

Automatische Erkennung von Schriftzeichen (Aufgabe A-72)

Das Problem für die Schriftenerkennung bei digitalisierten Dokumenten liegt darin, die vorliegenden Daten in ein vom Rechner erkennbares Format zu bringen. Diese Form der Erkennbarkeit wird für gedruckte Dokumente durch geeignete Texterkennungssoftware erreicht.

Eine spezielle Architektur, die für zerlegbare Probleme entwickelt wurde, sind die Tafelsysteme (blackboard systems). Hier wird ein globaler Datenspeicher (die Tafel) definiert, über den Daten ausgetauscht werden und auf die verschiedenen sogenannte Wissensquellen (knowledge sources) unabhängig voneinander zugreifen können. Diese Architektur ist speziell für grosse wissenbasierte Systeme konzipiert, die durch die Wissensquellen strukturiert werden.

Wenn auf dem Blackboard (d.h. in dem für alle sichtbaren Speicherbereich) Eingabedaten erscheinen, für die sich ein Agent zuständig fühlt, dann liest er sie, bearbeitet sie und schreibt seine Ausgabedaten danach wieder zurück auf das Blackboard. Dort können sie dann für einen anderen Agenten als Eingabedaten dienen. Es ist möglich, dass mehrere Agenten die gleichen Daten vom Blackboard lesen und bearbeiten, aber normalerweise ist die Kooperation über das Blackboard so geregelt, dass jeder Agent einen eigenen von den anderen disjunkten Aufgabenbereich übernimmt.

Ein Texterkennungssystem nimmt ein Dokumentabbild und wandelt die darin enthaltenen Textblöcke in eine vom Computer besser handhabbare, textuelle Repräsentation um. Diese Konversion erfolgt konzeptionell in mehreren, aufeinanderfolgenden Schritten:

1. Digitalisierung
2. Bildverarbeitungsoperationen
3. Layoutanalyse
4. Zeilensegmentierung
5. Zeichensegmentierung
6. Zeichenklassifikation
7. Postprocessing

Die integrierte Verarbeitung dieser Schritte führt im allgemeinen zu einer enormen Vergrößerung der Softwarekomplexität eines Texterkennungssystems, da die dabei beteiligten Prozesse nicht trivial sind. Des Weiteren kann jeder Verarbeitungsschritt auf mehrere Arten ausgeführt werden. Um alle Verarbeitungsschritte in integrierter Weise durchführen zu können und zur Verwaltung der verschiedenen Erkennungshypothesen ist daher eine geeignete Softwarearchitektur notwendig. Ein Blackboard-Modell ermöglicht eine flexible Verwaltung von Erkennungshypothesen und erlaubt in beliebiger Reihenfolge auszuführen, ohne dass sich die Softwarekomplexität vergrößert.

Das allgemeine Blackboard-Modell umfasst drei Komponenten: eine Datenstruktur namens Blackboard, mehrere Wissensquellen und eine Steuerkomponente. Das Blackboard speichert alle Hypothesen, die zu einem bestimmten Verarbeitungszeitpunkt verfügbar sind. Die Wissensquellen lesen und modifizieren diese Hypothesen, und die Steuerkomponente kontrolliert den Verarbeitungsablauf.

Das im Projekt von A.Garzotto, R.Sennhauser und P.Stucki verwendete Blackboard-Modell ist eine speziell für Texterkennungsanwendungen zugeschnittene Implementierung des allgemeinen Blackboard-Modells: ein blackboard-basiertes Texterkennungs-Rahmensystem. Es speichert temporäre Erkennungsergebnisse wie beispielsweise segmentierte Textblöcke, Segmentierungshypothesen, alternative Zeichenklassifikationen, etc. als Erkennungshypothesen

auf dem Blackboard. Die Wissensquellen umfassen alle Algorithmen, die für die Texterkennung notwendig sind, wobei jede einzelne Wissensquelle nur eine kleine Teilaufgabe des ganzen Erkennungsprozesses ausführt. Die Steuerkomponente wählt eine Erkennungshypothese und eine darauf anzuwendende Wissensquelle aus. Diese Auswahl erfolgt mittels eines dreiteiligen Regelsystems. Im ersten Teil werden die Wissensquellen für die Bildverarbeitung und die Segmentierung gerufen, im zweiten Teil für die Klassifikation und im dritten Teil diejenigen für des Postprocessing. Diese Teile wiederholen sich bis alle Hypothesen auf dem Blackboard bearbeitet wurden und die Änderungen stattgefunden haben.

Eine Besonderheit vom Blackboard-Ansatz und vor allem von Agentengestützten Systemen kommt eigentlich besonders zur Geltung, wenn einzelne Buchstaben nicht erkannt werden. Dann kommen nämlich Kontextanalyse-Agenten zum Einsatz, die im Dictionary nachschlagen und die zutreffendste Variante wählen, was die Fehlerrate erheblich verringert. Probleme entstehen mehr bei kurzen Wörtern als bei Längen. Dies ist auch logisch, da die Variationswahrscheinlichkeit von langen Wörtern auch geringer ist als von Kurzen.

[gestützt auf „Schlussbericht Forschungsprojekt Identifikation von Schriftzeichen in digitalisierten Dokumenten“ von A.Garzotto, R.Sennhauser, P.Stucki]

Abbild: Beispiel eines Agentengestützten Umfeldes im Blackboard-Ansatz

