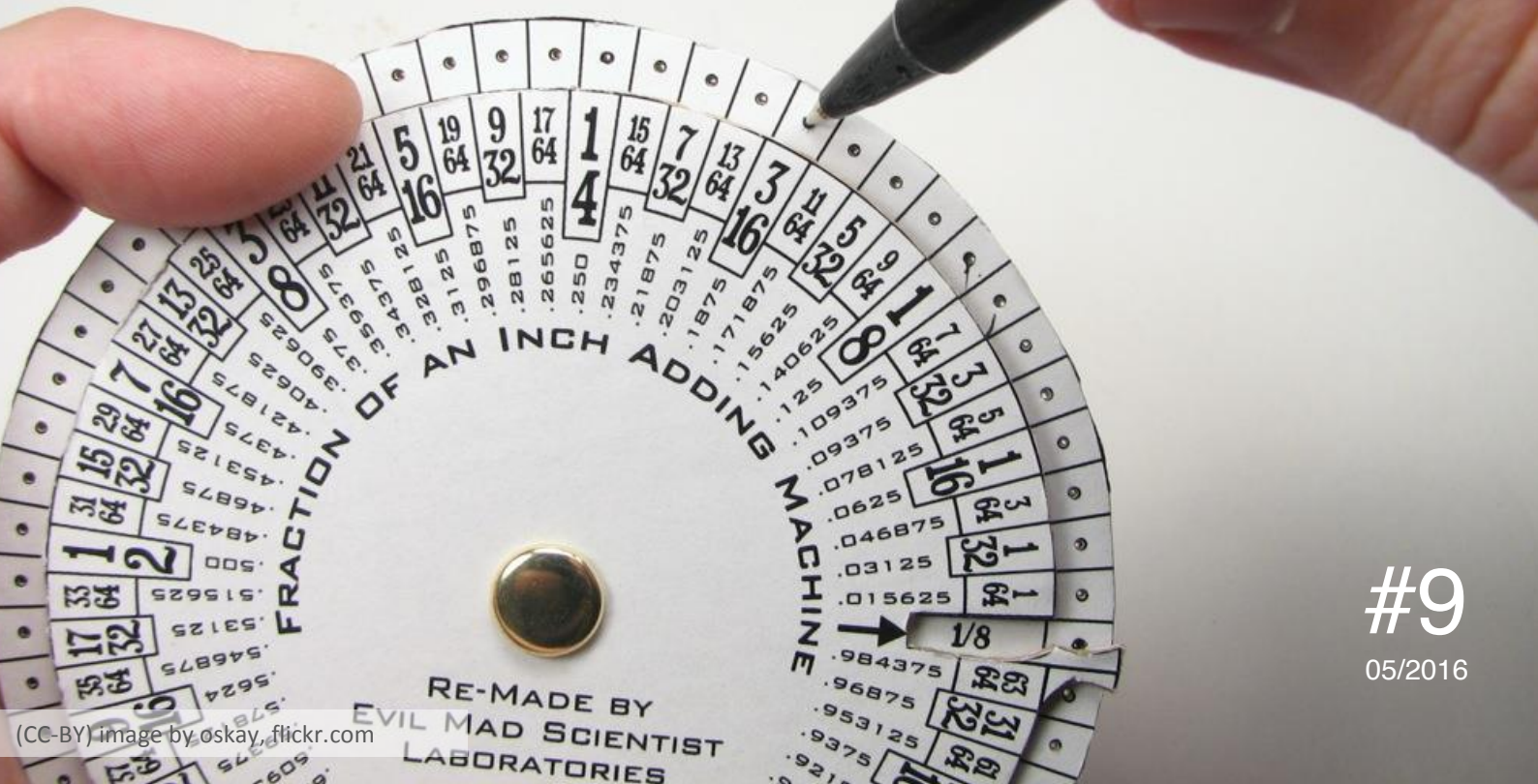




#9  
05/2016

# Zugriffszahlen und Reichweiten von Online-Medien analysieren

Relevante Webseiten filtern, Leser pro Artikel und die erzielte Gesamtreichweite ausrechnen.



„Zugriffe, Shares, Visits, Page-Impressions oder doch Leser pro Artikel? Echobot räumt auf mit den unklaren Erfolgsmetriken Ihrer Online-PR.“

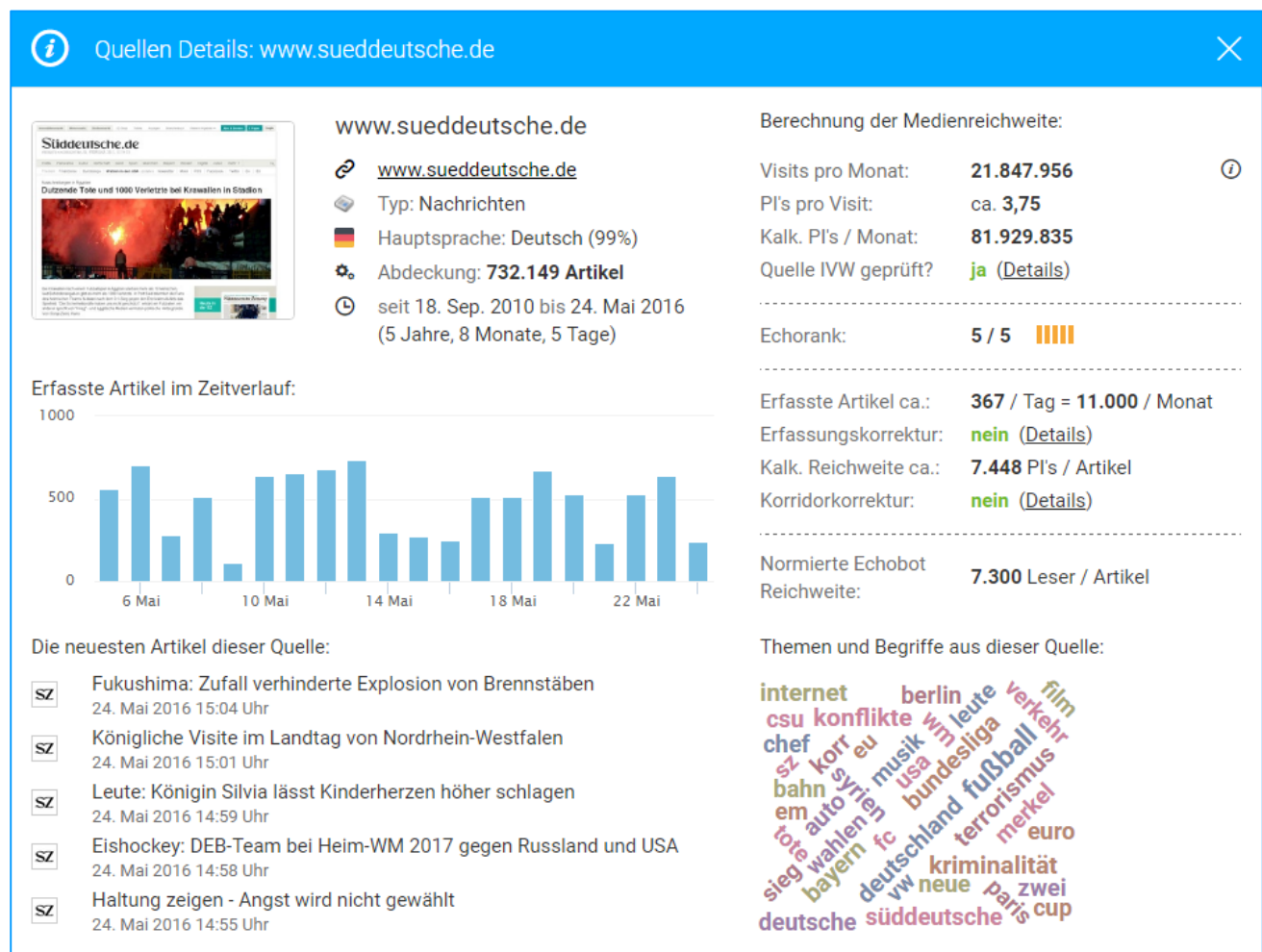
**Bastian Karweg** (Geschäftsführer Echobot)

# Online-Erfolge messbar machen

Wenn Ihr Unternehmen, Ihre Produkte und Marken auf Nachrichtenseiten, in Fachportalen, Foren oder Blogs erwähnt werden, so ist es oft von großem Interesse, einschätzen zu können, wie viel Reichweite die entsprechende Webseite erzielt hat oder besser noch, wie viele Leser der konkrete Artikel erreicht hat.

Diese Angaben zu Reichweite oder Leserzahl werden dann gerne genutzt, um sie im eigenen Pressespiegel aufzuführen, um im Krisenfall das Ausmaß des Problems zu quantifizieren oder um daraus den Mediawert der eigenen Pressearbeit zu berechnen.

Leider ist die Recherche und Abschätzung von Zugriffszahlen von Webportalen oft schwierig, da deren Reichweitenangaben stark variieren, sich Messmethoden unterscheiden oder von den Betreibern der Portale sehr unterschiedlich erhobene Mediadaten kommuniziert werden. Dieses Whitepaper schafft Abhilfe.



Unser Ergebnis: Die Echobot-Medienreichweite einer Online-Quelle

## Worum geht es in diesem Fachartikel?

Bei Echobot beschäftigen wir uns schon sehr lange mit der Thematik Online-Reichweiten, da unser Anspruch ist, zu jedem Treffer unserer Online-Medienbeobachtung zwei Angaben mit zu liefern:

1. **Reichweite der Webseite/des Portals** und
2. **geschätzte Leserzahl eines durchschnittlichen Artikels dieser Webseite.**

Basierend auf diesen Angaben berechnen wir zu jeder Online-Quelle den sogenannten:

### 3. **EchoRank**

Der „EchoRank“ ist dabei ähnlich wie der Google PageRank aufgebaut. Ein ganzzahliger Wert von 0 bis 5 mit dessen Hilfe Sie auf einen Blick erkennen, ob die Quelle/das Portal viele Leser hat oder nicht. In diesem Whitepaper möchten wir Ihnen unsere Überlegungen, Berechnungsmethoden und einige Hintergründe darlegen, die in die Kalkulation der oben genannten Werte mit eingeflossen sind. So können Sie besser nachvollziehen, wie sich die Reichweitenwerte zusammensetzen, und was Sie daraus ableiten können. Vorweg schon einmal so viel:

#### HINWEIS

*Wir haben es hier nicht mit einer exakten Wissenschaft zu tun. Es geht uns vor allem darum, die Größenordnungen abschätzen zu können – also ob ein Artikel 100 oder 1.000 Leser hatte – und so wichtige von unwichtigen Portalen zu trennen. Exakte Messwerte sind bei der großen Anzahl Quellen, gerade auch in der internationalen Betrachtung nicht möglich, wie noch gezeigt wird.*

## Worum geht es nicht?

- An dieser Stelle sei darauf hingewiesen, dass es in diesem Artikel nicht um die Berechnung von **Social-Media-Reichweiten** geht. Auch dort gibt es sehr unterschiedliche Werte: Facebook-Likes, Twitter-Follower und Retweets, YouTube-Views und Subscriber bis hin zu Google+-Circles. Diese werden wir zu einem späteren Zeitpunkt in einem weiteren Fachartikel behandeln.
- Sofern Sie unter Social Media jedoch auch **Blogs und Foren** betrachten, kann dieses Dokument hierfür sehr hilfreich sein. Da Forenbeiträge und Blogposts nicht in geschlossenen Netzwerken stehen, sondern in der Regel öffentlich über das Web auffindbar sind, können Reichweiten dieser genauso behandelt und berechnet werden wie bei normalen Nachrichtenportale.
- Die richtige Berechnung des **Mediawert oder Medienäquivalenzwert** ist ebenfalls nicht Teil dieser Ausführungen, da es auch dabei sehr unterschiedliche Herangehensweisen gibt. Lesen Sie dazu bitte den verlinkten [Fachartikel zum Medienäquivalenzwert](#).
- Als letzter Hinweis sei erwähnt, dass unabhängig von der Angabe der Reichweite für jeden Pressespiegel gerade im Online-Bereich auch eine **rechtliche Betrachtung** vorgenommen werden sollte. Auch hierzu haben wir ein entsprechendes [Whitepaper als Orientierungshilfe](#) erstellt.

# Reichweitenberechnung bei Print-Medien

Schon seit mehreren Jahrzehnten werden von Unternehmen und Presseabteilungen sogenannte Pressespiegel der erzielten Veröffentlichungen erstellt. Hierbei werden in der Regel Print-Ausschnitte der Publikationen erstellt und in einer Mappe gesammelt.

Schon bei Zeitungen und Zeitschriften war die Frage der erzielten Reichweite immer schwierig zu beantworten. Verlage unterscheiden zwischen der „gedruckten Auflage“, der „verbreiteten Auflage“ und der „verkauften Auflage“. Oftmals werden nämlich auch Tausende Exemplare kostenfrei in Flugzeugen oder Zügen verteilt. Auch wird regelmäßig von einer zusätzlich „erreichten Zielgruppe“ gesprochen, da man davon ausgeht, dass jedes Exemplar z.B. durch Haushalte, Bibliotheken, Cafés, Lesezirkel, etc. von mehr als einer Person gelesen wird.

Solche Reichweiten-Metriken beziehen sich im Print-Bereich immer auf die ganze Ausgabe und eben nicht auf den einzelnen Artikel im Medium. Sicherlich kann man einfach beweisen, dass ein Beitrag auf der Titelseite, der Seite 3 oder der Falt-Mitte häufiger gelesen wird als andere – aber wie steht es denn genau mit einem kleinen Ausschnitt auf Seite 17 in der unteren linken Ecke? Die wenigsten lesen eine Zeitung komplett von vorne bis hinten durch. Daher wurden für solche Bewertungen auch hier schon immer Schätzwerte eingesetzt.

# Metriken der Online-Reichweiten verstehen

Man könnte denken, dass es im digitalen Zeitalter und mit den Tracking-Möglichkeiten moderner Webseiten einfacher geworden ist, die Reichweiten zu messen und zu bewerten, das stimmt aber leider nur bedingt: Bevor man nämlich überhaupt daran denken kann, die Leser eines einzelnen Artikels zu ermitteln, muss man wie im Print-Bereich auch die „Brutto-Reichweite“ der Webseite oder des Portals erheben. Doch schon hier gibt es sehr unterschiedliche Messmethoden und Kennzahlen, die man verstehen sollte.

Zugriffs-Tracking-Systeme unterscheiden in der Regel zwischen drei allgemein gültigen Kenngrößen, deren Bezeichnungen jedoch je nach System und Hersteller variieren:

### 1) Eindeutige Besucher/Unique-Users/Unique Visitors

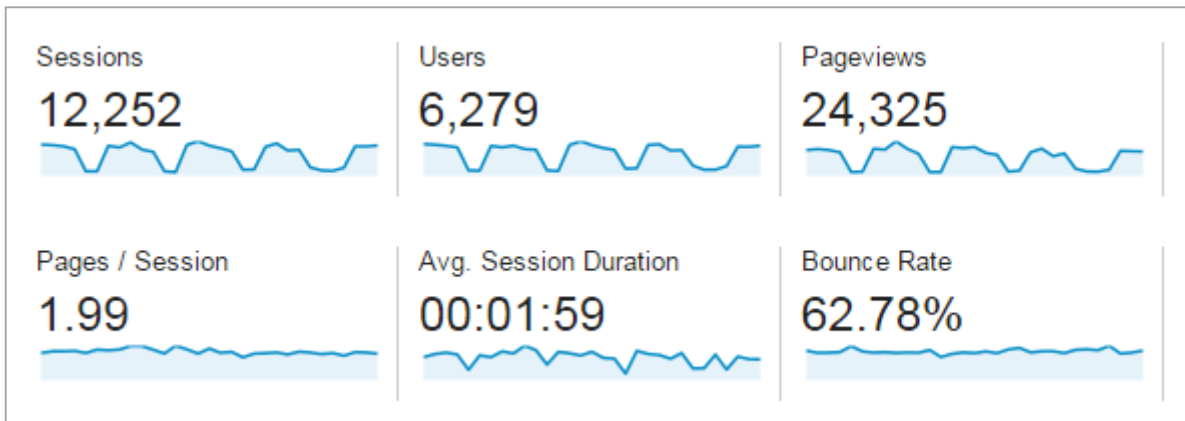
⇒ Wie viele Personen waren im betrachteten Zeitbereich (z.B. ein Monat) auf der Seite?

### 2) Visits/Sessions/Sitzungen

⇒ Wie oft waren diese Personen in Summe auf Ihrer Seite unterwegs?  
(z.B. einmal am Dienstagvormittag, dann wieder am Donnerstagnachmittag = 2 Sessions)

### 3) Page- Impressions = Page-Views = Seitenaufrufe

⇒ Wie viele Einzelseiten wurden während der Sitzungen in Summe angeklickt?



*Beispiel-Ansicht eines Web-Tracking-Tools (hier: Google Analytics)*

#### Konfigurationsfrage:

Je nachdem, welches Tracking-System Sie einsetzen, und wie dieses eingestellt ist, können hier sehr unterschiedliche Werte herauskommen. Beispielhaft im Folgenden einige Konfigurationsfragen:

- Wie lange ist eine Session gültig, bevor eine neue beginnt?
  - ⇒ Zählt mein Seitenaufruf um 14 Uhr zu einer neuen Sitzung, wenn ich morgens um 10 Uhr die Seite schon einmal besucht hatte?
- Zählt ein Besucher als neuer Besucher, wenn er den Browser wechselt oder Cookies blockiert?
  - ⇒ Die Zuordnung erfolgt in der Regel über Cookies. Sind diese blockiert oder wechselt der Nutzer den Browser, so beginnt jedes Mal eine neue Sitzung.
- Werden Aufrufe von Suchmaschinen und Crawlern mitgezählt?
  - ⇒ Schon über die Hälfte aller Webseiten-Aufrufe erfolgt durch automatische Crawler und Spider (z.B. auch von Echobot). Will ein Seitenbetreiber höhere Werte ausweisen, so kann er diese Abrufe von Suchmaschinen recht einfach als „Besucher“ mitzählen.

Dies sind nur drei Beispiele, die verdeutlichen sollen, dass sich erhobene Reichweitenangaben von Seite zu Seite sehr stark unterscheiden können. Mit den zunehmenden Abrufen über mobile Geräte und Apps wird das Thema noch einmal komplizierter.

#### ZWISCHEN- ERGEBNIS

*Es ist also sehr wichtig zu verstehen, dass man absolute Reichweitenangaben unterschiedlicher Quellen nicht einfach miteinander vergleichen kann. Viel mehr ist darauf zu achten, dass sie mit derselben Messmethode erhoben wurden und dann relativ zueinander betrachtet werden.*

# Messmethoden und -Systeme verstehen

Im Wesentlichen gibt es drei verschiedene Möglichkeiten, wie man an die Zugriffszahlen und die Reichweiten von Online-Portalen herankommt. Entweder über „exakte Messung“ oder „geschätzte Hochrechnung“ oder „Korrelation und Zurückrechnen“. So wird das in der Praxis umgesetzt:

- **Genaues Tracking über Zählpixel/-Scripts**

Die exakte Messung kann nur von den Webseiten-Betreibern selbst vorgenommen werden. Hierzu wird ein Zählpixel oder Scriptcode auf der Webseite integriert, der bei jedem Ladevorgang mit aufgerufen wird. Auch hier gibt es große Unterschiede je nach Software, Anbieter oder der konkreten Stelle der Einbindung des Zählsystems. Gängige Systeme sind z.B.:

➔ Google-Analytics, Piwik, Econda, Webtrakk, INFOnline, etc.

- **Hochrechnung anhand großer Nutzerpanels – z.B. mithilfe beliebter Browser-Plugins**

Wie bei der Hochrechnungen der Wahlergebnisse zur Bundestagswahl, kann man die Reichweite eines Portals sehr genau schätzen, wenn man nur eine genügend große Stichprobe von Personen analysiert. Hierzu wird von bestimmten Anbietern mittels Browser-Plugins deren Surf-Verhalten erfasst und anonymisiert analysiert.

Beispiel: Besuchen die beobachteten Personen 1.000 Mal [www.google.de](http://www.google.de) – im selben Zeitraum aber nur 100 Mal [www.web.de](http://www.web.de) – so kann man ableiten, dass Google ca. 10 Mal mehr Zugriffe hat als WEB.de.

Solche [Tracking-Software findet sich](#) in vielen populären Browser-Plugins, z.B. in der Erweiterung „Awesome Screenshot“. Bei deren Installation stimmt der Nutzer der anonymisierten Auswertung seines Surfverhaltens zu. Dank Millionen von aktiven Nutzern, sind solche Daten sehr akkurat.

➔ Gängige Anbieter solcher Daten sind z.B.: [www.similarweb.com](http://www.similarweb.com), [www.alexa.com](http://www.alexa.com)

## ZWISCHEN- ERGEBNIS

*Für unsere Analysen bei Echobot verlassen wir uns auf Daten von SimilarWeb. Unserer Meinung nach bietet dieser Anbieter die beste weltweite Abdeckung an Portalreichweiten. Zwar erhält man keine exakte Messung, durch die einfache und einheitliche Messmethode sind die Daten aber unverfälscht.*

- **Sharing Werte über Social Media**

Eine andere Methode ist, die Verbreitung von Inhalten über gängige Social-Media-Kanäle zu untersuchen und daraus Rückschlüsse auf die Portalreichweite zu ziehen. Zum einen misst man so Inhalte, die auch tatsächlich von Nutzern angesehen wurden, zum anderen erhält man Angaben pro geteiltem Artikel. Leider eignen sich diese Metriken nicht, für den Reichweitenvergleich, da das tatsächliche Sharing-Verhalten der Nutzer sehr stark davon abhängt, wie prominent die Sharing-Funktion integriert wurde und wie „viral“ die Überschrift oder das ausgewählte Bild ist.

➔ Anbieter, die Social-Media-Shares messen sind z.B.: XQVI, sharetally.co, socialbakers.com

# IVW, AGOF und INFOnline

Natürlich können Sie bei den einzelnen Portalen auch immer nach „Mediadaten“ suchen. Häufig werden Sie hier gerade bei seriösen Angeboten auch fündig. In den Informationen zu den Banner- und Anzeigenpreisen werden dort sogenannte TKPs (Tausenderkontaktpreise) ausgewiesen. Um einschätzen zu können, wie viel ein Banner im Monat kostet, ist die Angabe der Zugriffe nötig, die oft im direkten Umfeld zu finden ist.

Da das einzelne Abfragen der Mediadaten sehr mühsam ist, gibt es gerade in Deutschland die Angebote von IVW und AGOF, die sich als zentrale Instanz diesen Erhebungen verschrieben haben. Aber auch hier gibt es einiges zu beachten, wie wir im Folgenden näher beleuchten wollen.

Technisch umgesetzt werden die Reichweiten-Zählungen von der INFOnline GmbH in Bonn ([www.infonline.de](http://www.infonline.de)). Diese arbeitet mit der „Informationsgemeinschaft zur Feststellung der Verbreitung von Werbeträgern (e.V.)“ oder kurz „IVW“ zusammen, die regelmäßig die korrekte Integration der Zählskripte und deren ordnungsgemäße Kategorisierung und Verwendung überprüft. Daraus ergeben sich dann die monatlich aktualisierten Messergebnisse in der sogenannten „[Online-Ausweisung](#)“ der IVW.

Die „Arbeitsgemeinschaft Onlineforschung“ oder kurz „AGOF“ wiederum ist ein Zusammenschluss von Online-Vermarktern, die Studien zur Bekanntheit und Wahrnehmung von Marken und Webseiten durchführen. Durch die sogenannten „[internet facts](#)“ wurden vor allem auch Standards etabliert, die für die Einstellung der Messtools herangezogen werden. Neben der rein technischen Messung der tatsächlichen Aufrufe, bezieht die AGOF auch über Online-Umfragen weitere Internetnutzer in ihre Auswertung mit ein und geht mit ihren telefonischen Basisbefragungen auch der Frage nach, wie hoch die Bekanntheit von Marken und Webseiten in der Gesamtbevölkerung ist, die nicht direkt über das Internet ansprechbar waren.

## Wichtige Überlegungen:

IVW und AGOF können nur eingeschränkt zur Einschätzung der Online-Reichweite verwendet werden:

### QUELLEN-UMFANG

*Der größte Vorteil für den Webseiten-Betreiber ist, dass die Reichweitenangaben seiner Portale durch die IVW unabhängig geprüft und zertifiziert werden. Demnach können sich seine Werbekunden darauf verlassen, dass die bezahlte Reichweite auch tatsächlich vorhanden ist.*

*Diese Prüfung ist allerdings nicht ganz günstig, wodurch der Umfang der Portale, die sich das leisten wollen, natürlich eingeschränkt ist. Aktuell (Stand Geschäftsbericht 2015/2016) befinden sich in der IVW-Ausweisung **1.085 Online-Angebote**.*

Mit Echobot Monitoring beobachten wir aktuell ca. 25.000 deutschsprachige Online-Quellen (zusätzlich noch deutlich mehr internationale Portale und über 600.000 Firmenwebseiten). Daraus ergibt sich also, dass wir über die IVW nur zu einem sehr kleinen Teil der Quellenbasis überhaupt Angaben erhalten können.

Und das ist nicht das einzige Problem:

Auch muss man verstehen, dass das Ziel von IVW und AGOF die Angabe der Verbreitung von **WERBETRÄGERN** ist. Diese sind aber eben nicht mit Webseiten oder Domains gleichzusetzen. Oft besteht ein „Werbeträger“ sogar aus vielen Portalen, die zusammengefasst werden.

### Beispiel aus der aktuellen IVW-Online-Ausweisung

Alle mit „A“ gekennzeichneten Einträge fassen mehrere Domains/Webseiten zusammen:

|                                |   |            |            |
|--------------------------------|---|------------|------------|
| ▶ General-Anzeiger-Rhauderfehn | A | 141.786    | 141.786    |
| ▶ GEO.de                       | A | 1.902.862  | 1.399.956  |
| ▶ germanblogs.de               |   | 123.376    | 123.376    |
| ▼ GesünderNet                  | A | 10.723.367 | 10.723.367 |
| GesünderNet (Online)           |   |            |            |
| ▶ gesundheit.de                |   | 7.224.941  | 2.552.696  |
| ▶ Gießener Allgemeine          | A | 468.373    | 345.880    |
| ▶ Giga-Music.de                |   | 32.946     | 32.946     |

#### Angebotsbestandteilsliste

Einzelbestandteile des Digital-Angebots 'GesünderNet' und deren Anteil an der Gesamtanzahl der Pls im Monat April 2016

Die Anteile beziehen sich auf die Gesamtzahl der Pls aus der Online-Ausweisung sowie der Mobile-Ausweisung nach Kat 2.0.

Namensgebendes Angebot:  
gesuendernet.de: 0,39 %



Weitere Angebotsbestandteile:

dr-gumpert.de: 18,12 %  
sanego.de: 17,50 %  
bmi-rechner.net: 12,86 %  
klinikbewertungen.de: 9,00 %  
fettrechner.de: 7,64 %  
wunderkessel.de: 5,39 %  
psychic.de: 5,05 %  
kalorientabelle.net: 3,82 %  
medizinfo.de: 3,73 %

Vollständige Liste der Angebotsbestandteile

Fenster schließen

Zwar könnte man jetzt versuchen, diese sogenannten „Locallisten“ auch noch vollständig auszuwerten, aber auch diese werden bei der IVW nicht komplett mit den jeweiligen Prozentangaben ausgeliefert.

Speziell auf den Echobot Monitoring Use-Case bezogen, ergibt sich sogar noch ein weiteres Problem:

#### TRAFFIC-HERKUNFT

*Die Reichweite eines von der IVW geprüften Portals verteilt sich auf verschiedene Seitenbereiche. Diese Klicks müssen also nicht nur von redaktionellen Artikeln stammen, sondern können auch durch Bildershows, Kreuzworträtsel, E-Mail-Funktionen oder Dating-Bereichen der Seite generiert werden. Ebenso werden auch von Nutzern Inhalte erstellt, die nicht immer als „Artikel“ zu betrachten sind (z.B. Rezepte bei Chefkoch.de). Zwar gibt es in den IVW-Details eine Aufschlüsselung nach Inhaltstyp und nach Herkunft der Reichweite, diese Einstellung ist aber bei vielen Portalen leider oft nicht sehr aussagekräftig.*

*Beispiel-Ansicht Traffic-Details eines Portals aus der Online-Ausweisung der IVW*

| Erzeuger                                      |            |            |
|-----------------------------------------------|------------|------------|
| Redaktion                                     | 11.545.136 | 29.538.835 |
| User                                          | 198.928    | 1.370.103  |
| Unbekannt                                     | 8.824.244  | 78.579.860 |
| Homepage                                      |            |            |
| Homepage des Angebots                         | 6.489.591  | 10.682.130 |
| Keine Homepage                                | 10.465.045 | 98.806.668 |
| Homepage der Fremddomains bei Multi-Angeboten | ---        | ---        |

Zusätzlich wird das Ganze noch dadurch verkompliziert, dass immer mehr Reichweite auch durch Apps, mobilen Traffic und ePaper-Angebote erhoben wird. Auch haben heutzutage viele Nutzer in Ihrem Browser Plugins installiert, um Tracking-Skripte oder Anzeigen zu blockieren. Im Fall der Verlage wird gelegentlich sogar zwischen bezahltem Traffic (hinter einer Pay-Wall) und öffentlichem Traffic unterschieden.

#### Bitte nicht falsch verstehen!

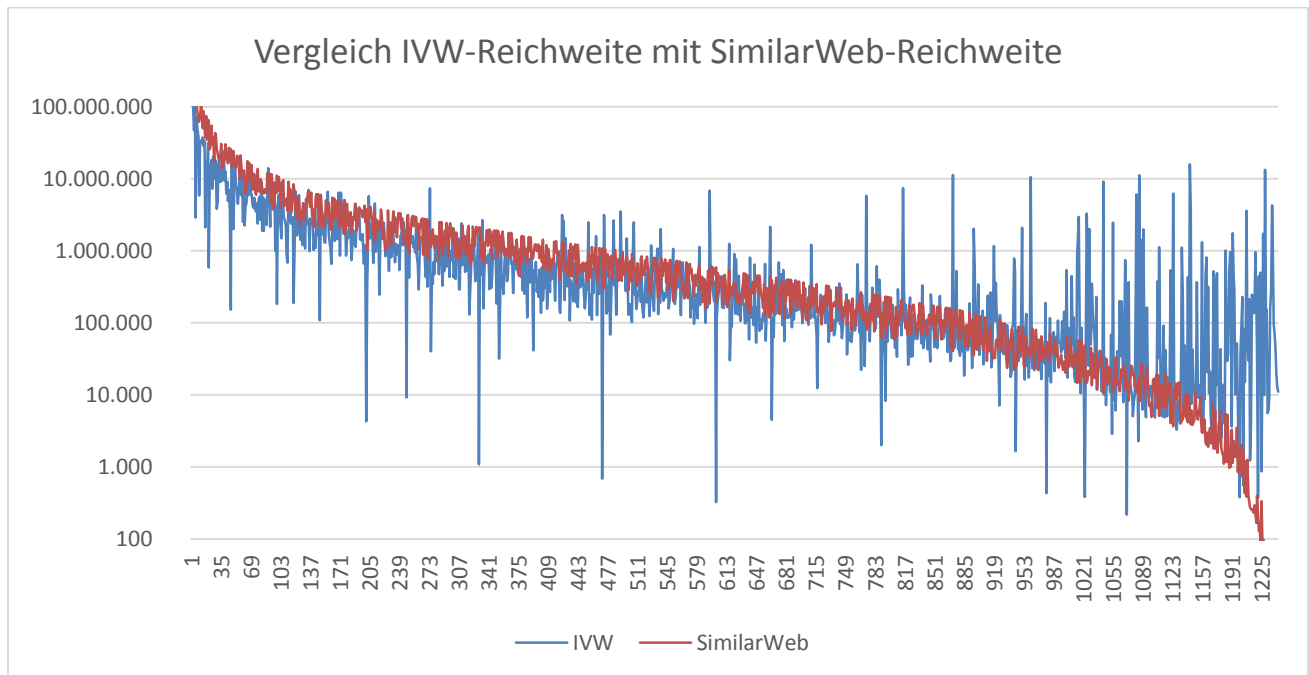
Für den Anwendungsfall, Reichweiten von Online-Werbeträgern und Vermarktungsgemeinschaften zu messen und daraus z.B. Bannerpreise zu bestimmen, funktioniert die IVW ganz hervorragend. Auch liefert sie einen tollen Benchmark und eine große Sicherheit und Referenz was die Portalgröße betrifft.

Um jedoch die tatsächliche Reichweite einer Domain oder deren Artikel zu erheben und daraus die erzielte PR-Reichweite oder einen Medienäquivalenzwert auszurechnen, gibt es bessere Möglichkeiten. Gerade auch im Hinblick auf internationale Quellen, kommt man mit rein deutschen IVW-Daten nicht sehr weit.

## Reichweiten-Vergleich IVW und SimilarWeb

Wie erwähnt, liefern die IVW-Zahlen trotz einzelner Abweichungen in der Hauptdomain, einen sehr guten Benchmark, um die Güte anderer Messverfahren zu prüfen. Dies haben wir mit dem von uns ausgewählten Anbieter SimilarWeb auch getan. Zu insgesamt 1.225 Portalen konnten wir in den letzten Monaten Vergleichswerte erheben. Zur Erklärung:

- Portale nach Reichweite absteigend (von links nach rechts) abgetragen
- Roter Wert = SimilarWeb-Reichweite der Domain
- Blauer Wert = IVW-Reichweite der Domain



### Erkenntnisse:

1. Generell korrelieren die Reichweiten-Angaben sehr gut. Die Ausreißer nach oben oder unten sind vor allem auf das oben beschriebene Problem der Multi-Domain-Ausweisungen zurückzuführen.
2. Bei größeren Seiten (Nr. 1 bis 1.000 von links auf der X-Achse mit mehr als 100.000 Zugriffen) ist die Übereinstimmung sehr deutlich. Hier scheint das IVW-Messverfahren eher noch einige Zugriffe herauszufiltern.
3. Bei kleinen Seiten (Nr. 1.000 bis 1.225, rechts auf der X-Achse mit weniger als 100.000 Zugriffen) sind oft sehr große Abweichungen festzustellen. Dies ist vor allem auf Domain-Weiterleitungen, Angebots-Konsolidierungen oder Subdomains zurückzuführen.

### Beispiel:

Die Portale [www.magnus.de](http://www.magnus.de), [www.connected-home.de](http://www.connected-home.de) und [www.color-foto.de](http://www.color-foto.de) werden seit einiger Zeit alle auf [www.pc-magazin.de](http://www.pc-magazin.de) umgeleitet.

Auch hier sind die Angaben von SimilarWeb verlässlich, da der Traffic der Zielseite gemessen wird.

## Fazit zur Brutto-Reichweite

Egal ob man nun Unique-User, Visits oder Page-Impressions misst, absolut exakte Werte eines Portals zu erhalten ist quasi unmöglich. Eine relative Vergleichbarkeit der Portalreichweiten wird dann hergestellt, wenn man dieselbe Messmethode anwendet. SimilarWeb liefert mit seiner Hochrechnung ein Vorgehen, das diese Vergleichbarkeit ermöglicht und hat zusätzlich den Vorteil, dass die Angaben genauer werden, je bekannter die Webseite ist. Dies lässt sich auch gut an der nach oben immer besser werdenden Korrelation mit den IVW-Daten ableiten. Da SimilarWeb die Daten auf Messungen im Browser basiert, werden Zugriffe von Suchmaschinen und Crawlern schon vom Systemaufbau her nicht berücksichtigt.

## EchoRank-Berechnung

Der EchoRank bietet als einfache Metrik eine konsolidierte Zusammenfassung der Reichweite eines Portals, damit Sie diese auf einen Blick erkennen und mit den Funktionen der Echobot-Medienbeobachtung leicht filtern können. Dabei stellt der EchoRank auf die Page-Impressions ab, einer Angabe, die sowohl die Besucher einer Seite mit einbezieht, als auch die Frage, wie viele Inhalte auf der Seite abgerufen werden (PIs/Visit).

Als Korrektiv gehen wir für unsere Berechnung dabei jedoch von maximal 5 PIs pro Aufruf aus, um Klickbetrug zu vermeiden, der z.B. entsteht, wenn vom Seitenbetreiber bei jedem Blättern einer Bildershow ein neuer Abruf produziert wird.

Als weitere Voraussetzung sehen wir an, dass die Seite auch redaktionell „aktiv“ ist. Dies machen wir an der Anzahl ihrer in einem Monat veröffentlichten Artikel fest. Ein schönes Beispiel ist [www.doodle.com](http://www.doodle.com): Der Terminvereinbarungsdienst erfreut sich großer Beliebtheit mit vielen Zugriffen, hat jedoch redaktionell nur sehr wenige Inhalte (ca. 1 Blog-Artikel pro Monat). Da hier die minimale Artikel-Zahl für einen höheren EchoRank nicht erreicht wurde, verbleibt die Seite trotz sehr hoher Zugriffe auf EchoRank „1“.

### Tabellarische Übersicht:

| EchoRank | Häufigkeit | Min. PIs pro Monat | Min. Artikel pro Monat | Möglicher Korridor für Leser/Artikel | Symbol                                                                                |
|----------|------------|--------------------|------------------------|--------------------------------------|---------------------------------------------------------------------------------------|
| <b>0</b> | 32%        | -                  | -                      | 1 bis 10                             |  |
| <b>1</b> | 32%        | 500                | 5                      | 1 bis 100                            |  |
| <b>2</b> | 20%        | 5.000              | 10                     | 10 bis 500                           |  |
| <b>3</b> | 10%        | 50.000             | 100                    | 100 bis 1.000                        |  |
| <b>4</b> | 4%         | 500.000            | 250                    | 250 bis 5.000                        |  |
| <b>5</b> | 2%         | 5.000.000          | 500                    | 500 bis 25.000                       |  |

Durch Anwendung dieser Berechnungsmethodik ergibt sich eine sehr gesunde degressive Verteilung, bei der jeder höhere Rank nur etwa halb so häufig vorkommt, wie der EchoRank zuvor. Seiten für die keine Werte erhoben werden konnten, oder die nicht mehr aktiv sind, rutschen automatisch auf EchoRank „0“.

Der EchoRank wird jeden Monat für alle Portale neu berechnet, alte Werte werden dabei überschrieben.

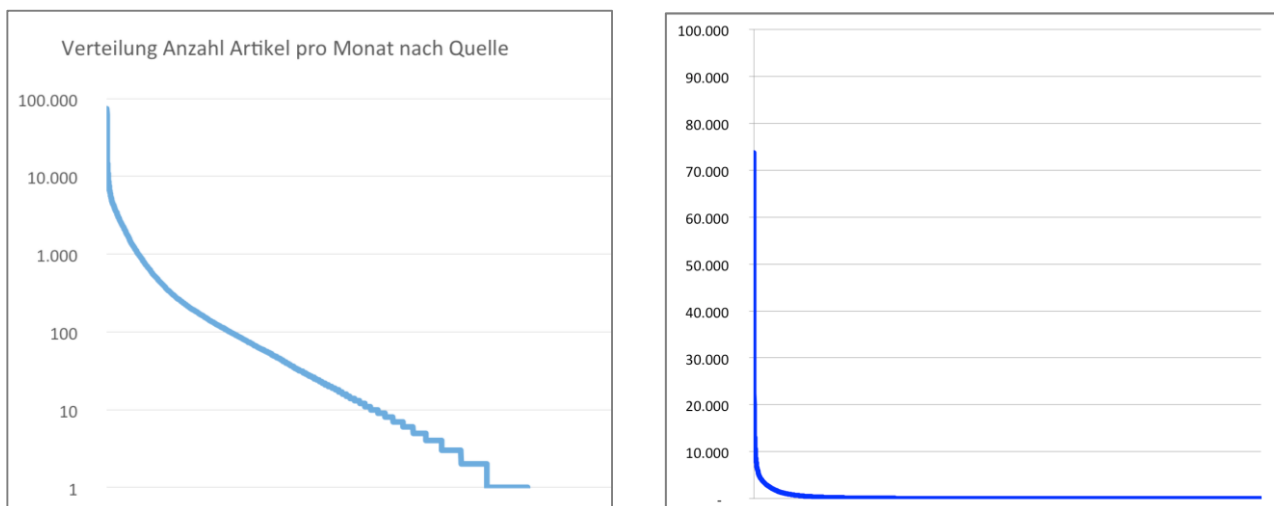
## Ermitteln der Anzahl „Leser pro Artikel“

Eine große Kunst ist es, die Anzahl der „Leser pro Artikel“ abzuschätzen. Denn gerade dieser Wert ist für unseren Anwendungsfall der PR-Reichweiten und Mediawert-Analyse von großer Bedeutung.

Natürlich ist es wichtig zu wissen, wenn ein Artikel über die eigene Firma auf einem großen Portal wie [www.web.de](http://www.web.de) erschienen ist. Auch die Tatsache, dass WEB.de ca. 85 Mio. Visits und über 400 Mio. Page-Impressions erreicht, beeindruckt sicher zunächst. Letztlich sind diese Werte jedoch unerheblich, wenn es darum geht, wie viele Personen den Artikel tatsächlich gelesen haben. Denn nicht jeder, der über WEB.de seine E-Mails abrufen, hat auch deren redaktionellen Bereich beachtet.

Umgekehrt kann aber auch ein kleineres Fachportal wie LaborPraxis vom Vogel-Verlag mit einer moderaten Reichweite von ca. 60.000 Page-Impressions pro Monat auch Hunderte oder gar Tausende Leser produzieren, die von ihrer Affinität zum Thema sicherlich noch einmal deutlich spannender sind.

Schauen wir uns also zunächst einmal die Verteilung der veröffentlichten Artikel nach Quelle während der letzten 30 Tage vor dieser Analyse an:



*Rechts die normale Skalierung, links logarithmisch für bessere Lesbarkeit.*

Wie man an der typischen „Power-Law“-Verteilung schön erkennen kann, sorgen relativ wenige aktive Webseiten, für den Löwenanteil der Artikel. Gerade Foren, Nachrichtenagenturen oder Presseportale schaffen es, teilweise Tausende Meldungen pro Monat zu veröffentlichen.

Beispiele für sehr hochfrequente Seiten mit über 20.000 Artikeln pro Monat sind z.B. [www.finanznachrichten.net](http://www.finanznachrichten.net), [www.ad-hoc-news.de](http://www.ad-hoc-news.de), [www.gutefrage.net](http://www.gutefrage.net) oder [www.stackoverflow.com](http://www.stackoverflow.com). Spitzenreiter nach Artikelanzahl bei den redaktionell geführten Medien aus Deutschland sind [www.derwesten.de](http://www.derwesten.de), [www.welt.de](http://www.welt.de) und [www.rp-online.de](http://www.rp-online.de) jeweils mit ca. 17.000 Artikeln im Monat.

Eine nach unseren Analysen „typische“ redaktionelle Webseite veröffentlicht etwa zwischen 250 und 2.500 Meldungen in einem Monat, wobei auch das natürlich stark von der thematischen Ausrichtung abhängt.

# Unschärfen der Berechnung

Unserem Ziel die „durchschnittliche Leserzahl pro Artikel“ kommen wir grob ein gutes Stück näher, wenn wir die vom Portal erzielten Page-Impressions durch die Anzahl der im Monat publizierten Artikel teilen:

$$\text{Naive Leserzahl} = \frac{\text{PageImpressions pro Monat}}{\text{Neue Artikel pro Monat}}$$

Leider gibt es bei diesem naiven Ansatz einige Unschärfen, die man berücksichtigen sollte:

- **Unschärfe 1: Verteilung der Zugriffe auf die Artikel einer Seite**

Nicht jeder Artikel eines Portals erhält gleich viel Reichweite. Ist der Artikel auf der Startseite verlinkt oder erhält Empfehlungen aus Sozialen Medien, so steigen auch seine Zugriffe. Ist der Artikel umgekehrt auf einer regionalen Unterseite vergraben, die erst über drei Klicks zu erreichen ist, so werden ihn weniger Leute finden.

- **Unschärfe 2: Leser ist nicht gleich Leser**

Ob ein Leser einen Artikel nur aus Versehen anklickt, diesen überfliegt oder sich sehr lange damit beschäftigt, fließt in diese Verteilung nicht mit ein. Ebenso wenig, ob der Leser ein Fachentscheider (z.B. auf einem Medizinportal) oder ein sportinteressierter Konsument ist. Auch daher nutzen wir gerne die Daten von SimilarWeb, da hier auch die „Verweildauer auf der Seite“ und die „Absprungrate“ (z.B. zurück zu einer Suchmaschine) mit einbezogen werden kann.

- **Unschärfe 3: Reichweite des Archivs und von anderen Seitenbereichen**

Viele Webseiten werden heute sehr stark für Suchmaschinen optimiert. Ziel ist dabei, dass einmal erstellter Inhalt über das stetig wachsende Archiv für Google auffindbar ist und dann bei Suchabfragen nach dem entsprechenden Schlagwort aufgerufen wird. Bei manchen Seiten ist der Anteil der Aufrufe über Suchmaschinen sehr hoch, wodurch auch ältere Inhalte in der Berechnung mit berücksichtigt werden müssten. Bei Portalen mit schlechter Google-Optimierung entfällt der Traffic hingegen eher auf neuere Inhalte.

Je nach Typ der Seite, ist das Verhältnis zwischen aktuellen Inhalten und Archiv-Inhalten auch von Natur aus unterschiedlich:

|                      |   |                                                                 |
|----------------------|---|-----------------------------------------------------------------|
| ○ Firmenseiten       | - | hoher Anteil statischer Inhalte, aktiver Newsbereich/Blog       |
| ○ Presseportale      | - | sehr aktiver aktueller Bereich, Archiv oft wenig frequentiert   |
| ○ Nachrichtenportale | - | abhängig vom Aufbau der Seite, aktuelle Inhalte überwiegen      |
| ○ Finanzseiten       | - | Börsenkurse stehen im Fokus, Ad-hoc-Meldungen dazu aktiv        |
| ○ Blogs              | - | großer Fokus auf aktuellem Inhalt, oft aber gutes SEO           |
| ○ Portalseiten       | - | Reichweite zu großen Teilen über E-Mail-Funktion, etc.          |
| ○ Presseagenturen    | - | extrem fokussiert auf Tagesgeschehen, Archive kaum aktiv        |
| ○ Fachportale        | - | Hauptanteil der Reichweite entfällt oft auf Rezepte, Jobs, etc. |
| ○ Foren              | - | aktive Artikel, die durch Nutzerantworten ergänzt werden        |

- **Unschärfe 4: technische Umsetzung**

Wie oben bereits dargestellt, spielen je nach Aufbau der Seite auch noch die technische Umsetzung der Seite (z.B. Abrufe von Bildershow) oder weitere Angebote auf der Domain eine Rolle wie z.B.: Dating-Bereiche, Kleinanzeigen, Kreuzworträtsel, Horoskope, Rezepte, E-Mail Angebot, etc.

Auch technische Umleitungen ([www.msn.de](http://www.msn.de) → [www.msn.com/de-de](http://www.msn.com/de-de)) oder Subdomains (z.B. bei [www.yahoo.de](http://www.yahoo.de) → <http://de.yahoo.com>) können zu Verfälschungen der ermittelten Reichweite führen, wenn man nicht genau arbeitet.

- **Unschärfe 5: Mobile Webseiten, Social Media, RSS-Feeds, Newsletter und Apps**

Letztlich ist die Webseite natürlich nicht der einzige Ort, an dem der Inhalt dargestellt wird. Von dort aus wird er über weitere Channels wie klassische Newsletter und RSS-Feeds aber auch moderne Reader-Apps und Social-Media-Netzwerke weiterverbreitet, was nicht immer zu zählbaren Aufrufen auf dem Ursprungsportal führt.

Gerade die extrem starke mobile Nutzung steuert mittlerweile große Teile der Reichweite bei und bietet mit „Facebook-Instant-Articles“ oder „Google accelerated mobile pages“ immer neue technische Möglichkeiten, deren Effekte auf die Reichweite und Messbarkeit derzeit noch gar nicht erforscht sind.

- **Unschärfe 6: Artikel ist nicht gleich Artikel**

Auch im „Nenner“ unserer obigen Division muss man sehr genau aufpassen, wodurch man eigentlich teilt, denn der reine Aufruf einer Webseite sagt noch nichts über deren Inhalt aus. So kann es sich um ein eingebettetes Video handeln, einen kurzen Hinweis oder einen mehrseitigen Kommentar, der mit großer Sorgfalt über Monate recherchiert wurde.

Bei Echobot geben wir uns daher größte Mühe nur Artikel zu berücksichtigen, die eine gewisse Mindestlänge aufweisen und thematisch auch als „Artikel“ durchgehen. Unser Vorgehen, um diese Entscheidung zu treffen, ist ein Mehrschrittverfahren:

- Mehrschrittverfahren zur Klassifikation von Artikeln bei Echobot
  1. Ist die Seite verfügbar, gibt es den Inhalt unter dieser URL oder handelt es sich um eine Weiterleitung (doppelte Seiten filtern und zusammenfassen)?
  2. Klassifikation, welche Bereiche eines Portals relevante Inhalte ausliefern:
    - Klassifikation über Menüstruktur, Sitemap und Einstiegspunkte
    - Erkennung von RSS-Feeds
    - Manuelle Qualitätsprüfung der Seitenbereiche bei wichtigen Domains
  3. Extraktion des Artikels: Welcher Textbereich auf der Seite ist tatsächlich der relevante Inhalt (Filterung der Navigation, Menüs, ähnliche Artikel, Kommentare)?
  4. Automatische Erkennung von Überschrift, Datum, eindeutiger URL und Autor

Hierbei sei darauf hingewiesen, dass das Ziel von Echobot ist, einen Qualitätsindex der werthaltigen Inhalte aufzubauen und keinen Kompletindex aller Unterseiten. Nur erstere werden berücksichtigt.

# Intelligente Korrekturfaktoren

Aufgrund der dargestellten Unschärfen haben wir in den letzten Monaten einige intelligente automatische Korrekturfaktoren entwickelt und in unsere Berechnungs-Formel eingebaut, die in Summe zu einer sehr viel besseren Abschätzung der tatsächlichen Leserzahl pro Artikel kommen.

Ein wichtiger Hinweis ist, dass es sich immer noch um **einen Schätzwert für die Reichweite eines durchschnittlichen Artikels der Webseite** handelt. Bessere Näherungswerte sind sicherlich möglich, erhöhen aber unnötig die Komplexität, sodass es für den Nutzer kaum mehr nachvollziehbar wird.

## Beispiel für aufgelöste Komplexität in der neuen Berechnung:

Da sich die Reichweite von Portalen mit der Zeit ändert, haben wir früher historische Werte für jedes Portal gespeichert. Dies hatte zur Folge, dass zwei Artikel desselben Portals unterschiedlich geschätzte Leserzahlen aufweisen konnten, wenn die Artikel in unterschiedlichen Monaten erschienen waren.

Das ist zwar sachlich korrekt, aber für den Nutzer weder offensichtlich noch einfach nachzuvollziehen. Im Gegenteil hat es eher für Verwirrung gesorgt und zur Meldung von vermeintlichen „Fehlern“ geführt. Durch den Verzicht auf historische Werte, entsteht jetzt jeder Report aus der aktuellen Perspektive zum Zeitpunkt seiner Erstellung. Die Reichweiten sind konsistent und nachvollziehbar immer auf dem neuesten Stand.

## Erfassungskorrektur

Dennoch wissen wir natürlich bei der Berechnung nie, wie viel Prozent des tatsächlichen Inhalts unseren scharfen Artikel-Kriterien (siehe oben) Stand gehalten haben. Das ist für unsere Formel vor allem ein Problem, wenn es auf einer Webseite sehr viele Inhalte gibt, die wir nicht erfassen.

Die „Erfassungskorrektur“ ist ein Algorithmus, der den tatsächlich gecrawlten (= erfassten) Content betrachtet und mit weiteren Inhalten und der Gesamtreichweite abgleicht. Ist das Verhältnis hier stark verschoben, wird dieses entsprechend korrigiert (z.B. die Artikelanzahl hochgesetzt).

So können wir zum Beispiel für die Seite [www.chefkoch.de](http://www.chefkoch.de) die Reichweite pro Artikel gut abschätzen, obwohl die ermittelten Zugriffe hauptsächlich durch bei uns nicht erfasste Rezepte entstehen.

## Korridorkorrektur

Doch auch mit der richtigen Artikel-Anzahl gibt es immer noch Portale und Webseiten, die über viel mehr Zugriffe verfügen als jemals durch reine Leser von Inhalten erzielt werden könnten. Ein gutes Beispiel hierfür sind Webseiten von großen Online-Shops wie [www.penny.de](http://www.penny.de) oder Portale von Universitäten wie [www.uni-trier.de](http://www.uni-trier.de). Zwar finden sich auch hier redaktionelle Inhalte, die von Echobot erfasst und verarbeitet werden – die gesamte Reichweite des Portals diesen zuzurechnen wäre aber sicherlich falsch, da die Besucher nicht vorrangig zum Lesen von Inhalten auf diesen Seiten unterwegs sind.

Daher haben wir die sogenannte „Korridorkorrektur“ entwickelt, die automatisch den berechneten Wert „Leser pro Artikel“ für jede betroffene Seite in einen für diesen vergleichbaren Größen-Korridor verschiebt.

Das exakte Vorgehen, wie dies abläuft, bleibt jedoch unser Betriebsgeheimnis. In den Details jeder Quelle können Sie immer transparent nachvollziehen, ob einer der Korrekturfaktoren angewendet wurde.

## Weitere Verbesserungen zur alten Version

Wenn Sie die alten Echobot-Reichweiten-Statistiken öfters verwendet haben, sind Ihnen sicherlich einige weitere Verbesserungen aufgefallen. Hier noch einmal ein vollständiger Überblick:

- **Anzeige der Abdeckung**

Wir zeigen jetzt an, seit wann Echobot eine Quelle überwacht, wie viele Artikel bereits indexiert wurden und wie viele neue Artikel pro Tag und Monat hinzukommen. So können Sie noch besser einschätzen, wie genau die Qualität des Echobot Monitorings bei für Sie wichtigen Quellen ist.

- **Darstellung neuester Inhalte und Themen**

Über die Darstellung der neuesten Artikel erhalten Sie eine schnelle Ansicht, was zuletzt von Echobot indexiert wurde und wie schnell die Indexierung für das Portal funktioniert. Die Tag-Cloud liefert einen thematischen Überblick über wichtige Schlagworte, die häufig im Inhalt der Seite auftauchen.

- **SimilarWeb statt IVW als Datenbasis**

Zwar führen wir die IVW-Reichweiten in den Quellen-Details noch auf, sofern diese verfügbar sind, jedoch basieren alle Berechnungen nur noch auf den besser vergleichbaren SimilarWeb-Daten.

- **Keine Reichweiten-Historisierung mehr**

Wie oben bereits ausgeführt, werden zu jedem Portal immer nur die aktuellsten Reichweiten-Informationen angezeigt. Die Perspektive zum Zeitpunkt der Analyse ist für Kunden einfacher.

- **Ausgewogene EchoRank-Verteilung**

Da der EchoRank jetzt nur noch auf der Summe Page-Impressions und der Anzahl Artikel beruht, ergibt sich eine viel ausgewogenere Verteilung. Nur noch 2% aller Domains fallen in EchoRank „5“.

- **Falsche, kaputte und unlogische Rankings korrigiert**

Einige Webseiten waren aufgrund von Subdomains oder Weiterleitungen leider oft mit falschen Werten belegt. Diese Fehlerfälle wurden durch die neue Berechnungsmethodik deutlich reduziert.

- **Kaum Ausreißer/unrealistische Werte**

Durch die Erfassungs- und Korridorkorrekturen, wird vermieden, dass einzelne Quellen bei der Summierung der Reichweiten zu sehr ins Gewicht fallen. Zwar opfert man gerade in den Grenzbereichen sehr großer oder sehr kleiner Portale die absolute Genauigkeit, hat aber in Summe einen geglätteten und viel homogenen Datenraum, der letztlich zu deutlich besseren Analysen und Hochrechnungen führt.

Wie man schon an der Länge dieses Fachartikels merkt, ist in diese Berechnung sehr viel Arbeit geflossen. Gerade auch deswegen würden wir uns über Ihr Feedback freuen. Testen Sie das Echobot Monitoring und die neue Reichweitenberechnung aus. Erstellen Sie einen Pressespiegel mit Reichweitensumme und berichten Sie uns von Ihren Erfahrungen ([service@echobot.de](mailto:service@echobot.de)).

Wir freuen uns drauf! Bastian Karweg (2016)

# Über Echobot

Am Anfang ging es uns vor allem darum, Abdrucke von Pressemitteilungen in der stetig wachsenden Zahl von Online- und Social-Media-Quellen aufzufinden. Heute ist Echobot zu einer sogenannten „Smart Data Intelligence Platform“ geworden, die Milliarden digitaler Inhalte analysiert und Nutzer bei über 750 Firmenkunden in Echtzeit mit geschäftsrelevanten Daten und Informationen versorgt. Wir richten den Blick unserer Nutzer nicht mehr nur auf die Frage: „Was wird über mein eigenes Unternehmen geschrieben?“, sondern viel mehr darauf: „Welche wichtigen Entwicklungen und Ereignisse gibt es bei meinen Kunden, Lieferanten und Geschäftspartnern?“. Für unsere innovativen Machine-Learning-Verfahren zur Signalerkennung haben wir auf der CeBIT 2015 den INNOVATIONSPREIS-IT gewonnen. Eine Erfolgsgeschichte, die täglich weiter geht.

## Noch Fragen?

Unsere Mission ist Ihr Erfolg. Dieses wie auch unsere weiteren Whitepaper und Case-Studies sollen Ihnen helfen, Echobot noch besser und zielgerichteter einzusetzen. Wenn Sie gerne noch weitere Beispiele lesen möchten oder vielleicht sogar selbst eine tolle Einsatzmöglichkeit für Echobot gefunden haben, freuen wir uns auf Ihren Anruf oder Ihre E-Mail.

### Herausgeber:

**Echobot Media Technologies GmbH**

Durlacher Allee 73

76131 Karlsruhe

Deutschland

*Wir helfen Ihnen bei Fragen gerne weiter!*

Telefon: **+49 (0) 721 - 500 57 500**

E-Mail: [service@echobot.de](mailto:service@echobot.de)

Blog: <https://info.echobot.de/blog>

Social Media: <http://www.facebook.com/echobot>

<http://www.twitter.com/echobot>



Bastian Karweg (Geschäftsführer)