

Prof. Dr.-Ing. Torsten Zesch:

**"Verbesserte Effizienz von veranstaltungsbegleitenden Literaturempfehlungen durch semi-automatisch generierte Leseverständnistests"**

## Ausgangslage und persönliche Motivation

In vielen Lehrveranstaltungen wird zusätzlich zu den Vorlesungen ergänzende und vertiefende Literatur bereitgestellt. Dieses Angebot wird, auch nach eigener Erfahrung, von den Studierenden jedoch oft nicht wahrgenommen. Leseverständnistests gelten als ein Weg, Studierende zur Beschäftigung mit der zu einer Vorlesung bereitgestellten Literatur zu motivieren (Howard, 2004). Die Studierenden können so eigene Wissenslücken frühzeitig erkennen und setzen sich mit dem Vorlesungsstoff zusätzlich auseinander. Darüber hinaus stellen sie auch dem Lehrenden Feedback über mögliche Verständnisprobleme und Wiederholungsbedarf bei den Studierenden zur Verfügung. In der digitalen Hochschullehre bieten sich nun zwei Ansatzpunkte zur Verbesserung:

- 1) **Verbesserte Bereitstellung** Damit Leseverständnistests überhaupt wirksam werden können, müssen diese erst einmal vorhanden sein. Der Zusatzaufwand für Lehrende zur Generierung solcher Tests ist jedoch erheblich, weswegen diese oft nicht oder nicht in ausreichendem Umfang eingesetzt werden.
- 2) **Verbesserte Integration** Die Tests müssen möglichst gut in den Lehrablauf integriert werden. Freiwillige oder vertiefende Angebote werden kaum genutzt, wenn diese als nicht prüfungsrelevant erscheinen.

Die geplante Lehrinnovation adressiert beide Punkte, indem zum einen ein System zur semi-automatischen Generierung von Leseverständnistests zur Verfügung gestellt wird, zum anderen eine nahtlose Integration in bestehende eLearning-Umgebungen sichergestellt und evaluiert wird.

### **Bereitstellung**

Nur wenn Leseverständnistests überhaupt vorhanden sind und sich an der tatsächlich bereitgestellten Literatur orientieren, kann überhaupt ein positiver Effekt für Lernergebnisse eintreten. Erklärtes Ziel des Projekts ist es, den Aufwand für die Erstellung von Leseverständnisfragen deutlich zu senken. Eine teil-automatisierte Erstellung soll den Einsatz von Leseverständnisfragen in Lehrbereichen ermöglichen, in denen der Aufwand sonst für zu hoch angesehen würde, so z.B. für den Einsatz von Leseverständnisaufgaben zur Vergabe von Bonuspunkten. In diesem Fall ist davon auszugehen, dass die Aufgaben schnell bekannt werden und daher jedes Jahr neu erstellt werden müssen.

Es soll ein Web-basiertes System zur Kuratierung von automatisch erzeugten Leseverständnisfragen erstellt werden. Das System wird über existierende Programmierschnittstellen an einen Moodle-Kurs, wie er in den meisten Veranstaltungen der UDE genutzt wird, angebunden

---

## **Semi-automatisch generierte Leseverständnistests**

---

und erzeugt automatisch einen Pool an möglichen Fragen. Der Lehrende wählt die passendsten Fragen aus, kann sie bei Bedarf nachbearbeiten und das System erstellt daraus ein einfach in Moodle zu importierendes Format. Eine Studie zur semi-automatischen Generierung von Multiple-Choice Tests hat gezeigt, dass es nur ein Viertel der Zeit braucht, um aus einem Pool von generierten Fragen zu einem Lehrbuchtext über Linguistik eine bestimmte Menge an geeigneten Fragen auszuwählen und gegebenenfalls nachzubearbeiten, als die gleiche Anzahl an Fragen eigenständig zu formulieren (Mitkov, Ha, & Karamanis, 2006). Durch die enge Kopplung der Fragengenerierung an die Ausgangstexte ist sichergestellt, dass die Leseverständnistests auch wirklich aus der Literatur heraus motiviert sind.

Zur Generierung der Leseverständnisfragen kann auf bereits entwickelte sprachtechnologische Algorithmen zurückgegriffen werden. Ich forsche seit Jahren in diesem Gebiet und bringe umfassende Expertise dazu in das Projekt ein. Die Herausforderung besteht eher in der nahtlosen Integration der Technologie in die digitale Hochschullehre und in der Erprobung verschiedener Testformate (siehe S. 3f.).

### ***Stand der Forschung – Generierung von Leseverständnistests***

Für die automatische Generierung unterschiedlicher Übungstypen kann auf verschiedene, teils von mir selbst entwickelte, sprachtechnologische Verfahren zurückgegriffen werden. Im Bereich der Freitextfragen können Faktenfragen durch syntaktische Transformation aus Deklarativsätzen gewonnen und automatisch nach Qualität gerankt werden (Heilman & Smith, 2010). Solche Generierungsmethoden finden beispielsweise im Sprachenlernen (Michael Heilman, Kevyn Collins-thompson, Jamie Callan, 2006) Anwendung, aber auch in Schreibassistenzsystemen im akademischen Bereich, bei denen eigene Schreibprodukte des Studierenden vom System hinterfragt werden (Liu, Calvo, & Rus, 2012).

Bei der Generierung von Multiple-Choice-Fragen werden relevante Konzepte im Text identifiziert, und Fragen sowie Distraktoren generiert (Mitkov et al., 2006). Eine besondere Herausforderung dabei ist die Erstellung von nicht-trivialen Distraktoren, d.h. falschen Antworten, die nicht auf den ersten Blick als solche zu erkennen sind (Zesch & Melamud, 2014).

Im Bereich von Lückentexten liegt die Herausforderung darin, Lücken zu finden, die tatsächlich nur eine mögliche Lösung haben. Ich habe dazu ein neues Übungsparadigma, die sogenannten gap-bundles, automatisiert. Dabei wird ein Kernbegriff in unterschiedlichen Sätzen durch eine Lücke ersetzt, so dass die Summe der Kontexte die Lösung eindeutig macht (Meyer, Wojatzki, & Zesch, 2016; Wojatzki, Melamud, & Zesch, 2016).

Es ist also zusammenfassend festzustellen, dass die notwendige Technologie zur semi-automatischen Generierung von Leseverständnisfragen bereits zur Verfügung steht. Es gilt nun, diese in einem echten Lehr/Lernsetting auch einzusetzen und zu testen, damit diese dann für die Allgemeinheit zur Verfügung gestellt werden können.

### ***Integration***

Maßgeblich für eine gelungene Integration der Leseverständnistests in eine spezifische Lehr-Lern-Situation ist immer, dass die Tests zum Thema, zur Lehrtradition der Disziplin und zum

## Semi-automatisch generierte Leseverständnistests

Lernziel passen. Das Projekt wird daher ein breites Spektrum an Testtypen zur Verfügung stellen. Die Testtypen zielen dabei vor allem deklaratives Wissen ab, da die Studierenden so in dem meisten Fällen sofort automatisiertes Feedback zu ihren Antworten erhalten können. Auf diese Weise sollen auch die TutorInnen entlastet werden, so dass in den Präsenzsitzungen mehr Zeit bleibt, Anwendungs- und Transferaufgaben zu diskutieren.

In den von mir durchgeführten Vorlesungen wird exemplarisch eine bestimmte Kombination erprobt und evaluiert. Mit Hinblick auf ein möglichst hohes Transferpotential der Ergebnisse werden aber auch andere Kombinationen bereits mit umgesetzt.

Im Folgenden gehen wir genauer auf die vorgesehenen Testtypen und deren Potential in der digitalen Lehre ein. Ausgangspunkt ist jeweils ein bestimmter Text, wie hier zum Beispiel ein Ausschnitt aus dem für die Vorlesung „Grundlagen der künstlichen Intelligenz“ eingesetzten Lehrbuch „Artificial Intelligence: A Modern Approach“ (Russell & Norvig, 2002).

### Beispiel für einen Lesetext

#### Acting humanly: The Turing Test approach

The **Turing Test**, proposed by Alan Turing (1950), was designed to provide a satisfactory operational definition of intelligence. Rather than proposing a long and perhaps controversial list of qualifications required for intelligence, he suggested a test based on indistinguishability from undeniably intelligent entities- human beings. The computer passes the test if a human interrogator, after posing some written questions, cannot tell whether the written responses come from a person or not. Chapter 26 discusses the details of the test and whether a computer is really intelligent if it passes. For now, we note that programming a computer to pass the test provides plenty to work on. The computer would need to possess the following capabilities:

- **natural language processing** to enable it to communicate successfully in English.
- **knowledge representation** to store what it knows or learns;
- **automated reasoning** to use the stored information to answer questions and to draw new conclusions;
- **machine learning** to adapt to new circumstances and to detect and extrapolate patterns.

Turing's test deliberately avoided direct physical interaction between the interrogator and the computer, because physical simulation of a person is unnecessary for intelligence. However, the so-called **total Turing Test** includes a video signal so that the interrogator can test the subject's perceptual abilities, as well as the opportunity for the interrogator to pass physical objects "through the hatch." To pass the total Turing Test, the computer will need

- **computer vision** to perceive objects, and
- **robotics** to manipulate objects and move about.

## Semi-automatisch generierte Leseverständnistests

- **Multiple-Choice-Fragen** eignen sich nach dem Lesen des Textes zur Überprüfung des Verständnisses. Dabei werden plausible falsche Antworten aus dem Text oder domainenspezifischen Korpora automatisch generiert.

- |   |  |
|---|--|
| 1) What is needed to pass the total Turing test?<br>a) human intelligence<br>b) computer vision<br>c) planning<br>d) psychological experiments  | 2) What is included in the total Turing test?<br>a) a video signal<br>b) a roboter<br>c) aerodynamics<br>d) intelligent entities |
| 3) Turing's test deliberately _____ direct physical interaction between the interrogator and the computer, because physical simulation of a person is unnecessary for intelligence.<br>a) avoids<br>b) provides<br>c) encourages<br>d) disregards |  |

- **Lückentexte** können ebenfalls linguistisch informiert generiert werden, so dass beispielsweise nur Substantive ausgewählt und aus dem restlichen Text zusätzliche Distraktoren extrahiert werden.

The \_\_\_\_\_, proposed by \_\_\_\_\_ (1950), was designed to provide a satisfactory operational \_\_\_\_\_ of \_\_\_\_\_. Rather than proposing a long and perhaps controversial \_\_\_\_\_ of \_\_\_\_\_ required for \_\_\_\_\_, he suggested a \_\_\_\_\_ based on \_\_\_\_\_ from undeniably intelligent \_\_\_\_\_ - human \_\_\_\_\_.

Wählen Sie aus folgenden mögliche Lösungen: *definition, beings, Turing Test, processing, indistinguishability, list, entities, Alan Turing, intelligence, information, qualifications, intelligence, test, reasoning*

- Faktenfragen mit **Freitextantworten** lenken den Fokus der Studierenden auf relevante Informationen im Text und erfordern aktive Formulierungsarbeit. Während bei den beiden vorgenannten Aufgabentypen sofort Feedback zur Richtigkeit der Antwort gegeben werden kann (hat der Studierende die richtige Antwort angekreuzt?), ist die automatische Korrektur von Freitextantworten schwierig, weil es oft sehr viele Möglichkeiten gibt, eine korrekte Antwort zu formulieren.<sup>1</sup>

<sup>1</sup> Das automatische Scoring von Freitextantworten ist nicht Bestandteil dieses Antrags, soll jedoch mittelfristig ebenfalls in das geplante System integriert werden. Dazu kann ich auf eigene Vorarbeiten zurückgreifen (Zesch, Heilman, & Cahill, 2015; Zesch, Wojatzki, & Scholten-Akoun, 2015).

## Semi-automatisch generierte Leseverständnistests

- What is a Turing test? (Ask for definitions of bold-print terms?)
- When was the Turing test proposed?
- Who proposed the Turing test?
- When does a computer pass the Turing test?
- Name capabilities needed by a computer to pass the Turing test.
- What is needed by a computer to use stored information to answer questions?
- What is a total Turing test?

- **Gap-fill Bundles** sind eine wenig bekannte, von mir erstmals automatisierte Übungsform (Wojatzki et al., 2016) mit der Kernbegriffe aus dem Lesetext (im Beispiel „Computer Vision“) aufgegriffen und in anderen Kontexten geübt werden können.

Finden Sie den Begriff aus dem Text, der in drei der 5 Lücken passt, und kreuzen Sie an in welche:

\_\_\_\_\_ is an interdisciplinary field that deals with how computers can be made for gaining high-level understanding from digital images or videos.

\_\_\_\_\_ is the subfield of computer science that, according to Arthur Samuel in 1959, gives "computers the ability to learn without being explicitly taught"

\_\_\_\_\_ is a science that develops models and methods for understanding, analyzing, acquiring and processing images

\_\_\_\_\_ is a science that develops models and methods for understanding, analyzing, acquiring and processing images

Reasoning is the ability to make inferences, and \_\_\_\_\_ is concerned with the building of computing systems that automate this process.

### Motivation

Mit dem Projekt möchte ich einen Beitrag dazu leisten, die von mir in meinen Veranstaltungen wahrgenommen Defizite in der Nutzung der Literaturempfehlungen durch die Studierenden zu verringern. Ich möchte ganz gezielt den Schritt von meiner derzeitigen methodisch getriebenen Forschung zur Fragengenerierung hin zum praktischen Einsatz der Algorithmen in einer spezifischen Lehr-Lern Situation machen.

Da es sich um ein Projekt mit großem Transferpotential handelt, verspreche ich mir vom Austausch mit den anderen Fellows die einzigartige Möglichkeit, Feedback aus vielen verschiedenen Disziplinen und Studiengängen auf die Probleme von Leseverständnistests zu gewinnen. Ich erhoffe mir, dadurch das geplante Generierungssystem noch besser Nutzer-zentriert entwickeln zu können und dadurch eine Nachnutzung auch außerhalb der engen Domäne meiner eigenen Vorlesungen zu ermöglichen. Außerdem hoffe ich, unter den Fellows weitere Personen zu finden, die das zu entwickelnde Werkzeug einsetzen möchten, und deren Erfahrungen in die Weiterentwicklung einfließen können.

## Semi-automatisch generierte Leseverständnistests

---

### Geplante Lehrinnovation: Operationalisierung und Evaluation

Für die Umsetzung der geplanten Lehrinnovation ist es notwendig ein leicht nutzbares System zur automatischen Erstellung und Kuratierung der Leseverständnisfragen zu entwickeln. Da wie oben dargestellt die entsprechenden Algorithmen zur Generierung bereits zur Verfügung stehen, muss der Fokus auf der Gebrauchstauglichkeit der **Nutzeroberfläche** sowie den **Schnittstellen** für den Import der Lesetexte und den Export der fertigen Tests liegen.

#### **Nutzeroberfläche**

Es wird eine Web-basierte Nutzeroberfläche entwickelt, da ein Plugin für Moodle aus folgenden Gründen nicht praktikabel erscheint:

- Eine direkte Integration von prototypischen Plugins in das Produktivsystem der Universität ist schwierig.
- Die Generierung der Leseverständnistests ist effizient möglich, jedoch ist der Bedarf an Rechenkapazität zu hoch für ein direkt im Kernsystem laufendes Plugin.
- Veränderungen in der Produktivversion des LMS bedingt durch schnelle Versions-sprünge erfordern ständige Änderungen an Plugins, die dadurch nicht nachhaltig wären. Eine Anbindung über Programmierschnittstellen erzeugt daher mehr Konstanz.
- Für das geplante Projekt bietet eine selbst entwickelte Nutzeroberfläche deutlich mehr Flexibilität und erlaubt eine Konzentration auf die Projektziele.

Wie wir weiter unten im Kapitel zu Nachhaltigkeit und Transfer noch detaillierter ausführen werden, planen wir die Nachnutzbarkeit durch ein auf Open-Source Entwicklung und auf die strikte Einhaltung von Programmierstandards ausgerichtetes Projektkonzept zu gewährleisten.

#### **Schnittstellen zum LMS Moodle**

Damit die Lehrenden möglichst wenig Aufwand für die Bereitstellung der Leseverständnistests haben, sollen die Literaturempfehlungen und die dazu gehörigen Textdateien direkt aus Moodle oder aus eigenen lokalen Literaturbeständen importiert werden. Entsprechende Schnittstellen stehen z.B. über den Materialiendownload in Moodle zur Verfügung.

Die vom Lehrendem ausgewählten Fragen werden direkt in ein von Moodle unterstütztes Format (<https://docs.moodle.org/30/en/Questions>) konvertiert und über eine der Import Schnittstellen von Moodle ins LMS integriert ([https://docs.moodle.org/30/en/Import\\_questions](https://docs.moodle.org/30/en/Import_questions))

Ein Export in die entsprechenden Formate für andere LMS ist aus zeitlichen Gründen in diesem Projekt nicht vorgesehen. Das System wird aber so modular gestaltet, dass dieses später leicht nachrüstbar ist (siehe Abschnitt zur Nachhaltigkeit).

---

## Semi-automatisch generierte Leseverständnistests

---

### **Zielveranstaltungen**

Die geplante Lehrinnovation ist prinzipiell für alle Veranstaltungen einsetzbar, in denen Literaturempfehlungen gegeben werden und Leseverständnistests eingesetzt werden können. Zumindest im Bereich der Informatik/Angewandte Kognitionswissenschaften, den ich aus eigener Anschauung kenne, trifft dies auf nahezu alle Vorlesungen und Seminare zu.

Im Rahmen dieses Projekts wird das System in den folgenden von mir angebotenen Vorlesungen eingesetzt:

- „Grundlagen der künstlichen Intelligenz“ (Angewandte Informatik/Komedia – Bachelor, Wahlpflicht, Sommersemester, 250 Studierende),
- „Wissensbasierte Systeme“ (Angewandte Informatik/Komedia – Master, Wahlbereich, Sommersemester, 30 Studierende), und
- „Sprachtechnologie“ (Angewandte Informatik/Komedia – Bachelor, Wintersemester, 50 Studierende)

In allen diesen Vorlesungen erhebe ich ab diesem Semester Daten zur Nutzung der Literaturempfehlungen, die als Vergleichswert dienen können. Die Erhebung erfolgt dabei in den Präsenzvorlesungen unter Nutzung des Live-Feedback-System PINGO „Peer instruction for very large groups“ (Reinhardt et al., 2012). Erfahrungsgemäß ist die Teilnehmerzahl in der letzten Sitzung vor der Klausur besonders hoch und umfasst ebenfalls Studierende, welche die Vorlesung eher unregelmäßig oder gar nicht besuchen.

### **Einsatzszenario**

Die Zielveranstaltungen sind jeweils Vorlesungen mit Übung und einer Klausur am Semesterende. Zu jeder Vorlesung werden Übungsaufgaben angeboten, die in wöchentlichen Tutorien von studentischen Hilfskräften mit den Studierenden besprochen werden. Die Teilnahme an den Tutorien und die Bearbeitung der Übungsblätter ist optional, jedoch kann man durch eine bestimmte Menge an richtig gelösten Übungsaufgaben Zusatzpunkte für die Klausur bekommen.

In jeder der Zielveranstaltungen werde ich das System einsetzen, um jeweils 10 Leseverständnistests pro Einheit zu generieren, die den Studierenden gemeinsam mit den weiteren Übungsaufgaben über Moodle zur Verfügung gestellt werden. Je nach Veranstaltung wird dabei ein bestimmter, dem Thema angemessener Mix an Testtypen eingesetzt. Um zur Bearbeitung des Tests zu motivieren wird: 1) die Wichtigkeit der Tests zur Überprüfung des eigenen Wissensstandes hervorgehoben und 2) werden in den Präsenzphasen die aggregierten Ergebnisse diskutiert und mögliche häufige Fehler noch einmal diskutiert. Auf diese Art erhalten die Studierenden zusätzliches Feedback über ihren Leistungsstand.

Es sei an dieser Stelle explizit darauf hingewiesen, dass im Projekt ein nicht aufzulösender Widerspruch besteht zwischen optimaler didaktischer Einbindung der Tests und dem Evaluationsziel, d.h. der Messung des Einflusses einer vereinfachten Bereitstellung auf das Nutzungsverhalten. Viele vorgeschlagene Möglichkeiten zur besseren Einbindung der Tests (z.B. über Anreizsystem wie Bonuspunkte oder einer verpflichtenden Bearbeitung) würden mit diesem Evaluationsziel interferieren. Dies sind jedoch spannende und sehr lohnende Zielsetzungen für hoffentlich mit dem hier zu entwickelnden System noch ausstehende Folgeforschung.

## Semi-automatisch generierte Leseverständnistests

---

### Evaluation

Eine Evaluation des Projekts soll auf zwei Arten erfolgen: Zum einen durch eine Usability-Studie, die die Nutzbarkeit des entstehenden Systems für die Dozierenden überprüft. Zum Zweiten soll die Effizienz und Akzeptanz der entstehenden Aufgaben bei den Studierenden gemessen werden.

Die Nutzbarkeit des Systems wird von mir laufend intern evaluiert und entsprechend verbessert. Die Gebrauchstauglichkeit eines nur von einer Person entwickelten Systems ist aber wenig belastbar. Daher sollen sogenannte **UX (User eXperience) Tests** durchgeführt werden, bei denen ein Lehrender vorgegebene Aufgaben im System ausführt und dabei vom Versuchsleiter beobachtet wird. Es ist vorgesehen 5 solche Tests in verschiedenen Stadien des Projekts durchzuführen, wenn möglich auch an anderen Fakultäten und Universitäten, um den unterschiedlichen Lehr-Lern-Kulturen Rechnung zu tragen.

Die Effizienz der Leseverständnistests kann indirekt durch das Abschneiden in der Klausur evaluiert werden. Je nachdem welcher Anteil der Studierenden die Tests überhaupt bearbeitet, ist die Aussagekraft der so gewonnenen Zahlen jedoch zweifelhaft. Es sollen daher weitere Evaluationskriterien erfasst werden, die direktere Schlussfolgerungen erlauben.

Inwieweit die Studierenden die nun zusätzlich bereitgestellten Fragen überhaupt beantworten, lässt sich direkt messen. Ob dadurch eine vertiefte Auseinandersetzung mit den Lehrinhalten erreicht wird, ist nur indirekt über Befragungen herauszufinden. Dazu soll eine **erste qualitativ-quantitative Studie** am Anfang zur Bestimmung der Ausgangssituation stattfinden. Wieviel werden Literaturempfehlungen genutzt? Wie stark korreliert dies mit einer Teilnahme in der Vorlesung? Wenn Nutzung, warum? Wenn keine Nutzung, warum?

Nach Einführung der Leseverständnistests Intervention erfolgt eine **zweite Studie**. Wie werden die Tests angenommen? Hat sich das Leseverhalten geändert? Wenn ja, warum? Wenn nein, warum nicht? Welche Arten von Fragen werden besonders positiv/negativ aufgenommen?

Die so gewonnenen Daten zur Nutzung der Literaturempfehlungen können mit Daten verglichen werden, welche am Ende des laufenden Semesters in den Zielveranstaltungen erhoben werden, um einen Referenzrahmen zu erzeugen.

### Ausblick: Transfer in die Breite und Nachhaltigkeit

Das Nachnutzungspotential des geplanten Systems ist außerordentlich hoch, da es sich nicht um ein inhaltlich auf einen Studiengang bezogenes Programm, sondern auf ein allgemeingültiges Prinzip zur Unterstützung der Lehrenden bei der Erstellung von Leseverständnistests handelt. Durch die verschiedenen unterstützten Fragetypen und die enge Integration ins prototypische LMS Moodle ist die Innovation leicht auf andere Veranstaltungen in fast allen Fachgebieten übertragbar.

---

## Semi-automatisch generierte Leseverständnistests

---

Bereits während der Projektlaufzeit werde ich versuchen andere Lehrende zum Einsatz des Systems zu animieren. Wie oben bereits erwähnt erhoffe ich mir vom Austausch mit den anderen Fellows möglicherweise bereits auch Nutzer von anderen Universitäten gewinnen zu können und vielfältiges Feedback über Verbesserungsmöglichkeiten zu bekommen.

### **Umfeld**

Ich bin im Rahmen diverser Projekte und Kooperationen eng in die E-Learning Aktivitäten an der Universität Duisburg-Essen eingebunden. Besonders relevant ist hier die Mitarbeit zum Thema e-Assessment im hochschulweiten Projekt „Bildungsgerechtigkeit im Fokus“ (BMBF, Qualitätspakt Lehre) wodurch ich ausgezeichnet mit allen Akteuren am Standort vernetzt bin und auch regelmäßig an den E-Learning Netzwerktagen teilnehme. Das geplante Projekt passt thematisch eng zur UDE-Lehrstrategie (<https://www.uni-due.de/qualitaet-der-lehre/strategien.php>) und zur Strategie zur Digitalisierung in Studium und Lehre.

Darüber hinaus bin ich als Sprecher des DFG Netzwerkes INDUS „Individuelles Sprachlernen“ in ein hoch interdisziplinäres Netzwerk an ForscherInnen eingebunden und kann daher auf vielfältige Expertise u.a. aus den Bereichen Bildungsforschung, Lernpsychologie, und Testentwicklung zurückgreifen.

### **Nachhaltigkeit**

Die geplante Lehrinnovation beruht zu einem wesentlichen Teil auf dem innovativen, webbasierten System zur Generierung und Kuratierung der Leseverständnistests. Damit eine Weiterentwicklung nach Projektende gewährleistet werden kann, werden alle Komponenten des Systems als Open-source Projekt entwickelt und direkt auf der GitHub Plattform veröffentlicht. Ein besonderer Fokus soll dabei auch auf einer lückenlosen Dokumentation der Komponenten und der notwendigen Schritte zur individuellen Weiterentwicklung gelegt werden. Dadurch soll eine Nutzergemeinschaft entstehen, die entsprechend neue Fragentypen oder Integrationen für andere LMS beisteuern kann.

Die Programmierung selbst wird sich an bewährten Programmierstandards orientieren, mit festgelegten Formatierungsrichtlinien, Zielwerten für die Testabdeckung, automatisch generierter Dokumentation, sowie strikter Versionierung und der Veröffentlichung von Releases nach jedem Meilenstein (siehe Zeitplan).

Dieses Vorgehensmodell wird vom Antragsteller im Rahmen der Entwicklung der DKPro Plattform zur sprachtechnologischen Verarbeitung von Texten (<https://dkpro.github.io>) bereits eingesetzt und kann hier einfach übernommen werden (Daxenberger, Ferschke, Gurevych, & Zesch, 2014; de Castilho & Gurevych, 2014). Es garantiert eine nachhaltige Entwicklung des Systems auch nach Ablauf des Förderzeitraums.

---

**Semi-automatisch generierte Leseverständnistests**

---

**Literatur**

- Daxenberger, J., Ferschke, O., Gurevych, I., & Zesch, T. (2014). DKPro TC: A Java-based Framework for Supervised Learning Experiments on Textual Data. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics. System Demonstrations*. Baltimore, MD, USA.
- de Castilho, R., & Gurevych, I. (2014). A broad-coverage collection of portable NLP components for building shareable analysis pipelines. In *Proceedings of the Workshop on Open Infrastructures and Analysis Frameworks for HLT* (pp. 1–11). Dublin, Ireland.
- Heilman, M., & Smith, N. A. (2010). Good question! statistical ranking for question generation. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics* (pp. 609–617).
- Howard, J. R. (2004). Just-in-time teaching in sociology or how I convinced my students to actually read the assignment. *Teaching Sociology*, 32(4), 385–390.
- Liu, M., Calvo, R. A., & Rus, V. (2012). G-Asks: An Intelligent Automatic Question Generation System for Academic Writing Support. *D&D*, 3(2), 101–124.
- Meyer, N., Wojatzki, M., & Zesch, T. (2016). Validating Bundled Gap Filling--Empirical Evidence for Ambiguity Reduction and Language Proficiency Testing Capabilities. In *Proceedings of the joint workshop on NLP for Computer Assisted Language Learning and NLP for Language Acquisition at SLTC, Umeå, 16th November 2016* (pp. 51–59).
- Michael Heilman, Kevyn Collins-thompson, Jamie Callan, M. E. (2006). Classroom success of an Intelligent Tutoring System for lexical practice and reading comprehension. In *Proceedings of the 9th International Conference on Spoken Language*.
- Mitkov, R., Ha, L. A., & Karamanis, N. (2006). A computer-aided environment for generating multiple-choice test items. *Natural Language Engineering*, 12(2), 177.
- Reinhardt, W., Sievers, M., Magenheimer, J., Kundisch, D., Herrmann, P., Beutner, M., & Zoyke, A. (2012). PINGO: Peer Instruction for Very Large Groups. In A. Ravenscroft, S. N. Lindstaedt, C. D. Kloos, & D. H. Leo (Eds.), *EC-TEL* (Vol. 7563, pp. 507–512). Springer.
- Russell, S. J., & Norvig, P. (2002). *Artificial intelligence: a modern approach* (International Edition).
- Wojatzki, M., Melamud, O., & Zesch, T. (2016). Bundled Gap Filling: A New Paradigm for Unambiguous Cloze Exercises. In *Proceedings of the Building Educational Applications Workshop at NAACL* (Vol. 11). San Diego, USA: ACL.
- Zesch, T., Heilman, M., & Cahill, A. (2015). Reducing Annotation Efforts in Supervised Short Answer Scoring. In *Proceedings of the Building Educational Applications Workshop at NAACL*.
- Zesch, T., & Melamud, O. (2014). Automatic Generation of Challenging Distractors Using Context-Sensitive Inference Rules. In *Proceedings of the 9th Workshop on Innovative Use of NLP for Building Educational Applications at ACL*. Baltimore, USA.
- Zesch, T., Wojatzki, M., & Scholten-Akoun, D. (2015). Task-Independent Features for Automated Essay Grading. In *Proceedings of the Building Educational Applications Workshop at NAACL*.