Creating and Exploiting the Mappings from Conference Review Forms to a Generic Set of Review Criteria

Vojtěch Svátek[1], Sára Juranková[1], Radomír Šalda[2], Petr Strossa[1], and Zdeněk Vondra[2]

[1] Dept. of Information and Knowledge Engineering, University of Economics, Prague
Nám. W. Churchilla 4, 130 67 Prague 3, Czech Republic
`{svatek|jurs02|petr.strossa}@vse.cz`
[2] Dept. of Multimedia, University of Economics, Prague
Nám. W. Churchilla 4, 130 67 Prague 3, Czech Republic
`{salr00|zdenek.vondra}@vse.cz`

Abstract. Conference papers are evaluated according to many criteria reflected in numerical scores, and the wording of the criteria differs among conferences. This makes the role of meta-reviewers tough when summarizing the evaluation across multiple criteria and reviewers. Based on a micro-study within semantic technology conferences, we conjecture that the criteria can, for particular fields, be mapped on generic metrics, and provide a provisional ontological representation for such a mapping and a set of metrics, as well as a manual mapping tool. Finally, we showcase an application exploiting the mappings: a graphics generator that aggregates the review data into a complex pictorial metaphor.

1 Introduction

Conference papers are often evaluated according to multiple criteria reflected in numerical scores. In large conferences with many reviews per paper this amounts to dozens of partial figures. This sheer number, and the fact that the wording of the criteria differs from one conference to another, make the role of meta-reviewers during the discussion periods difficult, and the effort invested into the detailed scoring may partly get lost.

In the research we first explored whether the criteria can be generalized *across events* within a field such as Semantic Technology (ST) to a small set of *review metrics*. Based on the positive outcome of this study, we designed a provisional *ontology* for representing the mappings between specific review forms and such generic criteria, and developed a simple *mapping authoring tool* and a *mapping execution* component. Finally, we developed a tool that demonstrates one possible way of exploiting the mappings: a *review visualizer* that assembles the metric values, for a set of reviews of the same paper, into a compound pictogram relying on the racing cars metaphor.

Contributions of this paper are thus both the small empirical study and a multi-part demo. In the demo we can demonstrate how: 1) a mapping from a form to the common set of metrics can be created and published, 2) the values for a concrete set of reviews can be manually entered, thus emulating an automatic input from a hypothetical component of a conference review system, and 3) the pictorial scene can be generated.

2 Review Criteria Micro-Study and Mapping to Generic Metrics

We analyzed the review forms of nine ST conferences, always for the latest edition we could access as author/s or reviewer/s. We semantically clustered the field labels (referring to the reviewer guidelines where in doubts), yielding seven partial *review metrics* that we named as in the first column of Tab. 2, plus two global metrics, *Confidence* and *Overall score*, present in all forms. The partial criteria converged well despite the varying wording, though some forms missed certain metrics at all. ISWC and SEMANTiCS were clearly influenced by one another, having the same set of fields. ESWC had two fields that we both ranged under 'Technical quality'. We do not list K-CAP, as it had no partial numerical field; this may be related to the 'workshop flavor' of this event.

3 Ontological Representation of Review Forms and Metrics

For a review of existing relevant ontologies we can refer to our up-to-date study of research-related ontologies in general [2]. The recently developed *FAIR* ontology[3] covers the overall review process (reviewers, reviews and venues). The associated *Review Measures* module of the *BIDO* ontology[4] contains, among other, a large collection of individuals corresponding to different rating/confidence scales and their values. None of these ontologies however addresses the semantics of partial review metrics. Therefore, we rapidly prototyped an ontology (not yet considering all best practices, thus likely subject to revisions in the future) that supports the publishing of metrics and their relationships to review forms. The ontology is online at `http://kizi.vse.cz/pictoreview/ontology/`, and contains the classes *ReviewMetrics*, *ReviewForm*, *ReviewFormField* and *F2M_Mapping* (for the field-to-metric mapping), plus the connecting properties. The proposed metrics set (applicable on ST conferences, and probably many other computing field's ones) is at `http://kizi.vse.cz/pictoreview/metrics/`. Finally, a sample mapping (that used in the example below) is at `http://kizi.vse.cz/pictoreview/map/semantics18/`.

4 Demo Suite

We developed a suite of four simple tools to demonstrate the whole concept. They are bundled by the web page `http://pictoreview.vse.cz/`. The source code for the first three tools is at `https://github.com/jurs02/PictoReviewDev`.

The first tool allows the user to create a *mapping* from the custom set of review form fields of a particular event to the proposed set of generic metrics. The mapping can be 1:1, 1:N or N:1. An example of a mapping (for the SEMANTiCS'18 Research Track) is in the first two columns of Tab. 1. The mapping can be currently stored as a JSON structure or as an RDF dataset described by our ontology from Sect. 3.

---

[3] `https://sparontologies.github.io/fr/current/fr.html`

[4] `https://sparontologies.github.io/bido-review-measures/current/bido-review-measures.html`

The second tool is a simple *mapping execution API*, which transforms a set of review form fillings of a specific conference (a JSON structure) to the generic metrics (also output in JSON), using the JSON mapping (valid for that conference) authored by the first tool. For the N:1 mapping (i.e., of multiple fields to the single metric), a numerical mean of the values is computed. Note that the first and second tool together provide (a baseline of) a general review data interoperability infrastructure, usable independent of the rest of the demo; for example, the reviewing emphases of different conferences could be compared based on the mappings.

The third tool emulates the role of a hypothetical plug-in to an off-the-shelf *review management system* (RMS). The user manually enters both JSON data structures expected by the second tool: the (saved) mapping, and the specific review form fillings, for example, those from the last three columns of Tab. 1. The data is then transformed to generic metrics (by the second tool) and passed to the fourth tool.

The fourth tool, the *pictogram generator*, eventually, converts the generic metric values to components of a complex pictorial metaphor. We identified 'car' as a relatively close metaphor to a research paper, and car components (plus other 'car race' features) as *visual variables* expressing the metrics values. In Fig. 1 we see the visual representation of the set of reviews from our SEMANTiCS'18 example, cf. Tab. 1. The whole picture encodes 27 numerical values: 9 metrics × 3 reviewers. For brevity let us only point out the 'good' and 'bad' scores. Reviewer 3 (R3) appreciated the novelty of the paper (big engine), its evaluation (solid wheels), and also presentation (smiling face). R1 valued the state of the art (shining headlamp) and technical quality (body style: cabrio as most suitable for a racing car), and also evaluation (wheels). R2, in turn, only praised the paper for its high relevance (this would be indicated by the track quality, however, the difference is too small;[5] with an even lower value, the track would change to dirt or even turf), while the presentation was poor in particular (frowning face). The *reviewer confidence* (lower for R1) does not measure the paper quality as such; therefore we use an orthogonal visual magnitude paradigm, the color saturation/salience. Finally, the cars are positioned on the track by their *overall* scores.

Table 1. Numerical scores of the example SEMANTiCS'18 paper

Original review field	Mapped to metric	Rev. 1	Rev. 2	Rev. 3
Appropriateness	Relevance	4	5	4
Originality / innovativeness	Novelty	3	3	4
Implementation and soundness	Technical quality	4	3	3
Related work	State of the art	4	3	3
Evaluation	Evaluation	4	3	4
Impact of ideas and results	Significance	3	3	3
Clarity and quality of writing	Presentation	3	2	5
Reviewer's confidence	Confidence	3	4	4
Overall evaluation	Overall score	0	-1	2

---

[5] For simplicity, all scores except the overall evaluation are mapped to three-valued visual variables only, thus 4 and 5 fall to the same interval.

Table 2. Proposed mapping between generic review metrics and form fields of KE conferences

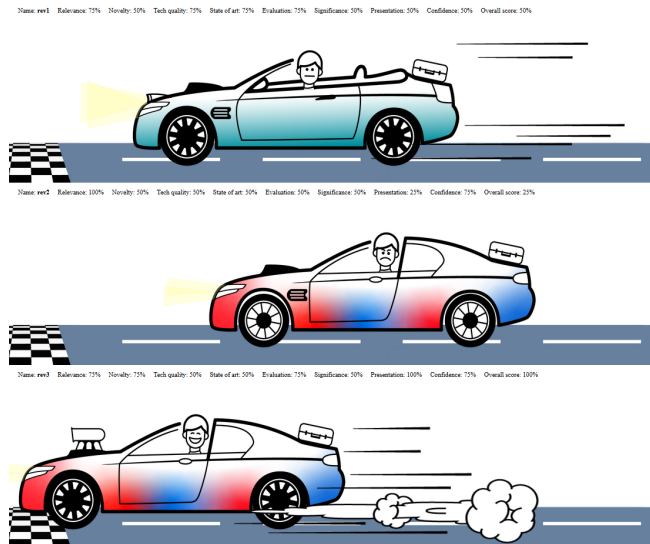| Review metric	ECAI (2016)	EKAW (2020)	ESWC (2018)	FOIS (2016)	IJCAI (2019)	ISWC, SEMANTiCS (2018)	KR (2014)
Relevance	Relevance	NA	Relevance to ESWC	NA	Relevance	Appropriateness	Relevance of the paper to KR
Novelty	Originality	Novelty	Novelty of the proposed solution	Novelty or innovation	Originality	Originality / innovativeness	Novelty of the contribution
Technical quality	Technical quality	Technical soundness and depth	Correctness and completeness of the proposed solution; Demonstration and discussion of the properties of the proposed approach	Scientific or technical quality	Technical quality	Implementation and soundness	Technical quality
State of the art	Scholarship	NA	Evaluation of the state-of-the-art	References	Scholarship	Related work	Discussion of related work
Evaluation	NA	NA	Reproducibility and generality of the experimental study	NA	NA	Evaluation	NA
Significance	Significance	NA	NA	NA	Significance	Impact of ideas and results	NA
Presentation	Presentation quality	Clarity and quality of writing	NA	Presentation	Clarity and quality of writing	Clarity and quality of writing	Quality of the presentation

Name: rev1 Relevance: 75% Novelty: 50% Tech quality: 75% State of art: 75% Evaluation: 75% Significance: 50% Presentation: 50% Confidence: 50% Overall score: 50%

Name: **rev2** Relevance: 100% Novelty: 50% Tech quality: 50% State of art: 50% Evaluation: 50% Significance: 50% Presentation: 25% Confidence: 75% Overall score: 25%

Name: rev3 Relevance: 75% Novelty: 75% Tech quality: 50% State of art: 50% Evaluation: 75% Significance: 50% Presentation: 100% Confidence: 75% Overall score: 100%

Fig. 1. Visual metaphor of three reviews of the example SEMANTiCS'18 paper

5   Future Prospects

The paper presents an initial proof of concept of a review form interoperability framework, plus a review pictogram generator on the top of it. To bring the concept closer to real usage, we have to undertake *experiments* determining whether and in what setting the pictograms provide an added value over numerical tables. Some of the visual variables adhere to metaphors studied by psychologists [1] (e.g., "linear scales are paths" for overall score, or "thought is motion" for originality) and might thus be relatively intuitive; however, others might require a longer adaptation period. As regards the semantic web aspects of the research, we plan to submit the current review metric *ontology* to a redesign process based on competence questions; review ontologies (such as FAIR and BIDO), and possibly even multimedia ontologies, are likely to be reused.

References

1. Lakoff, G.: The contemporary theory of metaphor. In A. Ortony (Ed.), Metaphor and thought Cambridge, MA: Cambridge University Press (1993).
2. Nguyen V. B., Svátek V., Rabby G., Corcho O.: Ontologies Supporting Research-related Information Foraging Using Knowledge Graphs: Literature Survey and Holistic Model Mapping. In: EKAW 2020. Springer LNCS, to appear.