



Program *Transkribus* i njegova upotreba u muzikologiji: primjer III. sveska Kuhačeve korespondencije

Sara Ries

HMD, 21. godišnji susret

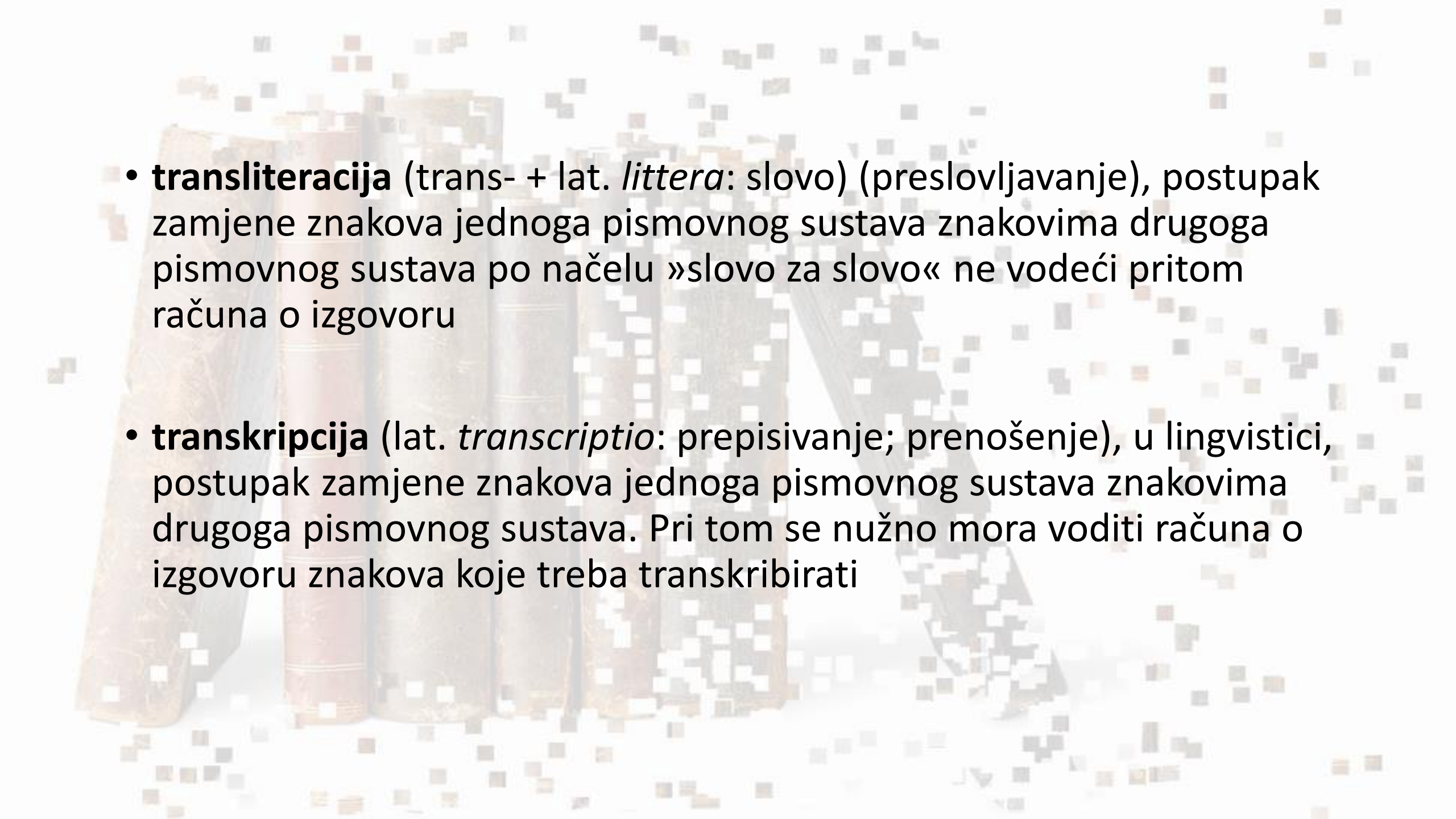
Zagreb 2020.

Transkribus



- optičko prepoznavanje rukopisnih znakova povijesnih dokumenata
- automatsko prepoznavanje slova rukopisa na temelju unesenoga skupa podatka kojima se koristi za trening te dodatno omogućuje pretraživanje sadržaja digitaliziranih rukopisa
- **HTR** (handwritten text recognition) training model \neq **OCR** (Optical Character Recognition)
- podržava i stare rukopise poput hebrejskoga, ćirilice i arapskoga, tj. svaki jezik i rukopis
- <https://read.transkribus.eu/category/success-stories/>



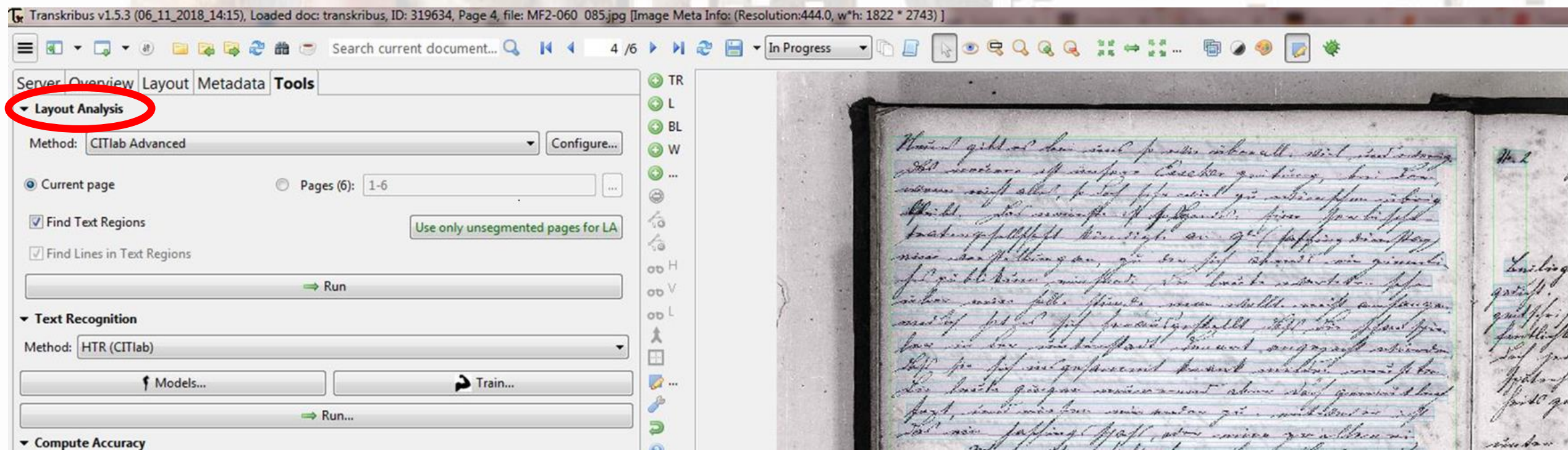
- 
- **transliteracija** (trans- + lat. *littera*: slovo) (preslovljavanje), postupak zamjene znakova jednoga pismovnog sustava znakovima drugoga pismovnog sustava po načelu »slovo za slovo« ne vodeći pritom računa o izgovoru
 - **transkripcija** (lat. *transcriptio*: prepisivanje; prenošenje), u lingvistici, postupak zamjene znakova jednoga pismovnog sustava znakovima drugoga pismovnog sustava. Pri tom se nužno mora voditi računa o izgovoru znakova koje treba transkribirati

Proces transliteracije na Kuhačevoj korespondenciji

- rukopisna gotica
- korespondencija je transliterirana kompletno i djelomično u par navrata
- Vinko Žganec 1950-ih: prve transliteracije (čuvaju se u strojopisu u Odsjeku za povijest hrvatske glazbe HAZU)
- Žgančeve transliteracije sadrže pogreške (krivo pročitane riječi) te pisma nisu doslovno transliterirana, odnosno nije se pazilo na tadašnju ortografiju koju je Kuhač koristio
- Ladislav Šaban je 1980-ih pokrenuo projekt u sklopu kojega je Vjera Katalinić načinila transliteracije pisama iz prvog sveska korespondencije (1860–1863)
- od 2017. projekt „Umrežavanje glazbom: promjene paradigmi u 'dugom 19. stoljeću' – od Luke Sorkočevića do Franje Ksavera Kuhača“ (NETMUS19): transliterirani su drugi (1864-1869) i treći svezak (1870-1874) korespondencije (Sara Ries)
- za transliteraciju trećeg sveska koristio se i program *Transkribus*

Važnost pravilne priprave dokumenta

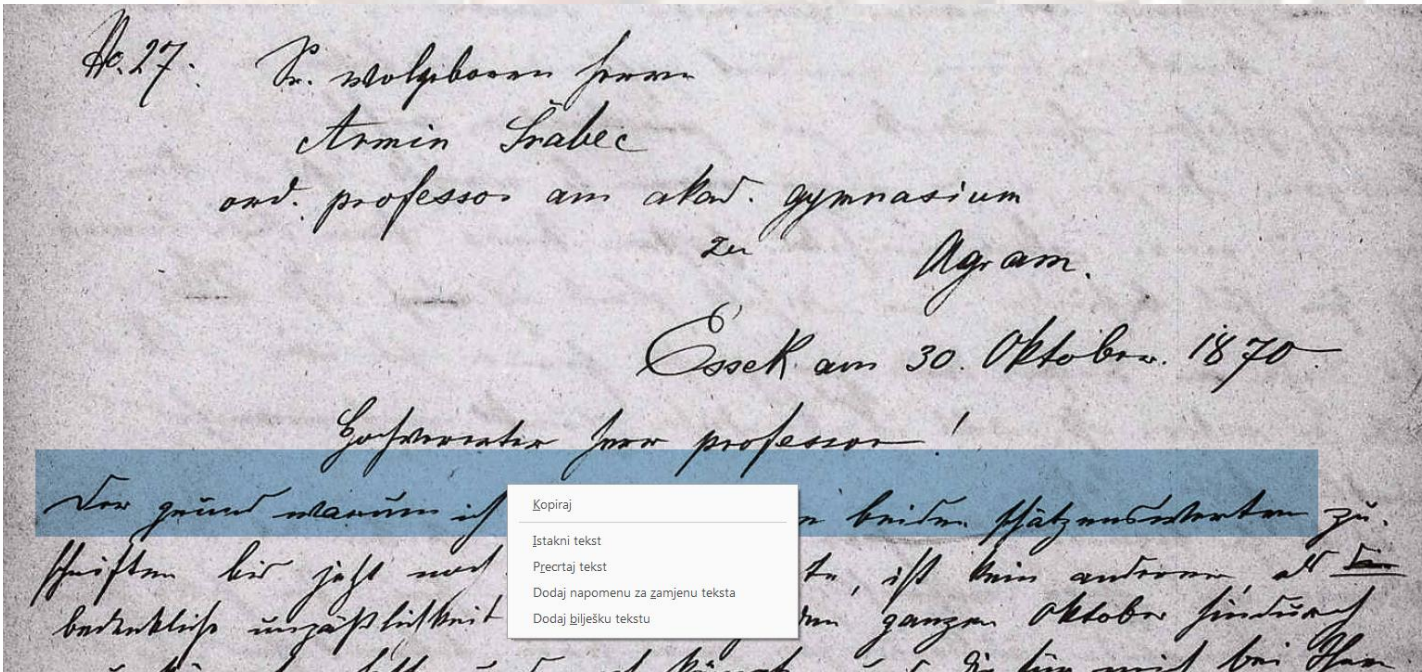
- nakon uvoza dokumenta potrebno je napraviti tzv. *layout analysis*, odnosno analizu rasporeda redaka
- problem ako slika teksta ima ispisane margine – budući da program prepoznaje retke, prepoznat će i retke ispisane na marginama → greške u transliteraciji



The screenshot displays the Transkribus v1.5.3 web interface. The top navigation bar includes 'Server', 'Overview', 'Layout', 'Metadata', and 'Tools'. The 'Layout Analysis' section is highlighted with a red circle. Below it, the 'Method' is set to 'CITlab Advanced'. The 'Current page' radio button is selected, and 'Pages (6): 1-6' is specified. The 'Find Text Regions' and 'Find Lines in Text Regions' checkboxes are checked. A 'Run' button is visible. The 'Text Recognition' section shows the 'Method' set to 'HTR (CITlab)', with 'Models...' and 'Train...' buttons. The 'Compute Accuracy' section is partially visible at the bottom. The main document view shows a page of handwritten text with a green bounding box around the main text area and a red bounding box around the margin text.

Izvoz dokumenta

- obrađene rukopise može se izvesti u različitim formatima poput datoteka Word, .txt, .xml te PDF-a
- u PDF dokumentu sa slikama rukopisa moguće je kopirati tekst
- opcija označavanja riječi koje su programu nečitke



Sr. Wolgeboren herrn
Armin Šrabc
ord. professor am akad. gymnasium
zu
Agram
Essek, am 30 Oktober 1870
Hochvererter herr professor!

Der grund warum ich Ihnen auf Ihre beiden schätzenswerten zu
schriften bis jetzt noch nicht antwortete, ist kein anderer, als den
bedenkliche unpässlichkeit, mit der ich den ganzen Oktober hindurch
zu kämpfen hätte, und noch kämpfe, und die für mich bei Ihnen
um nachsicht bittet. Es ist mit mir lieber professor so schlecht
bestellt, dass ich das schlimmste zu befürchten habe, kommt nicht vor
außer irgend welche hilfe.

A row of seven old, worn books with a pixelated effect. The books are arranged in a slightly curved line, with the one on the far right leaning against the others. The covers are made of brown leather and show signs of age and wear. The background is white with a pattern of small, scattered pixels in various shades of brown and grey, creating a digital or glitch-like aesthetic.

Hvala na pažnji!