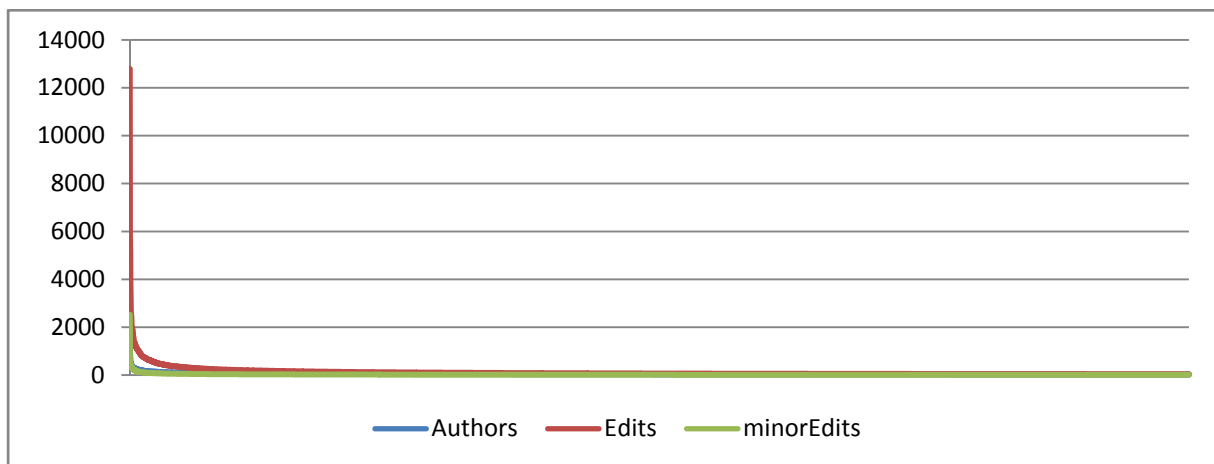


Abbildung 1: Verteilung der Variablen

Es lässt sich festhalten, dass alle Variablen nicht normal, sondern stark exponentiell verteilt sind und damit einer Annäherung des Power-Laws entsprechen. Die Verteilung der Variable „Größe in Byte“ ist ebenfalls exponentiell verteilt (auf die Darstellung wurde aus Gründen der Achsenhöhe verzichtet).

2.1.3. TOP-10 Diskussionsseiten nach Variable

Interessant für eine Einordnung könnte ggf. auch die Betrachtung der ersten zehn Diskussionsseiten sein. Diese sollen tabellarisch anhand der einzelnen Variablen aufgelistet werden:

Diskussionsseite (Titel)	Anzahl der beteiligten Autoren
Hauptseite	3253
Adolf_Hitler	972
Völkermord_an_den_Armeniern	804
Zeugen_Jehovas	774
Deutschland	766
Scientology	752
Tokio_Hotel	676
Die_Simpsons	648
Homöopathie	619
Fußball-Weltmeisterschaft_2006	567

Diskussionsseite (Titel)	Anzahl der Editiervorgänge
Hauptseite	12.784
Zeugen_Jehovas	7.012
Homöopathie	5.816
Völkermord_an_den_Armeniern	5.574
Joseph_Beuys	5.272
Scientology	5.237
Jesus_von_Nazaret	4.263
Junge_Freiheit	3.848
Libanonkrieg_2006	3.615
Neger	2.891

Diskussionsseite (Titel)	Anzahl der Minor Edits
Hauptseite	2.515
Joseph_Beuys	1.859
Völkermord_an_den_Armeniern	730
Kroatien	600
Zeugen_Jehovas	493
Sexismus	473
Homöopathie	465
Jahr_Null	458
Libanonkrieg_2006	443
Adolf_Hitler	411

Diskussionsseite (Titel)	Größe in Bytes
Homöopathie	3.960.358
Hauptseite	3.286.525
Scientology	3.259.174
Zeugen_Jehovas	3.171.377
Völkermord_an_den_Armeniern	2.183.276
Jesus_von_Nazaret	1.912.587
Atlantis	1.690.719
Kreationismus	1.595.335
Junge_Freiheit	1.534.944
Transzendente_Meditation	1.448.861

2.1.4. Vergleich der Rangfolge

Um der exponentiellen Verteilung gerecht zu werden, wurde untersucht, an welcher Stelle der Rangfolge die beiden Artikel für die jeweils einzelnen Variablen positioniert sind:

	Autoren	Edits	Minor Edits	Größe
Tsunami	489 (0,22%)	1.407 (0,65%)	1.901 (0,87%)	1.689 (0,77%)
Second Life	733 (0,34%)	1.500 (0,69%)	768 (0,35%)	2.490 (1,14%)

Tabelle 2: Rangfolge nach Variable

Die Prozentwerte hinter den Rangpositionen geben an, wie viel Prozent aller Wikipedia-Diskussionsseiten für die betreffende Variable einen höheren Rang erzielen.

Insgesamt befinden sich die beiden Diskussionsseiten in allen Werten (mit Ausnahme der Größe bei Second Life) in dem oberen 1 Prozentbereich aller Wikipedia-Diskussionsseiten. Die durchschnittliche Rangposition beider Diskussionsseiten ist interessanterweise nahezu gleich. Die Tsunami-Diskussion reiht sich in Position 1.371 ein, die SecondLife-Diskussion bei 1.372. Damit liegen 0,63 % aller Wikipedia-Diskussionen auf vorderen Plätzen – sind also als aktiver anzusehen als die beiden betrachteten Diskussionsseiten.

2.1.5. Fazit: Relativ hohe Aktivität

An der Tsunami-Diskussion sind besonders viele Autoren (105) beteiligt. Die Diskussion um den SecondLife-Artikel verzeichnet weniger Autoren und ist in Folge dessen auf einem niedrigeren Rang zu finden. Bei der Anzahl und Rangposition der Editiervorgänge liegen beide Diskussionen etwa gleichauf. Die Minor Edits erweisen sich als generell wenig, so dass hier keine Interpretation sinnvoll scheint. Die Größe der Diskussionsseite, gewissermaßen als Basis und Ergebnis einer Diskussion gleichermaßen, ist bei Tsunami etwas größer.

Bei den beiden Diskussionsseiten lässt sich folglich eine relativ starke Aktivität zu anderen Artikeln über den gesamten Zeitraum feststellen, da sie im oberen Prozentbereich platziert sind. Weitere Analysen, die einzelne Zeitabschnitte separat betrachten werden folgen.