

Patrick Reuter

Mail: patrick.reuter@campus.tu-berlin.de

21. November 2016

Expose: Bachelorarbeit über CRDTs

Implementierung einer Library

Kollaboratives Arbeiten ist mittlerweile ein wichtiger Bestandteil im Umfeld von verteilten Systemen. Bekanntestes Beispiel dafür ist Google Docs, welches es ermöglicht gemeinsam an Dokumenten, Präsentationen und Tabellen zu arbeiten. Die zugrundeliegende Technologie, Operational Transformation, ist allerdings schwierig zu implementieren und sehr komplex. Eine mögliche Alternative sind Conflict-Free Replicated Data Types (CRDTs). Diese Datentypen garantieren, dass am Ende alle Kollaborateure in einem verteilten System den gleichen Zustand haben. Gleichzeitig sind sie einfacher zu verstehen und umzusetzen.

Ziel der Bachelorarbeit soll es sein eine Software-Bibliothek zu schreiben, welche die Einbindung und den Einsatz von CRDTs ermöglicht.

Dazu sollen im ersten Schritt Objekte definiert werden, die technisch realisierbar sind (z.B. Counter, Register und Sets). Neben den eigentlichen Daten müssen die Objekte zudem über Funktionen verfügen, die es erlauben die Daten zu lesen und zu schreiben. Außerdem muss der aktuelle Zustand, also die eigene Version der Daten, bei Änderungen an die anderen Teilnehmer verschickt werden. Die Übermittlung kann entweder über einen Server oder mittels WebRTC direkt realisiert werden. Sind die Daten erst einmal beim Kollaborateur angekommen muss er über eine Funktion verfügen, die es ihm ermöglicht, dass er den erhaltenen Zustand mit seinem eigenen verschmilzt („merged“). So kann eine „strong, eventual consistency“ erreicht werden. Das bedeutet, dass am Ende alle Kollaborateure im Besitz einer identische Version des Objekts sind.

Im nächsten Schritt sollen Beispielanwendungen auf Basis der implementierten Library programmiert werden. Dabei soll es sinnvolle Beispiele für die Clients (also die Kollaborateure) geben. Diese sollen in Form einer Webapplikation mit einem Server kommunizieren, der entweder die Zustände zwischen den Clients versendet oder die Vermittlung für die Direktkommunikation via WebRTC übernimmt. Da WebRTC technisch

noch nicht ausgereift ist, ist hier noch nicht genau abzuschätzen, inwieweit eine konkrete Implementierung dieser Server-Variante möglich ist.

Sind die Library und die Beispielanwendungen fertig muss dann eine Dokumentation erstellt werden, um den Einsatz der CRDTs einfach und schnell zu gestalten. Dabei soll der Fokus auf Verständlichkeit gelegt werden, um die Hemmschwelle für die Nutzung dieser Datentypen zu senken und somit zu ihrer Weiterverbreitung beitragen zu können.

Im letzten Schritt bedarf es dann auch noch der Evaluation. Dabei kann analog vorgegangen werden zur Performance-Messung von Dang und Ignat¹, welche die Performance bei einer steigenden Zahl von Google-Docs-Nutzern untersucht haben. Zudem wird zu überprüfen sein, ob die Daten innerhalb des Systems konsistent sein werden. Dies kann erreicht werden, indem einige Tests durchgeführt werden, in denen das kollaborative Arbeiten simuliert und am Ende die Zustände der einzelnen Clients überprüft werden.

Abschließend lässt sich festhalten, dass CRDTs noch eine relativ geringe Verbreitung haben. Diese Bachelorarbeit soll dabei helfen die Nutzbarkeit dieser Datentypen zu fördern und Eintrittsschwellen abzubauen. So soll dazu beigetragen werden, dass auch zukünftig noch mehr kollaborative Webanwendungen entwickelt werden.

¹ Quang-Vinh Dang, Claudia-Lavinia Ignat. Performance of real-time collaborative editors at large scale: User perspective. Internet of People Workshop, 2016 IFIP Networking Conference, May 2016, Vienna, Austria. pp.548-553, 2016, Proceedings of 2016 IFIP Networking Conference, Networking 2016 and Workshops.