



## *Automatische Klassifikation*

PROSAR-AIDA klassifiziert strukturierte und semistrukturierte, aber auch formlose Dokumente ("Freiform") automatisch und extrahiert für jeden Dokumententyp die Daten, die für Ihre Geschäftsprozesse erforderlich sind. Paradatec stellt mit PROSAR-AIDA ein marktführendes Erkennungssystem bereit, das mit einem lizenzierten jährlichen Verarbeitungsvolumen von über 900 Millionen Images bei zahlreichen Kunden im Einsatz ist. Wichtige Erfolgsfaktoren sind höchste Leistungsfähigkeit sowie benutzerfreundliche Entwicklungswerkzeuge, die es den Kunden ermöglichen, eigenständig Ihre Erfassungslösung zu pflegen.

### **Klassifikationsverfahren**

Die Klassifikation von gemischtem Beleggut, wie es für Posteingang typisch ist, basiert auf dem Textinhalt der Belege. Dadurch sind auch äußerlich sehr ähnliche Dokumente leicht klassifizierbar. Die Klassifikation durch PROSAR-AIDA kann mit

- manuell erstellten Klassifikatoren (Regeln) sowie
- trainierbaren Klassifikatoren

erfolgen.

Trainierbare Klassifikatoren sammeln während des Trainings statistische Informationen aus dem Textinhalt der Dokumente jeder Klasse und wenden diese im späteren Betrieb für die Unterscheidung an. In einem Regelbaum können trainierte Klassifikatoren und manuell definierte Regeln über eine grafische Benutzeroberfläche nahtlos miteinander verbunden werden.

Beide Klassifikationsverfahren lassen sich auf Images ebenso anwenden wie auf bereits codiert vorliegende Daten, wie z.B. Emails. Für Images führt PROSAR-AIDA zunächst eine Volltexterkennung (OCR) durch, die auf künstlichen neuronalen Netzen basiert. PROSAR-AIDA erkennt in höchster Qualität und beispielloser Geschwindigkeit die maschinenlesbaren Schriftzeichen auf den Dokumenten. Der Text wird dann genutzt, um den Beleg einer Dokumentenklasse zuzuordnen.

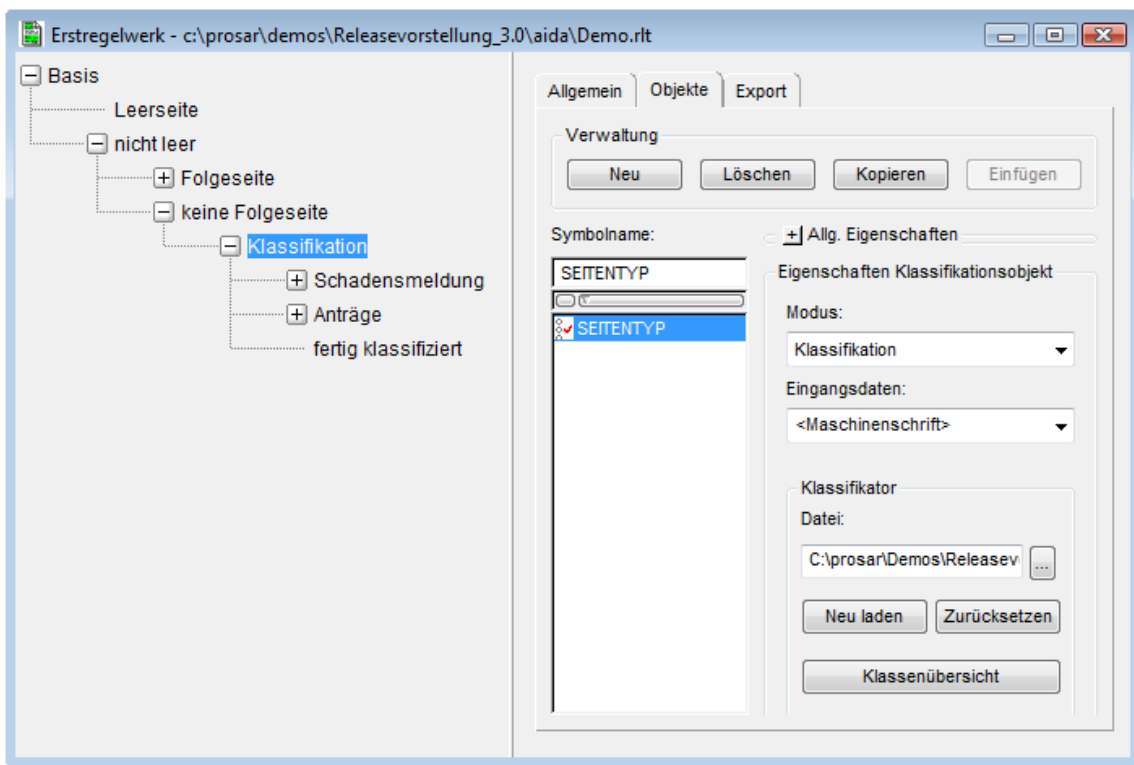
Je nach Anwendungsfall werden durch die Klassifikation auch in aufeinander folgenden Seiten Dokumentengrenzen erkannt um die Seiten automatisch zu Dokumenten zusammenfassen zu können. Alternativ erfolgt die Klassifikation für vorliegende Dokumente oder Vorgänge.

## Konfiguration

Um PROSAR-AIDA auf das Beleggut eines Kunden anzupassen bedarf es keiner Programmier- oder Skriptingkenntnisse. Über eine grafische Benutzeroberfläche wird ein sogenannter Regelbaum erstellt. In einem Regelbaum können ein oder mehrere trainierte Klassifikatoren verwendet und auch mit manuell definierten Regeln kombiniert werden. In der Praxis ergeben sich daraus vielfältige Szenarien, zum Beispiel:

- Durchscheinende Rückseiten können anhand einer Regel ausgesondert werden, bevor der trainierte Klassifikator die Texte analysiert.
- Wenn der trainierte Klassifikator einzelne Images aufgrund ihrer Inhalte nicht eindeutig zuordnen kann, lassen sich diesem manuelle Regeln nachschalten, die beispielsweise spezielles Fachwissen zur Unterscheidung berücksichtigen.

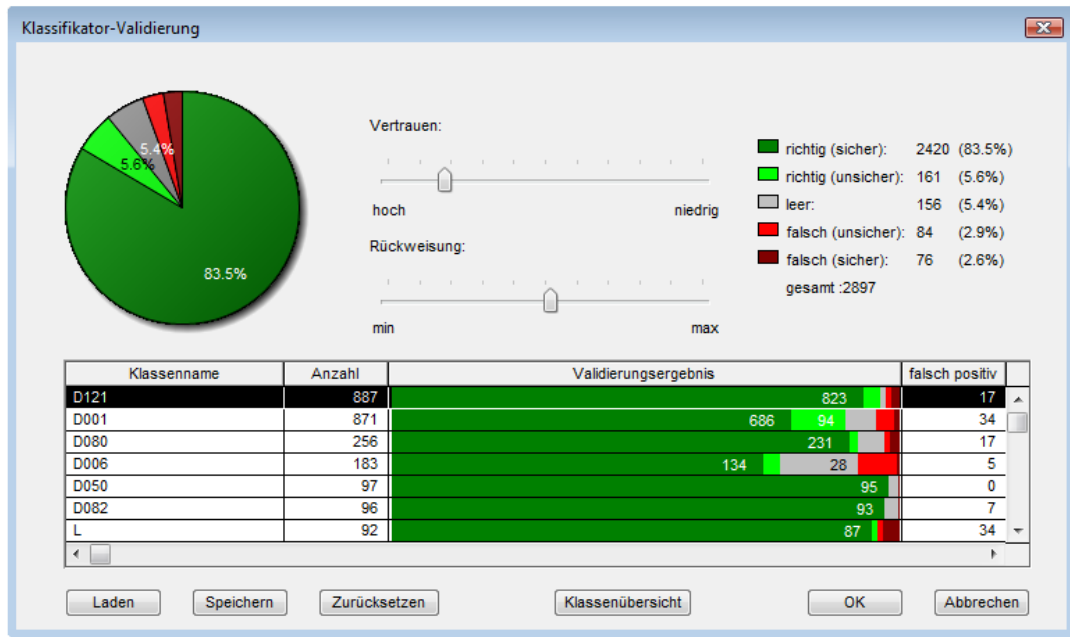
*Beispiel für ein PROSAR-AIDA Regelwerk mit einem trainierbaren Klassifikator*



## Werkzeuge

Training, Überprüfung der Trainingsergebnisse und Betrieb des automatischen Klassifikators werden in einer einheitlichen grafischen Oberfläche durchgeführt. Das senkt den Einarbeitungsaufwand. Integrierte Werkzeuge ermöglichen online die Analyse der Klassifikationsgüte. Im laufenden Betrieb lassen sich Schwellwerte anpassen, welche das „Vertrauen“ in die Ergebnisse des Klassifikators regeln.

## Konfigurationsoberfläche für einen trainierbaren Klassifikator



## Nutzen

Durch den kombinierten Einsatz trainierter und manueller Klassifikatoren im Projekt lassen sich die jeweiligen Vorteile beider Ansätze vereinen:

- Trainierte Klassifikatoren ermöglichen auf geeigneten Dokumenten Projektfortschritte innerhalb sehr kurzer Zeit.
- Nicht für alle Dokumenttypen ist die trainierte Klassifikation gleichermaßen geeignet oder wird die erforderliche Granularität der Klassifikation erreicht. Durch manuell definierte Regeln können auch feinste Klassifikationsentscheidungen getroffen sowie das Fachwissen der Mitarbeiter im System abgebildet werden.

Die Kombination der Verfahren ermöglicht eine effiziente automatische Verarbeitung von gemischtem Dokumentgut, während die manuelle Nachbearbeitung auf Ausnahmefälle reduziert werden kann.

## Unternehmen

Paradatec wurde 1988 mit Sitz in Braunschweig gegründet. Das Unternehmen hat sich systematisch zum Technologieführer für intelligente Dokumentenanalyse entwickelt. Software von Paradatec wird in unterschiedlichen Branchen im In- und Ausland eingesetzt. Paradatec ist in England und den USA mit Tochtergesellschaften vertreten.