Exploring Project

Rohrverbindungen automatisiert Schweißen

Ausgangssituation

Die Analyse von komplexen Texten hat in den letzten Jahren mit den großen Sprachmodellen enorme Fortschritte gemacht. Jedoch zeigt sich immer wieder, dass die Verarbeitung von Print-Dokumenten mit komplexen Layouts (Spalten, Tabellen, Grafiken usw.) eine Herausforderung darstellt. Oft stehen nur die finalen Dokumente im PDF-Format zur Verfügung, sodass die Erschließung der Dokumentinhalte eine Grundvoraussetzung für die weitere Nutzung für andere Zwecke darstellt.
Texte manuell zu extrahieren ist im Einzelfall mit hohem Aufwand verbunden und erlaubt es nicht, auf Marktentwicklungen in ausreichender Geschwindigkeit zu reagieren. Deshalb ist eine automatische Verarbeitung der Dokumente erforderlich, die die Inhalte (vor allem den Text) aus den Dokumenten schnell und in hoher Qualität zur Verfügung stellt.

Nutzen

Neben den als Ergebnis gewünschten Textinhalten (Überschriften und Fließtext) enthalten die Dokumente viele weitere Strukturen wie Kopf-/Fußzeilen, Textboxen, Abbildungen, Sonderzeichen, Erläuterungen und Legenden, die bei der Analyse berücksichtigt und vom eigentlichen Text unterschieden werden müssen.
Normale Text- und Bildregionen werden sehr zuverlässig erkannt. Eine Nachbearbeitung ist für bestimmte Strukturen wie Text-Artefakte (Sonderzeichen), Kopf-/Fußzeilen und Textblöcke in Bildregionen erforderlich. Besondere Aufmerksamkeit verdient die Lesereihenfolge (Reading Order) der Textregionen, die nur teilweise korrekt erfasst wird und eine eigene Nachbearbeitungs-Logik erforderlich macht.
Die Ergebnisse der Analyse und der automatischen Nachbearbeitung wurden anhand der Ground-Truth-Texte evaluiert. Dabei wurden Ähnlichkeitsmaße basierend auf Levenshtein Matching Blocks verwendet. Die Übereinstimmungen liegen je nach Testdokument zwischen 82 und 98 Prozent. Bei den Dokumenten mit den geringsten Übereinstimmungen lässt sich gut nachvollziehen, welche Dokumentstrukturen für die Extraktion problematisch sind. Konkret sind das z. B. aufwendig gestaltete Zeitleisten und große Bilder in mehrspaltigen Layouts. Diese Strukturen können sowohl in der visuellen Analyse als auch in der Nachbearbeitung noch besser berücksichtigt werden. Wichtig sind aber auch die Erkenntnisse darüber, welche Dokumentstrukturen zu Problemen führen, was in Zukunft bei der Erstellung der Dokumente berücksichtigt werden kann.

Lösungsidee

Bisherige Ansätze zur Dokumentenanalyse erfordern mehrere Arbeitsschritte, darunter die Layout-Analyse, mit der Dokumentregionen für Texte, Bilder und andere Inhalte unterschieden werden können, gefolgt von der eigentlichen Extraktion der Texte aus den Regionen. Aktuelle Tools und Frameworks fassen diese Arbeitsschritte zusammen und liefern mit einem Durchlauf die gewünschten Ergebnisse. Dabei kommen unter anderem Vision Language Models (VLMs) zum Einsatz.
Im Projekt wurden die Tools Docling, Qwen und YOLO eingesetzt, um die Inhalte aus Dokumenten mit komplexen Layouts zu erschließen. Neben der Ergebnisqualität wurden dabei auch die Ressourcenanforderungen betrachtet, die sehr unterschiedlich ausfallen können. Vor allem die Anzahl der GPUs und die Größe des VRAM sind hier entscheidend. Die Verarbeitungszeiten sind je nach Tool sehr unterschiedlich. Mit Docling konnten auf einem leistungsfähigen Server mit 4 Nvidia DGX A100 GPUs Verarbeitungszeiten zwischen 1 und 2 Sekunden pro Doppelseite erzielt werden.
Als Testset wurden 8 Dokumente mit insgesamt 30 Doppelseiten aus mehreren Reiseführern zusammengestellt. Die Seiten weisen unterschiedliche Layouts und Strukturen auf, sodass viele typische Dokument- und Layoutstrukturen abgedeckt sind. Als Ground Truth wurde der reine Fließtext aus den Dokumenten manuell erstellt, der anschließend für die Bewertung der Ergebnisqualität genutzt werden konnte.

In Zusammenarbeit mit

MAIRDUMONT

DEWE Stolz Energiesysteme GmbH

»Bereits heute sind digitale Plattformen für die Präsentation und den Vertrieb unserer Inhalte von großer Bedeutung und in Zukunft wird diese weiter steigen. Die hochwertige Erschließung der Inhalte aus unseren Publikationen ist die Voraussetzung für deren weitere Nutzung auf den verschiedenen Plattformen. Im Exploring Project mit dem Fraunhofer IAO sind wir dabei einen großen Schritt vorangekommen, konnten geeignete Tools und Vorgehensweisen für die Dokumentenanalyse kennenlernen und deren Leistungsfähigkeit einschätzen.«

Robert Krämer
MAIRDUMONT GmbH & Co. KG

Kontakt

Werner Kraus
werner.kraus@
ipa.fraunhofer.de

Kontakt:
info@ki-
fortschrittszentrum.de

Weitere Informationen:
www.ki-
fortschrittszentrum.de

Fraunhofer-Institut für Produktionstechnik und Automatisierung IPA

Nobelstraße 12
70569 Stuttgart

www.ipa.fraunhofer.de