

LeQua @ CLEF 2022: A Shared Task for Evaluating Quantification Systems

Andrea Esuli, Alejandro Moreo and Fabrizio Sebastiani

Istituto di Scienza e Tecnologie dell'Informazione, Consiglio Nazionale delle Ricerche, 56124 Pisa, Italy

Abstract

LeQua 2022 (<https://lequa2022.github.io/>) is a new shared task (a.k.a., “challenge”, or “competition”) for comparatively evaluating systems for learning to quantify in a tightly controlled environment. LeQua 2022 is one of several shared tasks organized under the umbrella of the CLEF 2022 conference (<https://clef2022.clef-initiative.eu/>), which will take place in Bologna, Italy, in September 2022. LeQua 2022 participants will be provided with a training set of labelled documents and a development (validation) set of samples of labelled documents, and will be asked to provide class prevalence estimates for a test set consisting of samples of unlabelled documents. The evaluation will use a dataset obtained from a large crawl of product reviews. The training set will be obtained via stratified sampling from the dataset, while the validation samples and the test samples will be obtained by sampling from the dataset using predetermined class prevalence distributions, where these distributions will be obtained by sampling uniformly at random from the set of all legitimate class prevalence distributions. LeQua 2022 will comprise two tasks, the *vector task* and the *raw-documents task*. In the vector task, participants will be provided with data already in vector form; this task will thus purely measure the effectiveness of (data-agnostic) quantification algorithms. In the raw-documents task, participants will instead be provided with the original textual documents, and will thus be able to evaluate end-to-end systems. Both tasks will allow two subtasks, a binary task of quantification by sentiment and a multiclass task of quantification by topic. A one-day workshop during CLEF 2022 will be held in order to allow participants to discuss their results and participating systems.

LQ 2021: 1st International Workshop on Learning to Quantify, Gold Coast, AU, November 1 and November 5, 2021.

✉ andrea.esuli@isti.cnr.it (A. Esuli); alejandro.moreo@isti.cnr.it (A. Moreo); fabrizio.sebastiani@isti.cnr.it (F. Sebastiani)

🆔 0000-0002-5725-4322 (A. Esuli); 0000-0002-0377-1025 (A. Moreo); 0000-0003-4221-6427 (F. Sebastiani)



© 2021 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)