DOI:10.1145/1810891.1810893

More Than One Way to Annotate Metadata

THE ARTICLE “MANAGING Scientific Data” by Anastasia Ailamaki, Verena Kantere, and Debabrata Dash (June 2010) explained that data generated by research projects is valuable only when annotated with metadata describing the data’s provenance, context, and meaning. However, a given data item can be annotated in more ways than one, for two reasons:

Provenance. A multidisciplinary project can track its progress with basic metadata indicating the provenance of its samples and their associated data. Each data item can also be annotated in a more detailed way through tools particular to the technique used to generate the data item; these annotations are themselves interpretable by people (and software) in the relevant discipline; and

Assumptions. By definition, a research field involves a basic set of concepts used to understand the field but that is not yet agreed upon. Annotations beyond where and when the data was recorded incorporate assumptions that may be contentious among experts.

Data storage and metadata should thus be decoupled. A data repository must be capable of returning any data item stored within it, along with a list of places needed to find the relevant metadata. A metadata repository must be capable of identifying the schema it adheres to and respond to queries about specific data items with relevant annotations.

Decoupling the architecture this way eases development of an ecosystem of repositories and annotation schemas.

Chris Morris, Warrington, U.K.

Authors’ Response:

Separate (and multiple) metadata stores are indeed essential for a number of scientific applications and should be available to user scientists as an option. However, because data queries likely need to combine information held in separate metadata stores, processing them requires appropriate mechanisms for distribution, access control,

and the merging and branching of the stores.

Anastasia Ailamaki, Verena Kantere, Debabrata Dash, Lausanne, Switzerland

No Straw Man in Empirical Research

In his Viewpoint “Is Computer Science Truly Scientific?” (July 2010), Gonzalo Génova would have made a stronger case if he used the words “theoretical” or “conceptual” instead of “speculative” to support his argument against the excessively empirical orientation of much of today’s CS research. The life cycle of scientific ideas generally progresses from the speculative phase in which many candidate ideas are pursued, with only a few surviving to be presented or published as theoretical contributions, often supported by robust analytical models. Journal editors are unlikely to summarily reject contributions making it to this stage because they provide the conjectures and hypotheses that can be tested through rigorous empirically oriented research.

Génova also set up a straw man when he railed against the excesses of verificationism and empiricism. Who would argue against the proposition that credible scientific advances need good empirical research experiments, simulation, proof-of-concept prototype construction, and surveys? Such research needs models and hypotheses that might have begun as speculative conjectures at an earlier point in time.

Naïve empiricism has no place in CS research. Moreover, purely speculative research without adequate analytical foundations is unlikely to help advance CS (or any other) research.

Joseph G. Davis, Sydney, Australia

Author’s Response:

Davis (“credible scientific advances need good empirical research”) and I (“experimentation without the guide of speculative thinking is worthless”) fundamentally agree. When I said “speculative thinking,” I meant “theoretical contributions supported by robust analytical

models,” not freely dancing ideas without purpose.

There may also be slight disagreement regarding empirical validation, the excesses of which I criticized. It is clear that theories about physical phenomena require empirical validation; theories about mathematical objects do not. Many areas in CS deal with conceptual or information objects more akin to mathematical objects than to their physical counterparts. Therefore, requiring empirical validation is out of place here.

Gonzalo Génova, Madrid, Spain

What CS Academics Think They Teach

Poul-Henning Kamp’s article “You’re Doing It Wrong” (July 2010) would have been considerably more valuable and effective if it had been written more professionally and, more important, avoided gross exaggerations. For example, Kamp said the computer architecture depicted in his Figure 7 “is totally bogus today.” Wrong. Though simplistic, it is entirely appropriate as a first architecture for beginning students, most of whom are unable to provide precise definitions even for words like “input” and “output.” Similarly, Kamp saying “It is the only conceptual model used in computer education” cannot be correct.

Alex Simonelis, Montréal

Author’s Response:

Reacting to the article, CS academics have taken offense, protesting the claimed educational deficiencies, while practitioners have confirmed them. I have seen only two reactions saying “We already learned that.” Students evidently do not learn what CS academics think they teach. But the proof is in the pudding; if graduates say “That’s news to me” when reading the article, then the CS academics are doing it wrong.

Poul-Henning Kamp,
Slagelse, Denmark

Communications welcomes your opinion. To submit a Letter to the Editor, please limit your comments to 500 words or less and send to letters@cacm.acm.org.

© 2010 ACM 0001-0782/10/0900 \$10.00