

Preface

The first Workshop on Negative or Inconclusive Results in Semantic Web (NoISE 2015) provided a forum for negative or inconclusive attempted approaches, methodologies, or implementations. NoISE aimed at breaking the taboo on negative results in Semantic Web and Linked Data research, by incentivizing researchers to share tests, applied methodologies or documented approaches that did not reach their goal. These results can now be studied and as a community we can discuss when and how negative or inconclusive should be published. This workshop addresses the way Semantic Web research deals with insufficient evidence and negative results. The NoISE workshop was a half-day workshop that took place on 1st June 2015 in Portoroz, Slovenia and was co-located with the 12th Extended Semantic Web Conference (ESWC 2015).

The workshop was organized in a series of alternative session formats. Prof. Dr. Maria-Esther Vidal opened with an inspirational *introduction talk*. She discussed the fundamentals of scientific method, theory and formal evaluations in computer science and their implications for negative results. The keynote was followed by the *Glorious Failure session* where a short and three extended papers with concrete and complete cases of contributions resulting in negative or inconclusive results were presented. Next, the *Confessions session* took place which consisted of an interview and two position papers. Jacco van Ossensbruggen interviewed Kjetil Kjernsmo regarding on how scientific methods provide guidance for Semantic Web Research and Development. Its transcript is included in these proceedings.

Last, the workshop concluded with a *Breakout session* on guidelines for reporting negative results, which are summarized as follows. A report of experimental results, either it is positive, but especially if it is inconclusive or negative, should consist of (i) the **Research Question (RQ)**, (ii) the **Hypotheses of the Evaluation (H)**, (iii) the **Experimental Evaluation** and, last, (iv) the **Analysis of the observed results**. Initially, the research questions which are targeted should be enumerated. Then, the hypotheses which are going to be evaluated should be clearly formulated by specifying the properties which determine whether each hypothesis is validated or not. The experimental evaluation should clearly state and justify the benchmark choices. To be more precise, the configuration setup, both regarding the data and queries, as well as for the computational equipment (infrastructure) and the operating systems should be mentioned. In respect to the data and queries used to validate the hypotheses properties, the characteristics of the data (e.g. their size, number of triples etc.), the queries and the use cases should be specified. Then, the methodology

followed and the metrics which are taken into consideration should be listed and aligned with the properties which they validate. Last, the statistical methods and tools used to analyze the results should be mentioned, as well as other parameters that could affect the evaluation. As soon as all the aforementioned are covered and the proposed approach has been validated, the analysis of the observed results should follow, described independently of their nature, namely either they contradict or confirm the hypotheses. In the case of inconclusive or negative results, instead of automatically rejecting the approach, it should be documented why it is of interest for the rest research community to know which properties of the original hypotheses behave in the case of the examined approach.

During the discussion some interesting issues on these guidelines and the review process were raised. First, what about qualitative research and user studies? What parts of the guidelines don't apply here or should others be added? What if the Research Question is not as clear as for a quantitative paper, or the paper is a model paper? Second, how are the guidelines different from positive papers and do they need to be? Third, would single or double blind reviewer dynamics be different for negative papers? Should there be an open review/research environment with rebuttal? And what are the incentives for the reviewer? Fourth, focus should be on the motives and the lessons learned of the research. The reason for trying this in the first place should be communicated. Also, a clear explanation on why this failure should be known to the community (e.g. amount of resources wasted on) is crucial.

Furthermore, possible venues for such results were discussed. There was a general consensus that native results help improve the quality of positive results. In addition, negative results in *hot* topics, where a lot of people are working on, should be published sooner, i.e. burned bridges. There are a lot of things that can go wrong in Semantic Web research, which may be worth a blog. But which are worth a conference or journal publication? Are good (e.g., IPython notebooks) blogs not sufficient? An experimental research track on an established conference would be low hanging fruit. It should explicitly welcome negative results, which are reviewed like positive results. A more informal *"Skill Sharing"* track was suggested. The program would include reports on (i) tried, but failed work with a recommendation to not try this technical solution; (ii) tried work with recommendations towards unforeseen evaluation issues. Additional suggestions were to publish (abstracts of) rejected conference papers and to invite famous people for negative result paper.

Portoroz, June 2015

Anastasia Dimou
Jacco van Ossenbruggen
Miel Vander Sande
Maria-Esther Vidal