

Konzepte zur Bewertung der Informationsqualität auf Plattformen mit nutzergenerierten Inhalten

Markus Specker, Alexander Krebs

Siemens AG
Siemens IT Solutions and Services (SIS), C-LAB
Fürstenallee 11, 33102 Paderborn
[markus.specker; alexander.krebs]@c-lab.de

Abstract: Dieser Beitrag stellt einen Ansatz vor, um Aussagen über die Informationsqualität (IQ) nutzergenerierter Inhalte auf Web 2.0-Plattformen machen zu können. Es wird ein Konzept einer nutzerbasierten IQ-Bewertung präsentiert, das im Rahmen eines laufenden Forschungsprojektes entstanden ist und mit Nutzern evaluiert wurde. In Kombination mit automatischen Messmethoden wird so eine solide Grundlage zur Bestimmung der IQ nutzergenerierter Inhalte geschaffen.

1 Einleitung/Hintergrund

Das Social Semantic Web ist in entscheidendem Maße auf nutzergenerierte Inhalte angewiesen. Mit ihnen steht und fällt der Erfolg entsprechender Plattformen. Ein wichtiger Aspekt ist dabei die Sicherstellung einer hohen Informationsqualität (IQ) solcher Inhalte. Denn wenn auf einer Plattform mit nutzergenerierten Inhalten die Beiträge insgesamt von einer hohen IQ sind und folglich den Nutzern in ihren individuellen Nutzungskontexten einen Mehrwert liefern, fördert dies die Attraktivität der Plattform. Dies wiederum ist eine Voraussetzung dafür, dass Nutzer regelmäßig zu der Plattform zurückkehren und diese langfristig weiterbenutzen. Daher sind entsprechende Verfahren zur Bewertung der IQ notwendig, mittels derer die Nutzer qualitativ hochwertige Beiträge von minderwertigen Inhalten unterscheiden können, und die als Ausgangspunkt für eine gezielte Verbesserung qualitativ minderwertiger Beiträge dienen.

Der Erforschung und Entwicklung derartiger Maßnahmen zur Messung und Verbesserung der IQ widmet sich Siemens IT Solutions and Services (SIS) im Rahmen des Projektes THESEUS [Th07]¹, einem vom Bundesministerium für Wirtschaft und Technologie (BMWi) initiiertem Forschungsprogramm, das das Ziel verfolgt, eine neue internetbasierte Wissensinfrastruktur zu entwickeln. Dabei geht es um eine Neuausrichtung und Weiterentwicklung des Internet zur besseren Erschließung und Nutzung von Wissen und seiner Verwertung in Form neuer Dienste. THESEUS ist in unterschiedliche Anwendungsszenarien unterteilt, die sich jeweils mit verschiedenen Fragestellungen

¹ Die nachfolgend dargestellten Forschungsarbeiten von SIS haben im Rahmen des THESEUS-Forschungsprogramms stattgefunden und werden vom Bundesministerium für Wirtschaft und Technologie (BMWi) unter dem Förderkennzeichen 01MQ07014 gefördert.

beschäftigen und allesamt an der Schnittstelle zwischen den Technologien des semantischen Web und der sozialen Vernetzungsphilosophie des Web 2.0 angeordnet sind.

Um auf nutzergeneriertes Wissen möglichst effizient und effektiv zugreifen und es nutzen zu können, existieren verschiedene Überlegungen zur automatischen Extraktion von semantischen Zusammenhängen in nutzergenerierten Beiträgen. Diese Ansätze dienen einerseits dazu, eine bessere Strukturierung des Wissens zu erhalten und so für Nutzer geeignete Instrumente zur Exploration von Informationen zur Verfügung zu stellen. Andererseits kann durch automatische Methoden die IQ der verschiedenen Beiträge gemessen und überwacht sowie in Verbindung mit technologiebasierten Verfahren eine qualitative Verbesserung der Inhalte sichergestellt werden.

Dennoch reicht (derzeit) die ausschließliche Anwendung automatischer Messmethoden nicht aus, um eine hohe Qualität der auf derartigen Plattformen verfügbaren Inhalte zu garantieren, da nach wie vor einzelne IQ-Kriterien existieren, die nicht oder nur sehr ungenau mittels algorithmischer Methoden erfasst werden können, z. B. weil regelmäßig ein hohes Maß an Subjektivität in die Bewertung einfließt. Es ist davon auszugehen, dass selbst mit neuesten Verfahren auf absehbare Zeit weiterhin der Mensch eine zentrale Rolle bei der Bewertung der IQ von nutzergenerierten Inhalten spielt, dass allerdings vermehrt algorithmenbasierte maschinelle Methoden unterstützend eingesetzt werden können und angesichts der immer weiter anwachsenden Menge an verfügbaren Informationen im Internet auch eingesetzt werden müssen. Daher postulieren wir wie auch [Pip02] einen ganzheitlichen Ansatz, der zur Beurteilung der IQ nutzergenerierter Inhalte beide Wege berücksichtigt: automatische algorithmenbasierte Messinstrumente sowie Bewertungsmethoden, in deren Rahmen eine Evaluation der Qualität von Informationen „händisch“ durch die einzelnen Nutzer erfolgt.

Ein weiterer Aspekt betrifft die Auswertung der IQ. Eine noch so gute Bewertungsmöglichkeit stiftet den Nutzern keinen Mehrwert, wenn diesen darüber hinaus nicht auch Instrumente zur Verfügung gestellt werden, mittels derer sie die Evaluationsergebnisse entsprechend auswerten können. Daher beschäftigt sich dieser Beitrag auch mit Fragen der Visualisierung und Interaktion für die Auswertung nutzergenerierter IQ-Bewertungen. Ungeachtet davon sind auch hier automatische Auswertungskonzepte von hoher Relevanz, da sie die Fragestellung eines semantischen Web adressieren.

2 Ein Konzept zur nutzerbasierten Informationsqualitätsbewertung

Um überhaupt Aussagen über die IQ von Nutzerinhalten treffen zu können, ist es notwendig, den Begriff „Informationsqualität“ genauer zu spezifizieren. Bisherige Plattformen (im Web 2.0-Bereich) bieten zwar ebenfalls verschiedene Verfahren zur Bewertung von Inhalten, diese sind in der Regel jedoch wenig ausdifferenziert und beschränken sich nicht selten auf die einfache Vergabe von Sternen, ohne dass der Nutzer genau weiß, welche Aspekte dieser bei seiner Evaluation berücksichtigen sollte (vgl. z. B. das Bewertungsverfahren auf www.ftd.de). Dies hat zur Folge, dass die entsprechenden Ergebnisse einer Bewertung nur einen geringen Aussagegehalt aufweisen, da hieraus nicht ersichtlich ist, warum ein Nutzer eine bestimmte Bewertung abgegeben

hat. So kann bei einer niedrigen Bewertung z. B. das in dem Artikel behandelte Thema nicht gefallen haben oder aber der Artikel kann fehlerhaft sein. Die genaue Ursache für die Bewertung bleibt anderen Lesern und dem Autor des Beitrags verborgen.

Zur Spezifizierung des Begriffs „Informationsqualität“ haben wir daher in einem ersten Schritt literaturbasiert Einzelfaktoren identifiziert, aus denen sich die Qualität einer Information zusammensetzt. Die für unseren Ansatz einer nutzerbasierte IQ-Bewertung herangezogenen Einzelkriterien beruhen dabei auf einer bei [Gr05] beschriebenen entscheidungstheoretischen Herleitung unterschiedlicher Dimensionen der IQ, wurden aber an den Kontext einer Bewertung der Qualität nutzergenerierter Inhalte auf Web 2.0-Plattformen angepasst. Bei den Kriterien, welche nachfolgend für eine Evaluation der IQ der verschiedenen Nutzerbeiträge zugrunde gelegt werden und eine differenziertere Aussage über die Qualität einer nutzergenerierten Information ermöglichen sollen, handelt es sich um die Kriterien Interpretation, Neuigkeit, Glaubwürdigkeit, Vollständigkeit, inhaltliche Relevanz und Aktualität. Für die genannten Kriterien wurde im nächsten Schritt ein Konzept entwickelt, das einerseits Benutzern das Bewerten der IQ von Beiträgen auf Basis dieser Kriterien ermöglicht, das ihnen andererseits aber auch das Auswerten und Interpretieren bereits abgegebener Bewertungen als Voraussetzung für die zielgerichtete Auswahl von einzelnen Beiträgen zur näheren Betrachtung erlaubt.

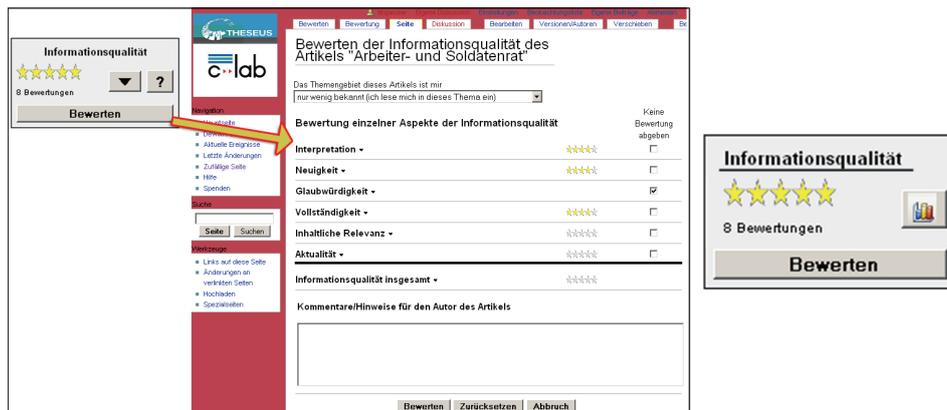


Abbildung 1: Konzept für eine nutzerbasierte IQ-Bewertung (links: Interaktionselemente auf Artikelseite, Mitte: detaillierte Bewertung, rechts: überarbeitete Interaktionselemente)

Das Bewertungssystem (vgl. hierzu Abbildung 1 links und Mitte) orientiert sich bei der Vergabe von Bewertungen an bekannten Bewertungssystemen auf Basis von Sternen, wobei ein Stern für eine schlechte Bewertung des jeweiligen Kriteriums steht und fünf Sterne der besten Bewertung entsprechen. Für die einzelnen Kriterien kann jeweils unabhängig eine bestimmte Anzahl an Sternen vergeben werden. Darüber hinaus können die Nutzer eine Bewertung der Gesamt-IQ des jeweiligen Beitrags vornehmen, die nicht automatisch auf Basis der im Hinblick auf die einzelnen Kriterien abgegebenen Bewertungen errechnet wird. Dieses Vorgehen liegt darin begründet, dass die Nutzer die Möglichkeit erhalten sollen, den Einfluss einzelner Kriterien auf die Gesamt-IQ eines Beitrags (falls gewünscht) nach individuellem Ermessen unterschiedlich zu gewichten.

Weiterhin kann ein Benutzer, falls dieser die Ausprägung eines bestimmten Kriteriums nicht einschätzen kann, dies entsprechend kennzeichnen und auf die Bewertung eines jeweiligen Kriteriums verzichten. So wird verhindert, dass Bewertungen zu einzelnen Kriterien (z. B. zum Kriterium Vollständigkeit) erzwungen werden, obwohl sich der Bewerter (bspw. angesichts seiner Unwissenheit in einem bestimmten Themengebiet) gar nicht in der Lage sieht, den Beitrag im Hinblick auf das jeweilige Kriterium zu evaluieren.

In der Regel interessiert sich ein Autor für die Gründe, die zu einer Abwertung seines Beitrags geführt haben, unter anderem auch deshalb, um gegebenenfalls seinen Beitrag entsprechend überarbeiten und korrigieren zu können. Die hier vorgestellten Einzelkriterien ermöglichen gegenüber einer ausschließlichen Bewertung der Gesamt-IQ zwar schon ein deutlich differenzierteres Bild, allerdings kann hierdurch bei einzelnen Kriterien, für deren Beurteilung mehrere unterschiedliche Aspekte zu berücksichtigen sind, noch immer keine ausreichende Detaillierung über die tatsächlichen Gründe einer niedrigen Bewertung erreicht werden. Daher sieht unser Bewertungskonzept im Falle einer Beurteilung eines Einzelkriteriums mit weniger als vier Sternen vor, auf freiwilliger Basis Gründe für das niedrige bis mittelmäßige Bewertungsergebnis aus einer Vorschlagsliste auswählen oder alternativ über ein Freitextfeld eingeben zu können.

Neben den vorgenannten Sachverhalten ist die Frage zu klären, wie die gesammelten IQ-Bewertungen zu den nutzergenerierten Inhalten anderen Nutzern als Informationsgrundlage für eine zielgerichtete Auswahl qualitativ hochwertiger Beiträge zugänglich gemacht werden sollen. Als Grundlage hierfür dienen einerseits Informationen über die durchschnittliche Gesamt-IQ sowie die Bewertungsergebnisse zu den einzelnen IQ-Kriterien, die z. B. im Rahmen der Suchergebnisliste angezeigt werden können. Neben einer Darstellung der durchschnittlichen Gesamt-IQ eines Beitrags kann sich der Nutzer eine grafische Veranschaulichung der Verteilung der Nutzerbewertungen zur Gesamt-IQ über die Sterne-Ausprägungen anzeigen lassen, die nach weiteren Kriterien, etwa nach dem Kenntnisstand des bewertenden Nutzers, differenziert werden kann. In einer weiteren Darstellung ist es zudem möglich, in einem Spinnennetzdiagramm die im Hinblick auf die einzelnen IQ-Kriterien abgegebenen Bewertungen anderer Nutzer zu analysieren.

3 Evaluation des Konzeptes zur nutzerbasierten Qualitätsbewertung

Um zu überprüfen, ob die oben genannten Kriterien von Nutzern zur Bewertung der IQ nutzergenerierter Inhalte akzeptiert werden, wurde eine Nutzerevaluation² durchgeführt. Dieser Evaluation lagen zwei Szenarien zugrunde: ein Szenario, in dem die Nutzer eine Beurteilung eines Artikels vornehmen und dabei das Bewertungskonzept ausprobieren und evaluieren sollten, und ein weiteres Szenario, in dem die verschiedenen Darstellungen der IQ-Bewertungen mit Nutzern evaluiert wurden. Das Konzept wurde je Szenario an fünf Probanden mit einer durchschnittlichen Dauer von ca. einer Stunde pro

² Für hilfreiche Kommentare sowie für die Unterstützung bei der Durchführung der Evaluation sind die Autoren ihrer Kollegin Therese Friberg zu besonderem Dank verpflichtet.

Proband getestet. Während der Evaluation wurden anhand eines Interviewleitfadens Fragen zu den verschiedenen Bereichen gestellt; zudem mussten die Probanden konkrete Aufgaben lösen.

Im ersten Szenario wurde ein modifiziertes Wiki benutzt, in welchem zu einem Artikel eine Bewertungsbox (vgl. hierzu Abbildung 1 links) eingeblendet wurde, die die Möglichkeit bietet, über einen „Bewerten“-Button zum Bewertungsfenster zu gelangen. Die befragten Personen wurden aufgefordert, den Artikel zu lesen und danach dessen IQ zu bewerten. Eine erste Hürde stellte die Bewertungsbox selbst dar. Einige Versuchspersonen interpretierten die Anzeige der bereits vergebenen Sterne als interaktives Tool und wollten hierüber auch ihre eigene Bewertung durchführen. Andere wiederum verstanden den Pfeil neben den Sternen als Bewertungsmöglichkeit per Dropdownbox, obwohl dieser eigentlich zu den detaillierteren Statistiken über die bisher abgegebenen Bewertungen führt. Auch das Fragezeichen, das eine Erklärung des Konzeptes der IQ liefert, war den Befragten nicht vollständig klar. Die genannten Erkenntnisse aus den Nutzertests flossen in eine überarbeitete Version der Bewertungsbox ein (vgl. Abbildung 1 rechts) (Pfeil durch Statistik-Symbol ersetzt; Fragezeichen durch Verlinkung des Begriffs „Informationsqualität“ ersetzt).

Auch zum Wording der einzelnen IQ-Kriterien lieferten die Probanden wertvolle Hinweise. Die Bewertungskriterien wurden den Nutzern in beiden Szenarien zunächst ohne weitere Erklärungen gezeigt. Die Probanden sollten beschreiben, was sie intuitiv unter den einzelnen Begriffen verstanden. Die Bezeichnung der verschiedenen Kriterien wurde dabei unterschiedlich gut verstanden. Der Begriff „Interpretation“ beispielsweise war den Nutzern nicht sofort klar. Auch war für die Nutzer der Unterschied zwischen Aktualität und Neuigkeit nicht immer eindeutig fassbar. Mit anderen Kriterien wie Glaubwürdigkeit oder Vollständigkeit hatten die Befragten indes kaum Probleme.

Anschließend wurden den Nutzern die entsprechenden Erklärungen zu den einzelnen Begriffen gezeigt. Die Testpersonen sollten beurteilen, wie verständlich diese Erklärungen sind und ob sie die Begriffe genau so verwenden würden, bzw. sie sollten alternativen Bezeichnungen vorschlagen. Durch die Ergänzung der Begriffe um die Erklärungen konnten die Nutzer den Sinn der einzelnen IQ-Kriterien deutlich besser verstehen. Dennoch sahen sie bei der Bezeichnung der Kriterien zum Teil noch Verbesserungspotenzial. Als Alternativvorschlag für das Kriterium Interpretation nannten einige Nutzer z. B. den Begriff „Verständlichkeit“, den andere wiederum als nicht trefflich empfanden. Diese tendierten eher zu Begriffen wie „(Inhaltliche) Struktur“, „Strukturelle Qualität“ oder „Dokumentenqualität“.

Der Bewertungsmechanismus selbst wurde positiv aufgenommen. Die Möglichkeit, bei einzelnen IQ-Kriterien keine Bewertung abzugeben, empfanden die Nutzer als sehr nützlich, denn es war ihnen nicht immer möglich, den Artikel im Hinblick auf sämtliche Kriterien zu bewerten. Allerdings wurde diese Option im Test teilweise übersehen, da sie relativ weit entfernt von der Bewertungsmöglichkeit platziert war. So gaben einige Nutzer als Ausweg eine mittlere Bewertung ab. Erst nach einem Hinweis, dass sie auch explizit wählen konnten, keine Bewertung abgeben zu wollen, machten sie von dieser Möglichkeit Gebrauch.

Ferner gab es weitere interessante Verbesserungsvorschläge, z. B. die Möglichkeit, die Bewertung der Gesamt-IQ zwar weiterhin von Hand durchführen zu lassen, aber zugleich einen automatisch auf Basis einer Gleichgewichtung der einzelnen Kriterien errechneten Vorschlagswert zu erhalten, den der Nutzer annehmen kann, aber nicht muss. Die Ergebnisse fließen zurzeit in die Überarbeitung des Bewertungsansatzes ein.

Im Ergebnis bestätigten die Nutzer insgesamt das vorgestellte Konzept zur Bewertung der IQ nutzergenerierter Inhalte. In überwiegender Mehrheit gaben diese an, dass sich der Ansatz aufgrund der vorhandenen Differenzierungen positiv von existierenden Ansätzen abhebe.

4 Fazit und Ausblick

Im Gegensatz zu den auf bisherigen Web-2.0-Plattformen angebotenen Instrumenten, mit denen Nutzer die verfügbaren Inhalte bewerten sollen, erlaubt das vorgestellte Konzept einer nutzerbasierten IQ-Bewertung ein deutlich differenzierteres Vorgehen. So können die Nutzer die Inhalte im Hinblick auf sechs verschiedene, bezüglich ihrer Relevanz für die IQ wissenschaftlich fundierte Einzelkriterien beurteilen. Für die Nutzer ist damit jederzeit ersichtlich, was genau sie im Rahmen der Bewertung beurteilen sollen. Dies führt zu zwei wesentlichen Vorteilen: Zum einen können die Nutzer die Ergebnisse der nutzerbasierten IQ-Evaluation dazu verwenden, Inhalte in Abhängigkeit von den ihnen subjektiv besonders wichtigen Kriterien (z. B. Glaubwürdigkeit) aus der Gesamtmenge an verfügbaren Informationen zielgerichtet zu selektieren. Zum anderen liefern die nach einzelnen IQ-Kriterien differenzierten Bewertungsergebnisse hilfreiche Anknüpfungspunkte für eine systematische Verbesserung der IQ der Beiträge, die sowohl manuell durch den Informationsbereinsteller als auch automatisch erfolgen kann.

Das vorgestellte Konzept einer nutzerbasierten IQ-Bewertung kann durch ein automatisches Verfahren zur Evaluation der IQ nutzergenerierter Inhalte sinnvoll ergänzt werden. Einerseits können auf diese Weise neue Inhalte, deren IQ von Nutzern bislang noch nicht bewertet wurde, mit einem Qualitätsurteil versehen werden. Andererseits wird es so möglich, bewusste Falschbewertungen von Beiträgen durch einzelne Nutzer, die nicht die tatsächliche IQ der Inhalte widerspiegeln, zu identifizieren, um die verantwortlichen Bewerter zu warnen oder gänzlich von der Plattform zu verweisen.

Literaturverzeichnis

- [Gr05] Gräfe, G.: Informationsqualität bei Transaktionen im Internet. Eine informationsökonomische Analyse der Bereitstellung und Verwendung von Informationen im Internet. Dt. Univ.-Verl., 1. Aufl, Wiesbaden, 2005.
- [Pip02] Pipino, L. L.; Yang, W. L.; Wan, R. Y.: Data quality assessment. Commun. ACM, Vol. 45 Nr. 4, ACM Press, New York, NY, USA 2002, S. 211-218.
- [Ta03] Tang, R.; Ng K. B.; Strzalkowski, T.; Kantor, P. B.: Automatically predicting information quality in news documents. Proceed. of the 2003 HLT/NAACL, Vol. 2, S. 97-99.
- [Th07] THESEUS-Beschreibung: <http://theseus-programm.de/was-ist-theseus>.