Formulation of FAIR Metrics for Primary Research Articles

Adam Craig and Carl Taswell

***Abstract*—Measuring the merits of a scholarly article only by how often other articles or social media posts cite it creates a perverse incentive for authors to avoid citing potential rivals. To uphold established standards of scholarship, institutions should also consider one or more metrics of how appropriately an article cites relevant prior work. This paper describes the general characteristics of the FAIR Attribution to Indexed Reports (FAIR) family of metrics, which we have designed for this purpose. We formulate five FAIR metrics suitable for use with primary research articles. Two measure adherence to best practices: number of correctly attributed background statements and number of genuinely original claims. Three measure specific deviations from best practices: number of misattributed background statements, number of background statements with missing references, and number of claims falsely indicated as original. We conclude with a discussion of plans to implement a web application for calculating metric values of scholarly works described by records in Nexus-PORTAL-DOORS System (NPDS) servers.**

***Index Terms*—citation metric, citation standards, plagiarism detection, scientometrics**

I. Introduction

Institutions increasingly use citation metrics as if they were objective, quantitative measures of the merits of research, researchers, and scholarly journals [1] [2], despite widespread consensus among the scientific community that such metrics fail as measures of the quality of scholarship or the value of research [3], [4], [5][6]. One problem with conventional citation metrics is that assuming the quality of a work is proportional to how many other works cite it creates a perverse incentive: Authors who diligently cite relevant prior work improve the scores of potential rivals, not their own (Figure 1). To bring balance to the system of incentives, institutions should also consider metrics of how appropriately, or fairly, a report of scientific work references prior reports as recorded in a citation index. In prior work, we present desired properties of such a FAIR Metric and possible approaches to the formulation of one [7]. Although authors use citations in varied and nuanced ways [8], we identified six possible relationships among pairs of works that represent the core function of attributing ideas to their proper sources [7]. The details of how to identify each scenario may depend on whether the work for which we are calculating the metric is a primary research article, a replication of a prior study, or a review article and whether the review article includes a meta-analysis. We here present five FAIR metrics that measure the presence or absence of the four scenarios relevant to primary research articles (Figure 2). We then outline an approach to implementation of a web application for computing values of the metric given a set of records of works.

Document created June 11, 2018, revised November 14, 2018; typeset November 14, 2018.

II. Methods

In prior work, we defined the essential features of a FAIR metric (Table I), the most essential being the ability to distinguish between appropriate presence of a citation, inappropriate presence of a citation, appropriate absence of a citation, and inappropriate absence of a citation [7]. For the current work, we determined what information would be necessary in order to compute such a metric and we would need to analyze it in order to compute a metric for each of the four scenarios relevant to a primary research article. The need to define whether one work should cite another based on *a priori* scholarly standards, not just which other works cite it, even when a field is divided into silos, calls for the use of content analysis to determine which statements in one work are equivalent to which statements in other works [9]. Since the goal of the calculation is to identify how a work fits into a larger set of works, inputs must be not only the work being evaluated but the entire set of descriptions of all potentially relevant works, including the network of citations between works. The need to consider domain-specific publication patterns and common knowledge calls for the use of data sets organized by problem domain and validated for relevance using concept-validating constraints, such as those used in the Nexus-PORTAL-DOORS System (NPDS) [10]. The first step of calculation must identify matches between content in an article and each prior article. For this step to be effective, it must be able to identify equivalence of concepts regardless of changes in wording, which we can best achieve with semantic analysis. Ideally, a natural language processing engine would reduce all statements to Resource Description Framework (RDF) triples using a consistent ontology, but, short of that, a human agent could create the RDF triples or rewrite the statements with standardized sentence structure and vocabulary for lexical comparison. NPDS supports this hybrid approach by managing both lexical metadata, including citation data, and semantic resource descriptions [11]. While some of our goals, such as reliably identifying what statements are common knowledge in a given field, may prove impractical, recent developments in semantic text analysis support the feasibility of the central goal of distinguishing legitimate reports of research from acts of plagiarism [12].

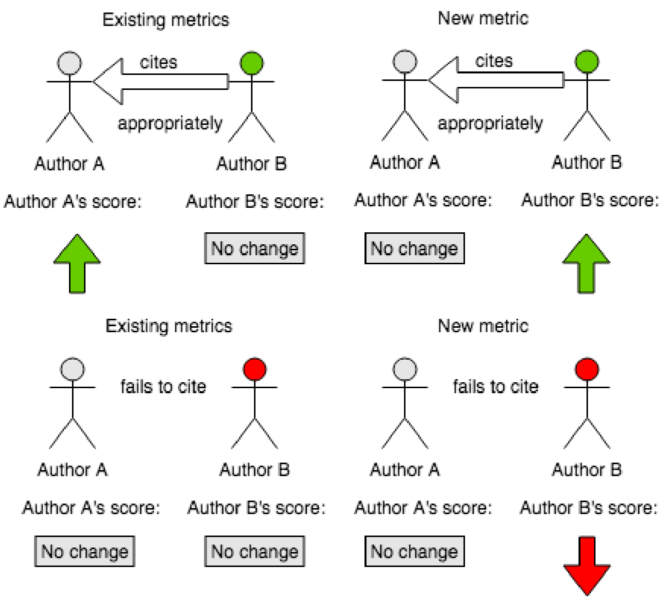

Fig. 1. Incentives in conventional vs FAIR metrics

TABLE I
Features of a FAIR Metric

Feature
distinguishes well-citing works from ones with errors
distinguishes plagiarism from minor errors
consistent even when bad practices are common
stable against attempts at obfuscation
allows comparison across problem domains
allows for common knowledge

III. Results

The resuls of the design process are a set of shared mathematical abstractions needed to define the metrics and a set of five FAIR metrics that capture the relevant features of the four key scenarios for primary research articles (Figure 2).

A. Core Concepts of FAIR Metrics

To measure the quality of citation practices in a scholarly work, we first need a description of the work, which we call a Report. We define the scope of the body of literature we are analyzing via the set of all Reports we consider, which we call the Index. In order to evaluate the appropriateness of a citation, we need to consider it not merely as an edge from one node to another in a citation network but as an attribution of some assertion made in one work to a prior work. This calls for the use of the concept of a Statement as any assertion made in a Report where S_r is the set of all Statements made in Report r. Since a work may cite multiple prior works sources for a given assertion, we define $A_{s,r,I}$ as the set of Reports in Index I to which Report r attributes statement s. In considering what works it is appropriate for a work to cite, we must also identify which works it could possibly cite. For this purpose, we define the set, $P_{r,I}$, as the set of all Reports in Index I made available for reading prior to the finalization of Report r. Whether a work should attribute a statement to a prior work hinges on whether the prior work includes an equivalent statement, so we define the equivalence relation $d(s_i, s_j)$ to be true if and only if Statements s_i and s_j express the same idea, regardless of exact phrasing and define $Q_{s_j,I} \equiv \{r \in I | \exists s_i \in S_r \text{ s.t. } d(s_i, s_j)\}$, the set of all Reports r in index I containing Statements equivalent to Statement s_j.

B. FAIR Metric Sub-Scores for Primary Research Articles

When describing an original primary research article, it is useful to divide the Statements found in such a Report r into two sets, the set of Statements claimed as original research results (C_r), the set of background Statements derived from prior work (B_r), and the set of background Statements that are common knowledge in the problem domain concerned (K_r). Since common knowledge statements do not require attribution, we can disregard them for the purpose of formulating the metrics. Fig. 2 illustrates the four situations among which a FAIR metric must differentiate in a primary research article. Define $g(r, I)$ to be the number of correct attributions in Report r of Background Statements to prior works that contain them:

$$g(r, I) \equiv \sum_{s \in B_r} |A_{s,r,I} \cap Q_{s,I}| \quad (1)$$

The Record may contain multiple correct attributions for the same Background Statement. Define $m(r, I)$ to be the number of mis-attributions of background statements to works from which they are absent:

$$m(r, I) \equiv \sum_{s \in B_r} |A_{s,r,I} \setminus Q_{s,I}| \quad (2)$$

The Record may contain multiple mis-attributions for the same Background Statement. Define $w(r, I)$ to be the number of background statements that are without attributions:

$$w(r, I) \equiv \sum_{s \in B_r} \begin{cases} 1, |Q_{s,I} \setminus A_{s,r,I}| = 0 \\ 0, \text{Otherwise} \end{cases} \quad (3)$$

Each unattributed statement only contributes to the total once. Define $n(r, I)$ to be the number of original claims that do not occur anywhere else in any prior Report:

$$n(r, I) \equiv \sum_{s \in C_r} \begin{cases} 1, |Q_{s,I} \cap P_{r,I}| = 0 \\ 0, \text{Otherwise} \end{cases} \quad (4)$$

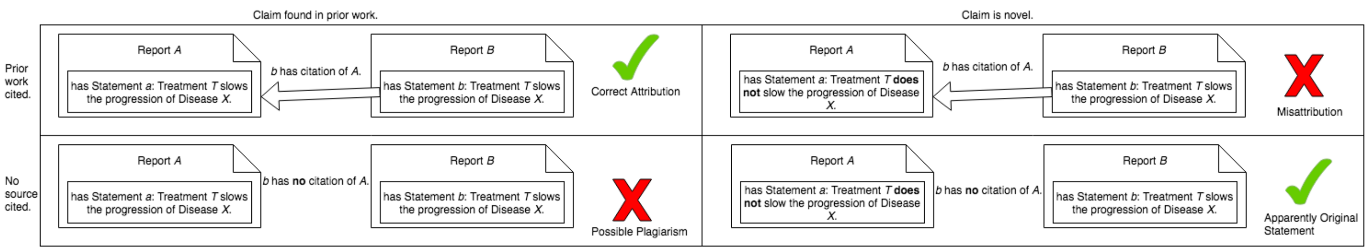

Fig. 2. Key scenarios a FAIR metric must differentiate within a primary research article (Note: a case where a Statement occurs in some prior work but not the one cited as its source would also be a misattribution.)

TABLE II
Correspondence between conventional and FAIR metrics at different levels

level	**conventional metric**	**definition**	**FAIR counterpart**
article	citation count	number of other articles citing this article	dependent on type of work
author	Hirsch Index	h s.t. author published h articles with citation count $\geq h$[13]	a s.t. author published a articles with scores $\geq a$
journal	Impact Factor	times cited in past 2 years / articles published in past 2 years[14]	mean score of articles published in past 2 years

Define $p(r, I)$ to be the number of purportedly original claims that do occur in at least one prior Report:

$$p(r, I) \equiv \sum_{s \in C_r} \begin{cases} 1, |Q_{s,I} \cap P_{r,I}| > 0 \\ 0, \text{Otherwise} \end{cases} \tag{5}$$

For example, manually parsing the abstact of retracted Report r ([15]) into Statements consisting of *subject-verb-object* triples revealed $n(r, I) = 7$ apparently original Statements and $p(r, I) = 22$ unattributed Statements equivalent to ones derived from the abstract of Report ([16]) using the same approach. Intuitively, these values suggest significant plagiarism occurred, reflecting identical experimental procedure and numerical data, despite differences in vegetable used, model fitted to the data, and time and place of the experiment.

IV. Conclusion and Future Work

The FAIR metrics provide ways of quantitatively describing how well a scholarly work adheres to established standards of attribution. The good practices scores count the number of instances where the work follows best practices: $g(r, I)$ counts the number of background statements that have appropriate references, and $n(r, I)$ counts the number of genuinely original claims. The bad practices scores count the number of instances where the work fails to follow best practices in a particular way: $m(r, I)$ counts the number of background statements mis-attributed to sources that do not contain the specified information. $w(r, I)$ counts the number of background statements without attributions. $p(r, I)$ counts the number of purportedly original claims also found in prior works and thus possibly plagiarized. In subsequent work, we plan to test these metrics over standard plagiarism detection data sets, such as those used in the International Competition on Plagiarism Detection [17]. From these data we hope to compute the distribution of values of each metric over legitimate primary research articles and for plagiarized ones, normalized for total number of background statements or purportedly original claims, as appropriate. This will allow us to perform statistical tests to determine which metrics are most informative for identifying plagiarism and possibly to develop a combined metric that can serve as more reliable warning sign that a work is plagiarized.

We also plan to implement software that can evaluate the metric with each selected using metadata stored in NPDS servers. The implementation will consist of a web form UI where the user will enter the title or entity label of an article, the type of article, and a list of names or entity labels of Nexus diristries to search [11]. The web form will use AJAX to send a request with this information to a web service, which will make an NPDS API request to each Nexus diristry (Figure 3). It will then use the information received in the response to compute each of the FAIR metrics and send them in a response to the web form, which will display them for the user (Figure 4).

The set of all entities with NPDS records of type "Publication" with records in any of these diristries will constitute the Index. Initially, we will need to rely on manual curators to embed in each entry a list of background statements and a list of original claims in a standardized format, either as RDF triples or as plain text written in a previously agreed on style in order to allow lexical comparison. Later, we hope to develop natural language processing methods that can extract these lists of claims automatically. Since each NPDS server keeps records relevant to a problem domain [10], we will associate with each such server a set of statements domain experts would regard as common knowledge not requiring attribution so that we can filter these from the set of background statements in a work prior to calculating metrics for it. We hope that this web interface will provide a useful tool for institutions wishing to uphold established norms of scholarship and proper attribution of work.

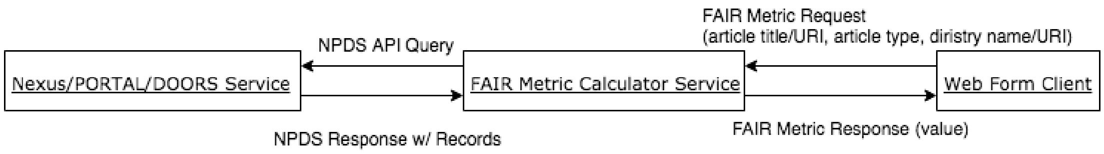

Fig. 3. Components of an NPDS-facing web service for computing the FAIR metric for each of a set of articles

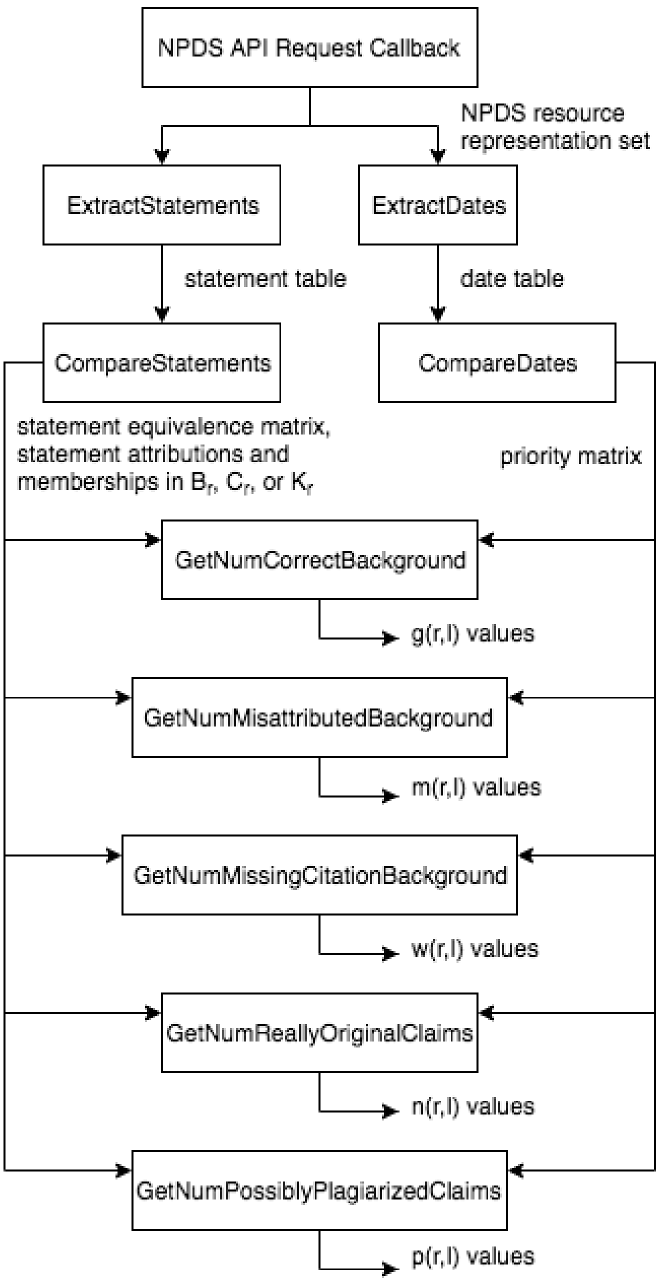

Fig. 4. Data pipeline for calculation of FAIR metrics from metadata in NPDS records

References

[1] A. Dix, "Evaluating research assessment: Metrics-based analysis exposes implicit bias in ref2014 results," *Impact of Social Sciences Blog*, 2016.

[2] H. F. Moed, *Citation analysis in research evaluation*. Springer Science & Business Media, 2006, vol. 9.

[3] D. Hicks, P. Wouters, L. Waltman, *et al.*, "Bibliometrics: The leiden manifesto for research metrics," 2015.

[4] R. Cagan, *The san francisco declaration on research assessment*, 2013.

[5] J. Wilsdon, *The metric tide: Independent review of the role of metrics in research assessment and management*. Sage, 2016.

[6] T. Tyler, "Citation metrics and impact factors fail as measures of scientific quality, in particular in taxonomy, and are biased by biological discipline and by geographic and taxonomic factors," in *Annales Botanici Fennici*, BioOne, vol. 55, 2018, pp. 185–191.

[7] A. Craig and C. Taswell, "The FAIR metrics of adherence to citation best practices," in *SIGMet*, 2018.

[8] D. Shotton, "Cito, the citation typing ontology," in *Journal of biomedical semantics*, BioMed Central, vol. 1, 2010, S6.

[9] Y. Ding, G. Zhang, T. Chambers, *et al.*, "Content-based citation analysis: The next generation of citation analysis," *Journal of the Association for Information Science and Technology*, vol. 65, no. 9, pp. 1820–1833, 2014.

[10] C. Taswell, "Concept validating methods for maintaining the integrity of problem oriented domains in the PORTAL-DOORS system," *AMIA*, 2010.

[11] ——, "A distributed infrastructure for metadata about metadata: The HDMM architectural style and PORTAL-DOORS system," *Future Internet*, vol. 2, no. 2, pp. 156–189, 2010.

[12] D. Gupta *et al.*, "Study on extrinsic text plagiarism detection techniques and tools.," *Journal of Engineering Science & Technology Review*, vol. 9, no. 5, 2016.

[13] J. E. Hirsch, "An index to quantify an individual's scientific research output," *Proceedings of the National academy of Sciences*, vol. 102, no. 46, pp. 16 569–16 572, 2005.

[14] E. Garfield, "The history and meaning of the journal impact factor," *Jama*, vol. 295, no. 1, pp. 90–93, 2006.

[15] F. Ullah, M. Kang, M. K. Khattak, *et al.*, "Retracted: Experimentally investigated the asparagus (asparagus officinalis l.) drying with flat-plate collector under the natural convection indirect solar dryer," *Food Science & Nutrition*, vol. 6, no. 6, pp. 1357–1357, doi: 10.1002/fsn3.603. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/fsn3.603. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/fsn3.603.

[16] S. Sansaniwal and M. Kumar, "Analysis of ginger drying inside a natural convection indirect solar dryer: An experimental study," *Journal of Mechanical Engineering and Sciences*, vol. 9, no. unknown, pp. 1671–1685, 2015.

[17] M. Potthast, A. Barrón-Cedeño, A. Eiselt, *et al.*, "Overview of the 2nd international competition on plagiarism detection," in *Proceedings of the 4th Workshop on Uncovering Plagiarism, Authorship, and Social Software Misuse*, 2010, pp. 1–14.