Third International Workshop on Capturing Scientific Knowledge (SciKnow2019): Summary Report

Daniel Garijo
Information Sciences Institute,
University of Southern California
dgarijo@isi.edu

Milan Markovic
University of Aberdeen
milan.markovic@abdn.ac.uk

Paul Groth
University of Amsterdam
p.groth@uva.nl

Khalid Belhajjame
University of Paris-Dauphine
Khalid.Belhajjame@dauphine.fr

Idafen Santana-Perez
Ontology Engineering Group
idafensp@gmail.com

ABSTRACT

From the early days of Artificial Intelligence, researchers have been interested in capturing scientific knowledge to develop intelligent systems. A variety of formalisms are used today in different areas of science, ranging from ontologies for organizing knowledge - particularly in biology and medicine; process representations to do qualitative reasoning in areas such as physics and chemistry; or probabilistic graphical models used by machine learning scientists.

Despite recent advances, scientific knowledge is complex and poses significant challenges for knowledge capture. The Third Workshop for Capturing Scientific Knowledge (Sciknow 2019)[1] provided a forum to discuss, envision and expand existing forms of scientific knowledge representation and dissemination, together with existing systems that use them.

SciKnow 2019 took place in Marina del Rey (Los Angeles) on November 19th, 2019; and had roughly 25 attendants during a half full day event. Attendees had mainly a computer science background, but had experience in bioinformatics, humanities and climate sciences. The workshop was the third of a series, which started at K-CAP 2015 with a full day event[2] and continued at K-CAP 2017.[3]

KEYWORDS

Scientific knowledge capture, SciKnow, K-CAP, scientific knowledge dissemination, scientific knowledge representation.

1 WORKSHOP SUMMARY

Sciknow 2019 started with an invited keynote from Dr. Yolanda Gil,[4] director of knowledge technologies at the Information Sciences Institute[5]. Dr. Gil presented her work on capturing scientific hypotheses and scenarios in scientific research, walking the attendees through her experience in different projects on bioinformatics and environmental sciences. In her keynote, Dr. Gil argued that AI systems should become science partners, i.e., active participants in scientific problems that focus on the repetitive tasks of science; leaving the creative parts to human researchers. For example, researchers can define a hypothesis and define the solution search space, while an automated system can launch executable experiments to test it and return the results. This becomes increasingly important as new data becomes available, since automated systems may re-assess previous hypotheses even if the original results have already been published.

The keynote also touched base on other topics that led to discussion in the workshop. One of the main challenges for interdisciplinary research is the social barrier between domain experts for knowledge transfer, especially when different parts of an experiment require validation from different experts. Experiments need proper machine readable metadata to ease these interactions and help authors converge towards an agreed solution.

The workshop was divided into 4 sessions, further described below:

The first session dealt with different aspects of knowledge capture in scientific workflows. Carrillo et al. [4] described an approach to create scientific workflows of carbon estimation models, through an assessment of the number of trees in an area. The approach included a series of geospatial transformation methods, which the authors tried to abstract to be reused in other domains. Markovic et al. [8] followed with an approach to represent advanced plans for scientific workflows at different levels of abstraction.

The second session focused on applications related to scientific knowledge capture. Sefit et al. [1] presented early work towards automating slide generation by creating appropriate summaries of scientific papers. Daga and Motta [3] continued with an approach to help scientists curate documentary evidence in digital humanities: according to their results, traditional entity extraction techniques tend to perform poorly because the entities of interest are often implicit in the analyzed text.

The third session emphasized scientific knowledge extraction for evaluation. Olleen et al. [9] described an approach towards comparing contributions of research papers in an open research knowledge graph. Their goal is to help users identify similarities and differences among different research papers, which is usually time consuming.

[1]https://sciknow.github.io/sciknow2019
[2]https://www.isi.edu/ikcap/sciknow2015/
[3]https://sciknow.github.io/sciknow2017
[4]http://www.isi.edu/~gil/
[5]http://isi.edu/

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/1122445.1122456

The workshop ended with a session on knowledge capture in scientific data. Garcia Silva et al. [5] discussed an approach to learn embeddings from scientific corpora using lexical, semantic and grammatical information. Their goal was to improve how to recognize multi-word expressions that are usually not well detected from scientific papers. Next, Iglesias Molina et al. [7] proposed to use spreadsheets as a meta-mapping language to automatically handle the construction of knowledge graphs. Garijo and Szekely [6] closed the session by introducing a decentralized approach for easily extending Wikidata with tabular data (while keeping the provenance of the process as part of the extension as well).

2 CHALLENGES AND NEXT STEPS FOR SCIENTIFIC KNOWLEDGE CAPTURE

In the report following the previous edition of the workshop [2], three main challenges for scientific knowledge capture were emphasized. The first one was *knowledge explainability*, i.e., the need for intelligent systems to produce results that are clear, unambiguous, explicit and accompanied by semantics to ease their usability. The second challenge was *knowledge transfer*, referring to the need for software, data and methods to be properly described with metadata so they could be used in different situations (e.g., used by a different researcher than the authors, used in other domains, etc.). The final challenge was *context capture*, referring to the need of representing in a machine-readable manner the research questions and hypotheses associated with a scientific result.

We believe these challenges are still relevant in the papers submitted to SciKnow 2019. For example, [6] and [8] highlight the need to capture provenance traces, either for easing the creation of knowledge graphs or allowing the inspectability of scientific workflows when executed in complex environments. Regarding knowledge transfer, [4] and [7] propose different methods to help capturing and translating knowledge in geosciences and facilitate data integration. Finally, [3] emphasizes the importance of linking evidence to support scientific research.

Inspired by the workshop presentations and discussion, we briefly describe below three additional challenges that we believe are becoming increasingly important to scientific knowledge capture.

The first challenge is *data translation and transformation*, which are critical for scientific knowledge transfer and consolidation. Even though these problems have been around for decades, we still do not have generic working solutions. As a result, we still rely on largely manual processes that suffer from numerous quality problems and are time consuming for researchers. Intelligent systems could provide metadata rich annotations of scientific results so data transformations could be handled by automated systems.

The second challenge is the need for approaches that help scientists *improve their productivity* by automating different parts of the scientific research cycle. Intelligent systems should become assistants that test experiments designed by scientists, help them compare their work against the state of the art (or even build automated benchmarks, inspired by [9]) or help them disseminate their results by automatically creating content from their publications (as shown in [1]).

Our third challenge refers to the need for intelligent systems that help scientists and developers of different disciplines to *address the knowledge gaps in their collaborations*. This challenge is a type of *knowledge transfer* that is currently a socio-technical problem, motivated by heterogeneous backgrounds and expertise from researchers in different fields. Intelligent systems should be able to capture enough context and metadata to effectively communicate a scientific result to researchers with different expertise.

ACKNOWLEDGEMENTS

The workshop editors would like to thank Dr. Gil and all SciKnow 2019 authors, participants and reviewers for their comments and contributions.

REFERENCES

[1] Athar Sefid, Jian Wu, P. M. C. L. G. Automatic slide generation for scientific papers. In *Proceedings of the Third International Workshop on Capturing Scientific Knowledge co-located with the 10th International Conference on Knowledge Capture (K-CAP 2019)* (2019), CEUR-WS, pp. 11–16.

[2] Badenes, C., Denaux, R., De Vos, M., Garijo, D., Gomez-Perez, J. M., Lawrynowicz, A., Lisena, P., Palma, R., Troncy, R., and Vila, D. K-cap2017 satellites: Workshops and tutorials. In *Proceedings of the Knowledge Capture Conference* (2017), ACM, p. 1e.

[3] Daga, E., and Motta, E. Challenging knowledge extraction to support the curation of documentary evidence in the humanities. In *Proceedings of the Third International Workshop on Capturing Scientific Knowledge co-located with the 10th International Conference on Knowledge Capture (K-CAP 2019)* (2019), CEUR-WS, pp. 17–20.

[4] Garcia, J. M. C., Garijo, D., Crowley, M., Carrillo, R., Gil, Y., and Borda, K. Semantic workflows and machine learning for the assessment of carbon storage by urban trees. In *Proceedings of the Third International Workshop on Capturing Scientific Knowledge co-located with the 10th International Conference on Knowledge Capture (K-CAP 2019)* (2019), CEUR-WS, pp. 1–6.

[5] García-Silva, A., Denaux, R., and Gomez-Perez, J. M. Learning embeddings from scientific corpora using lexical, grammatical and semantic information. In *Proceedings of the Third International Workshop on Capturing Scientific Knowledge co-located with the 10th International Conference on Knowledge Capture (K-CAP 2019)* (2019), CEUR-WS, pp. 27–32.

[6] Garijo, D., and Szekely, P. Wdplus: Leveraging wikidata to link and extend tabular data. In *Proceedings of the Third International Workshop on Capturing Scientific Knowledge co-located with the 10th International Conference on Knowledge Capture (K-CAP 2019)* (2019), CEUR-WS, pp. 37–39.

[7] Iglesias-Molina, A., Chaves-Fraga, D., Priyatna, F., and Corcho, O. Towards the definition of a language-independent mapping template for knowledge graph creation. In *Proceedings of the Third International Workshop on Capturing Scientific Knowledge co-located with the 10th International Conference on Knowledge Capture (K-CAP 2019)* (2019), CEUR-WS, pp. 33–36.

[8] Markovic, M., Garijo, D., and Edwards, P. Linking abstract plans of scientific experiments to their corresponding execution traces. In *Proceedings of the Third International Workshop on Capturing Scientific Knowledge co-located with the 10th International Conference on Knowledge Capture (K-CAP 2019)* (2019), CEUR-WS, pp. 7–10.

[9] Oelen, A., Jaradeh, M. Y., Farfar, K. E., Stocker, M., and Auer, S. Comparing research contributions in a scholarly knowledge graph. In *Proceedings of the Third International Workshop on Capturing Scientific Knowledge co-located with the 10th International Conference on Knowledge Capture (K-CAP 2019)* (2019), CEUR-WS, pp. 21–26.