

# Teilautomatisierte Begriffsanalyse zur Ermittlung zentraler Fachkonzepte im Bereich Eingebettete Systeme

Jörg Bader,<sup>1</sup> Mareen Przybylla<sup>2</sup>

**Abstract:** Physical Computing ermöglicht Schülerinnen und Schülern, im Informatikunterricht einfache eingebettete Systeme zu entwerfen. Dabei rücken fachliche Inhalte in den Fokus, die für viele Lehrpersonen noch unbekannt sind, infolge dessen werden die neuen Technologien derzeit häufig „nur“ als moderne Werkzeuge zur Vermittlung bereits etablierter Inhalte genutzt. Um das Potenzial der Themen stärker auszuschöpfen, arbeiten wir an der Entwicklung eines Kompetenzkatalogs für Physical Computing. Hierzu analysieren wir teilautomatisiert Fachliteratur zum Fachgebiet Eingebettete Systeme hinsichtlich zentraler Konzepte. In diesem Beitrag stellen wir die Analysemethode sowie erste Ergebnisse vor und geben einen Ausblick auf die nächsten Projektschritte.

**Keywords:** Physical Computing; Eingebettete Systeme; Konzepte; Dokumentenanalyse

## 1 Einführung und Motivation: Physical Computing in der Schule

Beim Physical Computing erwerben Schülerinnen und Schüler Kompetenzen des prototypischen Entwurfs einfacher eingebetteter Systeme. Neben der Programmierung rücken auch die Untersuchung von Eigenschaften eingebetteter Systeme sowie die Herausforderungen bei deren Entwurf in den Fokus des Unterrichts. Es gibt eine große Vielfalt an möglichen Schwerpunkten, sowohl inhaltlich als auch methodisch. Die Bewertung entstehender Projekte anhand transparenter und objektiver Kriterien ist nicht trivial und führt teilweise zu Unsicherheiten auf Seiten der Lehrpersonen. Insbesondere ergeben sich Zielkonflikte, wenn bei kreativen und gut gestalteten Produkten eine Diskrepanz zur Qualität der Informatikinhalte festgestellt werden kann [PR17]. Unabhängig davon, ob eine summative Beurteilung des Endprodukts oder eine Rückmeldung zum Lernprozess gegeben werden soll, ist es wichtig zu klären, welche Kompetenzen die Lernenden erwerben sollen. Unsere derzeitige und zukünftige Arbeit strebt daher einen kompakten und strukturierten Überblick über die Schlüsselkonzepte<sup>3</sup> des Physical Computings an, um die übergeordneten Informatikkonzepte des Bildungsbereichs [De04] zu verfeinern. Daraus wird ein Bewertungsraster für Physical Computing entwickelt, das Lehrpersonen darin unterstützt, ihren Fokus zu wählen und ihr Feedback auf möglichst objektive Kriterien zu stützen. In diesem Artikel stellen wir mit einer teilautomatisierten Dokumentenanalyse einen Arbeitsschritt in dieser Richtung vor.

---

<sup>1</sup> Pädagogische Hochschule Luzern, Fachdidaktik Medien und Informatik, Luzern, Schweiz joerg.bader@phlu.ch

<sup>2</sup> Pädagogische Hochschule Schwyz, Institut für Medien und Schule, Goldau, Schweiz mareen.przybylla@phsz.ch

<sup>3</sup> Als Schlüsselkonzepte werden zugrundeliegende Konzepte des Fachgebietes verstanden, die einen Zugang zum gesamten Fachgebiet darstellen und ein Verständnis der zentralen Themen und Ideen ermöglichen (vgl. [GR17]).

## 2 Ermittlung zentraler Konzepte

### 2.1 Teilautomatisiertes Vorgehen

Eine manuelle Zusammenstellung aller Schlüsselkonzepte, die den vielfältigen fachlichen Aspekten eingebetteter Systeme gerecht wird, ist schwierig, die objektive Gewichtung von Begriffen und Konzepten bezüglich ihrer Wichtigkeit fast unmöglich. Wir analysieren daher als Ergänzung zu einer manuellen Analyse teilautomatisiert verschiedene Texte, die sich mit eingebetteten Systemen befassen. Um unerwünschte Verzerrungen (Bias, Einseitigkeit, ungewöhnliche Begriffsbildungen) möglichst zu vermeiden, wird jeweils eine sorgfältig zusammengestellte Auswahl von Textdokumenten in Kategorien zusammengefasst (siehe Abschnitt 2.2). In jeder Kategorie werden die Dokumente gemeinsam auf häufig auftretende Wörter und Wortgruppen untersucht. Taucht ein Wort besonders häufig, in verschiedenen Literaturquellen und in mehreren der Kategorien auf, so ist dies ein Indiz dafür, dass es sich um einen Fachbegriff handeln könnte. Das grundsätzliche Vorgehen einer solchen Textanalyse ist etabliert und die vielfältigen Vorteile dargelegt [Ma04].

Für die automatisierte Textanalyse verwenden wir ein eigenes Pythonprogramm, das unter anderem auf das Python-Package spaCy zurückgreift, ein Werkzeug zur Verarbeitung natürlichsprachlicher Informationen („natural language processing“). Eine Tagging-Funktion von spaCy kann Wörter bezüglich ihrer syntaktischen Bedeutung im Satz einordnen und insbesondere für eine Vielzahl von Wörtern deren Wortart zuverlässig herausfinden. Verschiedene Sprachen sind in spaCy vorimplementiert, zum Sprachwechsel sind nur leichte Anpassungen in unserem Programmcode nötig. In den zu analysierenden Dokumenten führen wir eine Häufigkeitsanalyse der enthaltenen Substantive, Adjektive und derjenigen Wörter durch, deren Wortart spaCy nicht erkennt. Insbesondere typische Stoppwörter sind in unserer Häufigkeitsanalyse so nicht mehr enthalten. In einem zweiten Schritt filtern wir nochmals die Substantive und Adjektive aus dem Text und interpretieren alle Paare aufeinanderfolgender Wörter als Wortpaare. Häufig auftretende Wortpaare deuten auf zusammengesetzte Fachbegriffe hin.

### 2.2 Literatúrauswahl

Für unsere Analyse werden unter anderem Beiträge zur fachlichen und fachdidaktischen Forschung (Artikel von einschlägigen Konferenzen, Dissertationen) und Inhalte universitärer Bildung (Lehrbücher und Vorlesungsskripte) bezüglich der Häufigkeit enthaltener Begriffe verglichen. Als Testdatensatz für Forschungsbeiträge dient eine Auswahl der auf dem Dokumentenserver [arXiv.org](https://arxiv.org)<sup>4</sup> veröffentlichten Artikel zum Thema eingebettete Systeme. Es

<sup>4</sup> Bei [arXiv.org](https://arxiv.org) handelt es sich um eine Plattform zur Veröffentlichung wissenschaftlicher Dokumente (z. B. Preprints und andere unveröffentlichte Artikel). Hierbei findet zwar keine Qualitätskontrolle im Sinne eines Peer Reviews statt, es gibt jedoch eine Überprüfung bzgl. formaler Kriterien. Die Plattform bietet den Vorteil, dass dort abgelegte Artikel sehr aktuell sein können. Die eingeschränkte Qualitätskontrolle fällt für unseren Zweck weniger ins Gewicht, da eine Übersicht über relevante Themen und Begriffe nichtsdestotrotz gewonnen werden kann.

wurden alle Artikel zusammengestellt, deren Titel das Stichwort „Embedded System“ enthält und in deren Abstract „Hardware“<sup>5</sup> enthalten ist. Eingeschränkt auf die Erscheinungsjahre 2011 bis 2020 und das Themengebiet „Computer Science“ entspricht dies 46 Artikeln. Diese beziehen sich alle im engeren oder weiteren Sinn auf eingebettete Systeme. Die Bücherauswahl für unseren Testdatensatz beruht auf Literatur, die typischerweise im Hochschulunterricht verwendet wird (z. B. [Ac07; Bu04; Ca02; Ma03; Wo12]). Zusätzlich wurden die Konferenzbände der Reihe „Workshop on Embedded and Cyber-Physical Systems Education (WESE)“ aus den Jahren 2010 bis 2016 analysiert. Diese geben einen Einblick in fachdidaktische Forschung.

### 2.3 Erste Resultate

Die automatisierte Ausgabe enthält auch bei häufig auftretenden Wörtern und Wortpaaren noch eine Vielzahl an Begriffen (ca. 50%), die nicht zur Identifikation potenzieller Schlüsselkonzepte dienen und manuell entfernt werden müssen, während deren Anteil in englischsprachiger Literatur geringer ist als in deutschsprachiger. Wortpaare sind häufig nicht vollständig, liefern jedoch meist gute Hinweise auf die eigentlichen Begriffe (z. B. „time systems“, „internet things“). Beim manuellen Prüfen der extrahierten Begriffe zeigt sich auch, dass einige häufig auftretenden Wörter und Wortpaare aufgrund fehlenden Kontexts nicht oder nicht eindeutig einem Konzept zugeordnet werden können (z. B. „test“ – es ist unklar, ob sich dies auf den Test von Hard-/Software bezieht oder um eine Leistungsüberprüfung handelt).

Im Resultat ergeben unsere ersten Analysen Wortlisten, die einen Vergleich zwischen den Literaturkategorien zulassen. Es zeigt sich beispielsweise, dass in der Proceedings-Serie „WESE“ mehr übergeordnete Begriffe unter den häufig vorkommenden sind, während Forschungsbeiträge und Lehrbücher oft spezielle Themen im Bereich eingebettete Systeme fokussieren: Begriffe wie z. B. „execution time“, „power consumption“ oder „address space“ sind hier unter den zehn am häufigsten vorkommenden Begriffen. In der Liste der WESE-Proceedings tauchen äquivalente Begriffe erst auf den Plätzen 32, 110 und 209 auf. Hierbei muss beachtet werden, dass es sich bei WESE um eine Fachdidaktik-Konferenz handelt, während sich Beiträge zu fachwissenschaftlichen Konferenzen üblicherweise eher auf spezielle Teilgebiete konzentrieren. Dies deutet darauf hin, dass im Bildungsbereich die Schwerpunkte eher auf übergeordnete Konzepte gelegt werden, die für alle relevant sind.

## 3 Ausblick

Die teilautomatisierte Ermittlung der Begriffe aus verschiedenen Datenquellen hilft, die für verschiedene Autoren und Zielgruppen relevanten Fachbegriffe (potenzielle Schlüsselkonzepte) möglichst vollständig zu erfassen und dient vor allem dazu, die parallel stattfindende

<sup>5</sup> Das Stichwort *Hardware* wurde ergänzt, um Artikel zu z. B. biologischen eingebetteten Systemen auszufiltern. Die Suche wurde mit englischsprachigen Begriffen durchgeführt, da der Großteil der auf [arXiv.org](https://arxiv.org) angebotenen Artikel englischsprachig ist.

manuelle Ermittlung von Schlüsselkonzepten zu ergänzen und auf Vollständigkeit zu prüfen. In den nächsten Schritten werden zunächst manuell ermittelte Fachbegriffe in einem Begriffscluster kategorisiert und anschließend mit Begriffen abgeglichen, die aus einer erweiterten teilautomatisierten Analyse mit größerem Literaturkorpus stammen, um mögliche Lücken zu identifizieren und gleichzeitig die identifizierten Konzepte zu validieren. Ein ähnliches Verfahren wurde bereits für den Bereich Data Management vorgeschlagen und durchgeführt [GR17].

Im Anschluss wird mittels Unterrichtsunterlagen der Zielstufe untersucht, inwieweit die für Wissenschaft, Anwendung und die weitere Ausbildung relevanten Fachbegriffe schon in der Schule vermittelt werden. So kann auf mögliche Lücken im Curriculum hingewiesen werden. Ob jede dieser Lücken schon in der Schule gefüllt werden kann oder sollte, ist im Einzelfall zu entscheiden. Die Erkenntnisse des resultierenden Kompetenzkatalogs sollen in ein Bewertungsraster für Physical-Computing-Projekte einfließen, wie es von vielen Lehrpersonen gewünscht wird und für andere Bereiche bereits vorliegt [We20]. Das entstehende Bewertungsraster für Physical-Computing-Projekte an der Schule wird dann in Feldversuchen überprüft und nachgebessert.

## Literatur

- [Ac07] Aceto, L.; Ingolfsdottir, A.; Larsen, K. G.; Srba, J.: *Reactive Systems: Modelling, Specification and Verification*. Cambridge University Press, 2007.
- [Bu04] Buttazzo, G.: *Hard Real-Time Computing Systems: Predictable Scheduling Algorithms and Applications*. Springer, 2004.
- [Ca02] Catsoulis, J.: *Designing Embedded Hardware*. O'Reilly, 2002.
- [De04] Denning, P. J.: Great principles in computing curricula. In: *Proceedings of the 35th SIGCSE technical symposium on CS education*. S. 336–341, 2004.
- [GR17] Grillenberger, A.; Romeike, R.: Empirische Ermittlung der Schlüsselkonzepte des Fachgebiets Datenmanagement. In: *Proceedings der INFOS 2017*. Gesellschaft für Informatik, Bonn, S. 157–166, 2017.
- [Ma03] Marwedel, P.: *Embedded System Design*. Kluwer Academic, 2003.
- [Ma04] Mayring, P.: *Qualitative content analysis. A companion to qualitative research 1/2*, S. 159–176, 2004.
- [PR17] Przybylla, M.; Romeike, R.: The Nature of Physical Computing in Schools. In: *Koli Calling '17*. Association for Computing Machinery, S. 98–107, 2017.
- [We20] Weeda, R.; Izu, C.; Kallia, M.; Barendsen, E.: Towards an Assessment Rubric for EiPE Tasks in Secondary Education. In: *Koli Calling '20*. Association for Computing Machinery, 2020.
- [Wo12] Wolf, M.: *Computers as Components: Principles of Embedded Computing System Design*. Elsevier Science, 2012.