



## Axiomatic thinking for information retrieval: introduction to special issue

Enrique Amigó<sup>1</sup> · Hui Fang<sup>2</sup> · Stefano Mizzaro<sup>3</sup> · Chengxiang Zhai<sup>4</sup>

Received: 18 May 2020 / Accepted: 19 May 2020 / Published online: 25 May 2020  
© Springer Nature B.V. 2020

The dominant methodology of Information Retrieval (IR) research has so far been empirical, i.e., progress is guided by experimental results on various data sets. The availability of large data sets since the 90s, the growing computational power, and the possibility of automatizing experiments have accelerated empirical studies of IR in the last two decades, generating many empirical findings. However, a purely empirical methodology, mainly experimental and focused on effectiveness, does not match with the standard scientific procedures: hypothesis statement, definition of an experiment guided by the specific hypothesis, and result analysis. As a consequence, the IR community tends to produce solutions to a greater extent than knowledge. Moreover, the solutions found are often variants of similar ideas and thus do not add up to more progress as shown in Armstrong et al. (2009). It is clear that while empirical research is necessary, it must be accompanied by strong theoretic models as discussed in length by Norbert Fuhr in his Salton Award speech (Fuhr 2012). As one theoretical approach to IR research, axiomatic thinking has shown great promise when studying both retrieval models and evaluation measures, and this special issue is to report the most recent work in this direction.

Axiomatic thinking refers to a problem-solving strategy that is guided by axioms, and is closer to traditional methodologies in science. Generally speaking, when searching for solutions to a given problem, axiomatic thinking aims to find solutions that can satisfy all the axioms, i.e., all the desirable properties that a solution needs to satisfy. The explicit and clear articulation of the desirable properties makes such an approach not only theoretically appealing but also useful for suggesting interesting causal hypotheses to be tested in empirical experiments.

---

✉ Hui Fang  
hfang@udel.edu

Stefano Mizzaro  
mizzaro@uniud.it

Chengxiang Zhai  
czhai@illinois.edu

<sup>1</sup> Universidad Nacional de Educación a Distancia (UNED), Juan del Rosal 16, 28040 Madrid, Spain

<sup>2</sup> University of Delaware, 140 Evans Hall, Newark, DE 19713, USA

<sup>3</sup> University of Udine, Via delle Scienze 206, 33100 Udine, Italy

<sup>4</sup> University of Illinois at Urbana-Champaign, 201 N. Goodwin Ave, Urbana, IL 61801, USA

Axiomatic thinking has already been successfully applied to study of retrieval model, leading to both theoretical understanding of existing retrieval models and their relations and improvement of multiple models such as basic retrieval models (Fang et al. 2004; Fang 2007; Lv and Zhai 2011), feedback methods (Clinchant and Gaussier 2011), translation retrieval models (Karimzadehgan and Zhai 2012), and neural network retrieval models (Rosset et al. 2019). It has also been applied to study of evaluate measures, leading to deeper understanding of properties of evaluation measures and the introduction of better measures (Amigó et al. 2009; Busin and Mizzaro 2013; Moffat 2013; Sebastiani 2015), as well as similarity metrics (Lin 1998; Cazzanti and Gupta 2006).

Axiomatic thinking can be further applied more broadly to addressing many other problems in IR by both practitioners of IR in the industry and academia researchers. Moreover, the general ideas of using axiomatic thinking to study IR are potentially applicable to study of many other empirically defined tasks, notably many problems currently solved using statistical machine learning approaches, where axiomatic thinking may help addressing some difficult challenges such as optimal feature construction, optimal design of loss functions, interpretability of models. We hope this special issue will facilitate broader applications of axiomatic thinking in both IR and other related fields.

As discussed in Amigó et al. (2018), IR methodologies can be categorized along two dimensions. The first dimension includes two categories: theoretical and empirical. Theoretical approaches were often derived based on formal theories, while empirical approaches often by empirical observations made over evaluation data sets. Another dimension is bottom-up versus top-down. Bottom-up approaches are based on existing IR models and test cases driven by real scenarios, whereas top-down approaches often start with general axioms and synthetic test cases. There is no perfect methodology and different methodologies complement each other. Empirical experiments over test cases sampled from real scenarios provide quantitative results, statistical significance and evidence in terms of user satisfaction. When test cases are artificially developed (synthetic data), the results are more interpretable at the cost of user satisfaction and representativeness. Connecting and generalizing theoretical approaches provides universality, interpretability, and the possibility of

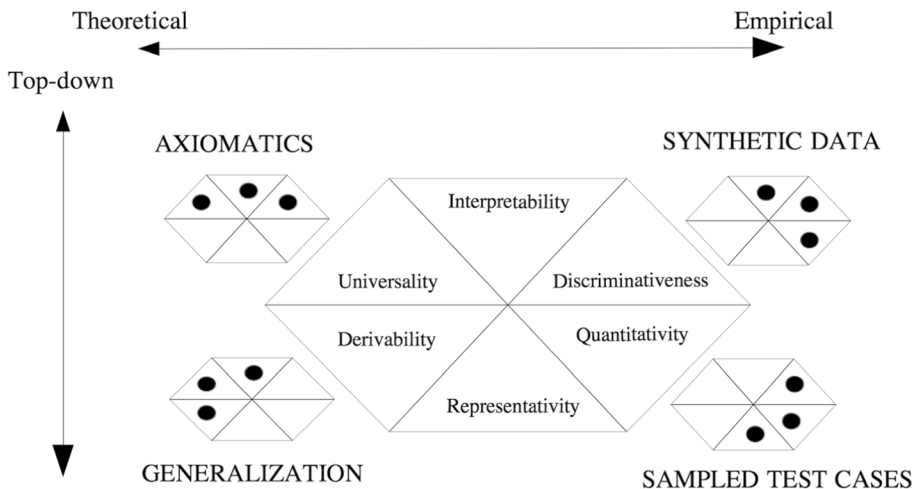


Fig. 1 IR methodology categorization [figure from Amigó et al. (2018)]

deriving new approaches. The fourth methodology is axiomatics, which poses a theoretical top-down perspective, in which unsuitable approaches can be discriminated on the basis of interpretable axioms without depending on the particularities of data sets (see Fig. 1). In this sense, the purpose of this special issue is also to complement current knowledge and advances with methodologies not so popular in the IR community.

The five papers selected for this special issue cover recent research efforts on applying axiomatic thinking to different problems including retrieval models, similarity functions, and evaluation metrics. These papers were all reviewed by the experts in the field and went through iterations of revision and further review. We now provide a brief summary of these papers and categorize them in the above framework.

Rahimi et al. (2020) delve into the axiomatics of retrieval models for corpus-based Cross-Language Information Retrieval (CLIR). The authors define a set of formal constraints and check whether existing CLIR methods satisfy them. Based on the defined constraints, they propose a hierarchical query modeling for CLIR which improves performance, compared to the existing methods. Therefore, the paper covers both the axiomatics and empirical bottom-up methodologies.

Amigó et al. (2020) tackle the notion of similarity. The authors analyze existing similarity axiomatic frameworks from other fields, such as Tversky's axioms (Tversky 1977) from the point of view of cognitive sciences, and metric spaces from the point of view of algebra. They observe that these frameworks do not completely fit the notion of similarity function in Information Access problems, and propose a new set of formal constraints. Under this framework, they categorize and analyze the properties of similarity functions applied to information access, and then introduce a similarity function which parameterizes the classical pointwise mutual information. This paper covers axiomatics, theoretical generalization, and a shallow study case over synthetic data.

The next three papers in the special issue focus on evaluation metrics but from different perspectives. Sebastiani (2020) performs a study on evaluation metrics for quantification, i.e., the task of estimating the prevalence of classes in unlabeled data. The paper presents a set of properties and discusses under what conditions each of these properties is desirable, and whether existing metrics satisfy or not the above properties. A significant result is that no existing metric satisfies all the properties identified as desirable. This work focuses exclusively on axiomatics. It identifies weaknesses in existing metrics and states a framework on which metrics can be improved at theoretical level.

The other two papers in this special issue both try to connect evaluation with measurement theory, but in different ways. In Ferrante et al. (2020), the evaluation process is interpreted as an effectiveness measurement process. Their starting framework states which IR evaluation measures can be considered as interval scales. In this study, they analyze how the scales of evaluation measures impact on statistical tests. Additionally, they analyze how incomplete information and pool downsampling affect different scales and evaluation measures. This work relies on axiomatic and empirical bottom-up methodologies, since axiomatic findings are corroborated via experimental benchmarks.

With a different approach, Amigó and Mizzaro (2020) interpret system outputs and gold standards (or ground truths) as measurements on a given scale. In this way, the task is determined by the measurement scale (classification/nominal, ranking/ordinal, etc.) This work aims to include all the possible information access tasks and states a general definition of evaluation metric which allows to derive most of the existing formal constraints in a particular task depending on the scale in which the definition is instantiated. In this sense, this paper is grounded on axiomatics, stating desirable properties on the

basis of measurement theory. In addition, the axioms and constraints defined in the literature for different tasks are generalized as properties of two metric definitions.

We hope that you will enjoy learning about these state-of-the-art studies in the field of axiomatic thinking for IR and will appreciate the power of axiomatic thinking in different applications.

## References

- Amigó, E., Fang, H., Mizzaro, S., & Zhai, C. (2018). Are we on the right track? An examination of information retrieval methodologies. In *Proceedings of ACM SIGIR'18* (pp. 997–1000). <https://doi.org/10.1145/3209978.3210131>.
- Amigó, E., Giner, F., Gonzalo, J., & Verdejo, F. (2020). On the foundations of similarity in information access. *Information Retrieval Journal*, this issue.
- Amigó, E., Gonzalo, J., Artiles, J., & Verdejo, F. (2009). A comparison of extrinsic clustering evaluation metrics based on formal constraints. *Information Retrieval*, 12(4), 461–486.
- Amigó, E., & Mizzaro, S. (2020). On the nature of information access evaluation metrics: A unifying framework. *Information Retrieval Journal*, this issue.
- Armstrong, T. G., Moffat, A., Webber, W., & Zobel, J. (2009). Improvements that don't add up: Ad hoc retrieval results since 1998. In D. W. -L. Cheung, I. -Y. Song, W. W. Chu, X. Hu, & J. J. Lin (Eds.) *CIKM* (pp. 601–610). ACM.
- Busin, L., & Mizzaro, S. (2013). Axiometrics: An axiomatic approach to information retrieval effectiveness metrics. In *Proceedings of the 2013 conference on the theory of information retrieval* (pp. 22–29).
- Cazzanti, L., & Gupta, M. R. (2006). Information-theoretic and set-theoretic similarity. In *Proceedings of the 2006 IEEE international symposium on information theory* (pp. 1836–1840).
- Clinchant, S., & Gaussier, E. (2011). Is document frequency important for PRF? In *Conference on the theory of information retrieval* (pp. 89–100). Berlin, Heidelberg: Springer.
- Fuhr, N. (2012). Salton award lecture: Information retrieval as engineering science. *ACM SIGIR Forum*, 46(2), 19–28.
- Fang, H. (2007). An axiomatic approach to information retrieval, dissertation. University of Illinois at Urbana-Champaign, 2007. <http://hdl.handle.net/2142/11352>.
- Fang, H., Tao, T., & Zhai, C. X. (2004). A formal study of information retrieval heuristics. In *Proceedings of the 27th annual international ACM SIGIR conference on research and development in information retrieval (SIGIR'04)* (pp. 49–56).
- Ferrante, M., Ferro, N., & Losiouk, E. (2020). How do interval scales help us with better understanding IR evaluation measures? *Information Retrieval Journal*, this issue.
- Karimzadehgan, M., & Zhai, C. (2012). Axiomatic analysis of translation language model for information retrieval. In *Proceedings of the 34th European conference on information retrieval (ECIR'12)* (pp. 268–280).
- Lin, D. (1998). An information-theoretic definition of similarity. In *Proceedings of the fifteenth international conference on machine learning, ICML'98* (pp. 296–304).
- Lv, Y., & Zhai, C. X. (2011). Lower bounding term frequency normalization. In *Proceedings of the 20th ACM international conference on information and knowledge management (CIKM'11)* (pp. 7–16).
- Moffat, A. (2013). Seven numeric properties of effectiveness metrics. In *AIRS'13* (pp. 1–12).
- Rahimi, R., Montazerlghaem, A., & Shakery, A. (2020). An axiomatic approach to corpus-based cross-language information retrieval. *Information Retrieval Journal*, this issue.
- Rosset, C., Mitra, B., Xiong, C., Craswell, N., Song, X., & Tiwary, S. (2019). An axiomatic approach to regularizing neural ranking models. In *Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval* (pp. 981–984).
- Sebastiani, F. (2015). An axiomatically derived measure for the evaluation of classification algorithms. In *Proceedings of the 2015 international conference on the theory of information retrieval* (pp. 11–20).
- Sebastiani, F. (2020). Evaluation measures for quantification: An axiomatic approach. *Information Retrieval Journal*, this issue.
- Tversky, A. (1977). Features of similarity. *Psychological Review*, 84(4), 327.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.