DIPF
Leibniz-Institut für Bildungsforschung und Bildungsinformation

EduTopics – Eine interaktive Web-App zur Exploration von Forschungsfeldern mit Textmining und Machine Learning

Alexander Christ

Ziele EduTopics

- Ziel: Kartierung des European Educational Research Landscapes
  - Aktueller Stand: Kartierung von Repräsentationen der Forschungslandschaft (done: ECER; WiP/under discussion: EARLI, WERA…)
- Interaktive, niederschwellige Bereitstellung der Ergebnisse

➢ Schaffung neuer Zugänge für Forschung und Forschungsinformation

Anwendungsbeispiel - EduTopics: ECER

Analyse und Kartierung der *n* > 30.000 Beiträge zur ECER von 1998 bis 2024, nach:

- Inhaltlichen Themen → **Topic Modelling**
- Autor:innen
- EERA-Netzwerken
- Affiliationsländer
- Jahr

… und allen Kombinationen dieser Variablen.

Visualisierung Erklärung Topic Modelling

https://dipf-lis.shinyapps.io/EduTopics2025beta/

>> Methods >> Topic Modelling (TM)

Topic Modelling: Zusammengefasst

Topic modelling ist ein computerlinguistisches Machine-Learning-Verfahren zur Identifikation latenter inhaltlicher Themen in einem Korpus (sog. Topics; siehe Blei, Ng & Jordan, 2003).

Annahmen:

- Im Korpus werden *k* Themen behandelt (hier *k* = 50).
- Die Themen können durch die Wörter bestimmt werden, die in den Dokumenten (gemeinsam) vorkommen.
- Jedem Wort kann pro Topic ein Relevanzgewicht β zugeordnet werden.
- Jedem Dokument kann pro Topic ein Relevanzgewicht γ zugeordnet werden.

➢ Anzahl der Topics *k* und bayesschen Prior können durch verschiedene Simulationsverfahren wie Gibbs-Sampling und qualitativer Beurteilung der Topic(-Cluster) bestimmt werden.

➢ Topic Modelling hat eine hohe Ähnlichkeit zum Vorgehen bei qualitativen Verfahren wie Grounded Theory (inkl. Mixed Membership), nur eben dass die Themen algorithmisch in > 1000 Iterationen identifiziert werden.

Beispiele: Forschungsfragen mit EduTopics

- Jegliche Art von bibliographischen Forschungsfragen zum ECER-Korpus
- Welche zeitlichen Trends lassen sich für Themen, Länder, EERA-Netzwerke identifizieren?
- Wie kooperieren Länder, Netzwerke, Autoren auf der ECER miteinander?
- Welche Unterthemen ergeben sich in dem übergeordneten Thema z.B. zu Metaresearch and Reviews? Wie ist deren zeitliche Entwicklung?
- Zu welchem Thema würde meine eigene Arbeit zählen? Welche Beiträge sind ähnlich dazu?

…. >> https://dipf-lis.shinyapps.io/EduTopics2025beta

Fazit, Ausblick etc.

- EduTopics ermöglicht eine eigenständige interaktive Exploration der Beiträge zur ECER seit 1998
- Dadurch werden neue Zugänge für Forschende geschaffen (z. B. Qualitative Analysen mittels Actor-Network-Theory)
- Die erste (Beta-)Version der App findet viel Anklang in der (ECER) Community und wird bereits in mehreren Forschungsprojekten genutzt.

Ausblick:

- Ggf. Integration weiterer Datenquellen (z.B. EARLI, WERA, **Fachportal Pädagogik**, **ERIC**, Institut français de l'éducation)
- Integration weiterer Funktionen (geplant: siehe App >> Currently WiP), z.B. Einlesen und Analysieren PDF

Herausforderungen und Probleme bei Erweiterung/Weiterentwicklung

- Mehrsprachigkeit der Dokumente (ggf. Jingle-Jangle-Fallacies, linguistische Besonderheiten)

⇒ Automatisierte Übersetzung: Details gehen verloren, aber: Lieber Infos haben als nicht.

- Länderspezifität vs. Verallgemeinerung

⇒ Bestimmung generelles Model, Anwendung auf einzelne Länder/Organisationen/Datenbanken, dabei weitere Topics erlauben (iterative Überprüfung notwendig).

- Unterschiedliche Menge (und Qualität) an Informationen: z.T. keine Autoren-Identifier; keine Abstracts oder Keywords

⇒ Hierarchisierende Analyse; erst Dokumente mit meisten Informationen, dann stufenweise auch Analyse der Dokumente mit schlechteren/geringeren Informationen

- Technische Limitationen (z.B. maximale Uploadgrößen, Rechenzeit, Grenzen der Visualisierbarkeit)

- Zeit

DIPF
Leibniz-Institut für Bildungsforschung und Bildungsinformation

Vielen Dank für ihre Aufmerksamkeit ʃつ•_•ʅつ!

Feedback, Fragen und Wünsche gerne an
a.christ@dipf.de