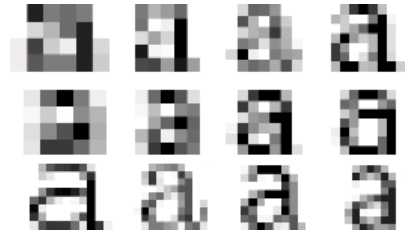


# Java-Wettbewerb 2009

## Aufgabenstellung

Die Aufgabe in diesem Jahr ist die Klassifizierung/Erkennung von einzelnen Buchstaben. Eine Grundgesamtheit von 15.600 Screenshot-Bildern einzelner Buchstaben wurde in zwei gleich große Mengen zerlegt. Die erste Menge (Lern-/Trainingsdaten) wird allen Teilnehmern ausgehändigt und dient dazu, Erkennungsalgorithmen zu entwickeln und zu testen. Die zweite Menge (Testdaten) wird von der Jury benutzt werden, um die abgegebenen Programme zu testen.



Das Ziel des Wettbewerbs ist es, eine Java Applikation zu entwickeln, die die Buchstaben möglichst gut erkennt. Sieger des Wettbewerbs wird der Teilnehmer oder die Teilnehmergruppe deren Programm die höchste Erkennungsrate besitzt.

### Hintergrund

Ein recht neues Forschungsgebiet im Bereich der Mustererkennung befasst sich mit der Erkennung von Screen-Rendered Text (SRT). SRT ist Text, der zur Darstellung auf Bildschirmen gerendert wurde. Aktuelle Renderingverfahren erlauben Text sehr klein zu rendern, was zu Text in sehr niedriger Auflösung führt, bei dem die einzelnen Buchstaben miteinander verschmelzen. Dies und weitere Umstände machen SRT für bisherige OCR Verfahren unlesbar, wenn die Auflösung sehr gering ist.



Die Erkennung von angezeigtem Text mit Hilfe von OCR ist für mehrere Anwendungen sehr wichtig. Übersetzungstools, Programme zur Indexierung von Web-Images, Bild-basierte Spamfilter und weitere Anwendungen sind auf OCR für gerenderten Text angewiesen.



### Daten

Die Daten für diesen Wettbewerb wurden im Rahmen einer Forschungsarbeit zu SRT erstellt und von Dr. Steffen Wachenfeld zur Verfügung gestellt. Die Nutzung der Daten im Zusammenhang mit anderen Zwecken, insbesondere die Nutzung für kommerzielle Zwecke, bedarf einer gesonderten Genehmigung.

Bei den Daten handelt es sich um systematisch erstellte Screenshot-Bilder von einzelnen Buchstaben. Die Buchstaben liegen in Form von 2x150 Sätzen (sogenannten Sets) vor. Ein Set besteht aus 52 Buchstaben (a-z, A-Z) in gleicher Größe und gleichem Stil. Die Sets umfassen Text verschiedener Schriftarten. Die Größe der Buchstaben ist mit einer x-Höhe von 4-8 Pixeln gering.

## *Implementierung*

Die Aufgabe ist in Form einer Java Applikation zu lösen, welche am Ende des Wettbewerbs in kompilierbarer Form abgegeben werden muss. Java ist in Version 1.5 oder höher zu benutzen.

Die Nutzung von fremdem Code und Toolboxen ist ausdrücklich erlaubt. Gefordert wird jedoch, dass die benutzten Tools verstanden wurden und die Funktionsweise der Tools auf der Abschlussveranstaltung erklärt werden kann. Sämtlicher zur Ausführung benötigter Code muss mit abgegeben werden (JARs, Zip-Datei).

Zur Erleichterung der Programmtests durch die Jury **muss** das Interface `SRCClassifier` implementiert werden. Dieses Interface sieht eine Methode `public char classifySRC(int[][] grayImage)` vor, die von der Jury zur Übergabe der Testdaten aufgerufen wird. Diese Methode soll Screenshot-Bilder von einzelnen Buchstaben als 2D int-Array entgegen nehmen und klassifizieren. Das Klassifikationsergebnis soll als `char` zurückgeliefert werden. Es gibt 52 zulässige Klassen (a-z, A-Z). Eine Rückweisungsklasse ist nicht vorgesehen und unzulässig.

## *Teilnahmebedingungen*

Die Teilnahme ist Einzelpersonen sowie Gruppen aus Personen gestattet, die Studierende der Westfälischen Wilhelms-Universität Münster sind und zu Beginn des Wettbewerbs noch nicht über einen Diplom- oder Masterabschluss verfügen.

Im Einzelfall entscheidet die Jury über die Teilnahmeberechtigung.

Abgabe mit Namen und Matrikelnummern aller Teilnehmer bis zum **15.04.2009** – 23:59h MEZ.

## *Materialien und Literatur*

Offizielle Seite zum Java-Wettbewerb (Downloads: `SRCClassifier.java` + Learning Data):

<http://www.viadee.de/unternehmen/java-wettbewerb.html>

Grundlagen zur Klassifizierung / Mustererkennung:

Skript zur Vorlesung Mustererkennung:

<http://cvpr.uni-muenster.de/teaching/ws08/mustererkennungWS08/>

Informationen zu Screen-Rendered Text Recognition:

<http://cvpr.uni-muenster.de/research/ScreenTextRecognition/index.html>

**Viel Erfolg und vor allem viel Spaß!!!**

*Das Wettbewerbsteam*