Why the Humanities?

Sayeed Choudhury

Johns Hopkins University

Computing Research Association

July 12, 2004
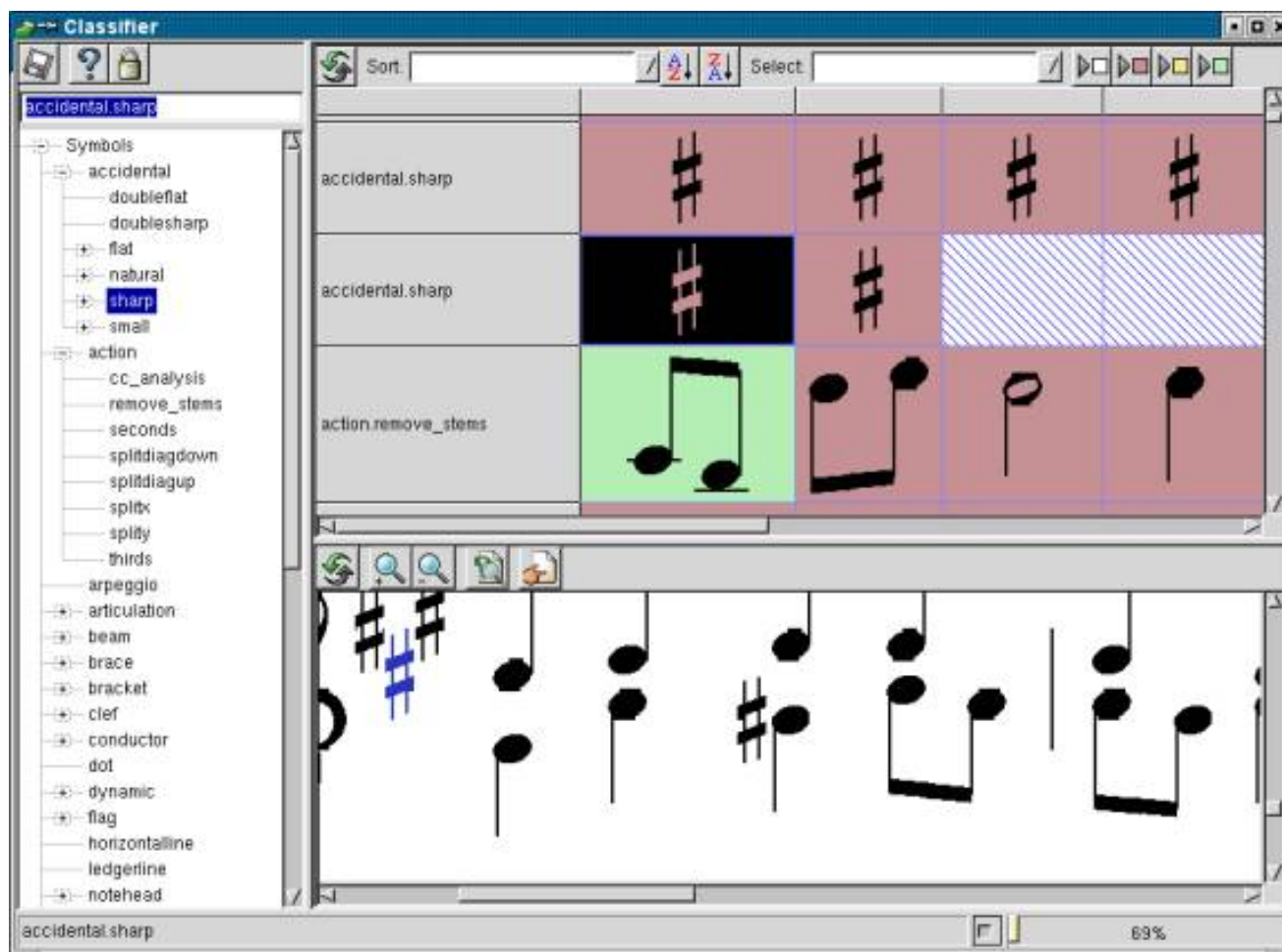
Concepts to Consider

- Scalability

- Complexity

- Context

- Interpretation

- Humanities offering challenges to computer science?

Optical Music Recognition

- Exemplar-based software to recognize, classify and interpret music and lyrics from Lester S. Levy Collection of Sheet Music -- AOMR

- Now at the point of considering application to collection, but also examining broader range of humanities collections (and problems) -- Gamera (funded by NSF ITR)

Training "session" with music

Other Humanities Collections

- Statistical Accounts of Scotland - Sir John Sinclair initiated a "census" in the 1790s and 1840 ("early modern english")

- Medieval french manuscripts such as the Roman de la Rose and Christine de Pizan documents

Scalability

- Humanities are not actually "data poor" but rather data needs to be extracted

- Still can't match the scale of large-scale, digital science projects such as Sloan Digital Sky Survey

Complexity

- Humanists have complex documents, and complex queries
- Rich tradition of dealing with context, interpretation, annotation, provenance…
- "Data are fuzzy…answers are approximate"

  - Alex Szalay, PI of the National Virtual Observatory Project

Contact information

- Sayeed Choudhury
- Sayeed@jhu.edu
- 410-516-4930

- http://dkc.jhu.edu/gamera