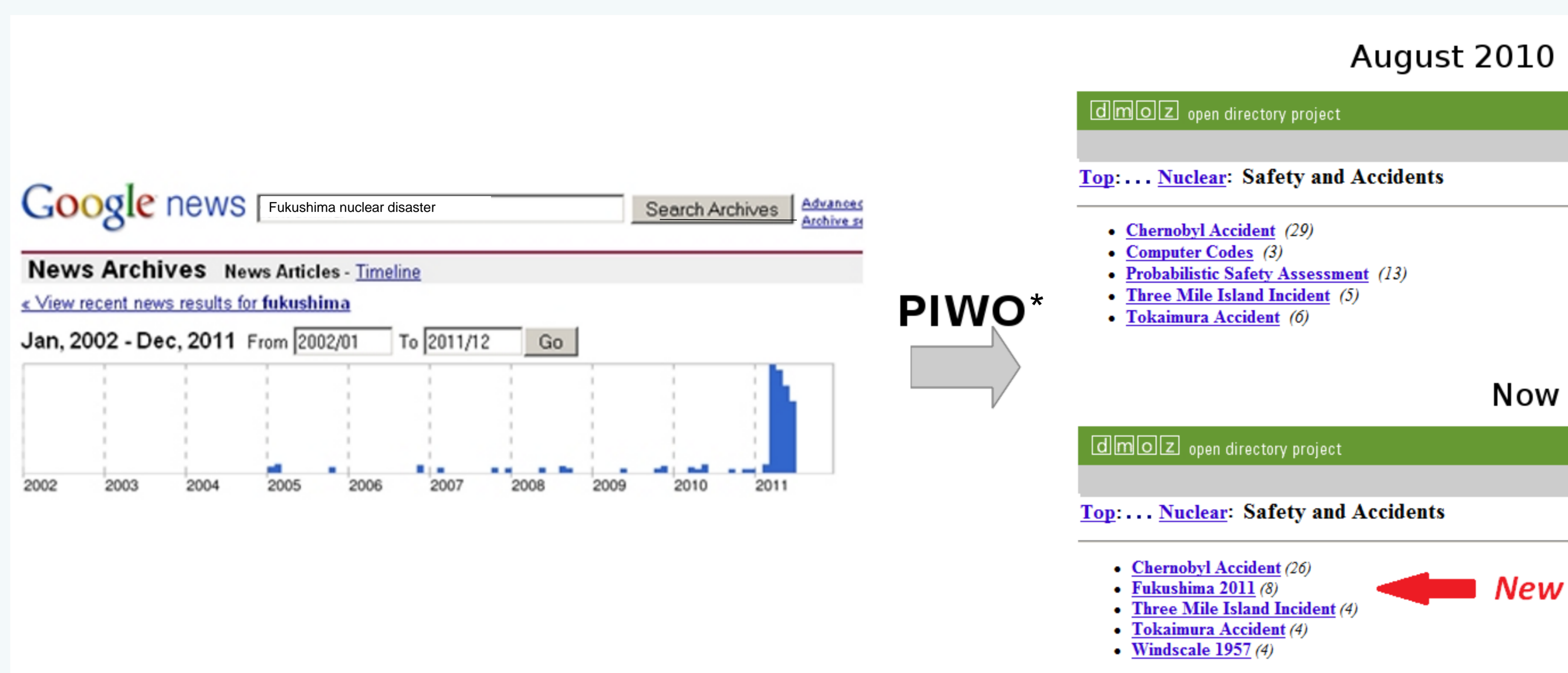


Predicting the Evolution of Collective Web Catalogues

Natalia Prytkova



News archive search for term *Fukushima* and changes in Web Catalogue
* PIWO = Predicting evolution In Web catalogues

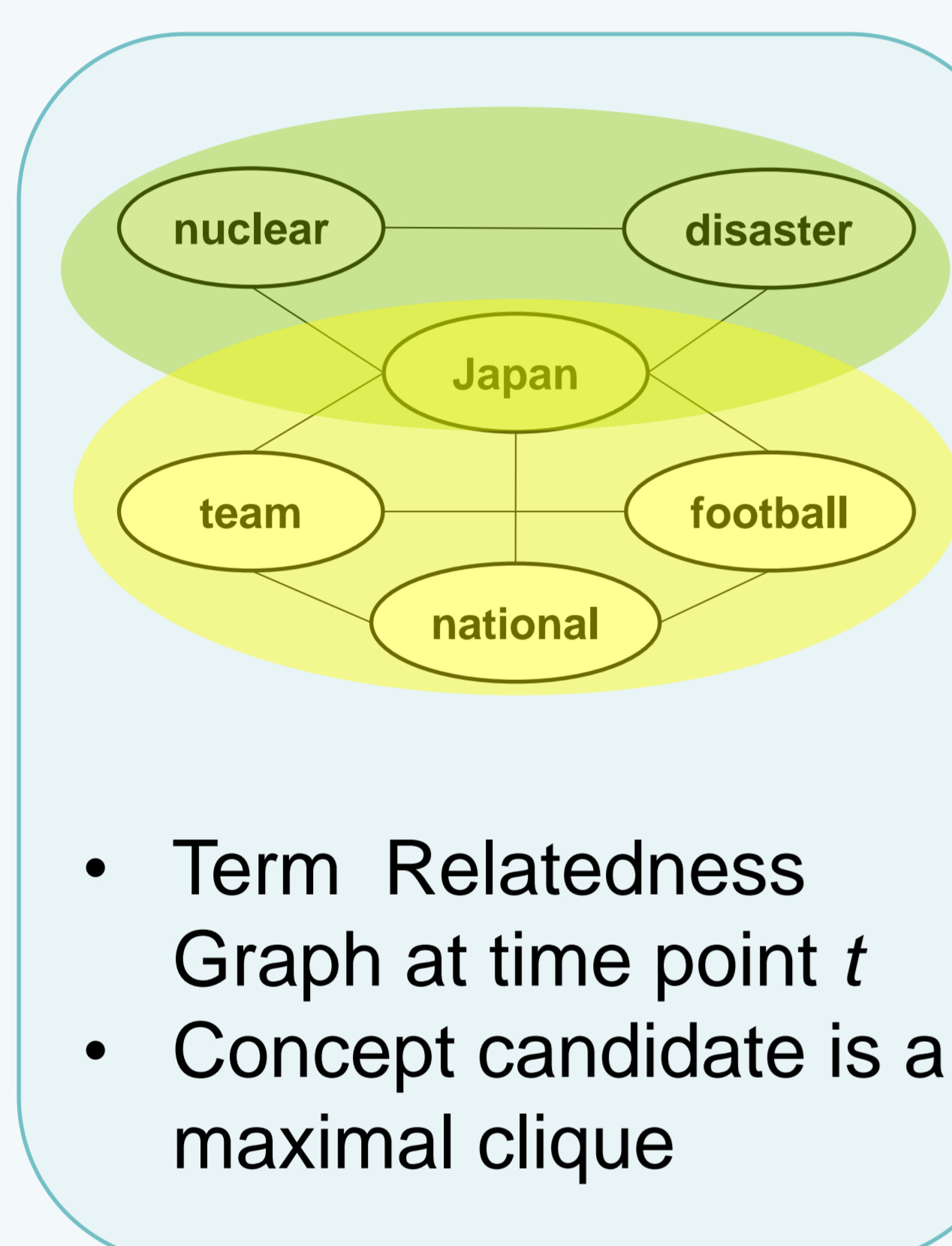
Goal

- Discover emerge of new concepts in news
- Predict structural changes in Web catalogues

Challenge

- What are concepts?
- What is concept dynamics?
- Which part of catalogue will change?

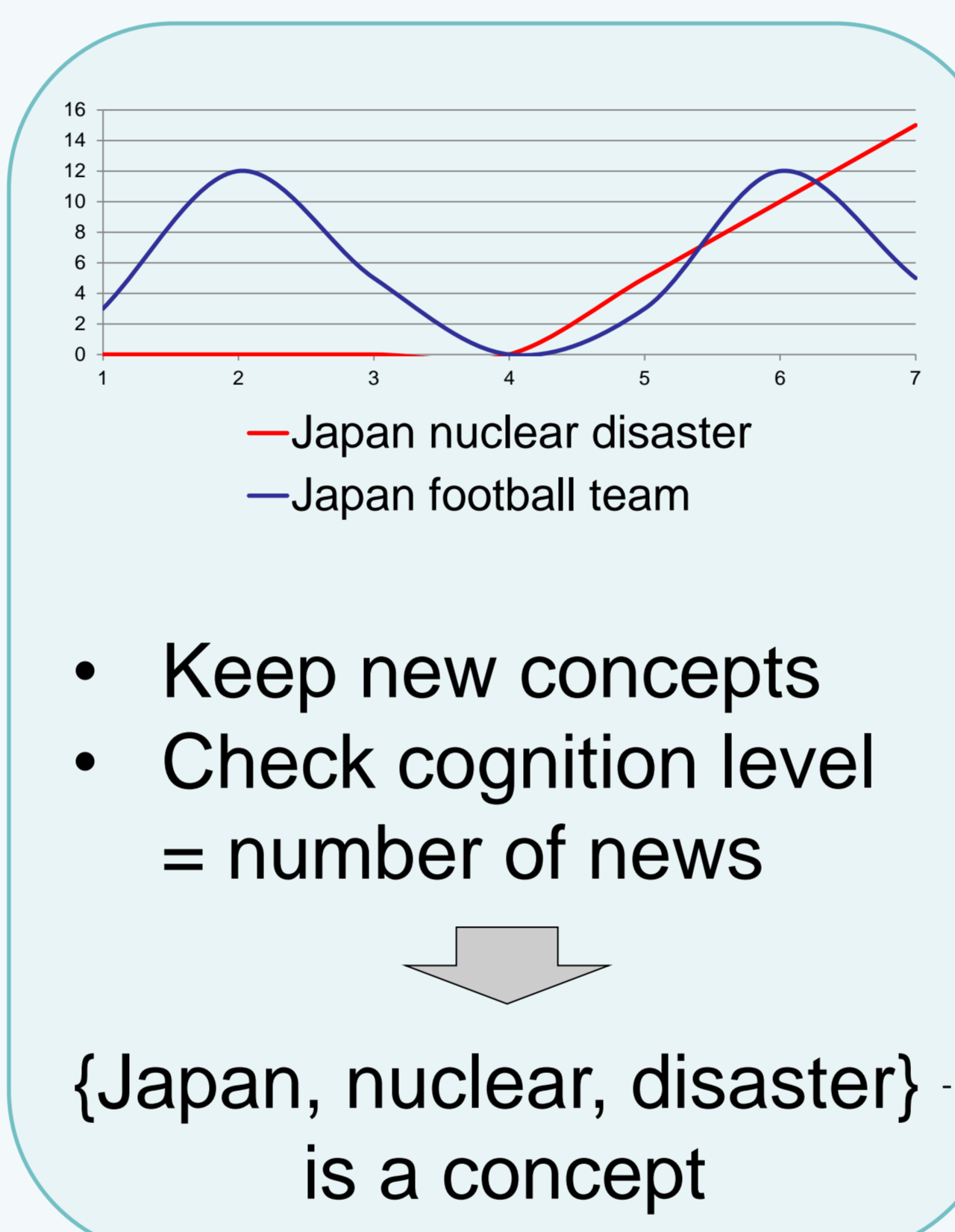
Find concept candidates



Set of candidates per time point

Distributed MaxClique Enumeration Algorithm In MapReduce

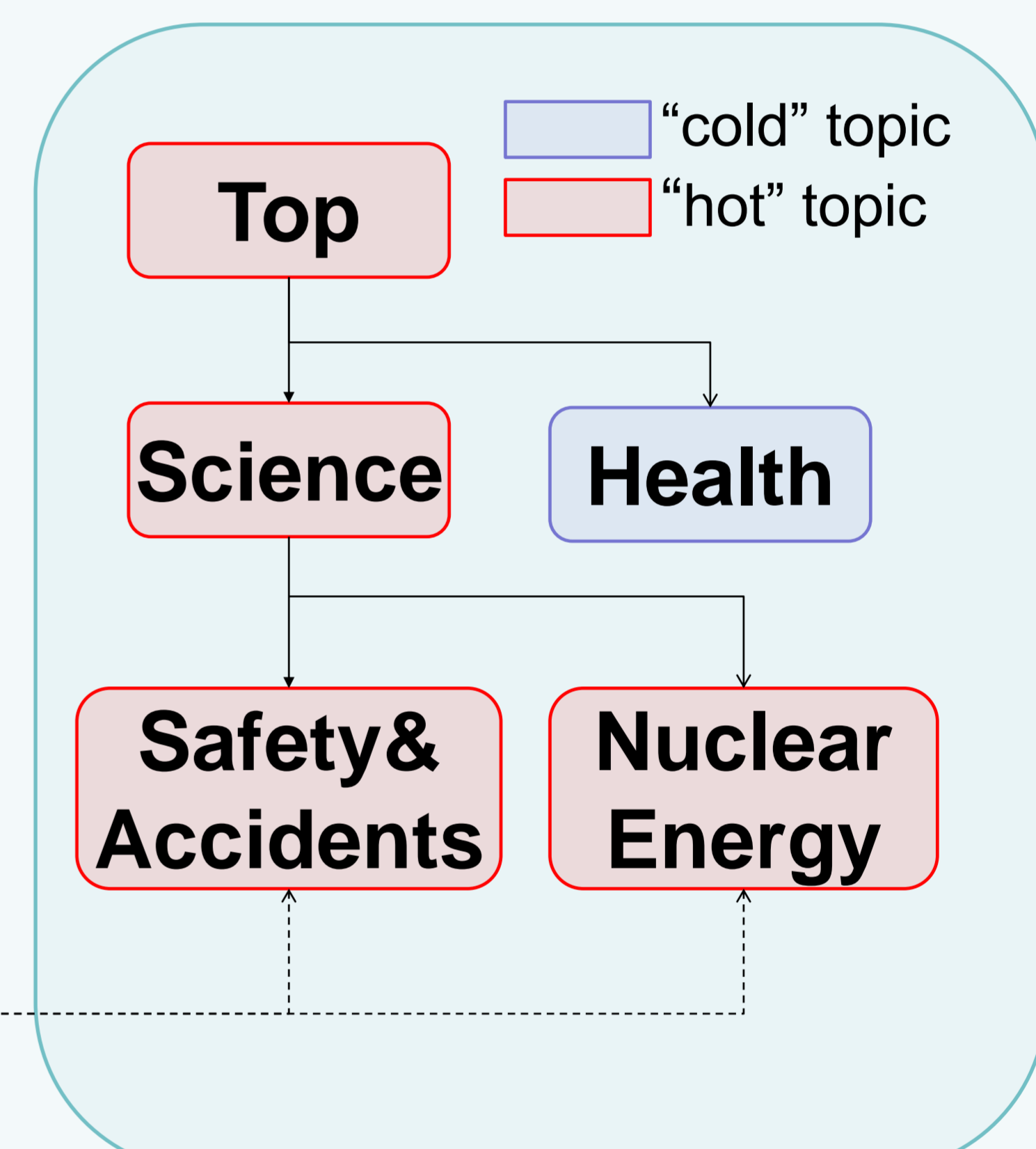
Trace concept dynamics



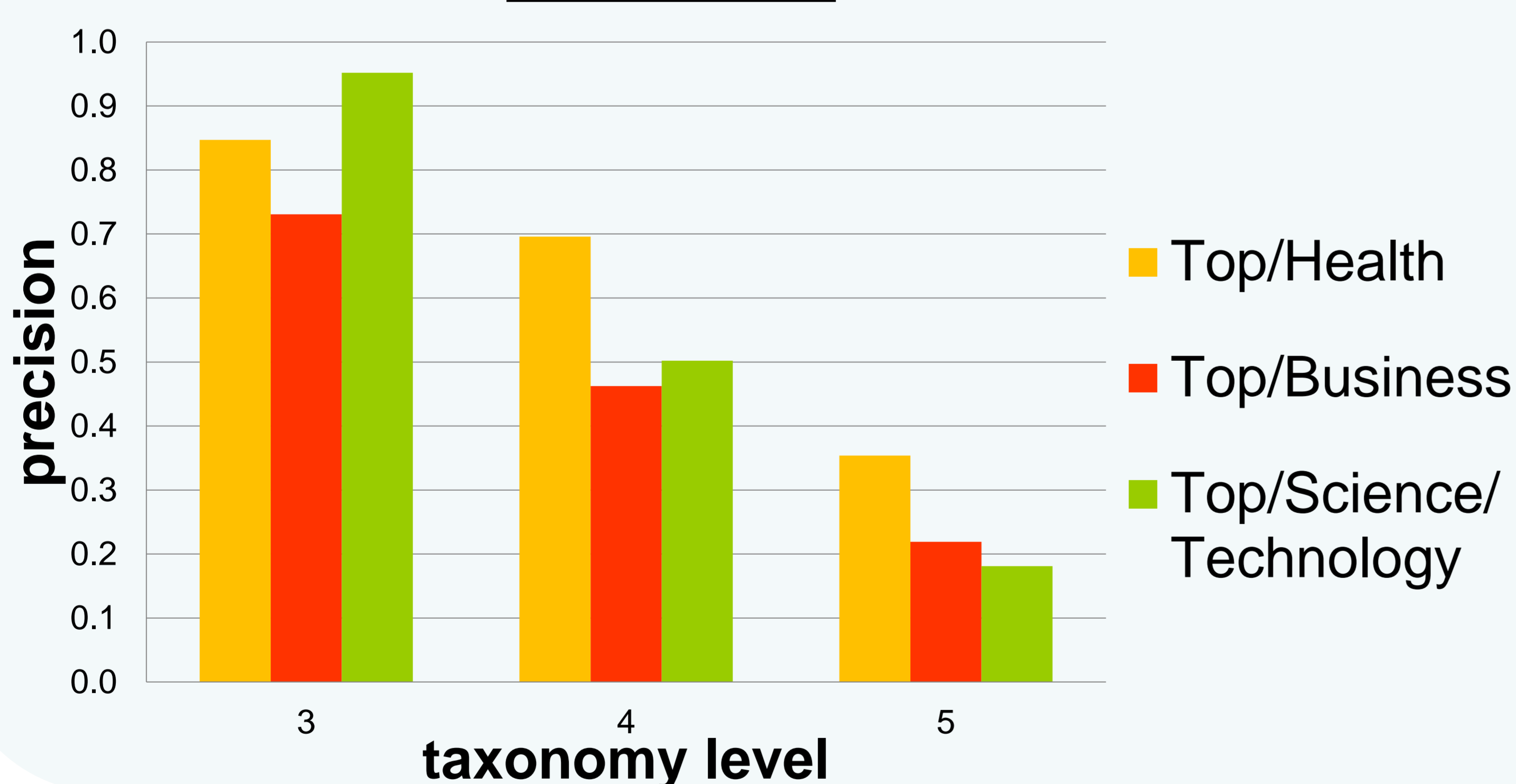
Concepts at time point t

match

Make prediction



Results



Conclusion

- Emerging latent concept detection
- High predictive power for taxonomy levels 3 and 4
- Distributed algorithm for large-scale experiments

Outlook

- Temporal taxonomy analysis
- Multilingual Wikipedia history analysis

References

N. Prytkova, M. Spaniol and G. Weikum Predicting the Evolution of Taxonomy Restructuring in Collective Web Catalogues, WebDB'12 (to appear)
Bin Wu et al. A distributed Algorithm to Enumerate All Maximal Cliques in Mapreduce, FCST'09

Acknowledgments

This work is supported by the 7th Framework IST programme of the European Union through the focused research project (STREP) on Longitudinal Analytics of Web Archive data (LAWA) under contract no. 258105.