

Aktuelles Schlagwort: Crowdsourcing

Tobias Hoßfeld, Matthias Hirth, Phuoc Tran-Gia
Universität Würzburg, Institut für Informatik, Lehrstuhl für Kommunikationsnetze
Am Hubland, 97074 Würzburg
Email: hossfeld@informatik.uni-wuerzburg.de

Von Machine Clouds zu Human Clouds

Seit der Öffnung des Internets für die Allgemeinheit Anfang der 90er Jahre hat eine rasante Entwicklung stattgefunden. Neue Paradigmen wie Peer-to-Peer (P2P), Web 2.0 oder Cloud Computing führen zu neuartigen Diensten und Anwendungen, welche bei den Anwendern längst etabliert sind und einen Großteil des Datenverkehrs im Internet ausmachen. Beispiele hierfür sind unter anderem P2P-Anwendungen wie BitTorrent zum Austausch riesiger Datenmengen, Skype für Sprach- und Videokonferenzen, Soziale Medien wie Facebook oder Twitter, Cloud Anwendungen wie DropBox als synchronisiertes Netzwerk-Dateisystem für verteilte Rechner oder Cloud Gaming.

Aktuell taucht ein neues Schlagwort im Internet auf: „Crowdsourcing“. Einige Aufgaben und Probleme, die für Menschen relativ einfach zu lösen sind, können derzeit selbst von modernen Machine Clouds noch nicht algorithmisch bewältigt werden. Hierzu zählen etwa Text- und Bilderkennung, das Verifizieren, Analysieren und Kategorisieren von Videoinhalten, das Schaffen von Wissen, das Verbessern und Kreieren von Produkten oder wissenschaftliche Forschung. Diese stellen Anwendungsgebiete von Crowdsourcing dar. Statt (oder zusätzlich zu) Machine Clouds wird die Masse der Internetnutzer in die Wertschöpfungskette eingebunden. Man spricht hier auch von **Human Clouds**.

Neben Sozialen Medien ist Crowdsourcing eine der wichtigsten aktuell aufstrebenden Technologien und Geschäftsmodelle im Internet, die die Zukunft des Arbeitens und der Arbeitsorganisation von Grund auf verändern wird. Die wirtschaftliche und gesellschaftliche Bedeutung von Crowdsourcing-Plattformen wächst ständig und fördert die Entstehung neuer Formen der Arbeitsorganisation. Jobs in Crowdsourcing-Plattformen besitzen eine viel kleinere Granularität im Vergleich zu denen im traditionellen Outsourcing bzw. Outtasking Bereich.

Das Aktuelle Schlagwort beleuchtet den Begriff „Crowdsourcing“ näher und wird zunächst wichtige Begriffe einführen, bevor die Anwendungsgebiete von Crowdsourcing sowie dessen Bedeutung in der Praxis und zukünftige Weiterentwicklung betrachtet werden.

Begriffe

Crowdsourcing: Der Begriff „Crowdsourcing“ wurde im Jahr 2006 von Jeff Howe eingeführt und stellt einen Neologismus aus den Wörtern „Crowd“ und „Outsourcing“ dar. Er bezeichnet *„the act of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an undefined, generally large group of people in the form of an open call“*.

Allgemein versteht man unter Crowdsourcing das Auslagern (Outsourcing) von Arbeiten und Leistungen an unbekannte Akteure (Crowd oder auch Human Cloud). Typischerweise handelt es sich bei dieser Human Cloud um eine große Anzahl von Internetnutzern.

Incentive Design: Um einen Akteur zum „Mitmachen“ zu motivieren, gibt es verschiedene Anreize, die in das Design der Aufgabenstellung integriert werden. Neben monetärer Bezahlung können das Weitervermitteln von Wissen, Einbringen von Kreativität, Spaß an der Arbeit, Freude am Problemlösen, Unterstützung Anderer oder Anerkennung durch Andere einen Anreiz darstellen. Der Einfachheit halber sprechen wir aber im Folgenden, unabhängig vom zugrundeliegenden Anreizmechanismus, von „Arbeitern“ und „Arbeitgebern“.

Crowdsourcing Plattform: Im Internet entsteht eine Vielzahl von Crowdsourcing-Plattformen. Diese bieten einen Marktplatz für die Vermittlung von Jobs von Arbeitgebern an Arbeiter und stellen zusätzlich Mechanismen bereit, die zur Abwicklung dieses Arbeitsverhältnisses wie z.B. Vergütung und Kontrolle der Qualität einer erbrachten Leistung dienen.

Open Innovation: Unter dem Begriff Open Innovation versteht man die Öffnung des Innovationsprozesses von Unternehmen oder anderen Institutionen, um Innovation innerhalb des Unternehmens zu schaffen. Aufgaben, die bislang intern bearbeitet wurden, werden an eine undefinierte, große Menge von Arbeitern vergeben. Der Unterschied zum Crowdsourcing besteht darin, dass ausschließlich Innovationen geschaffen werden sollen.

Open Call: Wird die zu bearbeitende Aufgabe oder die zu lösende Problemstellung öffentlich ausgeschrieben, spricht man von einem Open Call. Als Konsequenz daraus ergibt sich die Tatsache, dass die Arbeiter selbst entscheiden, welche Arbeiten sie annehmen.

Wie funktioniert Crowdsourcing eigentlich?

Crowdsourcing ist eine neue Form der Arbeitsauslagerung mit feinerer Job-Granularität. Anders als beim Outsourcing, bei dem Unternehmen Teilaufgaben an Drittunternehmen vergeben, werden beim Crowdsourcing die Aufgaben auf die Intelligenz und die Arbeitskraft einer großen Masse von unbekanntem Arbeitern im Internet ausgelagert. Die Unternehmen haben hierbei Zugriff auf eine Masse von Menschen, die ihnen im Unternehmen nicht zur Verfügung stehen würden. Aufgrund der häufigen Mikrobezahlung für Arbeitsaufgaben, die typischerweise im Bereich von einigen Cent bis einigen Euro liegt, spricht man in diesem Kontext auch bei den Arbeitern von **Microworkers** und von **Microtasks** für schnell zu erledigende Kleinstaufgaben (siehe Abbildung 1). Realisiert werden die hierfür benötigten Crowdsourcing-Plattformen mittels Web 2.0 Technologien.

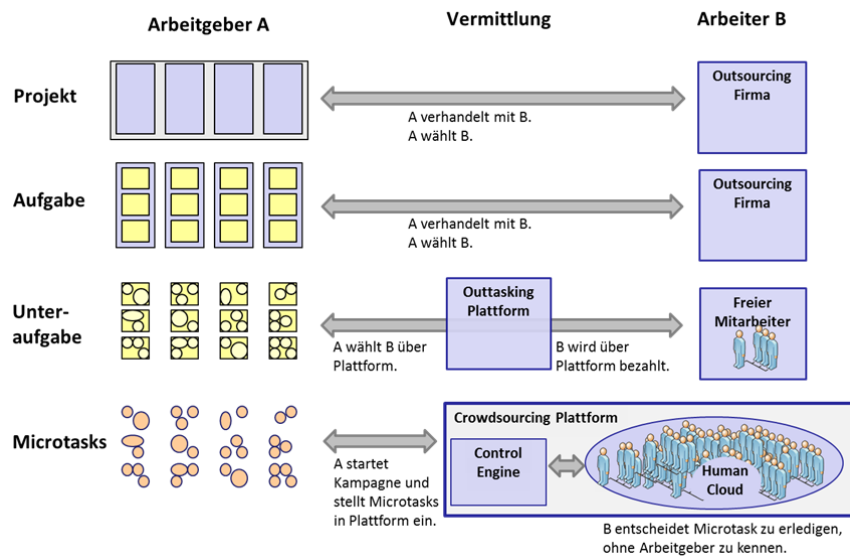


Abbildung 1. Evolution der Granularität von Arbeit und deren Organisation im Internet.

Der schematische Arbeitsablauf beim Crowdsourcing sieht wie folgt aus: Der Auftraggeber stellt einen Arbeitsauftrag in der Plattform online ein, der von einer Anzahl N unterschiedlicher Arbeiter durchgeführt werden soll. Dabei kann der Auftraggeber unterschiedliche Anforderungen an die anonymen Arbeiter stellen. Zum Beispiel kann für eine gezielte Meinungsumfrage das Herkunftsland oder die Muttersprache des Arbeiters relevant sein. Die Arbeiter, die diesen Anforderungen genügen, sehen die Arbeitsaufgabe in der Plattform und können diese Aufgabe bearbeiten, falls sie noch nicht von N anderen Arbeitern durchgeführt wurde. Sobald die Aufgabe von dem Arbeiter erledigt wurde, überprüft der Auftraggeber, ob die Aufgabe zufriedenstellend gelöst wurde, um dann die Zustimmung für eine erfolgreiche Arbeitsleistung zu geben.

Je nach Crowdsourcing-Plattform gibt es allerdings unterschiedliche Formen der Entlohnung für die Arbeit. Meist wird monetäre Bezahlung als Anreiz verwendet, um Arbeiter zum Arbeiten zu animieren. Andere Plattformen geben den Nutzern die Gelegenheit, ihr Wissen und Kreativität einzubringen, wofür diese in der Regel kein Geld, wohl aber Bekanntheit (z.B. Dell IdeaStrom) oder virtuelle Spielpunkte (z.B. Games with a Purpose) erhalten. Neben monetärer Entlohnung kann auch das Gefühl etwas Nützliches für die Allgemeinheit getan zu haben einen Anreiz bieten beim Crowdsourcing teilzunehmen, z.B. Wikipedia.

Was kann man mit Crowdsourcing machen? Welche Plattformen gibt es?

Die Aufgaben, für die Crowdsourcing-Plattformen genutzt werden können, lassen sich in drei unterschiedliche Kategorien klassifizieren: a) Routine-Aufgaben, b) komplexe Aufgaben und c) kreative Aufgaben. Die unterschiedlichen Kategorien lassen sich jedoch nicht immer eindeutig trennen, vielmehr gibt es einen fließenden Übergang zwischen ihnen.

Routine-Aufgaben: In die Kategorie der Routine-Aufgaben fallen Arbeiten, die mit geringem Aufwand, geringem Vorwissen und in der Regel mit wenigen Klicks im Internet erledigt werden können. Dazu gehören typischerweise die positive Bewertung von Videos, Artikeln, Blogbeiträgen, etc., um die Reputation des Auftraggebers im Internet zu steigern. Zu den Routine Aufgaben zählen auch bestimmte Aufgaben aus dem Bereich der Suchmaschinenoptimierung, deren Ziel es ist, bei Suchmaschinen wie Google höher gelistet zu werden. Hierzu verlinken Arbeiter die Webseite des Auftraggebers auf andere Webseiten und Blogs oder machen diese über Sozial Medien bekannt, um möglichst viele neue Webseitenbesucher zu werben. Ebenso gehört die manuelle Digitalisierung von Texten zu dieser Aufgabenkategorie, wie auch das Extrahieren von Daten aus Webseiten oder das Versehen von Bildern mit Schlagworten. Bekannte Plattformen, in denen vor allem Routine-Aufgaben angeboten werden, sind MicroWorkers.com und Mechanical Turk von Amazon.

Komplexe Aufgaben: Die zweite Kategorie umfasst die komplexen Aufgaben, z.B. das Generieren von Inhalten, Schreiben von Foren- oder Blogbeiträgen, Kommentieren von Inhalten, Schreiben von Produktrezensionen und die Teilnahmen an Nutzerumfragen. Auch das Testen von Webanwendungen und Software zum Finden von Fehlern und zur Verbesserung des Designs oder der Nutzbarkeit wird mit Hilfe von Crowdsourcing durchgeführt. Ein Beispiel für subjektive wissenschaftliche Studien, um die Wahrnehmung von Störungen bei Videoübertragungen zu untersuchen, findet sich in [Hoß11b]. Ein populäres Beispiel zum Generieren von Wissen ist das Online-Lexikon Wikipedia, welches von beliebigen Internetnutzern bearbeitet, erweitert oder verbessert werden kann. Die Nutzer erhalten für die Arbeiten keine monetäre Vergütung, sondern lediglich den Anreiz etwas Nützliches zu tun. Im Gegenzug ist Wikipedia für alle Internetnutzer frei zugänglich. Durch die freiwillige, kollektive Zusammenarbeit wird die Qualität der Beiträge nahezu immer gewährleistet. Allein durch die große Anzahl an Nutzern werden schlechte Beiträge, falsche Informationen oder Missbrauch des Dienstes schnell identifiziert und nahezu unmöglich gemacht. Man spricht hier von der Nutzung der kollektiven Intelligenz (**wisdom of the crowd**).

Kreative Aufgaben: Die dritte Kategorie der Arbeitsaufgaben umfasst die kreativen Aufgaben, wie Software- und Webentwicklung, das Lösen von komplexen Problemen und Forschungsfragen sowie die Bereitstellung kreativer Ideen. Eines der bekanntesten Beispiele ist die Crowdsourcing-Plattform Innocentive. Unternehmen können auf dieser Plattform Problemstellungen online einstellen, zu welchen dann von externen Experten Lösungsansätze vorgeschlagen werden. Unternehmen stehen dabei oft nicht der entsprechende Zeiträumen und das entsprechend spezialisierte Personal zur Verfügung, um solch ein Problem intern zu lösen. Eine andere Plattform ist Dell IdeaStorm, bei der Internetnutzer Vorschläge zu Dell-Produkten einreichen, bewerten und öffentlich diskutieren können. Die Internetnutzer werden hier zur Ideenfindung eingesetzt, damit das Unternehmen Verbesserungsvorschläge direkt von den Kunden erhält (Open Innovation). Als Anreizmechanismus wird ein Punktesystem eingesetzt, bei dem jeder Vorschlag von den Webseitenbesuchern bewertet werden kann. Eine monetäre oder anderweitige Vergütung des Nutzers findet nicht statt. Laut Dell konnten bisher über 300 Produktideen tatsächlich entwickelt werden und die Kosten für Marktforschung eingespart werden.

Crowdsourcing-Plattformen werden heute in verschiedensten Bereichen erfolgreich eingesetzt und auch von großen Unternehmen intensiv genutzt. Auf Plattformen, wie etwa Amazon Mechanical Turk, Microworkers.com oder Clickworker.com, die sich primär auf Aufgaben der Datenannotation, Übersetzungen, Klick-Jobs oder Umfragen fokussieren, werden täglich tausende Microjobs bearbeitet. Auf anderen Crowdsourcing Plattformen wie etwa tchibo ideas, die auf kreative Aufgaben spezialisiert sind, werden neue innovative Produkte entwickelt, die teilweise bereits nach wenigen Wochen im Handel erhältlich sind.

Wer versteckt sich hinter der Crowd?

Das Kapital einer Crowdsourcing-Plattform ist die zur Verfügung stehende Human Cloud. Eine erfolgreiche Plattform benötigt eine große und diverse Arbeitermasse, um einen dauerhaften rund-um-die-Uhr-Betrieb zu ermöglichen und vielfältige Aufgaben lösen zu können. Aber auch eine genügend große Anzahl von Arbeitgebern, die permanent neue Arbeit einstellen, ist notwendig. Untersuchungen zur größten derzeit existierende Crowdsourcing Plattform Amazon Mechanical Turk (MTurk) finden sich in [Ipe10, Ros09]. Allerdings ist MTurk sehr restriktiv bezüglich der zugelassenen Aufgabensteller und eine Auszahlung des erarbeiteten Betrages ist lediglich in den USA und Indien möglich. Ein Vergleich von MTurk und Microworkers.com, das sowohl internationale Arbeitgeber als auch Arbeitnehmer unterstützt, wurde in [Hir11] durchgeführt. Es zeigte sich, dass die Nutzer von Microworkers.com eine breitere Streuung über verschiedene Länder aufweisen als die von MTurk, obwohl auch hier 78% aller Nutzer aus lediglich 10 Ländern stammen. Die Entwicklungen der Anzahl registrierter Nutzer der Microworkers.com Plattform zeigt hierbei ein quadratisches

Wachstum über die Zeit [Hoß11], so dass der Crowdsourcing Boom in den nächsten Jahren noch zunehmen wird.

Welche wissenschaftlichen Herausforderungen stellen sich? Wie wird sich Crowdsourcing in der Zukunft weiterentwickeln?

Durch Crowdsourcing stellt sich eine Vielzahl wissenschaftlicher Herausforderungen aus unterschiedlichsten Disziplinen der Informatik. Beispielsweise werden Mechanismen zur **Qualitätssicherung** der durchgeführten Arbeiten benötigt. Die Qualität der Bearbeitung eines Jobs kann durch eine fehlende Qualifikation des Arbeiters, willentliches Fehlverhalten oder durch eine fehlerhafte Beschreibung von Jobs beeinträchtigt werden. Insbesondere komplexere Jobs lassen sich nur schwer klar und deutlich formulieren. Um die Verlässlichkeit von Ergebnissen zu erhöhen können bei einfachen Aufgaben Mehrheitsentscheide von Arbeitern verwendet werden [She08]. Bei komplexen Aufgaben eignen sich iterative Ansätze [Dai11, Lit09] oder spezialisierte Ansätze, bei denen beispielsweise die Arbeiter sich gegenseitig überprüfen [Kaz09] oder Expertenarbeiter hinzugezogen werden [Koc10]. Die Qualitätsgewinne und entstehenden Kosten werden in [Hir12] analytisch untersucht und die Anwendbarkeit der Ansätze auf verschiedene Aufgabenarten diskutiert. Eine andere Möglichkeit zur Qualitätssicherung liegt in einer abstrakten Beschreibungssprache für das (automatisierte) Design und Formulieren von Aufgaben. Des Weiteren sind geeignete **Anreizmechanismen** gefragt, um die Arbeiter zu guter Qualität zu motivieren, beispielsweise in Form von Spielen.

Die Zuordnung von Jobs zu Arbeitern erfolgt in aktuellen Plattformen, indem Arbeiter die Jobs nach verschiedenen Kriterien durchsuchen oder indem sie ihre Kompetenzen angeben und darauf basierend eine Filterung von angebotenen Jobs erfolgt. Um eine schnelle Bearbeitung von Kampagnen zu gewährleisten und die Benutzer im Finden von geeigneten Jobs zu unterstützen, sind **Empfehlungssysteme** wünschenswert.

Die zukünftige Entwicklung von Crowdsourcing geht unter anderem in Richtung **Mobile Crowdsourcing** oder **Ubiquitous Crowdsourcing** [Vuk10], wobei die Crowd als „mobiler Sensor“ agiert, um z.B. Umweltdaten einzusammeln. Die mobilen Sensoren können zusätzlich vernetzt sein, um sich untereinander zu koordinieren und gezielt Daten einzusammeln. **Real-time Crowdsourcing** betont, dass Aufgaben in Echtzeit erledigt werden, z.B. für Katastrophenmanagement. Ein anderes Beispiel ist die iPhone App VizWiz, die es blinden Menschen erlaubt, innerhalb kürzester Zeit Antworten auf Fragen zur momentanen Umgebung zu bekommen.

Literatur

- [Dai11] P. Dai, Mausam, D. Weld: Artificial Intelligence for Artificial Artificial Intelligence. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. San Francisco, USA, 2011.
- [Hir11] M. Hirth, T. Hoßfeld, P. Tran-Gia: Anatomy of a Crowdsourcing Platform - Using the Example of Microworkers.com. In: *Proceedings of the Workshop on Future Internet and Next Generation Networks (FINGNet)*, Seoul, Korea, 2011.
- [Hir12] M. Hirth, T. Hoßfeld, P. Tran-Gia: Analyzing Costs and Accuracy of Validation Mechanisms for Crowdsourcing Platforms. *Mathematical and Computer Modelling*, 2012.
- [Hoß11a] T. Hoßfeld, M. Hirth, P. Tran-Gia: Modeling of Crowdsourcing Platforms and Granularity of Work Organization in Future Internet. In: *Proceedings of the International Teletraffic Congress (ITC)*, San Francisco, USA, 2011.
- [Hoß11b] T. Hoßfeld, R. Schatz, M. Seufert, M. Hirth, T. Zinner, P. Tran-Gia: Quantification of YouTube QoE via Crowdsourcing. In: *Proceedings of MQoE 2011*, Dana Point, USA, 2011.
- [Ipe10] P. Ipeirotis: Analyzing the Amazon Mechanical Turk Marketplace. CeDER Working Papers, 2010.
- [Kaz09] G. Kazai, N. Milic-Frayling, J. Costello: Towards Methods for the Collective Gathering and Quality Control of Relevance Assessments. In: *Proceedings of the International Conference on Research and Development in Information Retrieval*. Boston, USA, 2009.
- [Koc10] S. Kochhar, S. Mazzocchi, P. Paritosh: The Anatomy of a Large-Scale Human Computation Engine. In: *Proceedings of Workshop on Human Computation*. Washington, USA, 2010.
- [Lit09] G. Little, L. Chilton, M. Goldman, R. Miller: TurkKit: Tools for Iterative Tasks on Mechanical Turk. In: *Proceedings of the Workshop on Human Computation*. Paris, France, 2009
- [She08] V. Sheng, F. Provost, P. Ipeirotis: Get Another Label? Improving Data Quality and Data Mining Using Multiple, Noisy Labelers. In: *Proceeding of the International Conference on Knowledge Discovery and Data Mining*. Las Vegas, USA, 2008.

[Vuk10] M. Vukovic, S. Kumara, O. Greenshpan: *Workshop on Ubiquitous Crowdsourcing* in 2010 and 2011.