

WBS 2004/05

Aufgabenblatt 7

Programme zum flexiblen Vergleich von Symbolfolgen besitzen zahlreiche Anwendungen z.B. zur Tippfehlerkorrektur, zum Vergleich von Gensequenzen oder der Evaluation von Mustererkennungsprogrammen. Ziel der Vergleichs ist die Ermittlung einer Editierdistanz, die die minimalen Kosten zur Umwandlung der einen Symbolfolge in die andere Symbolfolge beschreibt. Die Kosten werden für elementare Editieroperationen (wie Vertauschung eines Symbols in ein anderes) definiert und über alle erforderlichen Editieroperationen summiert.

1. Entwerfen Sie eine Menge von elementaren Editieroperationen für Symbolfolgen. Überlegen Sie, ob Ihre Menge minimal ist. Begründen Sie, warum Sie sich für eine minimale bzw. nichtminimale Menge von Operationen entschieden haben. Diskutieren Sie, inwieweit ihre Modellannahmen anwendungsspezifisch sind und welche alternativen Modellierungsansätze existieren.
2. Implementieren Sie ein Programm, das für zwei gegebene Symbolfolgen die minimale Editierdistanz berechnet. Nehmen Sie hierfür gleiche Kosten für alle Editieroperationen an.
3. Analysieren Sie das von Ihnen verwendete Suchverfahren. Was sind die Zustände des Suchraums? Ist der Suchraum endlich? Welche Rolle spielen die Kosten bei der Steuerung der Suche? Welche alternative Suchstrategien gäbe es? Worin bestehen die jeweiligen Vor- und Nachteile? Analysieren Sie den Ressourcenbedarf Ihres Programms theoretisch und empirisch.
4. Erweitern Sie Ihre Implementation für den Fall ungleicher Kosten. Begründen Sie, warum dies für manche Anwendungen vorteilhaft sein kann.