

Nutzung von GitHub für Open Educational Resources

Analyse zu Prozessen der Versionsverwaltung

Nadine Schröder¹, Peter Pfänder²

Abstract: Versionsverwaltung ist in der Softwareentwicklung weit verbreitet, wobei durchaus Potenziale für den Einsatz in anderen Bereichen bestehen, beispielsweise bei der Erstellung, Bearbeitung und Bereitstellung von Open Educational Resources (OER). Dieses Forschungsvorhaben untersucht in welcher Art und Weise GitHub als Plattform zur Versionsverwaltung zur Ablage und Nutzung von Lehr-/Lernmaterialien genutzt wird. Dabei finden insbesondere die Prozesse und Möglichkeiten der Versionsverwaltung Berücksichtigung, um der Frage nachzugehen, wie diese zu didaktischen Zwecken im Umgang mit Bildungsmaterialien genutzt und bei der kollaborativen Erstellung und Bearbeitung vorgegangen wird.

Keywords: Open Educational Resources, Infrastruktur, Versionsverwaltung, GitHub, Bildungsmaterialien, Lehr-/Lernmaterialien, Kollaboration

1 Einleitung und Hintergrund

In der Softwareentwicklung gilt Versionsverwaltung seit vielen Jahren als Standard, um Prozesse des gemeinsamen Arbeitens zu erleichtern und zu optimieren. Git ist ein etabliertes System zur verteilten Versionsverwaltung, auf dem die weit verbreitete webbasierte Plattform GitHub zur Bereitstellung von Softwarecode basiert. Versionsverwaltung dient zur Erfassung von Änderungen an Dokumenten und Dateien. Nachvollziehbarkeit ist vor allem bei kollaborativen Arbeitsprozessen innerhalb der Softwareentwicklung wertvoll, kann aber auch für die Erstellung und Bearbeitung von Lehr-/Lernmaterialien nützlich sein. Somit bietet sich GitHub auch als Plattform für die Ablage und den Austausch von Bildungsmaterialien an [Za15].

In diesem Forschungsvorhaben wird untersucht, in welcher Art und Weise GitHub für kooperative Erstellung, Bearbeitung und Bereitstellung von Lehr-/Lernmaterialien bereits genutzt wird. Ein Schwerpunkt liegt dabei darauf, inwieweit die Funktionalitäten und Workflows der Versionsverwaltung Anwendung finden, um Anforderungen für den Einsatz mit Bildungsmaterialien zu ermitteln. Die Analyse ist integriert in ein Forschungsvorhaben, bei dem untersucht wird, wie mit Versionen umgegangen werden kann, die bei

¹ Universität Duisburg-Essen, Learning Lab, Universitätsstr. 2, 45141 Essen, nadine.schroeder2@uni-due.de, <https://orcid.org/0000-0001-6650-4656>

² Universität Duisburg-Essen, Learning Lab, Universitätsstr. 2, 45141 Essen, peter.pfaender@uni-due.de

einer dezentralen und kooperativen Nutzung und Bearbeitung von Open Educational Resources (OER) entstehen. Das übergeordnete Forschungsprojekt beschäftigt sich mit einer verteilten Infrastruktur, um die Vernetzung und Verteilung von Open Educational Resources unter den deutschen Hochschulen zu fördern. Durch die Vernetzung von verschiedenen Hochschulsystemen über ein zentrales System und den Austausch von Metadaten sollen Auffindbarkeit und Zugriff auf Ressourcen standortübergreifend ermöglicht werden [Ke19].

2 Versionsverwaltung und Open Educational Resources

Die Idee von Open Educational Resources (OER) bezieht sich auf den Zugang zu Lehr-/Lernmaterialien nach dem Konzept von Open Science, der Forderung nach der Öffnung von Bildung und Wissenschaft, unter dem auch Open Source gefasst wird [FF14]. OER beruhen auf dem Konzept der 5 R, die verdeutlichen, welche Möglichkeiten offene Lizenzen beim Umgang mit Materialien bieten [Wi14]:

Retain – the right to make, own, and control copies of the content

Reuse – the right to use the content in a wide range of ways

Revise – the right to adapt, adjust, modify, or alter the content itself

Remix – the right to combine the original or revised content with other open content to create something new

Redistribute – the right to share copies of the original content, your revisions, or your remixes with others

Im Zusammenhang mit Versionen spielen vor allem die Aspekte der Nachnutzung (reuse), Änderung (revise) und Vermischung (remix) eine Rolle, indem durch die Bearbeitung von OER neue Versionen erstellt und kollaborativ weiterentwickelt werden können. Plattformen zur Versionsverwaltung, wie GitHub, umfassen mit den Möglichkeiten des Kopierens (retain) und Teilens (redistribute) von Materialien die weiteren Aspekte der 5Rs. Somit stellen OER einen Anwendungsfall für Versionsverwaltung dar.

2.1 Prozesse der Versionsverwaltung

Prozesse von Versionsverwaltungssystemen³, die in der Entwicklung von Software eingesetzt werden, können auf verschiedene Anwendungsszenarien von Lernressourcen übertragen werden, sodass sich Einsatzszenarien für OER und der kollaborativen Erstellung und Nutzung von Materialien in Lernumgebungen anbieten. Mit der *Fork*-Methode können Projekte kopiert und unabhängig weiterentwickelt werden. Damit können auch Änderungen und individuelle Anpassungen an Lernressourcen vorgenommen werden. Bearbei-

³ Hier am Beispiel von GitHub dargestellt

ter eines Forks können ihre Änderungen mittels eines *Pull Requests* an die Original-Ressource zurückspielen. Der Urheber entscheidet schließlich über die Annahme der Änderungen, die durch einen *Merge* mit dem Original zusammengeführt werden können. Darüber hinaus können Änderungen des Originals automatisch an Forks übermittelt werden, was einen wesentlichen Vorteil gegenüber der Weitergabe einer Datei darstellt. Kooperatives Arbeiten wird ebenso mit *Branches* unterstützt, indem unabhängig voneinander erstellte Elemente in ein Material integriert werden können. Die Speicherung von Änderungen in *Commits* und die Nachvollziehbarkeit durch eine Versionshistorie können darüber hinaus für das gemeinsame Arbeiten an Dokumenten förderlich sein.

2.2 Anwendungsfälle der Versionsverwaltung für OER

Bei der Versionsverwaltung von OER lassen sich verschiedene Szenarien unterscheiden, die mit der Anzahl der Ressourcen und der Ersteller bzw. Bearbeiter zusammenhängen. Neue Versionen entstehen bei der Arbeit von einem oder mehreren Urhebern an einem Material, wo vor allem Änderungskommentare und Versionshistorie zur Nachvollziehbarkeit bei der kollaborativen Arbeit hilfreich sein können. Andere Szenarien der Versionsverwaltung beziehen sich auf die Kopie einer Ressource durch einen Nachnutzenden bzw. Bearbeitenden, sodass zwei Ressourcen von unterschiedlichen Personen existieren. Dies kann mit einem Fork gleichgesetzt werden.

Auch in Kursumgebungen sind die kollaborativen Funktionalitäten von GitHub für die Interaktion zwischen Lernenden und Lehrenden einsetzbar. Zwar spiegelt GitHub als Plattform zur Ablage von Kursmaterialien die Funktion von Lernmanagementsystemen wider, geht aber insofern über das Teilen von Ressourcen hinaus, dass sowohl Studierende als auch andere Lehrende die Möglichkeit haben, die Materialien direkt zu nutzen und zu ändern. Durch Bearbeitungen oder Feedback können sie einen Beitrag leisten und werden in die Entwicklung von Materialien einbezogen. [Za15]

3 GitHub-Analyse

Um Möglichkeiten der Anwendung der Versionsverwaltung für Open Educational Resources auszuloten, wurde GitHub als Plattform mit aktiver Verwendung im Bereich der Softwareentwicklung ausgewählt. Bei der Analyse sollen die Prozesse der Versionsverwaltung untersucht werden, um der Frage nachzugehen, wie diese beim Umgang mit Bildungsmaterialien genutzt und wie häufig diese Prozesse angewendet werden, um Rückschlüsse auf das Nutzungs- und Interaktionsverhalten im Umgang mit OER zu ziehen.

3.1 Vorgehen

Zunächst wird auf das Konzept der GitHub Topics zurückgegriffen, bei dem Repositories von Projektbesitzern mit einem oder mehreren Topics gelabelt werden können, sodass eine

thematische Ausrichtung und inhaltliche Überschneidungen sichtbar werden [Gi20]. In der Analyse wurden 33 Topics ermittelt, die mit Lehr-/Lernmaterialien in Bildungskontexten in Zusammenhang stehen. Diese Topics umfassen Bildungsressourcen zu Kursen, Vorlesungen oder Workshops. So konnten 3325 zugehörige Repositories ermittelt werden. Diese wurden über die GitHub API gesammelt, bei der es sich um eine REST API handelt, die Zugriff aus einem Programm heraus ermöglicht, sodass Daten automatisiert erfasst werden können [Gi20b]. Es ist zu beachten, dass Topics von Projektbeteiligten individuell vergeben werden, sodass weitere Repositories existieren, die die Kriterien eines Bildungsmaterials erfüllen, aber ohne entsprechende Topics nicht Teil der Stichprobe sind.

Um Bildungsmaterialien zu identifizieren, wurden die Repositories aller Topics zunächst einer manuellen Analyse unterzogen. Hierzu wurde auf die README-Files der Repositories zurückgegriffen. Dabei handelt es sich um eine zentrale Informationsquelle auf GitHub, da jeder Ersteller eines Repositories dazu aufgefordert wird in einer autogenerierten Datei sein Projekt näher zu beschreiben. Kriterien bei der Auswahl der Repositories als Bildungsmaterialien umfassen einen thematischen Bildungs- bzw. Lehr/Lernschwerpunkt, Formate, wie Vorlesungen, Workshops, Kurse oder Selbstlernmaterialien sowie Materialarten für Lehr/Lernressourcen, wie Textdokumente, Präsentationsfolien, Lehrbücher oder Notebooks. Dabei konnten 1456 Repositories als Bildungsmaterialien identifiziert werden. Diese wurden näher nach ihrer Anzahl an Forks, Stars, Contributor, Branches, Commits, Issues und Pull Requests analysiert. Diese Daten konnten mit Hilfe eines Python-Skriptes gesammelt werden, welches das Package PyGitHub [Py20] nutzt, um die GitHub API in Python abzubilden sodass GitHub direkt aus einem Python-Skript gecrawlt werden konnte.

3.2 Datenauswertung

Um die Anwendung der Szenarien der Versionsverwaltung für OER zu untersuchen, wurde die Auswertung der Daten zunächst auf die Nachnutzung in Form von Forks, die Erstellung von Versionen mit Commits und die Kooperation von mehreren Personen fokussiert.

Die Anzahl der projektbeteiligten Mitarbeiter geben Hinweise darauf, inwieweit kollaboratives Arbeiten bei der Erstellung von Materialien auf GitHub stattfindet. Während 65% der Repositories von einer Person gepflegt werden, arbeiten bei 30% zwei bis zehn Personen an demselben Projekt. In welchem Umfang Änderungen (Commit) in einem Repository vorgenommen werden, hängt auch mit der Anzahl der Mitarbeitenden zusammen, die an einem Projekt arbeiten (Contributer). So beträgt die durchschnittliche Anzahl der Commits pro Contributor 43,8. Es zeigt sich, dass die Vorteile von GitHub Anwendung finden, da sowohl Materialien von mehreren Personen bearbeitet werden als auch Änderungen an den Materialien vorgenommen und Versionen erstellt werden.

Die Anzahl der Forks pro Repository geben Hinweise auf die Nutzung von Materialien von externen Personen. Abbildung 1 zeigt, dass 593 Repositories, also ca. 40% nie geforkt

wurden. Weitere 42% der Repositories wurden in geringem Umfang, 1- bis 10-mal, geforkt. Nur bei einzelnen beliebten Repositories existiert eine Anzahl von Forks im 1000er Umfang. Diese häufig geforkten Repositories weisen im Vergleich zu den zugehörigen Topics eine inhaltliche Gemeinsamkeit auf. Es handelt sich dabei überwiegend um Materialien, die mit Kursen an Hochschulen oder auf MOOC-Plattformen in Verbindung stehen oder auch Lehrbücher ergänzen.

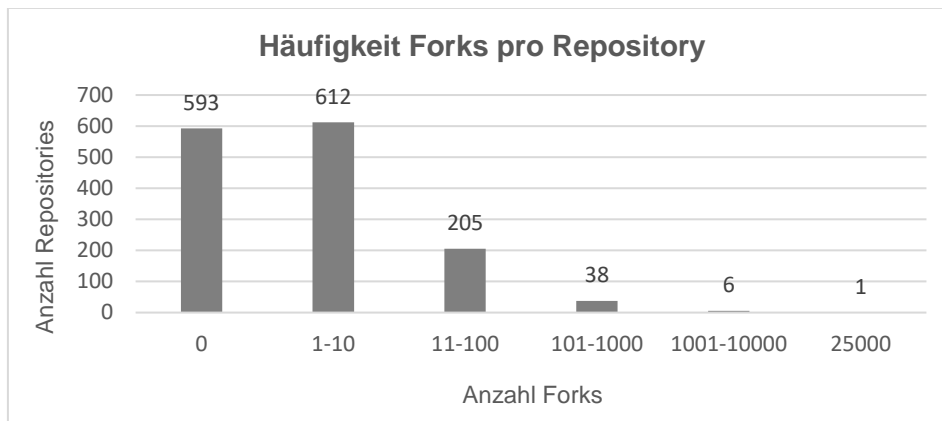


Abb. 1: Häufigkeit von Forks pro Repository

Bei diesen Ergebnissen lassen sich Parallelen zu Softwareprojekten insofern ziehen, dass es viele Projekte mit wenigen Forks und wenige Projekte mit vielen Forks gibt [LRM14]. Hieraus lässt sich allgemein die Schlussfolgerung ableiten, dass bestimmte Materialien besonders häufig von vielen Personen nachgenutzt werden.

4 Zusammenfassung und Ausblick

Die bisherigen Ergebnisse der GitHub-Analyse zeigen, dass die Plattform auch für Bildungsmaterialien verwendet wird, wobei jedoch überwiegend die Möglichkeit der Ablage und Bereitstellung von Ressourcen genutzt wird. Änderungen finden meist von einer Person, dem Projektbesitzer, statt, aber auch in Kooperation von mehreren Personen. Die weiteren Vorteile der Versionsverwaltung, wie Nachnutzung und Bearbeitung von Materialien durch externe Personen, werden in einem geringen Umfang eingesetzt.

Als ein weiterer Schritt bei der Analyse wird angestrebt, die Forks der analysierten Repositories zu untersuchen, um Hinweise auf den Umfang von Änderungen und Weiterentwicklungen sowie die Interaktionen mit dem Original, wie der Zurückführung von Änderungen zu erhalten. Um weitere Repositories zu ermitteln, bei denen es sich um Bildungsmaterialien handelt, muss über das Konzept der Topics hinausgegangen werden. Ein Ansatz besteht darin, die identifizierten Repositories als Grundlage zu nutzen, um einen Clas-

sifiziert nach [Pr18] zu trainieren, damit Repositories automatisch klassifiziert und als Bildungsmaterialien erkannt werden können. Als möglichen Ansatzpunkt für den Algorithmus bieten sich die Repositories an, die mit einer CC-Lizenz versehen sind, um festzustellen, inwieweit hier Bildungsmaterialien enthalten sind.

Die analysierten Bildungsmaterialien weisen einen inhaltlichen Schwerpunkt zu relevanten Themen der Softwareentwicklung auf, was hinsichtlich der Zielgruppe von GitHub nachvollziehbar ist. Daher kommt es bei nicht-technischen Anwendern häufig zu dem Hindernis, dass Funktionalitäten schwer verständlich sind [Za15]. Damit die Potenziale der Versionsverwaltung für die Anwendung von OER in Anlehnung an die 5 R genutzt werden, kann eine Lösungsoption sein, die Prozesse der Versionsverwaltung in eine Nutzeroberfläche zu überführen, die Anwendern außerhalb der Softwareentwicklung den Zugang erleichtert [Ov19]. Dieser Ansatz soll im weiteren Forschungsvorhaben fortgeführt werden.

Danksagung: Wir bedanken uns bei Johannes Wendt und Dominik Dyba für ihre Unterstützung bei der Prüfung der Repositories und des GitHub-Crawlings.

Literaturverzeichnis

- [FF14] Fecher, B.; Friesike, S.: Open Science: One Term, Five Schools of Thoughts. In (Bartling, S.; Friesike, S.): *Opening Science*. Springer, Cham, S. 17-47, 2014.
- [Gi20] GitHub.com: Classifying your repository with topics. <https://help.github.com/en/github/administering-a-repository/classifying-your-repository-with-topics>, Stand: 22.06.2020.
- [Gi20b] GitHub Developer: REST API v3. <https://developer.github.com/v3/>, Stand: 22.06.2020
- [Ke19] Kerres, M. et al.: EduArc. Eine Infrastruktur zur hochschulübergreifenden Nachnutzung digitaler Lernmaterialien. In: *Synergie. Fachmagazin für Digitalisierung in der Lehre* #07, S. 66 – 69, 2019.
- [LRM14] Lima, A.; Rossi, L.; Musolesi, M.: Coding together at a scale: GitHub as a collaborative social network. In: *Proc. ISWSM*, S. 295-300, 2014.
- [Ov19] Ovia, Steven: Addressing the Technical Challenges of Open Educational Resources. In: *portal: Libraries and the Academy* 19/1, S. 79-93, 2019.
- [Pr19] Prana, G.A.A. et al: Categorizing the Content of GitHub README Files. In: *Empirical Software Engineering*, 24, 1296–1327, 2019.
- [Py20] PyGithub. <https://pygithub.readthedocs.io/en/latest/>, Stand: 22.06.2020
- [Wi14] Wiley, D.: The Access Compromise and the 5th R. In: *Iterating Toward Openness* <http://opencontent.org/blog/archives/3221>, 2014.
- [Za15] Zagalsky, A et al: The Emergence of GitHub as a Collaborative Platform for Education. In: *Proc. CSCW '15*, S. 1906-1917, 2015.