



Freitag, 13.07.2018 / 10:44 0

## DSI Insights: Warum wir kollektive Intelligenz für Open Data brauchen

**Datenwissenschaftlerin und Gastautorin Cristina Sarasua über das Potential der Crowd.**

Die grossen Investitionen der Schweiz in Open Data haben bereits zu einer beeindruckenden Zahl von offenen Datensätzen geführt: von statistischen Daten von Städten über Mobilitätsplanung bis hin zu kulturellen Ereignissen. Wer eine App entwickeln oder eine Datenanalyse durchführen will, kann diese offenen Daten über Datenportale von öffentlichen Einrichtungen wie den statistischen Ämtern der Stadt und des Kantons Zürich abfragen, durchsuchen und herunterladen.



*(Bild: Universität Zürich)*

Allerdings sind die Datenaufbereitung und -veröffentlichung durch diese Institutionen meistens in mehrfacher Hinsicht problematisch, denn die Prozesse sind...

- nicht iterativ; d.h. die Daten werden nach der Veröffentlichung nicht weiter korrigiert, überarbeitet oder konzeptionell erweitert;
- unilateral; d.h. der einzige Akteur, der die Daten strukturiert, bereinigt, veröffentlicht und dokumentiert ist der Datenlieferant selbst,

der trotz seiner Kenntnis der Daten nicht alle Anwendungsfälle abdecken kann und damit die Sichtweise von nur einem Benutzer einbringt, was zu Datenverzerrungen führen kann;

- weitgehend automatisiert; d.h. wenn die Lieferanten der Daten diese mit anderen Datensätzen verknüpfen, verwenden sie dafür aus Zeit- und Kostengründen



automatische Werkzeuge.

Somit hat Open Data zwar bereits zu wertvollen Datensätzen geführt. Doch würde ein dynamischer und hybrider Prozess – also die Verschränkung von menschlicher und maschineller Intelligenz – nicht nur zu einer Verbesserung der Datenqualität führen, sondern auch zur Datenvielfalt beitragen. Maschinen sind gut darin, menschliche Einschätzungen nachzuahmen (z.B. beim maschinellen Lernen), Suchvorgänge effizient durchzuführen und

Routinen wiederholt auszuführen. Aber Menschen sind bei Aufgaben wie der komplexen Assoziation von Ideen, der Identifizierung von Qualität, der Bewertung von Relevanz und der Kreativität den Maschinen überlegen. Wenn wir also viele und unterschiedliche Personen in ein offenes Datenmanagement einbeziehen, können sie Fehler und Datenlöcher finden und korrigieren, den Bedarf an Datenerweiterung oder Daten-Reannotation erkennen, die Daten auf unterschiedliche Weise verbinden, verschiedene Interpretationen der Daten vorschlagen und neue Wege finden, die Daten zu verwenden.

Alle sprechen heute von "künstlicher Intelligenz". Doch vielleicht brauchen wir eine andere Art von KI – "kollektive Intelligenz" –, um einen offenen Datenmanagement-Lebenszyklus zu schaffen. Dieser Begriff wurde von Tom Malone und Kollegen geschaffen und meint eine "Gruppen von Individuen, die gemeinsam Dinge tun, die intelligent erscheinen".

Ein auf "kollektive Intelligenz" basierendes System zu entwerfen und zu betreiben, ist nicht so einfach wie ein Computerprogramm zu implementieren, seine Ausführung zu planen und auf die Ausgabe des Resultats zu warten. Grund dafür ist, dass wir bei einem Kollektiv von Menschen mit einer grossen Vielfalt bezüglich Motivation, kognitiver Fähigkeiten und Fehleranfälligkeit umgehen müssen. Um das volle Potenzial der kollektiven Intelligenz auszuschöpfen, muss man Mechanismen der Teilnehmerbindung und des Engagements entwickeln, die richtige Kombination von Menschen finden, eine gesunde soziale Dynamik stimulieren und verstehen, wie man Menschen helfen kann, so effektiv und effizient wie möglich zusammenzuarbeiten. Auch geht es darum, menschliche und maschinelle Berechnungen intelligent zu kombinieren und schliesslich Wege zu finden, die Qualität der Arbeit einer "kollektiven Intelligenz" zu prüfen.

Es gibt bereits viele erfolgreiche Beispiele für die Nutzung von kollektiver Intelligenz, von denen man lernen kann. Mit "CrowdLang" hat man beispielsweise ein System geschaffen, mit dem ein Buch in einer Stunde übersetzt werden konnte. Im Projekt "Zooniverse" analysieren Amateure in der Astronomie Bilder von Teleskopen. Ein weiteres Referenzbeispiel ist "Wikidata", eine kostenlose, mehrsprachige Wissensdatenbank, die jeder auf der Welt bearbeiten und abfragen kann und von Tausenden von Freiwilligen innert gut fünf Jahren geschaffen wurde.

Die Forschung hat also sowohl im akademischen als auch im privatwirtschaftlichen Kontext gezeigt, dass es möglich ist, nicht-professionelle "Crowds" in bezahlte oder auch ehrenamtliche Aufgaben einzubinden. Noch besser ist es, wenn Laien und Experten gemeinsam arbeiten – also wenn beispielsweise der Geograf eine räumliche Datenanalyse anleitet, die Statistikerin eine detaillierte explorative Datenanalyse entwirft oder der Bibliothekar sein Wissen bezüglich Katalogisierung einbringt. Fachleute können den Prozess leiten, die Nicht-Experten auszubilden oder die Aktivitäten des Kollektivs filtern, einordnen und organisieren.

Die Schweiz befindet sich in einer privilegierten Lage für den Aufbau kollektiver Intelligenzen. Bereits finden zahlreiche Hackathons (z.B. MakeZurich, Wikidata Zürich Datathon und Hackathon), Meetups, Initiativen (z.B. Open Data CH, Open Data Zürich) und technische Konferenzen statt, wo qualifizierte Entwickler und offene Datenenthusiasten Technologien entwerfen, implementieren und diskutieren. Um basierend auf dieser Grundlage einen iterativen, hybriden und kollektiven Datenmanagementprozess aufzubauen, muss man die Orchestrierung und aufgabenorientierte Zusammenarbeit konzipieren, kollektive Intelligenz für ein offenes Datenmanagement verwalten und die Fortschritte überwachen. Dafür schlage ich folgenden Massnahmen vor:

- Wir müssen die Datenportale erweitern, um strukturierte Diskussionen zu ermöglichen sowie Ideen und Feedback zu sammeln. Dies hilft, neue Daten zu identifizieren, die aktuelle Datensätze ergänzen oder erweitern sowie Probleme zu erkennen, die bei der Verwendung oder Analyse der im Datenportal gehosteten Datensätze auftreten. Das braucht unter anderem eine menschliche (oder halbautomatische) Moderation.
- Wir müssen den Austausch über Ergebnisse der Datenanalyse organisieren. Eine bestimmte Tatsache über einen Datensatz zu entdecken ist zwar nützlich, aber wir sollten Gespräche zwischen Datenanalysen über diese Ergebnisse fördern. Damit

wird man erkennen, dass verschiedene Personen Daten unterschiedlich analysieren und man wird die Ergebnisse, die zugrundeliegenden Daten und die Analysemethoden miteinander verknüpfen können.

- Wir müssen jene Leute identifizieren und fördern, welche aufgrund ihrer intrinsischen Motivation dem Wert des offenen Wissens besonders verpflichtet sind. Forschungen zeigen, dass solche "Power-User" in kollektiven Intelligenzen die meiste Arbeit leisten. Sie dienen dazu, andere für mehr Engagement zu motivieren.
- Wir dürfen soziale Aspekte der Kollaboration nicht vernachlässigen. Ein kollektives Intelligenz-System soll so entworfen werden, dass eine produktive Interaktion zwischen den Mitgliedern der Gruppe gefördert wird. Es gilt, die soziale Sensibilität der Gruppenmitglieder zu fördern und einen Raum für konstruktive Kritik zu schaffen, so dass verschiedene Menschen mit ihren alternativen Ideen aufwarten können.

Das Potenzial von wohlkonzipierter "kollektiver Intelligenz" ist gross – nutzen wir es!  
(Cristina Sarasua)

*Cristina Sarasua ist Informatikerin und arbeitet im Bereich Webdatenmanagement und Datenwissenschaft an der Universität Zürich. Ihre aktuellen Forschungsschwerpunkte umfassen unter anderem Human Computation, Crowdsourcing-Methoden und Collaborative Knowledge Engineering. An der UZH ist sie derzeit Dozentin für Social Computing. Sie ist die Hauptveranstalterin mehrerer von Wikimedia unterstützter Wikidata-Veranstaltungen in Spanien und der Schweiz, die den Wissenstransfer von der Wissenschaft zur Industrie und zu Bürgerinitiativen fördern.*

**Zu dieser Kolumne:** Unter "DSI Insights" äussern sich regelmässig Forscherinnen und Forscher der "Digital Society Initiative" (DSI) der Universität Zürich. Die DSI fördert die kritische, interdisziplinäre Reflexion und Innovation bezüglich aller Aspekte der Digitalisierung von Wissenschaft und Gesellschaft.

---

In diesem Zusammenhang... (Werbung)



Dieser Artikel wohnt in einem Virtual Datacenter von Cyberlink.

---

Technologie-Partner



Gold-Sponsor

