SciModeler: A Metamodel and Graph Database for Consolidating Scientific Knowledge by Linking Empirical Data with Theoretical Constructs

Citation for published version (APA):
Nuijten, R. C. Y., & Van Gorp, P. (2021). SciModeler: A Metamodel and Graph Database for Consolidating Scientific Knowledge by Linking Empirical Data with Theoretical Constructs. In S. Hammoudi, L. F. Pires, E. Seidewitz, & R. Soley (Eds.), *Proceedings of the 9th International Conference on Model-Driven Engineering and Software Development - Volume 1: MODELSWARD* (pp. 314-321). SciTePress Digital Library. https://doi.org/10.5220/0010315503140321

DOI:
10.5220/0010315503140321

Document status and date:
Published: 01/01/2021

Document Version:
Accepted manuscript including changes made at the peer-review stage

SciModeler: a metamodel and graph database for consolidating scientific knowledge by linking empirical data with theoretical constructs

Raoul Nuijten[1] [a] and Pieter Van Gorp[1] [b]
[1] *School of Industrial Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands*
{r.c.y.nuijten, p.m.e.v.gorp}@tue.nl

Keywords: Metamodel, Scientific Method, Theory-Building, Software Tool, Graph Database, Health Behavior Change

Abstract: An important purpose of science is building and advancing general theories from empirical data. This process is complicated by the immense volume of empirical data and scientific theories in some fields. Particularly, the systematic linking of empirical data with theoretical constructs is currently lacking. Within this article, we propose a prototypical solution (i.e., a metamodel and graph database) for consolidating scientific knowledge by linking theoretical constructs with empirical data. We conducted a case study within the field of health behavior change where the system is used to record three scientific theories and three empirical studies as well as their mutual links. Finally, we demonstrate how the system can be queried to accumulate knowledge.

1 INTRODUCTION

Over time, the scientific method has proven to be an efficient strategy for accumulating knowledge. This theory-building process serves to differentiate science from common sense (Reynolds, 1971). The process starts with an inductive phase, in which hypotheses are formed from observations and original theories. These hypotheses are evaluated empirically and then either accepted, or rejected. Subsequently, in a deductive phase, empirical data is interpreted to refine the original theory. Knowledge has accumulated, and the cycle can repeat itself.

Of course, in some domains, many different–but related–theories exist for explaining the same phenomenon. Still, if we assume the existence of a ground truth, these theories will at some point converge to the same equilibrium, if these theories continue to be refined according to the scientific method. However, this may be a rather time-consuming process, since especially the execution of empirical studies and the interpretation of empirical data to refine original theories tends to take quite some time.

The execution of empirical studies seems hard to accelerate, but the refinement of original theories based on empirical data can be advanced by interpreting the results of related empirical studies that were performed by others. Of course, scholars have refined their own theories based on related empirical data for ages. However, the process is highly redundant and scarce resources are often spent inefficiently in disconnected scientific communities. Especially since the volume of literature keeps growing at an increasing rate, there is a need to automate literature reviews. Furthermore, automated reasoning across theories is critical, not only to analyze results beyond one theory, but also to explore opportunities for merging and simplifying such theories.

Advances in Natural Language Processing (NLP) and Machine Learning (ML) have enabled the automated construction of semantic models from scientific articles (Tauchert et al., 2020). However, such approaches build models that are relatively close to the terminologies of scientific disciplines. At the same time, critical details regarding the experimental setups underneath empirical studies are often lost in the model-building process. That limits opportunities for reliably combining empirical studies into more generic theories across scientific communities. We do acknowledge that significant results have been achieved already, as evidenced also by commercial tools such as IBM Watson™ Insights for Medical Literature (IML) but to the best of our knowledge, it was not yet studied how one can encode in a transparent knowledge representation: 1) what exactly makes up a scientific theory, and 2) how exactly empirical studies support or refute one or more theories.

We have encountered this general problem in the specific discipline of health behavior change, where a

[a] https://orcid.org/0000-0003-0125-7708
[b] https://orcid.org/0000-0001-5197-3986

staggering number of theories have been developed within communities like behavioral economics and various subbranches of psychology. While promising work is being carried out to standardize on the terminology, the field still struggles with too large taxonomies, with too overlapping constructs (Michie et al., 2014; Davis et al., 2015; Eldredge et al., 2016). The result is that empirical studies are poorly coded, which hampers theory consolidation.

This article presents *SciModeler*, a metamodel and graph database to: 1) encode scientific theories in a semantic structure; 2) record the outcomes of empirical studies and their relation to theoretical constructs; 3) query across theories and empirical data to explore latent relationships; and 4) identify opportunities for simplifying theories via graph transformations.

The next section surveys literature on the key concepts in scientific theories and empirical studies in order to identify which elements need to be covered by the metamodel. The third section details the metamodel that addresses these requirements. In the fourth section we assess the potential value of the system in the field of health behavior change. We conducted a case study where the system was used to record three scientific theories and three empirical studies as well as their mutual links. Subsequently, we demonstrate how the system can be queried to accumulate knowledge. Finally, we discuss the principal results and weaknesses from this exercise, and provide guidelines for future work.

2 REQUIREMENTS

2.1 Components of a scientific theory

A theory comprises a set of abstract statements about reality (Reynolds, 1971). Hence, informal explanations, unfalsifiable statements and ideas are important but they are not scientific theories (Popper, 1959, p. 23). Instead, in a "theoretical system", "theoretical constructs" are introduced "jointly" (i.e., associated to each other) (Hempel, 1952, p. 32), such that a natural phenomenon and its antecedents are explained and their relations can be repeatedly tested and verified. For this study, we have assumed that a scientific *theory* comprises *constructs*, and the *relationships* between these constructs. While some related works proposed to model theories as *claims* within separate models of individual articles (Ciccarese et al., 2008; Clark et al., 2014), we explored a graph-based approach where theoretical elements are modeled centrally and supportive pieces of empirical data are linked to them.

2.2 Components of an empirical study

Several frameworks for developing and reporting empirical studies have proven valuable over time. Particularly, these frameworks have been used to assist the definition of research questions, as well as systematic literature surveys (Cooke et al., 2012). Especially the PICO framework is commonly used in evidence-based practice (e.g., in Evidence-Based Medicine). This framework suggests that a well-defined empirical study comprises: 1) a population; 2); an intervention; 3) a comparison; and 4) an outcome. Similar frameworks were coined to be applied to different research fields. For example, the PECO (i.e., Population, Exposure, Comparator, and Outcomes) framework was tailored for environmental, public and occupational health research (Morgan et al., 2018); the SPICE (Setting, Perspective, Intervention, Comparison, and Evaluation) framework was introduced to support qualitative research (Booth, 2006), as well as the SPIDER (Sample, Phenomenon of Interest, Design, Evaluation, Research type) framework (Cooke et al., 2012); and lastly the ECLIPSE (Expectation, Client group, Location, Impact, Professionals, ServicE) framework was introduced to support the field of health management (Wildridge and Bell, 2002). Across these frameworks, one can identify:

The Population refers to the community that is targeted within a study (e.g., Dutch high school students, or older adults at risk of being overweight, etc.). This concept is also referred to as the Patient group, Sample, Perspective, or Client group.

The Setting (or Location & Timing) describes when and where an intervention was evaluated (Booth, 2006).

The Expectation (from ECLIPSE, corresponding to the Outcome from PECO or the Evaluation of SPIDER) is the end point of interest. Once this dependent variable is known, the impact of studies addressing a similar outcome variable can be compared. Note that careful recording of this outcome variable is necessary, as a variable can sometimes be measured in different ways.

The Intervention (or Phenomenon of Interest, Professionals & Service) indicates the object that is studies and that is expected to cause a difference (e.g., the administration of a medical drug).

The Comparison (or PECO's Comparator, or SPIDER's Design) is measured against the interventionOften, the comparator is a different intervention or treatment, or alternatively the absence of an intervention or treatment.

The Impact (from ECLIPSE, corresponding to the Evaluation from SPICE) describes what results the evaluation yielded (Booth, 2006).

The Research type (from SPIDER) captures the study design that was adopted to evaluate the intervention (Cooke et al., 2012).

2.3 Deriving Modeling Requirements

From Section 2.1, one can conclude that theories consist of constructs and relations. While previous formalisms already support the encoding of claims of individual articles, it is worth representing theories as first-class modeling concepts, which can be linked from individual studies. Regarding the coding of empirical studies, various formalisms have already been proposed. However, the systematic linking of empirical data with theoretical constructs is lacking. In order to overcome these limitations, we propose a new metamodel that has two layers: The first layer should support the encoding of scientific theories (ST) and the second layer should support the encoding of empirical studies (ES). While the requirements for each distinct layer follows reasonably simply from Sections 2.1 and 2.2, we learned that especially the linking of the two layers is non-trivial.

For layer *ST*, we identified three information requirements, aimed at representing theories as graphs:

- *ST1* Record the name of the theory,
- *ST2* Record the primitive constructs of the theory,
- *ST3* Record the relations between these constructs.

For layer *ES*, our synthesis of Section 2.2 leads to:

- *ES1* Record the (characteristics of) the study population and study sample (i.e., to whom?)
- *ES2* Record the setting (i.e., place and time) of the study (i.e., where, and when?)
- *ES3* Record the expectation of the study (i.e., why?)
- *ES4* Record the interventions and comparison treatments (i.e., what? what else?)
- *ES5* Record the impact of the interventions and treatments on the study sample (i.e., how well?)
- *ES6* Record the research type (i.e., how?)

Regarding the interlinking of these two layers, one would ultimately like to see how specific elements of an empirical study relate to specific elements of a theory. Regrettably, many empirical studies only label interventions at the aggregate level of theories. From our modeling requirements point of view, we therefore need to support both ways of linking the empirical layer with the theoretical layer. Furthermore, concrete interventions in empirical studies can be coded differently according to one's point of view (even when aiming to minimize subjectivity). We will illustrate this challenging issue by means of a case study in Section 4 but regarding modeling requirements, we conclude here that there is a need to support competing classifications and leave it up to the scientific discourse to decide which classification is the best for a specific purpose.

- *ES→ST1* Record the relation between a theoretical construct and an actual intervention
- *ES→ST2* Record the argumentation for why this relation is appropriate
- *ES→ST3* Record the number of 'votes' for a suggested relation

The metamodel presented in the next section is a first attempt to satisfy all requirements that were identified thus far.

3 *SciModeler*: Metamodel and Tool

3.1 Metamodel

The *SciModeler* metamodel is distributed via Figshare (Nuijten and Van Gorp, 2020c). The colored rectangles in the background demonstrate what particular requirement is fulfilled by the rectangle's enclosed entities, attributes and associations.

The orange rectangle captures the entities, attributes and associations that were necessary to satisfy the requirements at layer *ST*. Particularly, to: 1) record the name of a theoretical framework using the *theory* entity [*ST1*]; 2) record the constructs within a theoretical framework using the *construct* entity [*ST2*]; and record the relations between the constructs of a theoretical framework via the *relation* entity [*ST3*]. The relation entity has a type attribute that can have the values: *has an influence on*, *has a positive influence on*, *has a negative influence on*, *is a component of*, and *is synonym of*.

The blue rectangles depict the entities, attributes and associations that were necessary to satisfy the requirements at layer *ES*.

First, the entities *population*, *sample*, *group*, *individual*, *demographic* and *characteristic* are necessary to record with whom a particular intervention was evaluated [*ES1*]. The *population* entity captures information about the audience that was targeted for a specific study. The *sample* entity records ho many subjects from this population have actually participated in the study. The *group* entity distinguishes the number of participants that were exposed to a specific treatment. The *demographic* entity can be used to collect additional information about these groups on different variables. For example, this entity, its attributes and associations, may be used to record that the average age of a sample was 27. In that scenario, *age* is the dimension of the variable associated to the demographic, the aggregation function of the demographic is *average*, and the value of the demographic is *27*. Note that the actual ages (i.e., recorded as *characteristics*) of the individuals within the sample may nevertheless be undisclosed.

Second, the *context* entity is used to record where and when a study was executed [*ES2*]. For example, a study may be executed at a high school (i.e., location) during the winter of 2018 (i.e., timing).

Third the *experiment* entity records the rationale behind a study [*ES3*]. Particularly, the point of interest, or outcome variable is recorded.

Fourth, the entities *treatment*, *treatment assignment*, *intervention* and *platform* are used to record what treatments were assigned, and how these compare to each other [*ES4*]. The *intervention* entity records particularities that are present within all treatments, whereas the *treatment* entity only records particularities that are unique for a specific treatment. The *platform* entity can be used to emphasize that a set of interventions relies on shared infrastructure. For example, a marketing intervention may be administered via a phone call, and different interventions may use similar infrastructure. As an example from the software engineering domain, the Eclipse framework could be a platform on which an empirical study on plug-in development could be based. Lastly, the entity *treatment assignment* can be used to assign a particular treatment to a group of participants.

Fifth, the entity *outcome* records the impact of a specific *treatment* [*ES5*]. Particularly, by capturing the treatment result and the significance of that result.

Sixth, the entity *source* is used to record the scientific article that describes the research method underpinning one or more experiments [*ES6*].

Finally, the yellow rectangle captures the entities, attributes and associations that are used to map empirical data onto theoretical constructs (i.e., linking layer *ES* and layer *ST*). The *classification* entity can be used to associate (parts of) a particular intervention or treatment with a theoretical construct [*ES*→*ST1*]. Since this step relies on interpretation, an explanation from a reviewer is required [*ES*→*ST2*]. Other reviewers can support a given *classification*, or commit their own [*ES*→*ST3*].

3.2 Recording data

To instantiate the class diagram and store data, we have adopted a graph-based approach. A graph-based approach was chosen, for its flexibility, and extensive coverage of database systems. Particularly, we have used Neo4j v4.1.3 to define the type graph, store example instances and evaluate queries, partly because Neo4j provides extensive tools for visualizing data and query results as an actual graph.

To record a scientific theory, or empirical study, a reviewer would have to examine the original research article presenting the theory or study. After a reviewer examined the article, she can write a set of statements (e.g., using Cypher, Neo4j‘s graph query language) to commit the theory or study to the database.

For documenting a scientific theory this exercise is generally relatively easy, as these theories are often already visualized as graphs with *constructs* and *relations*. Nevertheless, the exercise of extracting the correct information from an article presenting an empirical study may be somewhat more challenging, as the data is often presented in a mere text-based form. In order to reliably extract empirical data, we have established a workflow in which a reviewer can highlight a particular passage in the PDF-version of the article that details a certain attribute she wants to record (e.g., the sample size that was studied). These quotations are also recorded in the database such that the source of a piece of empirical data can easily be traced back to the original article. Additionally, the data sets that were obtained within an empirical study are typically not shared at the individual participant level. Hence, specific information about the *characteristics* of particular *individuals*, or the *impact* the intervention has had on a particular *individual* are mostly not revealed in scientific outlets. Note that therefore, the entities, attributes and associations that are displayed below the red dotted line in the class diagram (Nuijten and Van Gorp, 2020c) are included for completeness, but are known to be difficult to extract from most research articles. Then again, future articles on empirical studies may cite *SciModeler* instances as online attachments that document the study setup with greater precision.

4 CASE STUDY

4.1 Health behavior change as context

Particularly in the field of health behavior change many scientific theories are circulating. Still, there is no consensually agreed theory of health behavior change. Moreover, the process of developing a consensually agreed theory seems to be rather inefficient in this field, as empirical studies are originating from unique–but related–theories, without systematically contributing to each other.

The aim of this case study is to demonstrate how *SciModeler* can facilitate the more systematic development of scientific theory on health behavior change, by facilitating the interpretation of empirical data.

4.2 Method

This case study demonstrates the potential impact of our proposed system in the field of health behavior change. We portray how three defying theoretical frameworks within the more generic field of behavior change could be encoded in our system. Additionally, we illustrate how our system could record valuable information from three empirical studies on health behavior change. Subsequently, we discuss how the theoretical frameworks and empirical studies relate to each other, and how these relations could be represented by *SciModeler*. Finally, we highlight how the system could be queried to accumulate knowledge.

4.2.1 Selecting three theoretical frameworks

In the context of behavior change, theories seek to explain why, when and how a behavior does or does not occur, and the important sources of influence to be targeted in order to alter the behavior (Michie et al., 2014). Theories on behavioral change are prevalent: The book "ABC of behaviour change theories" reported 83 behavior change theories (Michie et al., 2014); a scoping review on theories of behavior change identified 82 distinct theories (Davis et al., 2015); and the book "Planning health promotion programs" discussed more than 40 behavior change theories (Eldredge et al., 2016). From these and other sources, we have compiled a list of 103 unique behavior change theories.

In an online survey, we have challenged behavioral scientists to express what theories they typically use in their behavior change initiatives. The survey was completed by 38 scientists who selected: 1) the Self-Determination Theory (Deci and Ryan, 1985, 16 mentions), 2) the COM-B system (Michie et al., 2011, 15 mentions), and 3) the Goal Setting Theory (Locke and Latham, 2002, 14 mentions) as the most useful theories of behavior change.

4.2.2 Selecting three empirical studies

To model three example empirical studies in the field of health behavior change reliably, we drew from our own collection of empirical studies. The examples have quite diverse study setups, demonstrating the expressiveness of *SciModeler* and providing a good basis for illustrating the power of the model as a foundation for query-based information retrieval.

4.3 Results: Proof of concept

In this section, we demonstrate how to record scientific theories, how to record empirical studies, how to map those to theories, and how to query the resulting graphs for extracting information relevant for accumulating knowledge as well as theory building.

4.3.1 Recording a theory

The data (i.e., constructs and relations between these constructs) that was captured for the three selected theories was taken from the descriptions of these theories in their original research articles. This section summarizes the content of each theory and highlights how some particularities for each theory were recorded within *SciModeler*.

The **COM-B System** is a theory that proposes that, in order for a behavior to occur, an individual must have the capability (i.e., physical or psychological) and opportunity (i.e., social or physical) to engage in the behavior, as well as the strength of motivation (i.e., 'reflective' or 'automatic') to engage in it must be greater than for any competing behaviors (Michie et al., 2011). The model emphasizes that components can interact: for example, motivation can be influenced by both opportunity and capability, which can in turn influence behavior. Behavior can then have a feedback influence upon a person's opportunity, motivation and capability to perform the behavior again. Online Supplement A1 (Nuijten and Van Gorp, 2020a) displays how the constructs within the COM-B system relate to each other, and how those relations could be captured within *SciModeler*.

The **Self-Determination Theory** (Deci and Ryan, 1985, SDT) provides a broad framework to study motivation, personality and behaviors. Central to the theory's explanation of behavior is the distinction between intrinsic motivation (i.e., motivation due to inherent interest or enjoyment) and extrinsic motivation (i.e., motivation due to external factors or controls), and people's basic need for autonomy, competence and relatedness (Deci and Ryan, 1985). The SDT is a meta-theory comprised of five mini-theories. The notion that theories can sometimes be composed of other theories can be recorded in our system using the recursive relationship the theory entity has with itself. Online Supplement A2 (Nuijten and Van Gorp, 2020a) shows how *SciModeler* enables to express such relationships between theories and mini-theories.

The **Goal Setting Theory** explains the mechanisms by which goals or intentions influence task performance (Locke and Latham, 2002). The theory's basic premise is that an individual's conscious ideas regulate her behavior (i.e., task performance). Additionally, performance can be moderated by a number of factors including the level of commitment, the importance of the goal, levels of self-efficacy, feedback and task complexity (Locke and Latham, 2002). Furthermore, Locke and Latham model impact of relationships between goals and their impact on satisfaction, as well as how goals act as mediators of incentives. Within the Goal Setting Theory, goal and intention are used as synonyms. The notion of equivalent constructs can be recorded in our graph using a *relation* of type synonym, see Online Supplement A3 (Nuijten and Van Gorp, 2020a).

4.3.2 Recording an empirical study

The information that was captured for the three empirical studies was recorded from the description of these studies in their original research articles (which are cited in the next paragraphs). Specifically, we annotated those articles and used a script to translate the annotation data into Neo4j data import scripts. The related prototypical infrastructure–including the example annotated PDF-files for the graphs in this section–is available online (Nuijten and Van Gorp, 2020b). This section summarizes the content of each study and highlights how some particularities for each study where recorded within *SciModeler*.

Study 1: TVC (Nuijten et al., 2019). This study evaluated two design elements of an mHealth solution – i.e., social proof and tangible rewards – and their impact on user engagement and argued that the introduction of a sufficiently meaningful, unexpected, and customized extrinsic reward can engage participants significantly. During a four-week campaign, a sample of 143 university staff members engaged in a health promotion campaign. Participants were randomly distributed over one of three treatment groups. Online Supplement A4 (Nuijten and Van Gorp, 2020a) displays how this information on the study's *population*, *sample*, and treatment *groups* could be recorded in *SciModeler* (i.e., by means of the purple, violet, and blue nodes). Additionally, the *demographic* information about the *sample* that the study disclosed was recorded as well (i.e., via the pink and grey nodes).

Study 2: VHC (d'Hondt et al., 2019). This study evaluated the impact of personalized motivational messages, as compared to randomized motivational messages, and argued that personalized messages are more appreciated than random messages, but also that personalized messages do not necessarily cause a change in long term behavior. Online Supplement A5 (Nuijten and Van Gorp, 2020a) demonstrates how the general *intervention* (i.e., motivational messages) and the two *treatments* (i.e., personalized messages, as compared to random messages) can be recorded in *SciModeler* (i.e., by means of the red-shaded nodes). Additionally, the *outcomes* are recorded in the yellow nodes, as well as the variables. Finally, the orange-shaded nodes in the top right of Online Supplement A5 (Nuijten and Van Gorp, 2020a) depict the *experimental* point of interest (and the *variable* that was explicitly measured for this purpose), as well as how the intervention was hosted on a particular *platform*.

Study 3: UCGS (Nuijten et al., 2020) This study evaluated social comparison as a driver of engagement with an mHealth application in preadolescents and argued that a team-oriented environment with involvement of a natural role model is more engaging than an individually-focused setting. To draw this conclusion, the authors designed a 12-week crossover experiment, in which they studied three approaches to implementing behavior change via social comparison. Every treatment *group* received their treatments in 2-week periods, and hence received every treatment twice. This advanced study design can be recorded in our graph as depicted in Online Supplement A6 (Nuijten and Van Gorp, 2020a). Particularly, note how treatment *groups* (i.e., blue nodes) are linked to the *treatments* (i.e., red nodes) through instances of the *treatment assignment* entity (i.e., brown nodes). The attribute order number on the entity *treatment assignment* is used to distinguish in what order the *treatments* were assigned to a particular treatment *group*.

4.3.3 Mapping theory and practice

The final exercise was to link (elements of) the interventions and treatments of our empirical studies onto theoretical constructs. We have ourselves coded our studies' interventions and treatments onto four theoretical constructs, see Online Supplement A7 (Nuijten and Van Gorp, 2020a).

4.3.4 Querying to accumulate knowledge

In this section we present three ideas for querying the graph that can be used to advance scientific theories.

First, one may query all interventions and treatments that address a particular theoretical *construct*. Then, one can evaluate the outcomes these interventions and treatments had on the target variables and check whether the theory under investigation would suggest that same outcome. For our case, we may query all interventions that were associated to the construct relatedness, see query 1a of Online Supplement A8 (Nuijten and Van Gorp, 2020a). We then find that there are two interventions associated with this construct, also see Online Supplement A7 (Nuijten and Van Gorp, 2020a). Now we can evaluate whether the outcomes are to be expected according to our theory on relatedness, and we may update our theories accordingly. Note that a user of this system may determine herself what theoretical constructs are interesting to evaluate: she can even jointly evaluate the empirical impact of multiple constructs, if she believes several constructs represent the same meaning, see query 1b of Online Supplement A8 (Nuijten and Van Gorp, 2020a).

Second, one may query all experiments targeting a particular *population*, or *context* to evaluate whether an *outcome* can be replicated within that *population* or *context*, see query 2 of Online Supplement A8 (Nuijten and Van Gorp, 2020a). Alternatively, one may query all *interventions* and *treatments* that address a particular theoretical *construct* (as suggested in the first example), to evaluate whether suggested theoretical *outcomes* also translate to other *populations* and *contexts*.

Third, one may query all experiments that have used the same *platform* to evaluate whether a theoretically suggested relationship is reported consistently with (probably) similar interventions and *treatments*. Using following statement one can find all *interventions* and *treatments* that were hosted using a similar *platform*, see query 3 of Online Supplement A8 (Nuijten and Van Gorp, 2020a).

Lastly, *SciModeler* enables its users to perform automated updates on the graph structures. This paves the way towards graph transformation systems that automatically explore which variations of existing industries support the results of previously coded studies most naturally, in careful consideration of heuristics like Ockham's razor (Hoffmann et al., 1996).

5 CONCLUSION & OUTLOOK

We have demonstrated the potential value of *SciModeler* by means of a case study. Even though the example queries were relatively simple, they could deliver information which would be very hard to obtain reliably when only reasoning about the original manuscripts. We have also suggested that this basic infrastructure paves a way towards automating the simplification and merging of theories. Still, the setup in which *SciModeler* was demonstrated has various limitations, that call for future improvements.

First, populating the *SciModeler* database is relatively cumbersome such that in its current form it would suffer from adoption problems. Hence, we aim to explore the use of ontological languages to ease the process of recording scientific theories. For encoding empirical studies however, we have already proposed a process that uses PDF annotations to extract data. Still we aim to explore how existing scientific tools for data annotation (Van Gorp et al., 2012) can potentially ease this process, as we envision a future where authors submit *SciModeler* data as a direct supplement to their articles. Until then, one may also want to leverage Natural Language Processing techniques for automatically mining *SciModeler* models. Regrettably, these algorithms will also suffer from the fact that many scientific publications are incomplete and ambiguous. Particularly, from a peer review of 313 research studies it was observed that over half (54%) of the studies did not report on the four PICO components (Thabane et al., 2009). Regardless of whether studies are labeled by their original authors, by another scientist, or by an artificially intelligent agent, one may want to collect community feedback on the quality of a *SciModeler* model. We have anticipated that by allowing users to review and 'up-vote' each other's *classifications* of experimental *interventions* and *treatments* as theoretical *constructs*. Future revisions should support that at the level of other entities and attributes too, such that the truthfulness of a particular attribute value can be measured by the degree to which reviewers agree on the information.

A second limitation is that we do not yet provide an interface for querying the graph, and for 'up-voting' specific classifications. To also allow possible non-expert end users to use the system, we plan to provide an interface, for instance with a set of default queries.

A third limitation is that we do not yet share a substantial database of *SciModeler* models. We did already invest significant efforts in the coding of 37 empirical studies on health behavior change. In fact, those efforts were based on more primitive scientific tools such as online spreadsheets and ultimately they have driven us to the development of *SciModeler*. Our aim is to revisit that initial exercise and demonstrate to the behavior change community how the model-based approach reported here can be used to develop a more unified theory for that field. At the same time, we aim to validate the current metamodel with other researchers, especially from the field of health behavior change. This may yield improvement directions for our metamodel. For example, researchers may express the need to actually discuss *classifications*, instead of only being able to 'up-vote' them.

Finally, at the level of the *SciModeler* metamodel, future work is to decompose the text-based node attributes into more fine-grained sub-graph structures. That would for example enable the query-based retrieval of studies that are recorded within the context of a high-school, with a duration of at least eight weeks per intervention. Until then, the Neo4j's query language fortunately offers support for regular expressions on node attribute values.

REFERENCES

Booth, A. (2006). Clear and present questions: formulating questions for evidence based practice. *Library hi tech*.

Ciccarese, P., Wu, E., Wong, G., Ocana, M., Kinoshita, J., Ruttenberg, A., and Clark, T. (2008). The SWAN biomedical discourse ontology. *Journal of Biomedical Informatics*, 41(5):739 – 751. Semantic Mashup of Biomedical Data.

Clark, T., Ciccarese, P. N., and Goble, C. A. (2014). Micropublications: a semantic model for claims, evidence, arguments and annotations in biomedical communications. *Journal of Biomedical Semantics*, 5(1):28.

Cooke, A., Smith, D., and Booth, A. (2012). Beyond pico: the spider tool for qualitative evidence synthesis. *Qualitative health research*, 22(10):1435–1443.

Davis, R., Campbell, R., Hildon, Z., Hobbs, L., and Michie, S. (2015). Theories of behaviour and behaviour change across the social and behavioural sciences: a scoping review. *Health psychology review*, 9(3):323–344.

Deci, E. L. and Ryan, R. M. (1985). Intrinsic motivation and self-determination in human behavior. *New York: Plenum Publishing Co*.

d'Hondt, J. E., Nuijten, R. C., and Van Gorp, P. M. (2019). Evaluation of computer-tailored motivational messaging in a health promotion context. In *International and Interdisciplinary Conference on Modeling and Using Context*, pages 120–133. Springer.

Eldredge, L. K. B., Markham, C. M., Ruiter, R. A., Fernández, M. E., Kok, G., and Parcel, G. S. (2016). *Planning health promotion programs: an intervention mapping approach*. John Wiley & Sons.

Hempel, C. G. (1952). *Fundamentals of concept formation in empirical science*. University of Chicago Press.

Hoffmann, R., Minkin, V., and Carpenter, B. (1996). Ockham's razor and chemistry. *Bulletin de la Societe Chimique de France*, 133.

Locke, E. A. and Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American psychologist*, 57(9):705.

Michie, S., Van Stralen, M. M., and West, R. (2011). The behaviour change wheel: a new method for characterising and designing behaviour change interventions. *Implementation science*, 6(1):42.

Michie, S., West, R., Campbell, R., Brown, J., and Gainforth, H. (2014). *ABC of behaviour change theories*. Silverback publishing.

Morgan, R. L., Whaley, P., Thayer, K. A., and Schünemann, H. J. (2018). Identifying the peco: a framework for formulating good questions to explore the association of environmental and other exposures with health outcomes. *Environment international*, 121(Pt 1):1027.

Nuijten, R. C. and Van Gorp, P. M. (2020a). SciModeler@MODELSWARD: Online Supplements. DOI: 10.6084/m9.figshare.13347239.

Nuijten, R. C. and Van Gorp, P. M. (2020b). SciModeler database v1.0.1. DOI: 10.6084/m9.figshare.13160141.

Nuijten, R. C. and Van Gorp, P. M. (2020c). SciModeler metamodel. DOI: 10.6084/m9.figshare.13347275.

Nuijten, R. C., Van Gorp, P. M., Borghouts, T., Le Blanc, P. M., Van den Berg, P. E., Kemperman, A. D., Hadian Haghighi, E., and Simons, M. (2020). Different implementations of social comparison as drivers for health behavior change: Evaluating engagement levels of preadolescent students with an mhealth intervention. *Journal of Medical Internet Research*. Under Review.

Nuijten, R. C., Van Gorp, P. M., Kaymak, U., Simons, M., Kemperman, A. D., and Van den Berg, P. E. (2019). Evaluation of the impact of extrinsic rewards on user engagement in a health promotion context. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 3600–3604. IEEE.

Popper, K. (1959). *The logic of scientific discovery*. Routledge.

Reynolds, P. (1971). A primer in theory construction.

Tauchert, C., Bender, M., Mesbah, N., and Buxmann, P. (2020). Towards an integrative approach for automated literature reviews using machine learning. In *53rd Hawaii International Conference on System Sciences, HICSS 2020, Maui, Hawaii, USA, January 7-10, 2020*, pages 1–10. ScholarSpace.

Thabane, L., Thomas, T., Ye, C., and Paul, J. (2009). Posing the research question: not so simple. *Canadian Journal of Anesthesia/Journal canadien d'anesthésie*, 56(1):71.

Van Gorp, P., Vanderfeesten, I., Dalinghaus, W., Mengerink, J., van der Sanden, B., and Kubben, P. (2012). Towards generic mde support for extracting purpose-specific healthcare models from annotated, unstructured texts. In *International Symposium on Foundations of Health Informatics Engineering and Systems*, pages 213–221. Springer.

Wildridge, V. and Bell, L. (2002). How clip became eclipse: a mnemonic to assist in searching for health policy/management information. *Health Information & Libraries Journal*, 19(2):113–115.