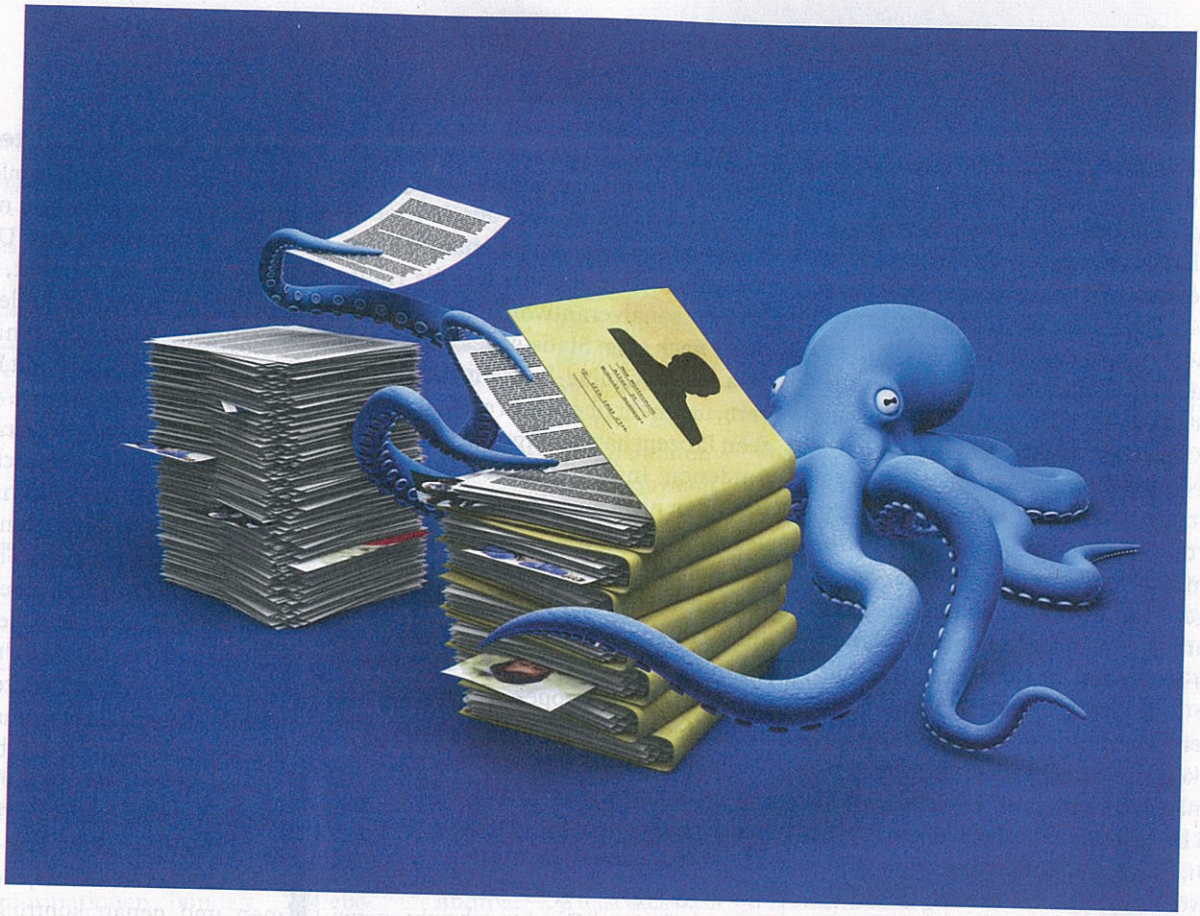

Sammelleidenschaft

Wie und wo Facebook seine Daten zusammenträgt

Klar ist: Facebook sammelt sehr viele Daten über seine knapp zwei Milliarden Mitglieder, und auch der Rest der Menschheit gerät in den Gravitationsstrudel dieses schwarzen Datenlochs. Aber wie genau kommt Facebook an die Informationen?

Von Herbert Braun

Facebook erfährt innerhalb der eigenen Plattformen auf fünf Weisen Neues: Indem es erstens auswertet, was Nutzer aktiv über sich angeben, was sie zweitens indirekt durch ihre Aktionen wie liken oder posten hinterlassen und was drittens Nutzer über andere verraten, etwa durch Taggen von Fotos. Außerhalb seiner Plattformen sammelt Facebook viertens fleißig jene Spuren, die Nutzer beim Surfen oder beim Einsatz von Apps hinterlassen – und zwar mithilfe von Codeschnipseln in Websites und Apps. Als Fünftes kommen noch aufgekaufte externe Datenbanken hinzu.

Die meisten aktiv eingegebenen Inhalte finden sich auf der „Info"-Seite jedes Profils: Ausbildung, Beruf, Lebensstationen, Adresse, Kontaktdaten, Familienmitglieder, Profilfoto. Hier hat der Nutzer die Kontrolle, was er preisgibt. Der Übergang zur zweiten Kategorie, den impliziten Daten, ist fließend. Ein hochgeladenes Foto enthält nicht nur Pixel, sondern in der Regel auch Metadaten über Kamera sowie Zeit und Ort der Aufnahme.

Durch Bilderkennungsalgorithmen lassen sich auch die Inhalte maschinell auswerten. Dass Facebook routinemäßig solche Techniken einsetzt, zeigt sich schon an der Bitte, den meist korrekt ermittelten Gesichtern auf dem Foto Namen zuzuweisen. Facebook versucht bereits seit Längerem, die Gesichter auf Fotos automatisch zu erkennen und zu kennzeichnen („photos that look like you"). Dies passiert bislang nach unserer Kenntnis nur in den USA.

Bei Status-Updates, geposteten Links und den standardmäßig nicht Ende-zu-Ende-verschlüsselten Chats im Messenger analysiert Facebook vermutlich die

Schlagworte – und sei es auch, indem es sie für allgemeine Auswertungen à la Google Trends aggregiert. Der mobile Client fügt jedem Posting Standortinformationen hinzu, solange man das nicht abstellt.

Unabhängig von Beiträgen jeder Art kann Facebook die Aktivitäten seiner Nutzer umfassend nachverfolgen, zum Beispiel das reine Leseverhalten. Beim Herumsurfen auf der Seite lassen bereits die Standard-HTTP-Header erkennen, wann und wie lange ein Benutzer Zeit für den Newsfeed hat, welche Beiträge er öffnet und ob er das vom stationären Rechner oder vom Mobilgerät aus tut.

Eine grobe Ortung lässt die IP-Adresse zu. Auf dem Mobilgerät haben Facebook- und Messenger-App in der Regel Zugriff auf die Ortungsfunktion. Damit könnten die Apps zu jeder Zeit den Standort des Nutzers nach Hause funken, auch wenn die App nur im Hintergrund läuft. Den Standort soll zumindest der Messenger nach Angaben von Facebook allerdings nicht verraten. Andererseits erklärt Facebook in seiner Data Policy explizit, Standortdaten zusammen mit Geräte-, Browser- und Verbindungsinformationen aufzuzeichnen.

Likes, Shares und Kommentare verbinden einen Nutzer sowohl mit anderen Profilen oder Seiten als auch mit bestimmten Themen – äußerst wichtig für Facebooks Werbekunden. Wenn ich ständig die Katzenfotos einer Freundin like, kann Facebook daraus sowohl mein Interesse an Katzen wie an der Posterin erschließen. Indizien dafür liefert aber bereits das Ausklappen abgekürzter oder zusätzlicher Beiträge, das Öffnen von Links beziehungsweise Fotos in der Einzelansicht oder das Einblenden verborgener Kommentare.

Im Zentrum jedes sozialen Netzwerks stehen die Verbindungen zu anderen Mitgliedern. Kein Wunder also, dass sich Facebook alle Mühe gibt, bestehende Kontakte auf seiner Plattform abzubilden. So erklären sich die teilweise penetranten Versuche, Nutzer zum Upload ihres Adressbuchs zu nötigen. Auf diese Weise gelangen auch Daten über Nichtmitglieder zu Facebook. Melden diese sich irgendwann einmal an, finden sie sich sofort in ihr soziales Umfeld eingebettet.

Weniger bekannt ist, dass Facebook als Internet-Provider agiert.

Daten-Zuträger

Nicht alle Informationen, die Facebook über eine Person besitzt, stammen von dieser selbst. Ein bekanntes Beispiel dafür ist das Taggen von Fotos – auf diese Weise kann das Netzwerk auch manche Personen erkennen und mit anderen in Verbindung bringen, die kein eigenes Facebook-Profil haben.

Mitunter scheint es zu genügen, dass sich zwei Mitglieder am selben Ort ungefähr zur selben Zeit aufhalten, damit sie einander als Freunde vorgeschlagen werden. Dass sich Facebook für das Freunde-Finden mit der räumlichen Nähe auseinandergesetzt hat, geht zumindest aus einem Patent des Unternehmens aus dem Jahr 2012 hervor. Und es würde auch die mysteriösen Freunde-Empfehlungen aus dem Artikel auf Seite 72 erklären.

Such- und Freundschaftsanfragen können als Indiz gelten, dass zwischen zwei Personen Kontakt besteht. Das gleiche gilt für Erwähnungen von Nutzern, Beiträge in der Timeline eines Profils oder wiederholtem Austausch, etwa in den Kommentaren zu Beiträgen oder im Messenger. Viele Lebensereignisse bieten die Möglichkeit, eine Person einzubeziehen, etwa „in Beziehung mit“; der oder die Betreffende kann dieses Ereignis dann der eigenen Timeline hinzufügen. Ein gegenteiliges Signal sendet das Stummschalten von Freunden, das Entfreunden oder Blockieren. Einladungen zu Veranstaltungen, geschlossenen Gruppen, Apps oder Spielen bringen ein Profil mit dem jeweiligen Thema in Verbindung.

Auch von Firmen erhobene Daten können bei Facebook landen. Angenommen, ein Unternehmen möchte bereits bestehenden Kunden auf Facebook Werbung für ein neues Produkt zeigen. Dazu speist es seine Kundendatenbank oder die Teilnehmer eines Preisausschreibens mit Namen, Mail-Adressen und Telefonnummern in eine sogenannte „Custom Audience“-Liste ein. Diese gleicht Facebook mit bestehenden Profilen ab, um ihnen gezielt die Werbung zu zeigen. Nach Facebooks Angaben läuft dieser Abgleich jedoch nur über Hash-Werte ab, die anschließend gelöscht werden.

Wachsendes Imperium

Facebook sammelt nicht nur in den eigenen Mauern Daten, auch die Tochterunternehmen steuern viele Informationen bei. WhatsApp zum Beispiel überträgt die Telefonnummern und die Nutzungszeiten seiner Anwender. Und das 2012 aufgekaufte Instagram – mit zirka 500 Millionen aktiven Benutzern ein weiterer Web-Gigant – liefert ebenfalls fleißig Daten an den Mutterkonzern. Instagram räumt sich mit seinen Datenschutzbedingungen das Recht ein, „Cookies, Protokolldateien, Gerätekennungen, Or-

Facebook erhebt nicht nur selbst Daten, sondern ermöglicht es auch Dritttunternehmen, Kontaktlisten hochzuladen.

tungsdaten und Nutzungsdaten“ an Facebook zu übertragen.

Weniger bekannt ist, dass Facebook in zunehmendem Umfang auch als Internet-Zugangsprovider agiert. In derzeit 53 Entwicklungsländern in Afrika, Ostasien und Lateinamerika bietet es im Rahmen von internet.org sein „Free Basics“-Paket fürs mobile Surfen an. Aktuell nutzen 40 Millionen Menschen diesen kostenlosen Internetzugang.

Ein Auszug aus den Datenschutzbedingungen von Free Basics: „Wir sammeln beschränkte Geräte-, Browser- und Nutzungsinformationen, wenn du Free Basics verwendest. Insbesondere sammeln wir die Art des von dir verwendeten Geräts bzw. Browsers und Betriebssystems, deine App-Version, App-ID und Geräte-ID, die Zeit und das Datum deiner Verbindung, deinen Mobilfunkanbieter, deine IP-Adresse, deine Telefonnummer, deine Batterie- und Signalstärke, dein Land, deine Spracheinstellung und die Dienste Dritter, nach denen du in Free Basics suchst bzw. die du dort nutzt.“ Ähnlich dicht an der Quelle sitzt die Facebook-Tochter Onavo, die für Android und iOS kostenloses VPN anbietet und deren Datenschutzbedingungen ganz ähnliche Klauseln enthalten.

Die anderen Facebook-Akquisitionen spielen in Sachen Datenaufkommen nicht in der gleichen Liga, sollten aber dennoch nicht übersehen werden. Bekannt ist vor allem das VR-Unternehmen Oculus. Die Fitness-App Moves begleitet die Nutzer beim Joggen und Radeln. Dazu kommen noch diverse Apps, die Facebook unter eigenem oder unter Instagrams Namen herausgibt. Bei den meisten ist der Bezug zum Netzwerk offensichtlich, aber etwa bei der beliebten Spaß-Video-App MSQRD („Masquerade“) oder dem Animationsgenerator Boomerang nicht unbedingt.

Überall mitgeloggt

Bei der Datensammlung Facebooks außerhalb der eigenen Plattform und der Töchter wird jeder als Erstes an den allgegenwärtigen Like-Button denken. Daneben gibt es aber noch diverse andere solcher sogenannter Social Plug-ins, die Website-Betreiber gerne einbinden. Mit den Share- und Quote-Plug-ins etwa teilt man Links oder Textauszüge; über das Send-Plug-in geht das auch via Messenger. Mit „Save“ speichert man Inhalte in einem privaten Bereich auf Facebook, mit dem Follow-Plug-in abonniert man die Updates eines bestimmten Accounts.

Mit weiteren Social Plug-ins lassen sich Facebook-Inhalte in externe Seiten einbinden – Videos, einzelne Beiträge, Kommentare oder eine Übersicht über eine Facebook-Seite (Page-Plug-in). Das Comments-Plug-in schließlich nutzt Facebooks Infrastruktur, um Leser für Kommentare zu authentifizieren und diese mit deren Accounts zu verbinden.

Bei allen seinen Social Plug-ins schlägt Facebook vor, diese über ein JavaScript einzubinden, das direkt vom Facebook-Server kommt. Auf diese Weise erfahren die Server in Menlo Park von jedem Aufruf der betreffenden Seite. Die meisten Menschen dürften nicht wissen, dass Facebook so ihre Klicks auf vielen Sites verfolgen kann. Auf größeren deutschen Websites trifft man den Like-Button in seiner ursprünglichen Form nicht mehr überall an – möglicherweise hat das „Shariff“-Projekt der c’t hier etwas bewegt (siehe Seite 80). Dieses sorgt dafür, dass die Facebook-Server erst kontaktiert werden, wenn der Nutzer den Button anklickt.

Das heißt aber nicht, dass alles gut wäre. Unter Webmastern mangelt es bei der Einbindung von Fremdservern an Datenschutzsensibilität – es gibt populäre Websites, die Inhalte von mehr als einhundert Domains einbinden. Da überrascht es nicht, dass gerade facebook.com beim Aufruf der meisten größeren Websites angefunkt wird. Weniger bekannt ist, dass auch viele Apps Inhalte von Facebook einbetten, zum Beispiel Spotify, Shazam, Tinder und Blendle.

Facebook Login ist eine praktische Authentifizierungslösung. Facebook-Mitglieder können damit ihren Account nutzen, um sich auf Drittanbieter-Webdiensten oder in Apps einzuloggen. Sie müssen sich also ein Passwort weniger merken. Mehrwert für Facebook: Das Unternehmen erfährt, welche Dienste und Apps seine Mitglieder nutzen, und wann. Mit dem „Account Kit“ eröffnet Facebook auch Menschen, die keinen Facebook-Account haben, diese Möglichkeit. Zur Registrierung benötigt Facebook deren Mobilfunknummer oder E-Mail-Adresse.

Kekspackung

Weitere Informationen kommen durch Facebooks Werbebanner außerhalb des Netzwerks zusammen. Das „Audience Network“ ist für Mobilgeräte gedacht und kommt in Apps wie in Mobil-Websites zum Einsatz. Zur Erfolgskontrolle und Konversionssteigerung hält der Konzern das Zählpixel „Facebook Pixel“ und seine Analytics-Lösung für Apps bereit. Die

Bild: Facebook

Aus allen Daten das Maximum herausholen: Für Facebook sind Bilder längst mehr als Pixelhaufen. Es kann sie analysieren und per Text beschreiben.

technische Basis für die zielgruppengerechte Werbung über alle Geräte hinweg kaufte Facebook 2013 von Microsoft: Atlas Solutions, einen Adserver mit Reichweitenmessung und Facebooks Gegenstück zu Googles DoubleClick.

Seit Mai dieses Jahres sammelt Facebook explizit auch von Nichtmitgliedern Daten, um maßgeschneiderte Banner ausliefern zu können. Eine gerichtliche Auseinandersetzung mit der belgischen Datenschutzbehörde über diese Praxis gewannen die Kalifornier im Juni. Bei diesem Prozess ging es um Cookies, mit denen Facebook auch außerhalb der eigenen Plattform Werbung ausliefern und Nutzerdaten analysieren kann.

Auch in Deutschland sammelt Facebook so Daten ein, sowohl für Mitglieder der Plattform als auch für Nichtmitglieder. Nichtmitglieder erhalten allerdings viel weniger Cookies als Mitglieder: Bei einigen Stichproben mit Seiten, die Facebook-Widgets nutzen, schickte uns Facebook im ausgeloggten Zustand fast nirgends ein Cookie; nur bei bild.de fingen wir uns eines namens „fr“ ein – zusammen mit gut 200 Cookies von über 80 anderen Domains.

Wer eingeloggt ist, wird in seinem Browser ein Dutzend kleine facebook.com-Cookies mit überwiegend langer Laufzeit finden, deren genaue Funktionsweise unserem Wissen nach nicht öffentlich dokumentiert ist. Am häufigsten aktualisiert werden „presence“ und „fr“, die mit etwa 160 beziehungsweise 80 Zeichen auch die umfangreichsten sind. „c_user“ enthält die Benutzer-ID und soll zusammen mit „datr“ und „xs“ eingeloggte Benutzer authentifizieren.

Nach eigenen Angaben verwendet Facebook die Cookies außerdem, um Identitätsdiebe oder Spammer zu erkennen, um (etwa in den Plug-ins) ortsbezogene Inhalte auszuliefern oder Einstellungen zu berücksichtigen sowie für die Verbesserung der Performance. Zusätzliche Einstellungen speichert Facebook im localStorage unter www.facebook.com, offenbar aber nicht zum Zweck der Identifizierung.

Datenkauf und -verschmelzung

Um Werbung noch gezielter auf seine Mitglieder zuzuschneiden, als das mit den eigenen Profilen ohnehin schon möglich ist, arbeitet Facebook mit externen Datenhändlern zusammen. Deren Informationen stammen teils von Werbenetzwerken, teils aber auch aus Quellen außerhalb des Internet – Personendaten, Einkaufsgewohnheiten, Autoregistrierungen, Beruf, Familie, Finanzdaten, Urlaubswünsche und vieles mehr.

Gefällt mir: Nur ein kleiner Knopf, aber er ermöglicht es Facebook, mitzuschneiden, wer die Seite besucht.

Facebook bezieht in Deutschland Daten von Acxiom und Datalogix, weltweit arbeitet es außerdem auch mit BlueKai, Epsilon und Quantium zusammen. Diese stellen auf Anfrage Zielgruppenlisten zur Verfügung, die Facebook mit seinen Nutzerprofilen abgleicht. Bei Acxiom und Datalogix geschieht das über eine gehashte Mail-Adresse. Facebook erhält also nur Daten über Menschen, die es bereits kennt. Und es gibt nach eigenen Angaben an seine Partner keine personenbezogene Informationen weiter.

Es ist kein Zufall, dass bei externen Daten immer wieder E-Mail-Adressen und Mobilfunknummern eine Rolle spielen. Sie sind die Schlüssel, mit denen Facebook Daten aus fremden Quellen mit seinen eigenen Informationen verknüpft: Die WhatsApp-Nutzungszeiten etwa kann Facebook über die Mobilfunknummer in aller Regel einem bestimmten Nutzer des sozialen Netzwerks zuordnen.

Facebook begnügt sich nicht damit, Daten aus verschiedenen Quellen zu sammeln und zu möglichst genauen Profilen zu aggregieren. Das Unternehmen betreibt vielmehr eine große Datenanalyse-Abteilung, die versucht, den vorhandenen Daten neue Informationen abzuringen. Ab und zu öffnen Facebooks Datenwissenschaftler die Tür zu ihrer Hexenküche einen Spalt und geben Einblick in ihre Forschungen.

Eine komplette semantische Analyse der Beiträge zum Beispiel ist mit dem DeepText-Projekt auf dem Weg. Es soll Facebook-Posts, Kommentare und Nachrichten in über 20 Sprachen verstehen. Dabei soll es neben dem grundsätzlichen Thema auch die Stimmung eines Textes sowie die Ereignisse, Personen und Orte identifizieren, um die es darin geht.

Im Zweifel aggressiv

Ihnen wird sicherlich aufgefallen sein, dass wir in diesem Artikel an etlichen Stellen defensiv formulieren. Die vorsichtigen Formulierungen haben den Grund, dass wir es einfach nicht genauer wissen. So groß das Interesse Facebooks an den Informationen über Dritte sein mag, so zurückhaltend ist es darin, Dinge über sich preiszugeben.

Wir haben Facebook eine Liste mit Fragen geschickt, um genauer zu erfahren, welche Daten das Unternehmen an welchen Stellen einsammelt und was es damit macht. Erhalten haben wir aber nur sehr vage allgemeine Informationen. Ein Stück weit kann man das auch nachvollziehen: Ein Werbekonzern wie Facebook will der Konkurrenz sicher nicht aufs Brot schmieren, wie man Datenberge über Menschen effektiv aufbereitet.

Für Sie bedeutet das, dass Sie unsere Erläuterungen interpretieren müssen. Und dabei sollten Sie eher davon ausgehen, dass Facebook aggressiv alles aufsaugt, was es kriegen kann und damit alle beschriebenen Methoden auf seine Daten loslässt – und womöglich noch ein paar mehr, von denen auch wir noch nichts gehört haben. Daten sind ein wertvoller Rohstoff in der Werbebranche und Facebook ist ein Meister darin, diesen Rohstoff zu raffinieren. *(jo@ct.de)* c't

Weiterführende Informationen:
ct.de/y29a