Saving Time, Saving Money

Or, how to do metadata curation using LLMs locally.

Aerith Y. Netzer

December 10, 2024

Abstract

In the past century, the well-documented consolidation of the academic publishing industry created the economic conditions for universities to publish their own articles. While these journals are necessary and good, they are not as well-funded or well-staffed as corporate players. These preconditions necessitate university libraries and presses to use all resources at their disposal to make their publishing operations run at peak efficiency. A large opportunity cost incurred by the university-owned publisher is machine-actionable metadata creation, leading to a cropping of for- and non-profit players in the industry to provide these services. We present three insights [there is a better word than insights here] - that reference metadata curation using LLMs is possible, that we can run these LLMs on local hardware, and we end with some tutorials for replication at other university-owned presses.

University-owned presses operate under far tighter economic constraints — direct and opportunity — and therefore must solve the same problems of corporate academic publishers with a fraction of the resources available. One of these problems is reference metadata, i.e. machine-actionable references that are then used to count citations of articles. The act of counting citations accurately is foundational to data-driven research assessment for universities and other research institutions. Rarely, even in libraries, are these considerations discussed, either due to lack of time, money, or knowledge. But they must be addressed in a university-owned press setting. Previous work in our institution discussed creating metadata using static-site services [1]

---

[1] Diaz (2018)

Bibliography

Diaz, Chris. 2018. "Using Static Site Generators for Scholarly Publications and Open Educational Resources." *The Code4Lib Journal*, no. 42 (November).