“Turkologischer Anzeiger / Turkology Annual” (TA):
An indispensable systematic bibliography for Turkology and Ottoman Studies

- contributions by experts from all over the world
- structured by broad categories
- references to books, articles, reviews, and conferences
- in more than 20 mostly non-Western languages like Russian, Turkish, Arabic, Japanese
- titles in less common languages (e.g. Hungarian, Arabic) are translated
- even single entries may contain chunks in several different languages

Our aims and objectives: Contributing to a modern research infrastructure

- digitizing all TA volumes and re-publishing the entries in an online database
- adding value by offering new, efficient search options in a multi-lingual user interface enhanced with an editing environment
- establishing a workflow for future digitizations of multi-lingual bibliographies
- demonstrating the chances of Open Access to scientific knowledge

Turkology Annual Online: A pilot project pointing out new challenges

- accordingly fine-tuned, state of the art Optical Character Recognition (OCR) software produced high quality results:
  - few remaining errors, mostly irrelevant for database reliability, only very few grave OCR errors
  - developing automatic OCR correction software is not a primary task
  - need for small scale manual corrections always remains:
    projects targeted at large or particularly committed user communities may also use online tools for manual corrections by users (“crowdsourcing”)

- crucial role of syntax analysis (parsing):
  - only implicit marking of different entry types and their data structures, some of them changing from volume to volume: each entry needs to be classified and accordingly parsed for storage in the database
  - cross-references contained in the entries need to be transformed into database relations
  - major features of the online bibliography like consolidation of information and hyperlinks depend on syntax analysis
  - high demands on parsing robustness: human-made entries may contain structural errors

- highly adaptive syntax analysis software needs to be tailored on the project using advanced language technology like machine learning and named entity recognition