10. Koraljka Golub (*Linnaeus U.*):
Automated Subject Indexing
Automatic Subject Indexing for Cultural Heritage

Koraljka Golub, Associate Professor, Head of the Institute

IMLS Workshop at the iConference 2019:
Developing a Computational Framework for Library and Archival Education
Definition and uses

- Subject indexing = assigning topics to documents (controlled, keywords, tags)
- Automatic subject indexing = here assigned indexing from KOS

- Possible CH uses:
  - Improve precision and recall
    - Additional access points in digital LAM collections
    - Interoperability
  - MAI (Golub 2016)
  - Support hierarchical browsing (Golub 2011)
  - Enhance social tagging (Golub, Lykke & Tudhope 2014) and author keywords creation
Approaches and challenges

• Approaches (Golub 2017)
  • Machine learning: supervised and unsupervised
  • String matching (Golub 2011)

• Challenges
  • Humanities harder than natural sciences
  • Multimedia harder than text
  • Evaluation issues (Golub et al. 2016)
    • Context of IR and relevance
    • Test-collection model
      – Exhaustivity / specificity
    – Bias:
      • Algorithm / training data
    • KOS
References


Computational thinking in ilnstitute education

- Courses:
  - Introduction to Programming for Digital Humanities
  - Computational Thinking (new course in Digital Learning Master)
  - Basic Programming for Teachers of Mathematics and Technology

- More planned as part of Digital Humanities Master
Institute current and *planned* future programs

**Undergraduate**
- B.Sc. in Library and Information Science
- B.A. in Interactive Media and Web Technologies
- B.Sc. in Information Logistics
- M.Sc. in Data Science
- M.Sc. in Information Management

**Graduate**
- M.A. in Cultural Sciences
- M.A. in Digital Humanities
  - Submitted to Faculty Board, may start Aug. 2020
- M.Sc. in Social Media and Web Technologies
- M.Sc. in E-Health
- M.Sc. in Information Systems
- M.A./M.Sc. in Digital Learning
  - Submitted to Faculty Board, may start Aug. 2020

**Doctoral**
- Ph.D. in Computer and Information Science
- Ph.D. in Contract Archeology (GRASCA)
- Ph.D. in Biomedicine
- Ph.D. in Library and Information Science
  - Discussion started