

Exposé zur Bachelor Thesis

Im Rahmen des Studiums der Wirtschaftsinformatik an der Hochschule Fulda muss jeder Student eine Bachelorarbeit verfassen. Das Thema kann dabei auf den Erfahrungen der vorher erbrachten Praxisphase aufbauen.

Das Thema der Thesis lautet:

Digitale Tabellenerfassung - Tabellenextraktion durch Bildverarbeitung
und künstliche Intelligenz

Thema der Thesis

Daten, welche durch eine computergestützte Auswertung erschlossen werden sollen, liegen vielfach in Tabellenform vor. Diese Tabellen sind ggf. in ein Seitenlayout von Text und Abbildungen eingebettet und einer unmittelbaren Verarbeitung nicht zugänglich.

Es gibt bisher keine universellen Verfahren, Tabellen aus Dokumenten zu extrahieren und ihre Inhalte zu erschließen.

Aufbauend auf den Arbeiten der Praxisphase, ist das Ziel dieser Thesis der Entwurf und die Validierung eines Systems zur automatischen Extraktion von Tabelleninhalten mit der Hilfe maschinellen Lernens.

Schwerpunkte sind dabei:

- Beschreibung, Vergleich und Evaluierung von bestehenden Lösungsmethoden
- Planung und Entwicklung eines geeigneten Systems auf Basis von neuronalen Netzen und Methoden der Bildverarbeitung
- Evaluieren von Trainingsdaten
- Training und Beurteilung des Systems

Zeitplanung

Für das Schreiben der Thesis sind 12 Wochen vorgesehen.

Woche	Aufgabe	Meilensteine
1	Recherche und Entwicklung geeigneter Algorithmen und entsprechender Software	
2		
3		
4		
5		M1: Abschluss Recherche
6	Vorbereitung fürs Erstellen der Thesis	M2: Skizze für die finale Thesis ist fertig
7	Schreibphase	
8		M3: Alle Texte wurden geschrieben
9	Konsolidierung	
10	Finale Schreibphase	M4: Alle gefundenen Fehler wurden korrigiert
11	Letzte Korrektur und Abgabe	