

# Literatur-Verwaltung

Karsten Weicker

11. Oktober 2017

## 1 Zielbestimmung

Beim wissenschaftlichen Arbeiten sammelt man unweigerlich eine große Mengen an wissenschaftlichen Veröffentlichungen anderer Autoren, die einen Bezug zur eigenen Arbeit haben. Eine solche Sammlung muss verwaltet werden, damit man sie sinnvoll benutzen kann.

## 2 Produktbeschreibung

- Die Software muss Meta-Daten von wissenschaftlicher Literatur verwalten
  - Typen: Buch, Tagungsband, Tagungsbeitrag, Zeitschriftenbeitrag, Technischer Bericht, Masterarbeit, Doktorarbeit, Sonstiges
  - alle relevanten Felder (vgl. Bibtex) sind zu verwalten
  - Abstract und Schlüsselworte sind zu erfassen
- Es soll nach Autoren, in den Abstracts und nach (mehreren) Schlüsselworten gesucht werden können.
- Export der bibliographischen Informationen als Bibtex-Eintrag.
- Zu einem Eintrag sollen ähnliche Dokumente suchbar sein. Dabei kann der Benutzer der Software wählen wie stark die Ähnlichkeit der Schlüsselworte, die direkten Referenzen aufeinander sowie die Überlappung der Autoren in den Literaturreferenzen benutzt werden soll.
- Zu einem Dokumente sollen alle anderen Dokumente der (oder eines der) Autoren anzeigbar sein.
- Die Daten sollen direkt im Dateisystem (evtl. unter Nutzung einer integrierten Datenbank wie H2) abgelegt werden.
- PDF-Dateien (Dokumente) sollen in der Software abgelegt und auch wieder exportiert werden.
- Die Meta-Daten sollen entweder über Eingabemasken eintragbar sein oder per Import als Bibtex-Eintrag erfasst werden – Bibtex-Einträge müssen einer Plausibilitäts- und Konsistenzprüfung unterzogen werden.
- automatischer Upload von PDF-Papern soll möglich sein – dazu gehört:

- Abstract und Schlüsselworte sollen automatisch aus der PDF-Datei extrahiert werden – dies soll die primäre Kernfunktionalität des Systems darstellen. Bezüglich der Schlüsselworte sollen alle nicht-trivialen Worte des Systems erhoben werden, per Unschärfe z.B. Singular und Plural oder einfache Konjugationen bei Verben aufeinander abgebildet und die 20 wichtigsten Begriffe des Texts erhoben werden.
- Ebenfalls sinnvolle wäre eine automatisierte Erkennung von Literaturreferenzen. Hier ist das Ziel Verweise auf andere in der Software abgelegte Dokumente zu erkennen.
- die weiteren bibliographischen Informationen müssen von Hand gepflegt werden
- Die Darstellung der Schlüsselworte als Tag-Cloud zu einem Dokument ist sinnvoll.