Überlegungen zu Tagging der DHA-Zotero-Bibliothek

**Umgang mit Über- und Unterkategorien**

* corpus
  + historical corpus
  + large corpus
* freedom
  + freedom of speech and expression
* data
  + big data
  + text data
* census
  + american census
* literature
  + english literature
  + australian literature
* collaboration
  + collaborative interpretation
  + collaborative text analysis
  + collaborative text annotation
  + collaborative text editing
* students
  + college students
* books
  + physical book
  + altered books
  + adress book
  + comic books
  + electronic book
  + textbooks
  + artist books
* periodicals
  + newspapers
* möglichst abstrahieren oder möglichst kleinteilig? Überkategorie immer mitangeben, wenn Unterkategorie vorkommt?
* wo ist Unterscheidung wichtig und wo darf verallgemeinert werden?
  + z.B. vorangestellte Adjektive „digital“, „social“, „collaborative“
  + z.B. digital text analysis, collaborative text analysis, text analysis
* hierarchisches System auch innerhalb der Kategorien?
* in Zotero grundsätzlich Suchen nach mehreren Tags möglich, insofern eventuell auf Kombination verschiedener Tags setzen, anstatt alle Informationen in ein einzelnes Tag zu geben?

**Endentscheidung Über- und Unterkategorien**:

* ein Ansatz: Splitten auf zwei einfachere Tags, wo möglich
* Hierarchisierung im Hintergrund: in Excel eigene Spalte mit nächsthöher Kategorie vermerken
  + nutzbar für Suche auf dha-Seite
* Entscheidung mit Interannotator-Agreement/im Plenum: was ist zu spezifisch?

**Endentscheidung Präfixe**: keine Präfixe

Antwort:

* Mut zum Löschen
* zu spezifische Sachen löschen
* Balance wichtig
* Counts miteinbeziehen
* eher in mehrere Tags aufsplitten: collaboration + text analysis
* „digital“ entfernbar

**Sprache der Tags**

* alles Fremdsprachliche (= nicht Deutsch und nicht Englisch) ausgeschlossen
* Deutsch und Englisch (alles übersetzen) vs. Deutsch oder Englisch (beide Teile behalten) vs. vereinheitlichen (Mehrheit auf Englisch)?

**Endentscheidung Sprache**:

* Englisch als primäre Sprache
* Spalte „german\_varient“ neu 🡪 hier deutsche Varianten (on-the-go ergänzen)
* wenn nur deutsch: englische Variante hinzufügen

**Grammatische Standards bei Tagbezeichnungen**

* was soll als Norm angenommen werden?
* Groß- und Kleinschreibung?
* woran bei Rechtschreibung orientieren? (z.B. open-source vs. open source, analyze vs. analyse)
* Abkürzungen immer auflösen?
* Verben: Infinitiv vs. Imperativ
* Nomen: Singular vs. Plural
* mitunter abhängig von Sprachwahl

**Endentscheidung Grammatik:**

* möglichst Kleinschreibung
* British English
* Referenzwerk auswählen (z.B. Oxford Dictionary, Merriam-Webster)
* Abkürzungen: nicht immer auflösbar, gängige Abkürzungen belassen und als prefLabel behandeln
* Orientierung an Tadirah: -ing-Endungen (Gerundium), z.B. analysing
* Singular (außer Fälle, die nur im Plural auftreten)

**Umgang mit Wiederholungen/Dopplungen in verschiedenen Kategorien**

* blog – blogging
* archival studies – archives – archiving – archivists
* libraries – librarians – library science
* nebeneinander bestehen lassen (auch wegen unterschiedlicher Gruppen) vs. vereinheitlichen und auf abstraktes Überkonzept zurückführen?

**Endentscheidung:** nebeneinander bestehen lassen

* im Idealfall am Ende: jeder Zotero-Eintrag durch zwei Dimensionen (Disziplin-Method/Activity vs. Object-Type) getaggt

**Umgang mit Eigennamen/Entitäten**

* valide Taggingkategorie oder ausschließen?
* falls Aufnahme: Frage nach Vorgehen bei Normalisierung
* z.B. Personennamen:
  + Clinton, Bill vs. Bill Clinton
  + Geburtsdatum in Klammer dahinter beibehalten?
* z.B. Ortsnamen:
  + Spezifizierungen in Klammern behalten und falls ja, standardisieren?
  + z.B. Graz, AT; California (southern)
* z.B. (historische) Sprachen/Sprachstufen:
  + middle english, 1100-1500 🡪 Zeitraum dabei lassen?

**Endentscheidung Kategorien**:

* Entitäten weglassen
* Hintergrund: oftmals Wiederholung in Titel

**generell: Umgang mit Klammerausdrücken**

* modernism (literature)
* minnesota, university of (minnesota population center)
* rosenzweig, roy (obituary)
* information systems applications (incl.internet)
* technology (applied sciences)
* textbooks (review article)
* text processing (computer science)
* internet (review article)
* löschen, als eigenes Keyword oder in Text von Keyword integrieren?

**Endentscheidung Klammerausdrücke:**

* Sammeln und im Plenum besprechen (einzeln zu lösen)

**Ziehen von Grenzen bei Auftrennen einzelner Keywords in mehrere Keywords**

* s. Beispiele in Exceltabelle
* study/learning/teaching etc. 🡪 zusammenführen mit Überbegriff, z.B. education

**Endentscheidung Splitting**:

* Splitting, wo möglich
* Löschen, wenn völlig unklar