

# ReqCheck: Automatisierte Qualitätssicherung von Anforderungsdokumenten mit Hilfe linguistischer Regeln

Alexander Klaus, Matthias Koch  
Fraunhofer IESE, Kaiserslautern

[Alexander.Klaus@iese.fraunhofer.de](mailto:Alexander.Klaus@iese.fraunhofer.de), [Matthias.Koch@iese.fraunhofer.de](mailto:Matthias.Koch@iese.fraunhofer.de)

Ursula Reuther, Gernot Hensberg  
IAI, Saarbrücken

[ursula.reuther@iai.uni-sb.de](mailto:ursula.reuther@iai.uni-sb.de), [gernot.hensberg@iai.uni-sb.de](mailto:gernot.hensberg@iai.uni-sb.de)

## Beitrags-Kategorie

Forschungsergebnisse

## Zielgruppe des Beitrags

RE-Anwender, RE-Interessierte

## Motivation der behandelten Fragestellung

Schriftliche Dokumentation wird traditionell in der Software-Entwicklung verwendet. Insbesondere natürlichsprachliche Dokumentation findet in der heutigen Praxis nach wie vor weitverbreitete Anwendung [1]. Jedoch besteht gerade bei dieser Form der Anforderungsdokumentation die Problematik, dass qualitative Mängel zusätzliche Nacharbeiten verursachen, etwa wenn Anforderungen nicht eindeutig spezifiziert sind, und daher Interpretationsspielraum herrscht. Solche Mängel sorgen durch die Notwendigkeit von Nachfragen mittels fortwährender mündlicher Kommunikation für einen erhöhten Aufwand und für höhere Kosten gegenüber qualitativ hochwertigen Anforderungen. Eine manuelle Prüfung der Dokumentation ist jedoch gerade bei großen Dokumenten sehr zeitintensiv und damit nur mit großem Aufwand möglich.

## Behandelte Fragestellung oder Problem

Vor diesem Hintergrund wurde untersucht, durch welche Maßnahmen die Überprüfungs- und Lektoratsaufwände reduziert und der Prozess der Qualitätssicherung vereinfacht werden können. Ein maschinelles Lektorat auf Basis einer linguistisch intelligenten Analysemethode liefert in jedem Fall einen größeren Datendurchsatz, doch die Frage stellt sich, ob die Ergebnisse verlässlich, korrekt und verwertbar sind. Weiterhin ist offen, wie ein solches maschinelles Lektorat dem Anwender zugänglich gemacht und in den Prozess der Qualitätssicherung sinnvoll eingebunden werden kann.

## Lösung und Ergebnisse

Das Institut für Angewandte Informationsforschung (IAI) und das Fraunhofer-Institut für Experimentelles Software Engineering (IESE) haben daher Wissen aus der Prüfung von technischer Dokumentation [2], [3], [4] und über die Qualität von Anforderungen gebündelt,

und eine Lösung erarbeitet, die automatisiert natürlichsprachliche Dokumente auf ihre Qualität prüft.

Einen Grundpfeiler der technischen Umsetzung der automatisierten Anforderungsprüfung bildet eine vom IAI entwickelte und vertriebene Software, die Dokumentationen hinsichtlich sprachlicher Regeln überprüft und Befunde markiert. Dabei wird das Dokument auf Basis einer Interpretation der Dokumenttypdefinition (DTD, Regeln zur Strukturierung von Dokumenten) in prüfungsrelevante Einheiten aufgeteilt. Diese werden linguistisch analysiert und die fehlerhaften Stellen im Text markiert. Die DTD-Interpretation ermöglicht eine je nach Strukturelement unterschiedliche Analyse und Korrektur der einzelnen Prüfeinheiten.

Typischerweise liegen Anforderungsdokumente im Microsoft Office Word-Format (docx) vor, da Office-Werkzeuge wie Word nach wie vor die am weitesten verbreiteten Werkzeuge im RE darstellen [5]. Damit das System zur Prüfung auf sprachliche Qualität ein solches Word-Dokument im Kontext der Anforderungsspezifikationen verarbeiten kann, muss es zunächst in ein HTML-Dokument konvertiert werden. Dieses wird im Anschluss vom System analysiert und um Tags angereichert, die die Befunde markieren. Diese Tags enthalten unter anderem auch die Information, welche Regel an dieser Stelle angeschlagen hat und erlaubt es somit, genaue Erläuterungen zum mutmaßlichen Fehler sowie Hinweise, wie dieser behoben werden kann, einzufügen. Diese Informationen werden als Popup im HTML-Dokument angezeigt, sobald sich die Maus über eine entsprechend markierte Befundstelle bewegt. Des Weiteren gibt es eine Möglichkeit, eine druckbare Version des Dokuments zu generieren, in der die Befunde getrennt vom eigentlichen Inhalt aufgelistet werden. Die Zuordnung erfolgt dann über eindeutige Bezeichner.

Die bestehende, vom IAI entwickelte, Plattform zur Prüfung technischer Dokumentationen macht Gebrauch von linguistischen Regeln, mittels derer problematische Textpassagen erkannt werden. Um die Prüfung von Anforderungsdokumenten zu realisieren, ist zunächst die Anpassung des von diesem System verwendeten Regelwerks notwendig. Für diesen Zweck wurden aus einer Menge allgemeiner Verständlichkeitsregeln mit Hilfe

von Experten diejenigen Regeln extrahiert, die für das RE von entscheidender Bedeutung sind. So werden unter anderem der Satzbau und die Komplexität, der Schreibstil, die Wortwahl und mögliche Auslassungen von Wörtern einer Prüfung unterzogen.

Die Untersuchung von Anforderungsspezifikationen wird somit insbesondere mit Hilfe der Regeln durchgeführt, die beispielsweise auf die Eindeutigkeit und Verständlichkeit abzielen. Diese Kriterien zählen laut ISO/IEC/IEEE 29148 [6] zu den Qualitätskriterien für einzelne natürlichsprachliche Anforderungen. Zudem werden Substantive ohne Bezugsindex oder unvollständig spezifizierte Bedingungen als problematisch eingestuft. Einer Untersuchung von Kandt zufolge sind die am weitesten verbreiteten Probleme im Zusammenhang mit natürlichsprachlichen Anforderungen „weak words“, die Nichteinhaltung der Singularität von Anforderungen sowie lange und kompliziert formulierte Anforderungen [7]. Folgerichtig werden auch diese Problematiken durch die Prüfung angegangen.

Durch Konfiguration eines entsprechenden Profils für das System wird sichergestellt, dass die zuvor extrahierten Regeln zur Prüfung von Anforderungsdokumenten angewendet werden.

Zur Evaluierung der Prüfregeln und der durch die Regeln monierten Fundstellen wurde eine Fallstudie mit 9 Dokumenten aus unterschiedlichen Branchen mit insgesamt über 53.000 Wörtern durchgeführt. Die Fallstudie ergab 542 Treffer, von denen lediglich 16 bei einer zusätzlich manuell durchgeführten Prüfung durch vier RE-Experten und Linguisten als false positive eingestuft wurden. Dies entspricht einer Precision von 97,05%. Der Recall war im Rahmen dieser Fallstudie kein Untersuchungsgegenstand.

### Was ist neu?

Neu bei dem hier vorgestellten Ansatz zur Qualitätssicherung von Anforderungsdokumenten ist der Einsatz von linguistischen Verarbeitungsmethoden, die auf einer tiefen morpho-syntaktischen Analyse [8] beruhen. So werden Verstöße gegen die Vorgaben der RE-Spezifikationen [6] wie beispielsweise Eindeutigkeit, Verständlichkeit und Konsistenz erkannt, die weit über die Überprüfung wortbasierter Regeln (weak words) und einfacher Kookkurrenzmuster hinausgehen.

Weiterhin erlaubt die linguistische Analyse auf Basis der DTD-Interpretation eine spezifische Überprüfung unterschiedlicher Dokumentarten sowie eine Tag-abhängige Überprüfung unterschiedlicher Textelemente.

Ebenfalls als neu anzusehen ist die Möglichkeit, die Prüfung von Anforderungsdokumenten ohne Installation eigenständiger Software auf Unternehmensrechnern durchführen zu lassen: Das System zur Prüfung von Dokumenten ist vollständig über eine Webseite ansprechbar, d.h. ein Dokument kann über eine gesicherte Verbindung an den Server übermittelt werden. Das Ergebnis wird wahlweise direkt auf der Webseite angezeigt mit der Möglichkeit, die HTML-Seite zu speichern

oder eine PDF-Datei mit den Befunden herunterzuladen, oder die entsprechenden Dateien können automatisiert per Mail an den registrierten Nutzer versendet werden.

### Bisherige Arbeiten und Vorträge über das Thema

Die Vorarbeiten des IAI beinhalten die Entwicklung des Systems zur Prüfung technischer Dokumentationen hinsichtlich sprachlicher Qualität. Insbesondere hinsichtlich linguistischer Forschungsarbeiten sind Publikationen auf nationalen [9] und internationalen [10] Konferenzen veröffentlicht worden. Das IESE beschäftigt sich in Forschungsarbeiten mit dem Thema der automatisierten Prüfung natürlichsprachlicher Anforderungen [11].

### Quellen

- [1] Adam, S., Wüsch, C. & Koch, M. 2013: Ergebnisbericht "RE-Kompass 2013", Fraunhofer IESE Report.
- [2] Reuther, U. & Schmidt-Wigger, A. 2000: Designing a Multi-Purpose CL Application, in: Proceedings of CLAW 2000, Seattle.
- [3] Reuther, U. & Rösener, C. 2007: CLAT (Controlled Language Authoring Technology) - Qualität von Anfang an! In: MDÜ. Fachzeitschrift für Übersetzer und Dolmetscher, Herausgeber: Bundesverband der Dolmetscher und Übersetzer e.V. (BDÜ), Heft Nr. 2/2007, Berlin.
- [4] Hertel, H. & Arandan Yamchi, A., 2010: Möglichkeiten der Sprachstandardisierung – Ein Erfahrungsbericht über Methoden, Anwendungen und Prozesse bei der Volkswagen AG im Bereich After Sales <http://2010.reconf.de/wissenschaftstrack/volkswagen/> aufgerufen am 23.09.2015
- [5] Adam, S., Wüsch, C. & Seyff, N. 2015: RE Kompass 2014/2015 Ergebnisbericht.
- [6] ISO/IEC/IEEE 29148:2011. Systems and software engineering -- Life cycle processes -- Requirements engineering.
- [7] Kandt, R. K. 2003: Software Requirements Engineering: Practices and Techniques. JPL Document D-24994. SQI Report R-3. Jet Propulsion Laboratory. California Institute of Technology.
- [8] Maas, H.-D. / Rösener, C. / Theofilidis, A. 2009: Morphosyntactic and Semantic Analysis of Text: The MPRO Tagging Procedure. In: Mahlow, Cerstin; Piotrowski, Michael (Hg.): State of the art in computational morphology. Workshop on systems and frameworks for computational morphology, SFCM 2009, Zurich, Switzerland, September 4, 2009. proceedings. 1st ed. New York: Springer (Communications in computer and information science, 41), S. 76–87.
- [9] Lehrndorfer, A. & Reuther, U. (2008). Kontrollierte Sprache - standardisierte Sprache? In: J. Muthig (Hrsg.), Standardisierungsmethoden für die Technische Dokumentation, tekom Hochschulschriften Nr.16, Lübeck.
- [10] Reuther, U. (2003). Two in One – Can it Work? Readability and Translatability by means of Controlled Language. In: Proceedings of EAMT-CLAW 2003, Dublin.
- [11] Klaus, A. (2014). Automatisierte Prüfung der Qualität natürlichsprachlicher Anforderungen. Vortrag: Software Technologie Initiative e.V. Kaiserslautern. 10.04.2014.