Arts and Graphs

Um Daten zu veranschaulichen, braucht man kein Programmierer mehr zu sein. Verena Dauerer stellt leicht bedienbare Visualisierungstools und interessante Projekte vor

Mit dem Tool Gephi bildete André Pannison die schneeballartige Verbreitung der Twitter-Nachricht von Mubaraks Rücktritt in Echtzeit nach *(www.youtube.com/panisson)*

■ **Am 11. Februar** arbeitete Dynamic-Network-Analyst André Pannison an einem Twitter-Visualisierungsprojekt, das die Universität Turin zusammen mit der ISI Foundation und der Indiana Universität durchführt. Ein selbst entwickelter Python-Webserver dient dem Informatik-Doktoranden dabei als Brücke, um Daten der Twitter-API in Echtzeit anzuzapfen und mit der Open-Source-Visualisierungssoftware Gephi in eine Animation umzusetzen. Ziel ist es, sichtbar zu machen, wie oft ein bestimmtes Hashtag retweetet wird. Dabei wird jeder User als Knotenpunkt dargestellt und mit den folgenden Retweets verknüpft, sodass nach und nach ein feingliedriges Fadengespinst entsteht.

André Pannison testete verschiedene Hashtags, bis er auf das recht beliebte „#jan25“ stieß. Und auf einmal gab es einen wahren Ausbruch in seiner Animation. Der Grund wurde ihm schnell klar: Das Twitter-Schlagwort „#jan25“ steht für den Beginn der Revolution in Ägypten, und Mubarak hatte gerade seinen Rücktritt bekannt gegeben. „Dass ich diese Ereignisse innerhalb unseres Projekts aufzeichnen konnte, ist purer Zufall“, so Pannison. Mithilfe solcher animierten Datenvisualisierungen kann man die Bedeutung von für sich allein genommen wenig aussagekräftigen Ausgangsdaten deutlich machen – in diesem Fall das Interesse und Engagement der Twitter-Gemeinschaft an einem geschichtlichen Ereignis (siehe auch PAGE 06.11, Seite 44 ff.).

Neben dem von André Pannison verwendeten Gephi gibt es noch weitere spannende Tools für die Datenvisualisierung, die wir Ihnen hier vorstellen (siehe Seite 92 ff.). Gemeinsam ist all diesen Lösungen, dass sie keine Vorbildung als Programmierer verlangen. Keine Visualisierung ohne Daten: Dabei liegt der größte Datenstrom jederzeit vor uns – das Internet mit seinen sozialen Netzwerken. Am Beispiel verschiedener Online-Projekte zeigen wir, wie sich durch kreative Visualisierungen die tägliche Flut an Informationen und die Welt der sozialen Interaktionen erforschen, strukturieren und verstehen lässt.

Conductor U-Bahn-Linien hörbar machen

Mit „Conductor“ visualisiert Alexander Chen, wie die New Yorker U-Bahnen ihre Strecken abfahren – jede Überschneidung der Linien erzeugt einen Sound. Und der User kann wie auf einem Seiteninstrument Klänge erzeugen

■ Conductor ist eine interaktive Anwendung, die im Zeitraffer darstellt, wie die New Yorker U-Bahnen ihre Linien abfahren. Dabei werden die Strecken zu den Saiten einer Gitarre, die immer dann erklingt, wenn zwei Bahnen sich kreuzen. Außerdem kann der Nutzer mit der Maus an einer Linie zupfen. Für die Programmierung der Anwendung nutzte Alexander Chen, Interactive Designer und Künstler aus New York, HTML5 und JavaScript. Die umfangreichen Daten der einzelnen Strecken und Haltestellen entnahm er der öffentlichen API der New Yorker U-Bahngesellschaft MTA.

Das Design zeichnete er zunächst in Illustrator und exportierte es dann in HTML5. Für die Erzeugung der Klangkulisse war das aber nicht ausreichend, da mehrere Töne simultan ausgegeben werden müssen, wenn sich unterschiedliche Linien gleichzeitig kreuzen. „HTML5 kann nicht allzu viele Audioelemente verarbeiten“, erklärt der Designer. Diese Aufgabe übernimmt deswegen eine Flash-Anwendung im Hintergrund, die sich mit der visuellen HTML-Schicht über einen Java-Script-Sound-Manager abstimmt. Jede einzelne U-Bahn-Linie hat einen eigenen Klang, sodass insgesamt 38 unterschiedliche MP3-Samples erforderlich waren. Der grundlegende Aufbau der Flash-Anwendung ist ganz einfach: Die Web-App lädt die Sounds beim Start vorab und weist jeder Linie ein Sound-Channel-Objekt zu. Für das Abspielen der unterschiedlichen Samples sorgt dann der Flash-Befehl *Sound.play()*.

↗ http://mta.me/; http://blog.chenalexander.com

Web Trigrams Wortkombinationen visualisieren

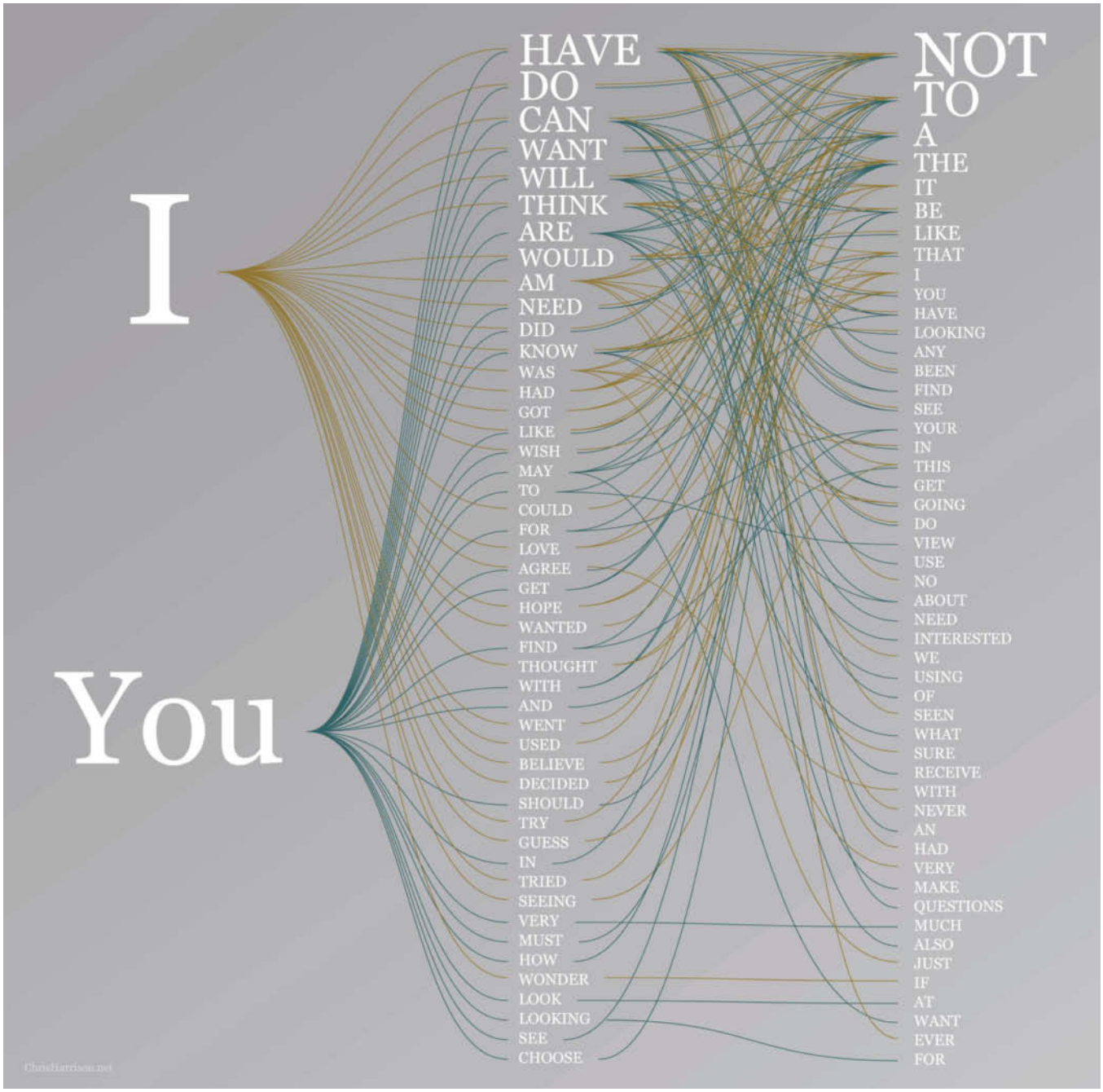
■ Chris Harrison, Doktorand am Human-Computer Interaction Institute der Carnegie Mellon University, verwendet keine Visualisierungstools, sondern programmiert lieber selbst. Als Wissenschaftler beschäftigt sich der Entwickler mit neuen Interaktionstechniken, seine Leidenschaft aber ist die Datenvisualisierung. „Ich liebe die in den Daten enthaltene Komplexität. Die daraus entstehende Schönheit kommt nicht vom Designer, sondern stammt von den Daten selbst. Sie verleiht ihnen eine eigne Stimme."

Für das Projekt „Web Trigrams" nutzte Harrison die von Google 2006 zur freien Verwendung in wissenschaftlichen Einrichtungen veröffentlichten Textanalysedaten. Die 100 Gigabyte umfassende Datensammlung enthält N-Gramme und Angaben über die Häufigkeit, in der sie auftreten. Ein Unigramm besteht aus einem, entsprechend ein Trigramm aus drei Wörtern (n=3). Beispiel eines Trigrams: I like food.

Mit 200 Zeilen Java entwickelte Chris Harrison seine „Web Trigams"-Visualisierungen. Diese zeigen zum Beispiel besonders häufig verwendete Wortkombinationen, die mit „I" und „You" beginnen oder mit „He" und „She". Dabei wird deutlich, wo sie sich überlappen und wo sie auseinanderdriften. Tatsächlich fallen bei den Top-120-Trigrams viele Überschneidungen ins Auge, interessanterweise sind aber bestimmte Kombinationen lediglich bei einem einzigen Subjekt zu finden. So wird „He" zum Beispiel ausschließlich mit „He argues" verbunden, „She" dagegen mit „She loves".

↗ *www.chrisharrison.net*

Auf Basis von Googles Textanalysedaten visualisiert der Entwickler Chris Harrison in seinem Projekt „Web Trigrams" die Häufigkeit von Wortkombinationen auf Webseiten



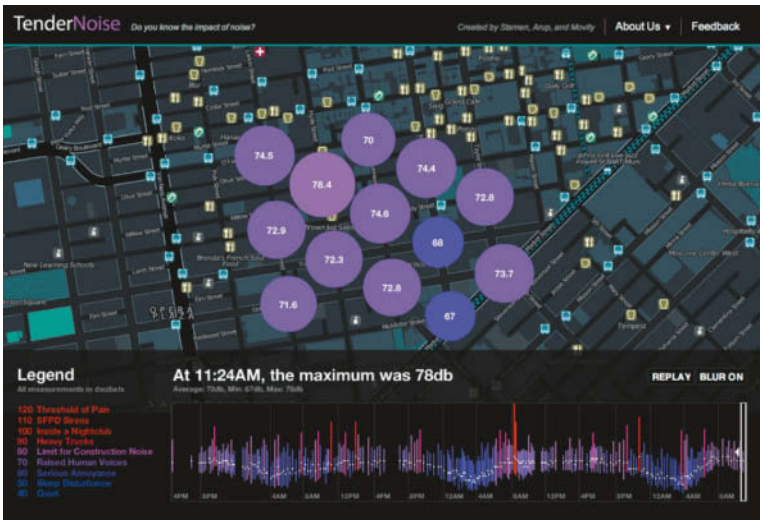
TenderNoise Lärm visualisieren

■ Das Start-up Movity, die Städteplaner von Arup und der Visualisierungsexperte Stamen entwickelten das Projekt „TenderNoise“, um die tägliche Lärmbelästigung für die Bewohner des Stadtteils Tenderloin in San Francisco darzustellen. Die Flash-Anwendung visualisiert den Lärmpegel innerhalb von 24 Stunden. Die Lautstärke springt durch pulsierende Farbpunkte ins Auge. Deren Größe und Farbe hängen von den gemessenen Dezibelwerten ab.

Jeder Kreis entspricht einem der Messgeräte, die das Team an Laternen und Strommasten in Tenderloin befestigt hat. Ein kleiner dunkelblauer Kreis steht für Stille (maximal 40 Dezibel). Über Lila, der Obergrenze für Baulärm (80 Dezibel), wechselt die Farbe zu Rot (120 Dezibel), lauter als eine Feuerwehrsirene. Der User kann die Kreise weichzeichnen lassen, so wirkt die Animation organischer. Oder er wählt einen beliebigen Punkt in der Zeitleiste, um die Werte dieses Moments zu sehen. Ohne große Erläuterung macht „TenderNoise“ schnell klar: Der Stadtteil hat ein Lärmproblem – das Visualisierungsprojekt hat das Problembewusstsein geschärft und das subjektive Lärmempfinden objektiviert.

Bei der grafischen Umsetzung hat das Team komplett auf Flash gesetzt, das sich für solche Animationen ganz besonders eignet, zumal es Effekte wie das Weichzeichnen beliebiger Grafikelemente ermöglicht. Ganz ohne ActionScript-Programmierung ließ sich die Animation zwar nicht umsetzen, der Code ist jedoch recht einfach gehalten. Die Messwerte stammen aus einer simplen TSV-Textdatei, bei der die Datenfelder durch Tabulatoren getrennt sind. Die Textfunktionen in Flash wandeln die Ausgangsdaten in einem Rutsch in ein Array um, das effektiv in einer Schleife durchlaufen wird.

↗ *http://tendernoise.movity.com/*



Das „TenderNoise“-Projekt visualisiert den täglichen Lärmpegel für die Anwohner im Stadtteil Tenderloin in San Francisco, um die Belästigung zu verdeutlichen

Datenvisualisierungssoftware

Diese Tools lassen sich praktisch ohne Programmierkenntnisse nutzen

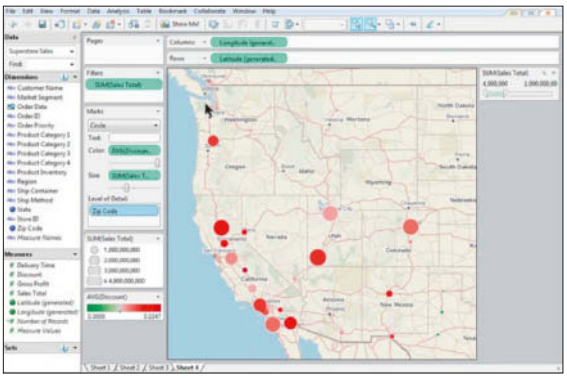
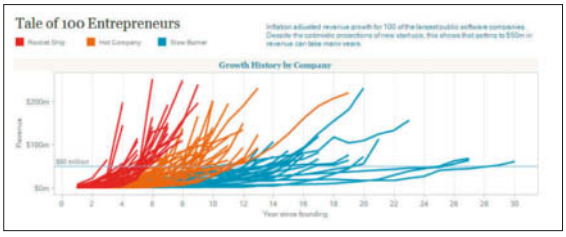


Tableau zeigt die Tabellendaten in einer Liste auf der linken Seite an. Per Drag-and-Drop lassen sich diese in die obere Fensterleiste schieben und werden grafisch umgesetzt



Visualisierung der Frachtkosten von Anbietern im Vergleich zu ihren Lieferverzögerungen

Tableau

■ Von Tableau Software, einem Spinoff der Universität Stanford, kommt der gleichnamige Webdienst für Datenvisualisierungen. Ursprünglich wurde Tableau für Wirtschaftsdaten aller Art konzipiert. Längst eignet es sich aber auch für andere Ausgangsdaten. Der Startpunkt ist eine Datei, beispielsweise eine tabellarische Textdatei oder ein Exceldokument, die sich bequem per Drag-and-Drop ins Programmfenster legen lassen. Die Daten werden dann automatisch übernommen. Tableau kann Daten interaktiv über einen bestimmten Zeitraum darstellen und im Fall von Ortsdaten diese zum Beispiel mit Google Maps koppeln. Es werden aber auch andere Map-Server unterstützt, sofern sie das WMS-Protokoll beherrschen.

Der Anwender kann aus verschiedenen Gestaltungsvorlagen für seine Visualisierung wählen und mit Filtern zusätzlich die Farben und Formen beeinflussen. Die Umsetzung erscheint ohne Verzögerungen. Das Ergebnis wird als interaktive Anwendung, die verschiedene Ansichtsmodi erlaubt, im Web veröffentlicht. Die auf dem Tableau-Server gespeicherten Daten lassen sich aber auch völlig problemlos in eigene Seiten einbinden, beispielsweise mit dem iframe-Tag. Der attraktive Output inklusive Interaktivität wird ganz ohne Flash realisiert.

Tableau steht in unterschiedlichen Editionen zur Verfügung. Neben Tableau Server, die etwa der IT-Nachrichtendienst ReadWriteWeb, die UNESCO oder auch der Sportsender CBS nutzen, gibt es zum Beispiel Tableau Public für Blogger oder Tableau Desktop, das nur auf Windows-Rechnern läuft. Interessierte können sich auch erst mal die Testversion ansehen.

Fazit: Obwohl es ausgefeilt und leistungsstark ist, lässt sich Tableau sehr einfach bedienen.

↗ *www.tableausoftware.com*

→

Many Eyes

■ Das von den Datavisualisierungs-Experten Fernanda Viégas and Martin Wattenberg bei IBM entwickelte browserbasierte Programm arbeitet mit Java und lässt sich deshalb von jedem Betriebssystem aus nutzen. Nach der Anmeldung auf dem Many-Eyes-Portal kann der User Daten hochladen und Darstellungen daraus generieren. Voraussetzung ist, dass beispielsweise eine Exceltabelle bestimmten Regeln entspricht. Per Klick kann der User zwischen verschiedenen Designs wählen, beispielsweise Graphen, Punkten oder auch Tag-Clouds. Für Letztere eignet sich Many Eyes besonders und kann dabei auch ganze Texte auswerten.

Die fertige Visualisierung ist sofort auf dem Portal einsehbar. Sie lässt sich als statisches Bild oder als Live-Visualisierung in jede Webseite integrieren und ist über Facebook, Twitter oder E-Mail versendbar. Da Many Eyes als soziales Netzwerk konzipiert ist, dürfen User Einträge kommentieren. Und sie können sie auf die hinter den Visualisierungen liegenden Daten zugreifen und damit neue Darstellungen kreieren.

Fazit: Many Eyes ist kostenlos und ein gutes Experimentierfeld für Einsteiger – man kommt spielerisch voran.
↗ *www-958.ibm.com*

Als Grundlage für die Visualisierung von Nutzerin Elena H. analysiert Many Eyes einen Text über Notunterkünfte im Katastrophenfall auf die Häufigkeit bestimmter Worte

Gephi

■ Oftmals verwenden Wissenschaftler die Open-Source-Plattform Gephi zur Darstellung ihrer Forschungsergebnisse. Mit der kostenlosen Anwendung lassen sich kompliziertere Datenverläufe in Echtzeit visualisieren. Besonders eignet Gephi sich für die Visualisierung von Hierarchien sowie für die Kartografie von Netzwerken bis zu 50 000 Knotenpunkten. Im sogenannten Layout legt der Nutzer die Algorithmen fest, die die Daten umsetzen. Zur genauen Justierung sind eine ganze Anzahl von Einstellungen bei der Rangfolge der Daten oder etwa bei den Farbverläufen der Knotenpunkte nötig. Die Funktionen sind vielschichtig und nicht auf den ersten Blick ersichtlich, aber die Online-Bedienungsanleitung nimmt auch Einsteiger bestens an die Hand.

Gephi ist ideal für Projekte, die in Echtzeit Daten auswerten, wie beim Twitter-Projekt von André Pannison gesehen (siehe Seite 90). Nur so war es möglich, einen Film von der Entwicklung der Retweets zu capturen. Das Tool bietet neben dem Echtzeit- auch einen statischen Renderer, der eine für den Offsetdruck geeignete Ausgabequalität liefert. Die Gephi-Architektur ist modular und basiert auf Plug-ins. Jeder Developer kann eigene Erweiterungen hinzufügen. Dabei benötigen die User selbst keine Programmierkenntnisse. Laut André Pannison ist das erklärte Ziel von Gephi, ein Standard für Graphen zu werden, so wie es Photoshop für die Bildbearbeitung ist.

Als Open-Source-Tool unterstützt Gephi, das unter Mac OS X, Windows und Linux läuft, Dateiformate aus anderen Visualisierungsprogrammen wie GEXF, GDF und GraphM. Und natürlich sind auch einfache CSV-Dateien aus Excel erlaubt. Die fertiggestellte Visualisierung lässt sich als vektorbasierte SVG-Datei exportieren und etwa in Illustrator weiterbearbeiten.

Fazit: Gephi ist ein sehr gutes Visualisierungswerkzeug, bedarf allerdings einer gewissen Einarbeitung.
↗ *http://gephi.org/*

Bei Gephi bestimmen Layoutalgorithmen das Aussehen des Visualisierungsgraphen. Hier kamen Multi-Level-Algorithmen zum Einsatz

Im Preview-Modus zeigt Gephi die fertige Visualisierung. Die Grundeigenschaften muss der User zwar vorher festlegen, doch Feinheiten wie Liniendicken, Fonts oder Farben kann er hier noch justieren