30_11_2022

Backward snowballing (Outgoing references), 2017 onwards

In order to further extend the selection of examined work starting from the set of the 65 selected papers, backward snowballing is performed.
Backward snowballing refers to the activity of extracting relevant work from the references appearing in papers that have already been deemed relevant. In other words, this means finding yet undiscovered (cited) work that should be included in the final review.
This activity of backward snowballing can be summarised in the following steps:

1. Go through the selected list of papers and, for each document, extract the list of references.
2. Filter the selected references with regards to the imposed time constraints.
3. Gather the initial set of publications resulting from this reference extraction process.
4. Apply the previously constructed search query to filter down the gathered publications.
5. Inspect the content of each remaining document and determine its relevance by applying the same set of inclusion/exclusion criteria utilised for the main selection.

Initial results

For the purpose of this activity, it has been decided to apply the same time-related constraints that have been used for the main selection (i.e. work released from 2017 onwards). On the other hand, no limitation on the origin venue (journal / conference) of the extracted work has been applied. Using this selection criteria, an initial corpus of **738 items** is extracted.
No filtering is applied with regards to the venue to which the extracted work belongs to. This means that forward snowballing allows to extend the scope of this review beyond the set of initially selected journals and conferences.
On this broad selection, the previously described query containing the 20 terms proximity operator (`"topic label*" OR ("topic model*" AND ("label*" NEAR "topic*"))`) is applied.
Side note: This query is applied on the set of locally stored papers using the FoxTrot Professional Search tool.
The corpus resulting from this filtering steps returns a total of **160 items** (231 if the proximity constraint is removed).

B.S. selection (selected 30 out of 160)

All papers contained in the corpus resulting from the initial reference extraction (and filtering) phase are individually inspected and evaluated against the established selection

criteria.

From the initial set of 160 papers, a final selection of 30 was found to meet the imposed requirements and therefore added to the systematic review.

The set of publications, together with their host venue, is presented below:

- Evaluating Visual Representations for Topic Understanding and Their Effects on Manually Generated Topic Labels, Transactions of the Association for Computational Linguistics
- Evaluating Topic Representations for Exploring Document Collections, Journal of the Association for Information Science and Technology
- United We Stand: Using Multiple Strategies for Topic Labeling, Natural Language Processing and Information Systems
- Transfer Topic Labeling with Domain-Specific Knowledge Base: An Analysis of UK House of Commons Speeches 1935–2014, Research & Politics
- In Search of Coherence and Consensus: Measuring the Interpretability of Statistical Topics, The Journal of Machine Learning Research
- Using structural topic modeling to identify latent topics and trends in aviation incident reports, Transportation Research Part C
- An Ontology-Based Labeling of Influential Topics Using Topic Network Analysis, Journal of Information Processing Systems
- Global Surveillance of COVID-19 by mining news media using a multi-source dynamic embedded topic model, BCB
- Free associations of citizens and scientists with economic and green growth: A computational-linguistics analysis, Ecological Economics
- Recommendation System for Knowledge Acquisition in MOOCs Ecosystems, SBSI
- Managing the Boundaries of Taste: Culture, Valuation, and Computational Social Science, Social Forces
- Scientific Evolutionary Pathways: Identifying and Visualizing Relationships for Scientific Topics, Journal of the Association for Information Science and Technology
- Full-Text or Abstract? Examining Topic Coherence Scores Using Latent Dirichlet Allocation, DSAA
- Topic modeling and sentiment analysis of global climate change tweets, Social Network Analysis and Mining
- Applying LDA topic modeling in communication research: Toward a valid and reliable methodology, Communication Methods and Measures
- The Politics of Scrutiny in Human Rights Monitoring: Evidence from Structural Topic Models of US State Department Human Rights Reports, Political Science Research and Methods
- Narratives of the Refugee Crisis: A Comparative Study of Mainstream-Media and Twitter, Media and Communication

- Towards Autoencoding Variational Inference for Aspect-based Opinion Summary, Applied Artificial Intelligence
- Modelling Research Topic Trends in Community Forestry, Small-scale Forestry
- Twitter speaks: A case of national disaster situational awareness, Journal of Information Science
- Toward understanding 17th century English culture: A structural topic model of Francis Bacon's ideas, Journal of Comparative Economics
- What Are MOOCs Learners' Concerns? Text Analysis of Reviews for Computer Science Courses, DASFAA
- Characterizing diabetes, diet, exercise, and obesity comments on Twitter, International Journal of Information Management
- Identifying topical influencers on twitter based on user behavior and network topology, Knowledge-Based Systems
- A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism, Tourism Management
- An Investigation of Brand-Related User-Generated Content on Twitter, Journal of Advertising
- Analyst Information Discovery and Interpretation Roles: A Topic Modeling Approach, Management Science
- Unsupervised Multi-Topic Labeling for Spoken Utterances
- A Knowledge-based Topic Modeling Approach for Automatic Topic Labeling
- Topic labeling towards news document collection based on Latent Dirichlet Allocation and ontology

Additionally, the following 15 papers that were already part of the initial selection of publications were retrieved once again following the backward snowballing procedure:

- Multimodal Topic Labelling, EACL
- Topic Model or Topic Twaddle? Re-evaluating Semantic Interpretability Measures, NAACL
- Automatic Generation of Topic Labels, SIGIR
- BART-TL: Weakly-Supervised Topic Label Generation, EACL
- Labeling Topics with Images using a Neural Network, ECIR
- Neural Models for Documents with Metadata, ACL
- W2VLDA: Almost unsupervised system for Aspect Based Sentiment Analysis, Expert Systems With Applications
- Document-based topic coherence measures for news media text, Expert Systems with Applications
- Detecting and predicting the topic change of Knowledge-based Systems: A topic-based bibliometric analysis from 1991 to 2016, Knowledge-Based Systems

- TaxoGen: Unsupervised Topic Taxonomy Construction by Adaptive Term Embedding and Clustering, KDD
- Topically Driven Neural Language Model, ACL
- A Topic Augmented Text Generation Model: Joint Learning of Semantics and Structural Feature, EMNLP
- Hierarchical Topic Mining via Joint Spherical Tree and Text Embeddin, KDD
- Measuring service quality from unstructured data: A topic modeling application on airline passengers' online reviews, Expert Systems With Applications
- CoRel: Seed-Guided Topical Taxonomy Construction by Concept Learning and Relation Transferring, KDD

B.S. venues (ongoing)

The following is the list of previously unexplored venues to which the work selected by the backward snowballing task belongs to:

Conferences

- ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (BCB)
- Database Systems for Advanced Applications (DASFAA)
- IEEE International Conference on Data Science and Advanced Analytics (DSAA)
- Natural Language Processing and Information Systems (NLDB)
- Sustainable Business and Social Impact Conference (SBSI)
- International Conference on Humanized Computing and Communication (HCC)
- International Conference on Informatics and Computational Sciences (ICICoS)

Journals

- Applied Artificial Intelligence
- Communication Methods and Measures
- Ecological Economics
- International Journal of Information Management
- Journal of Advertising
- Journal of Comparative Economics
- Journal of Information Processing Systems
- Journal of Information Science
- Journal of the Association for Information Science and Technology
- Management Science
- Media and Communication
- Political Science Research and Methods
- Research & Politics

- Small-scale Forestry
- Social Forces
- Social Network Analysis and Mining
- The Journal of Machine Learning Research
- Tourism Management
- Transactions of the Association for Computational Linguistics
- Transportation Research Part C
- International Journal of Advanced Computer Science and Applications

Forward snowballing (Incoming citations)

After having obtained the relevant (outgoing) references from the selected set of papers following the backward snowballing activity described in the previous section, the similar process of forward snowballing was carried out in order to to capture the relevant (incoming) citations.
The process follows a similar set of steps as the ones described for backward snowballing, with the two major differences being:

- The fact that an external repository (SemanticScholar) is used to obtain the citations since, unlike references, they cannot be extracted directly from the paper that is being examined.
- No filtering step based on publication time needs to be applied (since all citation will inherently respect the imposed time constraints).

Initial results

Executing forward snowballing on the set of 65 initially selected papers resulted in **1161** extracted citations. Once again, the query (and proximity operator) are applied to the extracted citation. This generates a set of **358** items (590 without the proximity operator) to which the inclusion/exclusion criteria need to be applied to.

Next steps

- Generate the final selection for forward snowballing
- Depending on the size of the final selection (after snowballing)
  - Increase/Decrease the time-frame / selected venues
- Build paper graphs using Pajek
- Start to rewrite the notes into the “Methods” section (i.e. transcribe the work done so far on Overleaf)
- After December 11th, gather papers from emnlp 2022

- Establish details of data collection process

Personal notes

DevonThink Search operators · GitHub

#Thesis/Weekly notes#