

NAEP Glossary of Key Terms

Attribute

A characteristic of a log, case, or event. Logs and cases minimally have an identifier attribute, while events must minimally have identifier and timestamp attributes. In NAEP, the ‘PseudoID’ attribute is the case identifier, while ‘BlockCode’, ‘AccessionNumber’, ‘ItemTypeCode’, ‘ObservableType’, ‘ExtendedInfo’ and ‘Timestamp’ are event attributes.

Case

A unique execution of a process. In NAEP, each case corresponds to all the events a single student executed while completing a particular test form.

Event

An electronic record of an action taken in a computer system by an agent while executing a process. In NAEP, events can be executed by students (e.g., clicking on an item option) or by the system itself (e.g., raising a time left alert).

Log

A collection of events recorded by a computer system supporting the execution of a process. In NAEP, logs are created at the form level, but can be shared at the form, block, or even individual item levels, depending at which level the process of interest is defined.

Log file

A serialization, or implementation, of an event log into some file format. In NAEP, logs are originally collected as specially tailored XML files. Logs can be shared using different types of file formats.

Identifier

An attribute that uniquely distinguishes each log, case, or event type. In NAEP, the ‘PseudoID’ attribute is the case-level identifier, as each case corresponds to a single student’s events. The ‘ObservableType’ attribute identifies the type of event executed by the student (e.g. ‘Click Choice’ vs. ‘Horizontal Item Scroll’).

Process

An ordered sequence of observed events, and their related attributes, executed in the pursuit of a well-defined goal, through a computer system. In NAEP, processes are defined at the form level for the purpose of data capture. Each student's log corresponds to the process of completing their assigned test form.

Time stamp

A character encoding of the moment in which an event was executed. In NAEP, time stamps are collected in ISO 8601 format, with millisecond resolution (e.g., '2020-06-01T09:10:20.123Z'). Time stamps are a sensitive attribute, as they can implicitly reveal operational details. Shared, semi-public data may have their time stamps reset to a random starting point, while maintaining their relative distance, or be replaced by cumulative elapsed time, at an arbitrary resolution (seconds or milliseconds).

Trace

The sequence of events effectively executed for any individual process instance, or case. Traces have a randomly variable number of events, as any process can be executed in a variety of ways. In NAEP, the process of completing a Grade 8, Mathematics test form results in traces that are on average 1100 events long, with a standard deviation of 300.

Sources

Mannhardt, F., de Leoni, M., Reijers, H. A., & van der Aalst, W. M. P. (2017). Data-driven process discovery: revealing conditional infrequent behavior from event logs. In E. Dubois & K. Pohl, *Advanced Information Systems Engineering: Proceedings of 29th International Conference, CAiSE 2017* (pp. 545–560). Springer. https://doi.org/10.1007/978-3-319-59536-8_34