Combining Machine Learning and Semantic Web
A Systematic Mapping Study[1]

Originally by

Anna Breit, Semantic Web Company;et al.

Presented By

Brandon Dave

CS7820

SPR2024

Instructor: Dr. Cogan Shimizu

[1]: https://dl.acm.org/doi/pdf/10.1145/3586163

Timeline

1. Key Terms
2. Challenges and Contributions
3. SMS Methodology
4. SWeML SMS Results
5. Conclusion

Key Terms

- Systematic Mapping Study (SMS)
- Semantic Web Machine Learning (SWeML) Systems
- SW Knowledge Structure

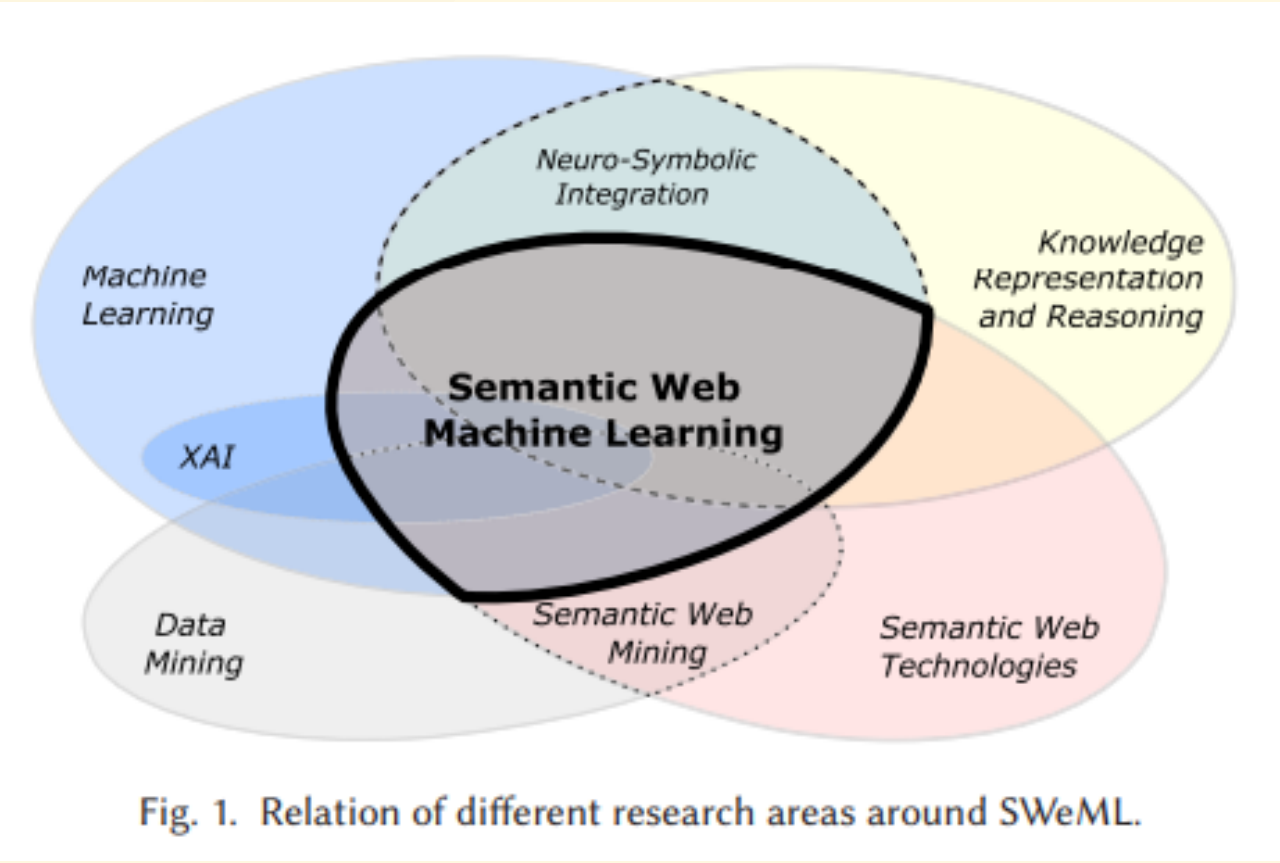


Fig. 1. Relation of different research areas around SWeML.

Key Terms

- **Systematic Mapping Study (SMS)**
- Semantic Web Machine Learning (SWeML) Systems
- SW Knowledge Structure

> **Systematic Mapping Studies** (also known as Scoping Studies) are designed to provide a wide overview of a research area, to establish if research evidence exists on a topic and provide an indication of the quantity of the evidence. [2]

[2]: Kitchenham, B., and Charters, S. Guidelines for performing systematic literature reviews in software engineering. 2007.

Key Terms

- Systematic Mapping Study
- **Semantic Web Machine Learning (SWeML) Systems**
- SW Knowledge Structure

> “ SWeML Systems are the result of **combining SW technologies** and an **inductive** model.[1] ”

Key Terms

- Systematic Mapping Study
- SWeML Systems
- **SW Knowledge Structure**

> “ [A] symbolic representation of a conceptual domain model and data complying with such domain models.[1] ”

RQs

1. What are the state of the art and trends related to systems that combine SW and ML components?
2. How can these systems be classified into a systematic taxonomy?

Challenges

1. "...keeping up with the main trends in the field has become unfeasible..."
2. "...the lack of a standardized way to report SWeML Systems..."

Contribution

1. A trends landscape
2. A classification system for SWeML Systems

Challenges

1. "...keeping up with the main trends in the field has become unfeasible..."

> [T]here is a need for a survey that adopts a solid review **methodology to complement current insights** with evidence-based findings.

Challenges

2. "...the lack of a standardized way to report SWeML Systems..."

“ [A]uthors of SWeML Systems would benefit from **a structured way to describe their system and its key characteristics**. ”

“ Readers...would benefit from **a structured way of interpreting such systems** ”

Contribution

1. A trends landscape
2. A classification system for SWeML Systems

SMS Methodology

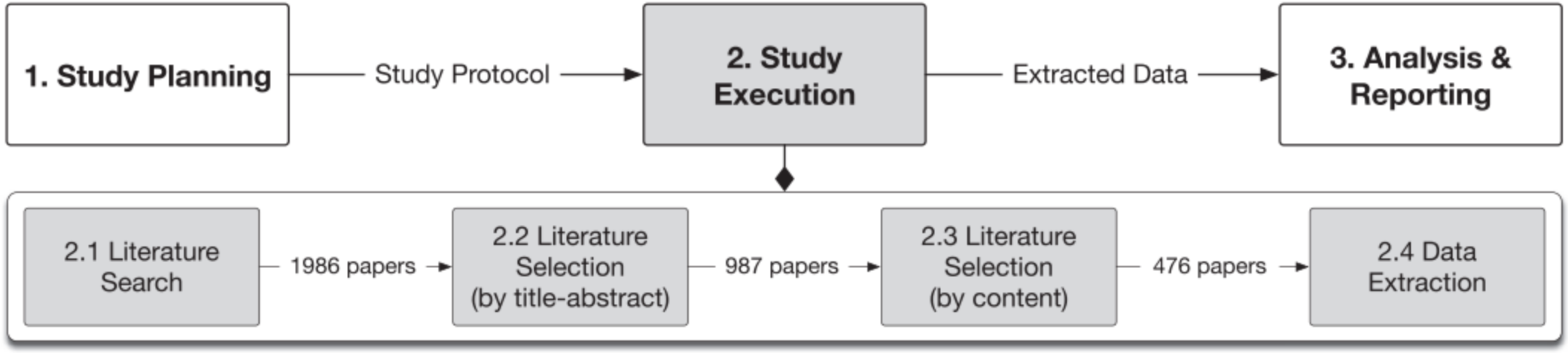


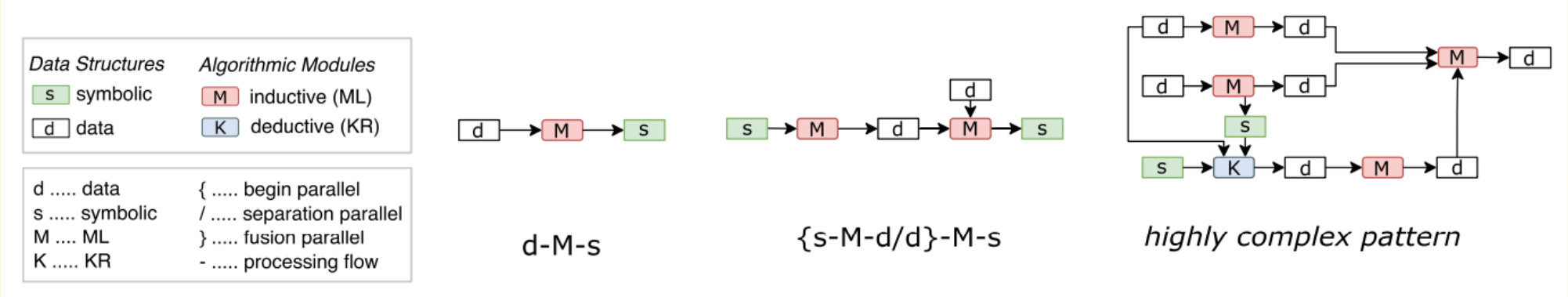
Fig. 2. Overview of the SMS process.

RQ1

“ RQ1 Bibliographic characteristics:
How are the publications temporally and geographically distributed? How are the systems positioned, and which keywords are used to describe them? ”

RQ2

“ RQ2 System architecture:
What processing patterns are used in terms of inputs/outputs and the order of processing units? ”



RQ3

“ RQ3 Application areas:
What kind of tasks are solved (e.g., text analysis)? In which domains are SWeML Systems applied (e.g., life sciences)? ”

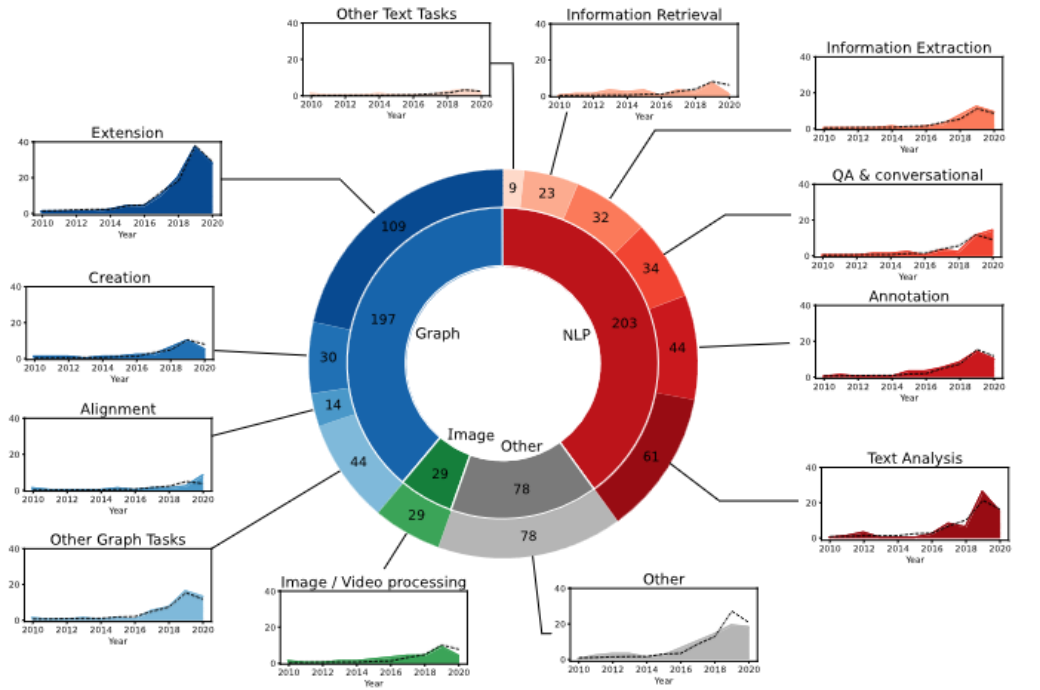


Fig. 14. Distribution of papers per targeted task. The dashed line represents the expected distribution $E(\mathbf{x}_T) = \frac{N_T}{N} * \mathbf{x}_T$ for task T, where N is the total number of papers published, N_T is the total number of papers published that target task T, and x_T represents the publications per year that target task T.

RQ4 and RQ5

- “ RQ4 Characteristics of the ML module: What ML models are incorporated (e.g., SVM)? Which ML components can be identified (e.g., attention)? What training type(s) is used during the system training phase? ”
- “ RQ5 Characteristics of the SW module: What type of SW structure is used (e.g., taxonomy)? What is the degree of semantic exploitation? What are the size and the formalism of the resources? Does the system integrate semantic processing modules (i.e., KR)? ”

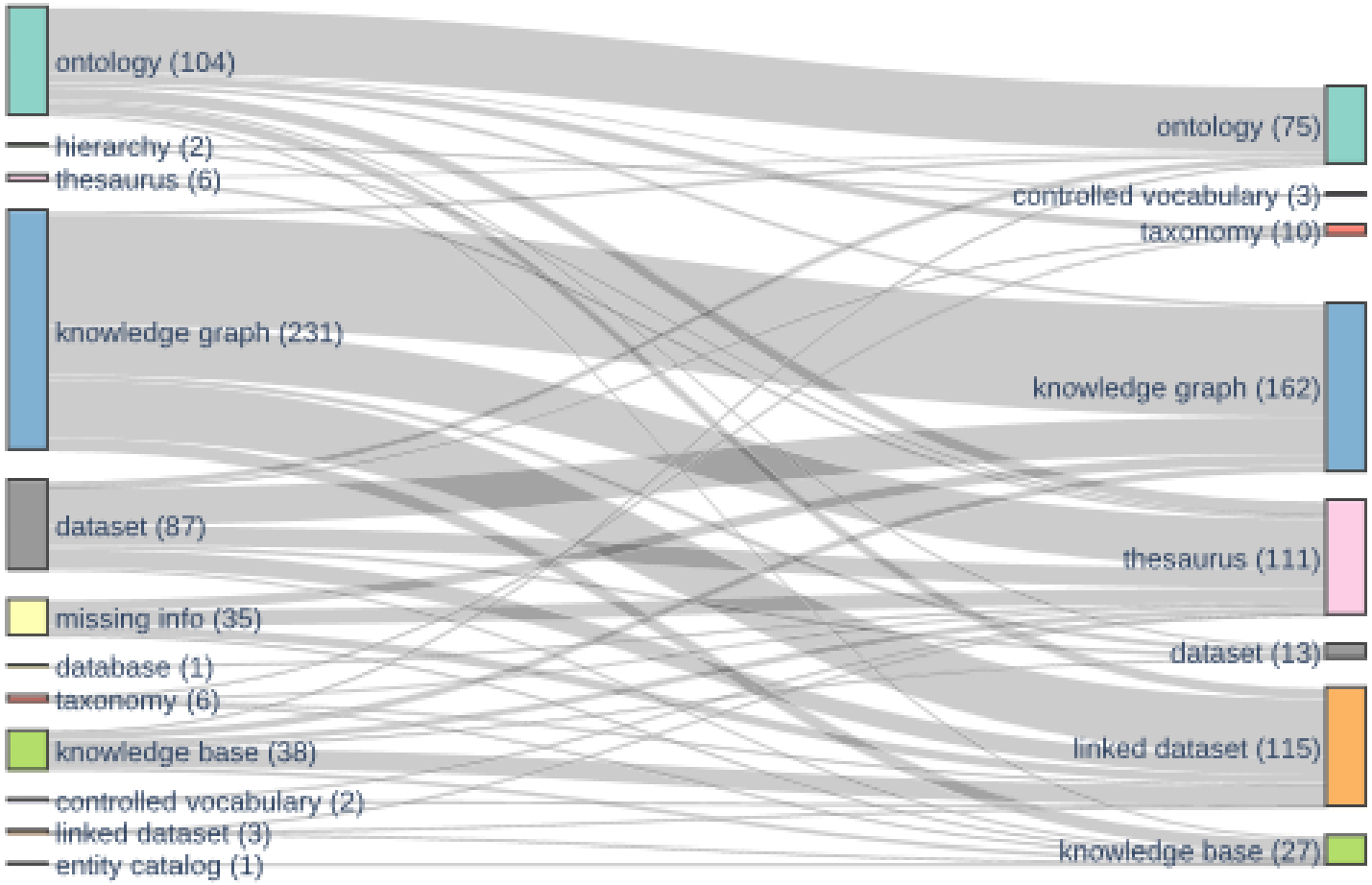


Fig. 26. Comparison of semantic resource types as assigned by authors (left) and the study team (right).

RQ6

“ RQ6 Maturity, transparency, and auditability:
What is the level of maturity of the systems? How transparent are the systems in terms of sharing source code, details of infrastructure, and evaluation setup? Does the system have a provenance-capturing mechanism? ”

Summary Contributions

- Formal definition of SWeML systems
- In-depth analysis of SWeML
- Classification System as an ontology
- Facilitation of documentation

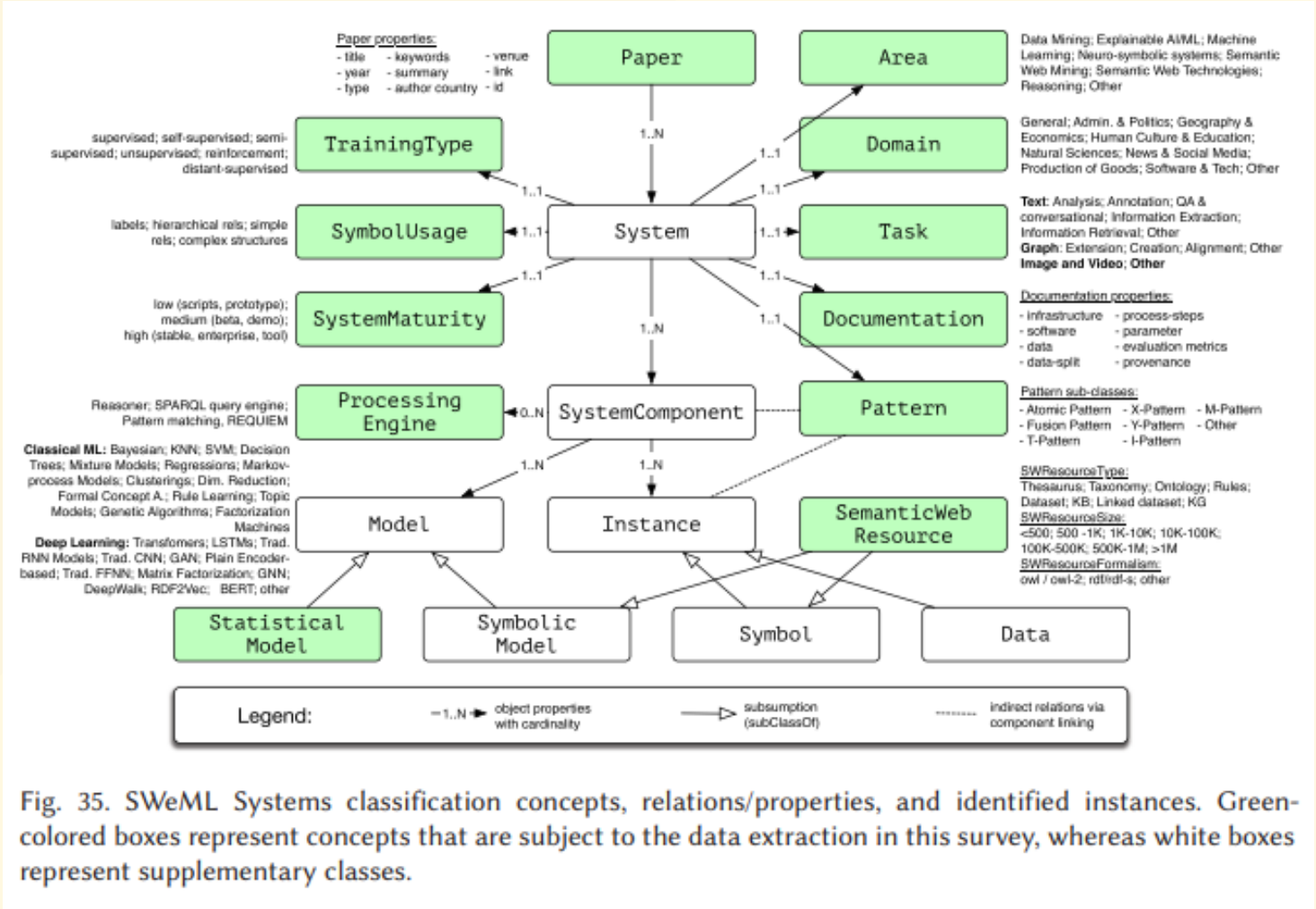


Fig. 35. SWeML Systems classification concepts, relations/properties, and identified instances. Green-colored boxes represent concepts that are subject to the data extraction in this survey, whereas white boxes represent supplementary classes.

Confusions

1. I got bamboozled by the name of this paper.

Interests

1. In AI, there is a demand for model transparency and through documentation we can achieve that to some degree
2. Merging of frameworks for a comprehensive documentation standard