

# Development of a user-friendly app for exploring and analyzing research topics in psychology

André Bittermann

*abi@leibniz-psychology.org*

ZPID – Leibniz Institute for Psychology Information, Universitätsring 15, D-54296 Trier (Germany)

## Background

Keeping track of the developments in a scientific field can be challenging. Regarding increasing numbers of publications, summarizing the contents of hundreds of thousands of scientific publications on specific topics is necessary to gain insights into the processes of a scientific field.

Many databases offer classifications, i.e., broad subject headings for categorizing the publications' contents. Past research has treated these categories as research topics (e.g., Krampen, 2016), but regarding the level of detail, topicality, and flexibility this approach has been criticized (e.g., Bittermann & Fischer, 2018).

Techniques for automated content analysis represent a promising approach for getting insight into large text corpora. Topic modeling (e.g., Blei, Ng, & Jordan, 2003), in particular, is gaining in popularity in scientometrics. In their well-known paper, Griffiths and Steyvers (2004) demonstrated how to find scientific topics by applying topic models to a corpus of scientific abstracts.

A topic-guided and user-friendly interface for databases of scientific literature can open publication trends to a broader audience with various user scenarios: exploring the current “hot topics,” investigating the ups and downs of topic popularity over time, or comparing publication trends concerning societal processes (e.g., the increasing trend of a topic referring to refugees and emotional trauma in psychological publications from the German-speaking countries after 2015).

## Aim

The goal of this project was to develop a user-friendly web-based application for exploring and analyzing research topics in psychology. This app is considered as an entry point to further research of scientific literature by informing the user about past and current developments of publication topics. To this end, the topics are directly linked to search queries in a database for psychological literature.

## Method

### Data

The psychological research topics were derived from PSYINDEX – the comprehensive database

containing references for German- and English-language publications in psychology and closely related disciplines from the German-speaking countries. It is developed and hosted by the Leibniz Institute for Psychology Information (ZPID; Trier, Germany). In April 2019, there were more than 350,000 psychological articles, book chapters, reports, and dissertations indexed in PSYINDEX. In the development of the app, documents published between 1980 and 2017 were included ( $N = 329,240$  in early 2019).

The PSYINDEX editorial staff assigns controlled terms from the *Thesaurus of Psychological Index Terms* published by the American Psychological Association (Tuleya, 2007). This standardized vocabulary of keywords the input for topic modeling. Main advantages compared to abstract texts are, inter alia, the direct usability for efficient literature search for this topic, the avoidance of stemming, stop words, and synonyms, as well as faster computation time (Bittermann & Fischer, 2018).

### Software

All analyses were conducted in R version 3.5.1 (R Core Team, 2018). For inference of research topics, the package *topicmodels* 0.2-8 (Grün & Hornik, 2011) was employed. The user interface was built as a *Shiny app* using *shiny* 1.2.0 (Chang, Cheng, Allaire, Xie, & McPherson).

### Topic Modeling

Topic modeling based on *latent Dirichlet allocation* (Blei et al., 2003) was applied. Following the best-practice recommendations by Maier et al. (2018), several candidates for the alpha hyperparameter (0.0001, 0.0005, 0.001) and the number of topics  $k$  (250–550) were examined. Delta was fixed to 0.01. The final model (alpha = 0.0005 and  $k = 325$ ) was selected regarding interpretability and document–topic assignments. Finally, only reliable topics with stability across multiple inference runs were included to increase the robustness of the results.

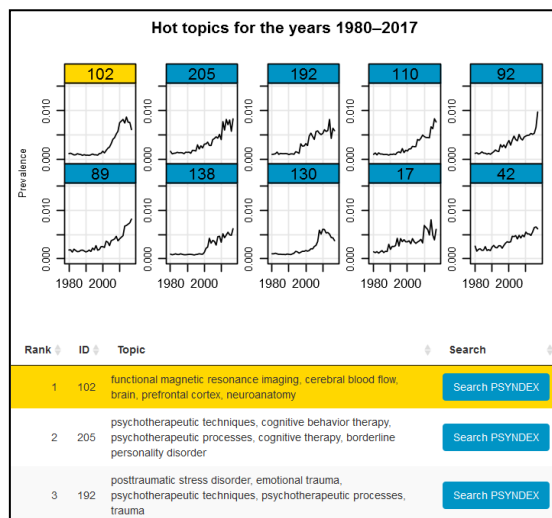
## Results

### Topic Model

The final model comprised 213 topics. The five terms with highest probabilities were included in the app. The topic with the highest prevalence overall was “psychoanalysis, psychotherapeutic processes, psychotherapeutic transference, counter-transference, psychoanalytic theory.” The most strongly increasing trend over the whole range of years was shown by the topic “functional magnetic resonance imaging, cerebral blood flow, brain, prefrontal cortex, neuroanatomy.” The “hottest” topic during the last three years was “posttraumatic stress disorder, emotional trauma, refugees, trauma, war.”

### Features of the App

Users can set the range of years from which popularity and trends are dynamically calculated. For exploring the topics, they can switch between “popular topics,” “hot topics” (see Fig. 1), “cold topics,” and an overview of “all topics.” Each topic entry has a search button that forwards a search query to PSYNDEX for literature relevant to this topic. For optimizing results, the terms in this query are weighted according to the term probability in the topic model.



**Figure 1. Hot topic view showing topics with most strongly increasing trends from 1980–2017.**

A demo version of the app can be accessed for free via <https://abitter.shinyapps.io/psychtopics/>

### Conclusions and Future Developments

Initial user experiences confirm the app’s ease of use. The implemented search queries help to clarify the topics’ contents and offer a low-threshold starting point to literature search. Topic inference is data-driven and independent from prior knowledge about a database’s contents. Since the standardized vocabulary used in PSYNDEX is updated on a

regular basis, the topic model can be updated as well and kept up to date with low maintenance efforts.

Future features will include current developments from our research group. For instance, topics with a high/low degree of empirical evidence can be shown, which may be of interest for research synthesis or explorative research. Using forecasting techniques, the observed trends can be compared to expected courses over time and help to quantify sudden increases and decreases in publication numbers. Author information could be included for investigating topical author networks.

The app can be applied to other databases with only few modifications necessary. The code is available for free on PsychArchives (<http://dx.doi.org/10.23668/psycharchives.2410>).

### References

- Bittermann, A. & Fischer, A. (2018). How to identify hot topics in psychology using topic modeling. *Zeitschrift für Psychologie*, 226(1), 3–13. <https://doi.org/10.1027/2151-2604/a000318>
- Blei, D. M., Ng, A. Y. & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022. <https://doi.org/10.1162/jmlr.2003.3.4-5.993>
- Chang, W., Cheng, J., Allaire, J. J., Xie, Y., & McPherson, J. (2018). shiny: Web Application Framework for R. R package version 1.2.0 [Computer software].
- Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences*, 101(Suppl. 1), 5228–5235. <https://doi.org/10.1073/pnas.0307752101>
- Grün, B., & Hornik, K. (2011). Topicmodels: An R package for fitting topic models. *Journal of Statistical Software*, 40(13). <https://doi.org/10.18637/jss.v040.i13>
- Krampen, G. (2016). Scientometric trend analyses of publications on the history of psychology: Is psychology becoming an unhistorical science? *Scientometrics*, 106, 1217–1238. <https://doi.org/10.1007/s11192-016-1834-4>
- Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., ... & Schmid-Petri, H. (2018). Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. *Communication Methods and Measures*, 12(2–3), 93–118. <https://doi.org/10.1080/19312458.2018.1430754>
- R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. [Computer software].
- Tuleya L. G. (Ed.). (2007). *Thesaurus of psychological index terms* (11th ed.). Washington, DC: American Psychological Association.