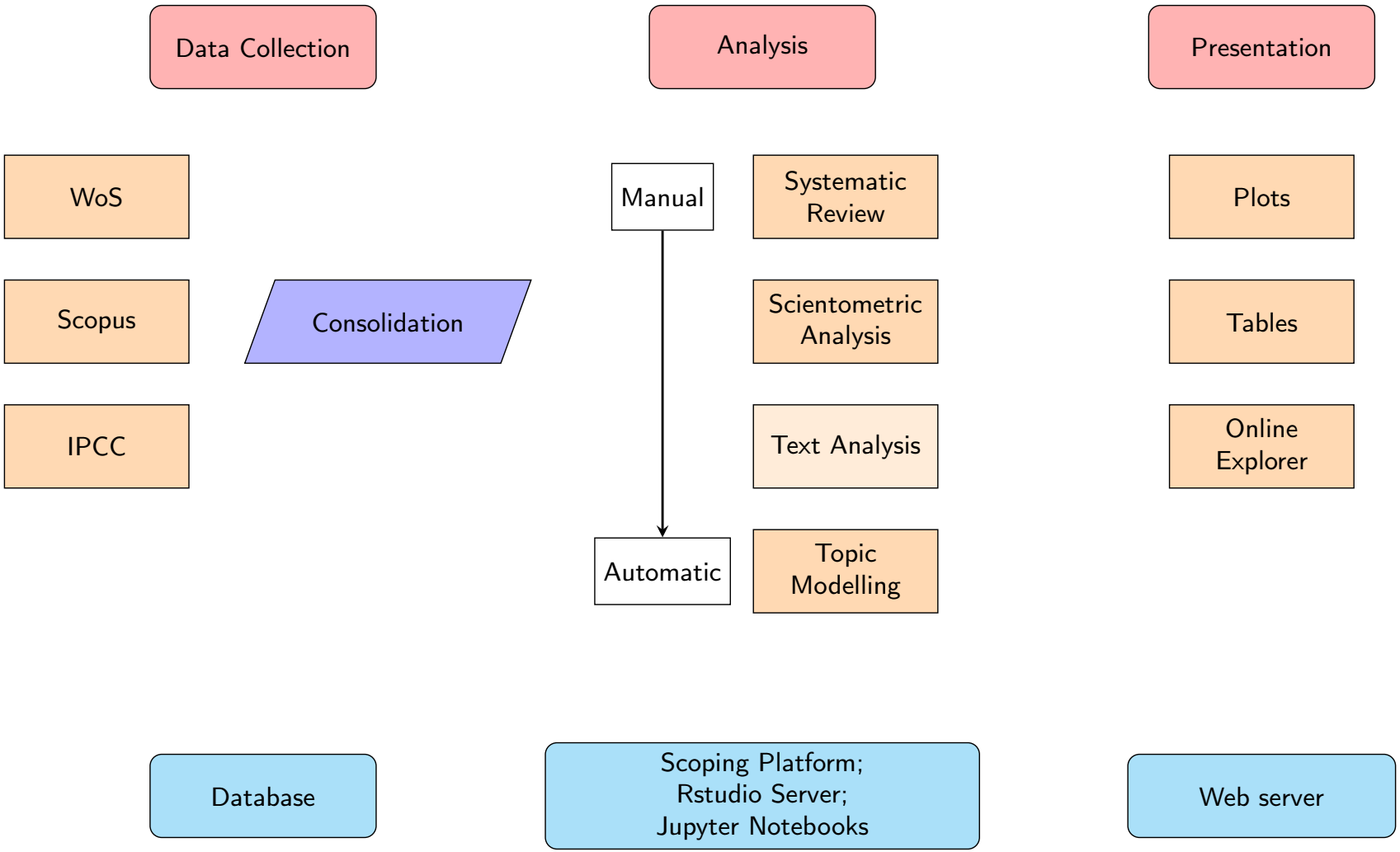
Apsis - Big literature tools and methods

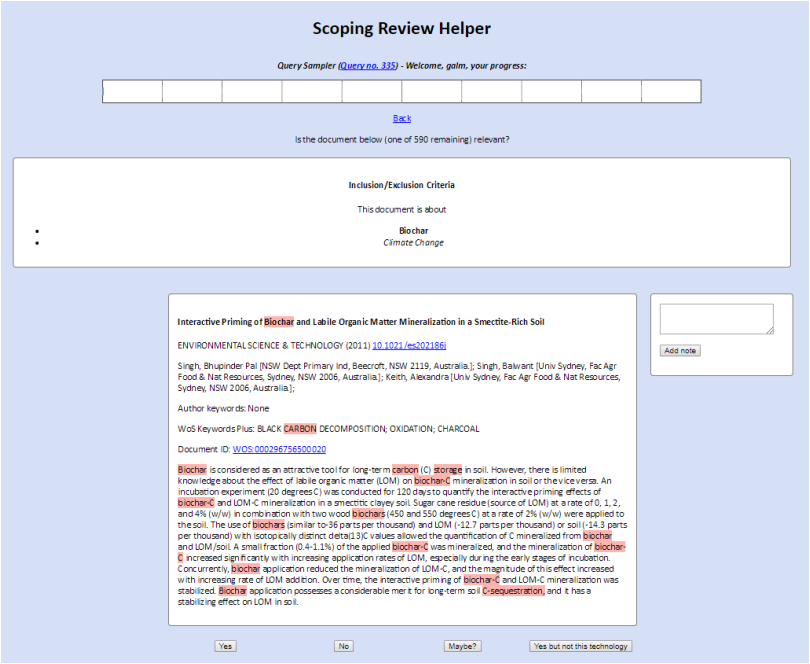
Max Callaghan

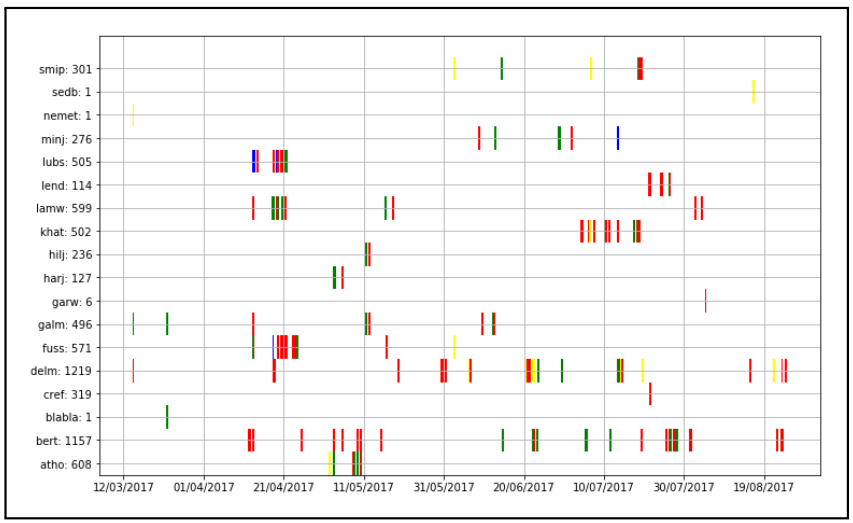
December 12, 2017

Infrastructure



Online Review Tool



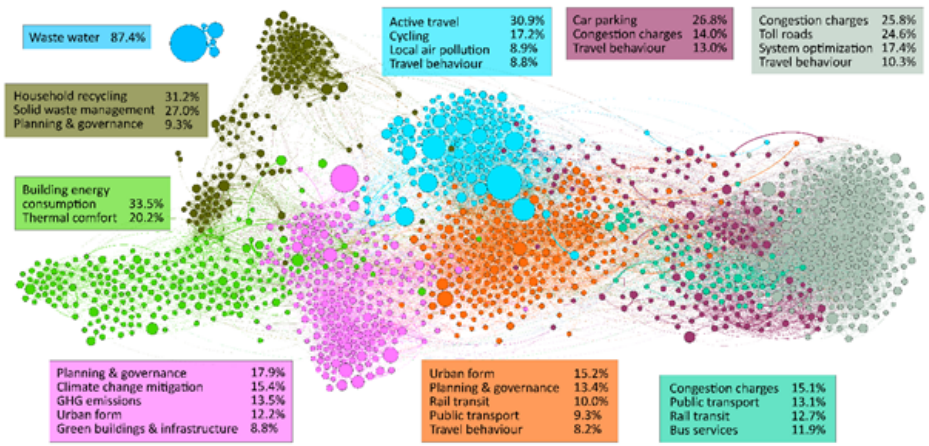


- A tool for teams to manage documents downloaded from online databases
- Teams can screen documents and mark them as relevant or not

Extensions

- Snowballing (getting more documents by looking through references and citations)
- Automatically emailing authors in the database and asking them to enter missing information

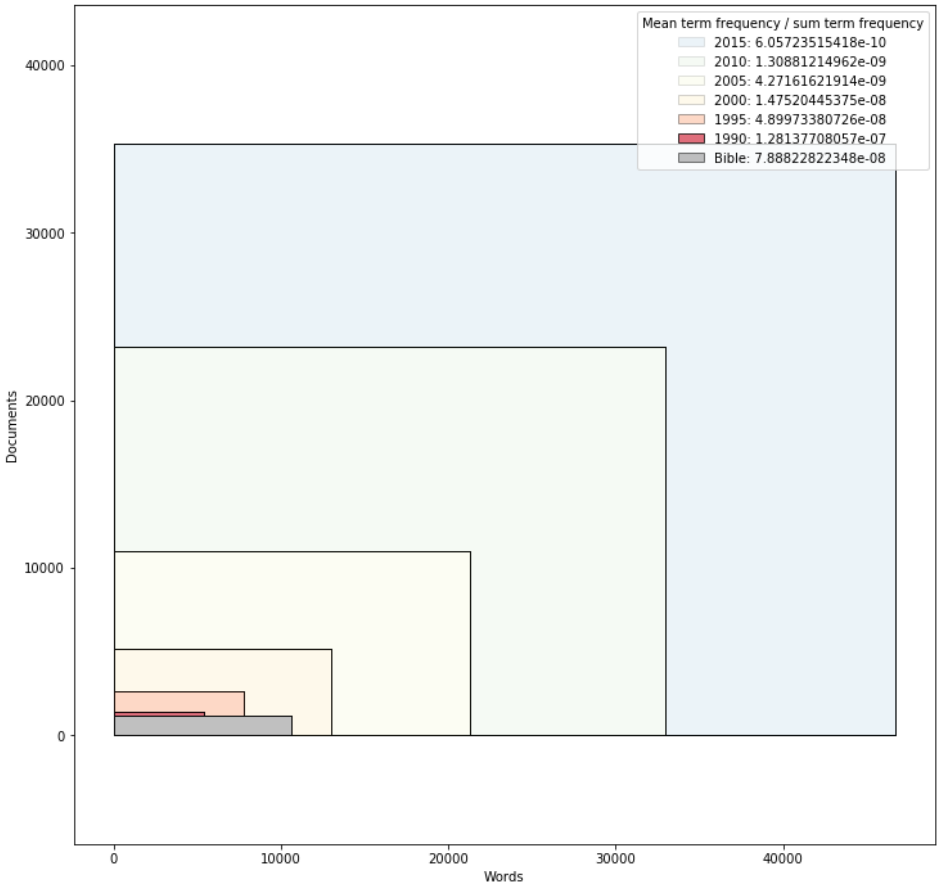
Scientometric Analysis



Scientometric Analysis uses the reference information of articles to answer questions about an area of literature:

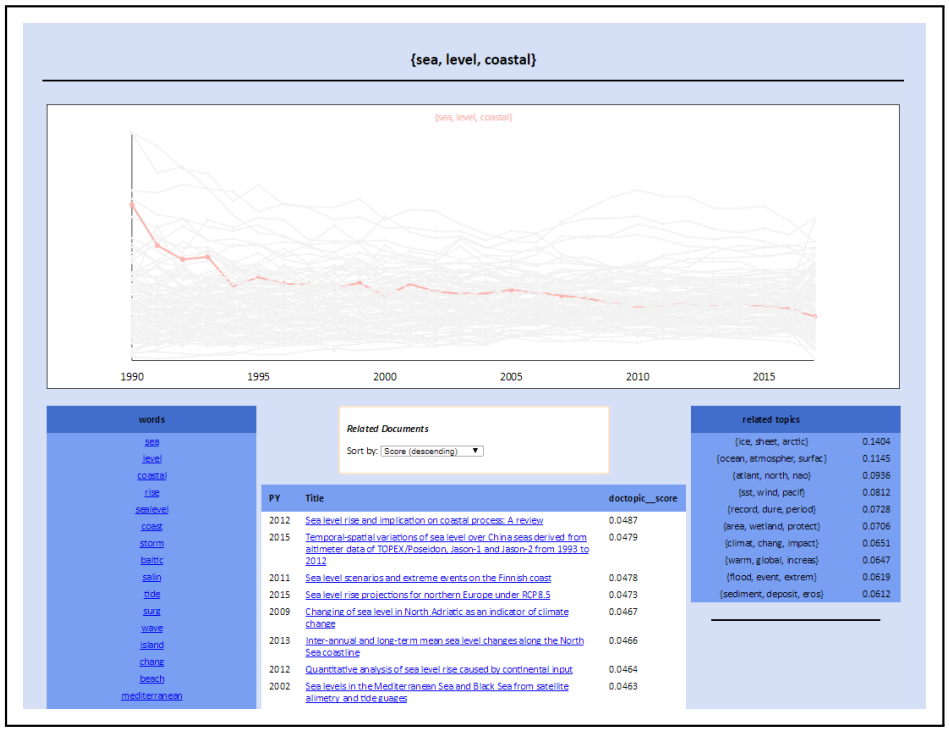
- What authors/institutions/articles play an important role?
- What is the community structure of a network of papers?
- How has an area of literature evolved? What are the roots of ideas, and how can we trace their development?

Topic modelling

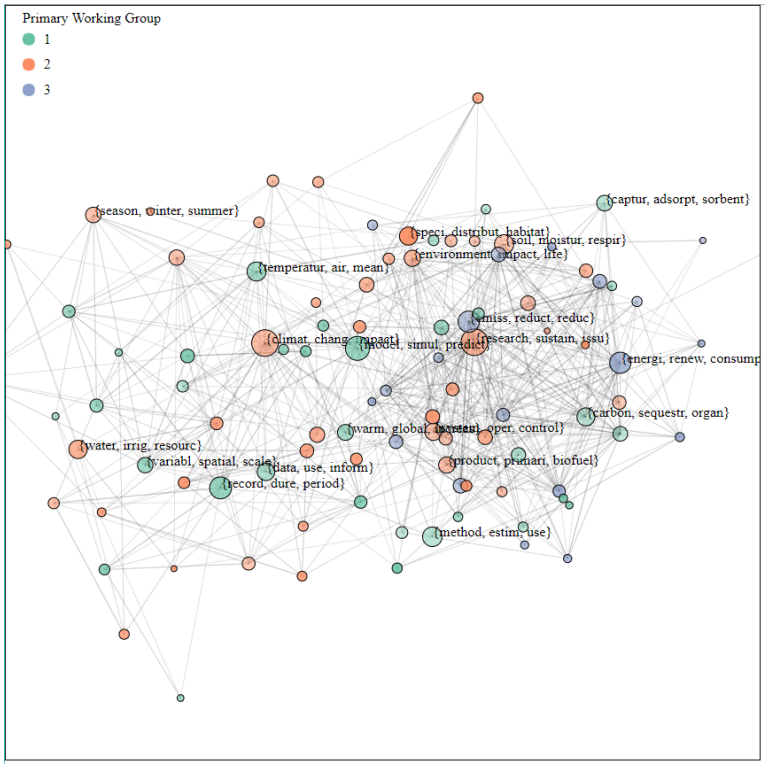


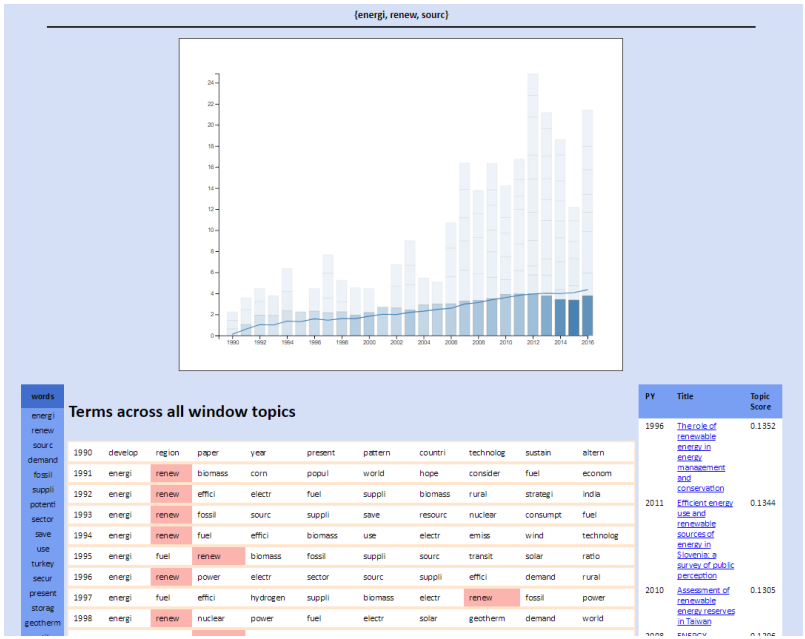
- Topic modelling is a way of reducing the dimensionality of a corpus of documents
- A large matrix of documents x words is factorised by a matrix of topics x words and a matrix of topics x documents (Lee and Seung, 1999)
- Topics describe the latent structure of the document corpus
- There are a number of different models, with different assumptions, and different approaches to estimation. A very good introduction to the concept is found in Blei et al. (2012)

Topic modelling



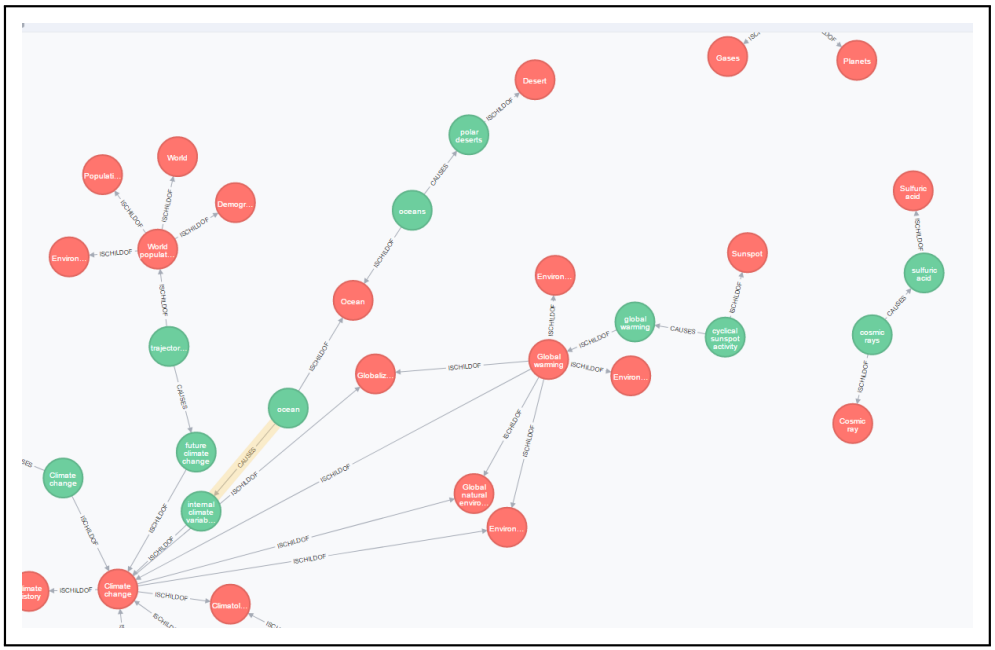
Topic modelling





- How do topics relate to one another?
- How do topics change over time? (Greene and Cross, 2016)
- How can we incorporate other information into the topic model? (Tvinnereim and Fløttum, 2015; Roberts et al., 2014)
- Topic models of speech (El-Assady et al., 2016)

Other Text Analysis



- Causaly collect causal statements from literature
- They aim to quantify and aggregate the strength of claims

Applications?

- Do we get more consolidated knowledge about causal relationships over time (in some WGs over others)?
- What can we learn about co-benefits and side-effects of different negative emission technologies?

Blei, D., Carin, L., and Dunson, D. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4):77–84.

El-Assady, M., Gold, V., Acevedo, C., Collins, C., and Keim, D. (2016). ConToVi: Multi-Party Conversation Exploration using Topic-Space Views. *Computer Graphics Forum*, 35(3):431–440.

Greene, D. and Cross, J. P. (2016). Exploring the Political Agenda of the European Parliament Using a Dynamic Topic Modeling Approach. pages 1–47.

Lee, D. D. and Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–91.

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B., and Rand, D. G. (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science*, 58(4):1064–1082.

Tvinnereim, E. and Fløttum, K. (2015). Explaining topic prevalence in answers to open-ended survey questions about climateÂăchange. *Nature Climate Change*, 5(8):744–747.