

jInfer Runner Module Description

Michal Klempa, Mário Mikula, Robert Smetana, Michal Švirec, Matej Vitásek
Advisors: RNDr. Irena Mlýnková, Ph.D., Martin Nečaský, Ph.D.

Praha, 2011

Target audience: developers willing to extend jInfer, specifically hack (or expand) the inference process.

Responsible developer	Matej Vitásek
Required tokens	<code>cz.cuni.mff.ksi.jinfer.base.interfaces.inference.IGenerator</code> <code>cz.cuni.mff.ksi.jinfer.base.interfaces.inference.SchemaGenerator</code> <code>cz.cuni.mff.ksi.jinfer.base.interfaces.inference.Simplifier</code> <code>org.openide.windows.IOProvider</code>
Provided tokens	none
Module dependencies	Base
Public packages	<code>cz.cuni.mff.ksi.jinfer.runner</code> <code>cz.cuni.mff.ksi.jinfer.runner.properties</code>

1 Introduction

Runner is the module responsible for inference process. The fact that the inference consists of exactly 3 steps as described in [KMS⁺a], is hard-wired in jInfer via this module.

2 Structure and operation

The main class responsible for the inference run is the `Runner` in `cz.cuni.mff.ksi.jinfer.runner` package. During its construction it loads the properties of the currently running project (via the `RunningProject` class) to find out which modules are selected for the inference. These modules are looked up and remembered - each new inference run should therefore use a new instance of `runner`.

The only public method in `Runner` is, unsurprisingly, `run()`. This method will start the first step of the inference process by invoking the selected *IGenerator's* `start()` method, as described in [KMS⁺a, section 3.5.1]. Callback methods `finishedIGenerator()`, `finishedSimplifier()` and `finishedSchemaGenerator()` are responsible for invoking the following stages of inference, or in the latter case for presenting the resulting schema to the user and terminating the inference process.

Invocation of every step in the process is encapsulated in a NetBeans *task*: this is the responsibility of `runAsync()` method. First of all, this means that all the work is done in an asynchronous thread independent from the GUI. Second, NBP presents each such task as a progressbar in the bottom right corner of the window, and allows the user to cancel it. Should this happen, the currently running module detects this by checking for `Thread.interrupted()` and responds by throwing an `InterruptedException`. `Runner` catches this exception, terminates the inference and informs the user (`interrupted()` method).

Furthermore, should any unexpected exception occur while running one of the modules, this will get caught in `Runner` again. Inference will be interrupted and user will be notified - this is the responsibility of `unexpected()` method.

Finally, the generated schema is annotated in the end with a comment stating the current date and time, and the modules used in the process. This is the responsibility of `getCommentedSchema()` method.

2.1 Settings

Runner has NetBeans-wide settings determining what should happen after the schema is inferred and which rule displayer should be used. The options panel along with all the logic is in the `cz.cuni.mff.ksi.jinfer.runner.options` package.

2.2 Preferences

Runner naturally has project-wide preferences for selection of inference modules. The panel and its logic is in `cz.cuni.mff.ksi.jinfer.runner.properties` package.

3 Extensibility

For certain inference algorithms it might be necessary to completely change the number or order of inference modules. For example, it might be necessary to have a dynamic inference with arbitrary number of iterations over the same module (modules). In this case, *Runner* can serve as a template: adding e.g. a cleaner between *Simplifier* and *SchemaGenerator* would mean just copy-pasting members and methods currently associated to *Simplifier*.

References

- [Aho96] H. Ahonen. *Generating grammars for structured documents using grammatical inference methods*. PhD thesis, Department of Computer Science, University of Helsinki, Series of Publications A, Report A-1996-4, 1996.
- [BNST06] Geert Jan Bex, Frank Neven, Thomas Schwentick, and Karl Tuyls. Inference of concise dtds from xml data. In *Proceedings of the 32nd international conference on Very large data bases, VLDB '06*, pages 115–126. VLDB Endowment, 2006.
- [BNV07] Geert Jan Bex, Frank Neven, and Stijn Vansummeren. Inferring xml schema definitions from xml data. In *Proceedings of the 33rd international conference on Very large data bases, VLDB '07*, pages 998–1009. VLDB Endowment, 2007.
- [Bou] Ronald Bourret. Dtd parser, version 2.0. <http://www.rpbourret.com/dtdparser/index.htm>.
- [gra] Graph visualization software. <http://www.graphviz.org/>.
- [HMU01] John E. Hopcroft, Rajeev Motwani, and Jeffrey D. Ullman. *Introduction to Automata Theory, Languages, and Computation (2nd Edition)*. Addison-Wesley, 2001.
- [HW07] Yo-Sub Han and Derick Wood. Obtaining shorter regular expressions from finite-state automata. *Theor. Comput. Sci.*, 370(1-3):110–120, 2007.
- [JAX] Java architecture for xml binding. <http://jaxb.java.net/>.
- [jun] Java universal network/graph framework. <http://jung.sourceforge.net/>.
- [KMS⁺a] Michal Klempa, Mário Mikula, Robert Smetana, Michal Švirec, and Matej Vitásek. *jInfer Architecture*.
- [KMS⁺b] Michal Klempa, Mário Mikula, Robert Smetana, Michal Švirec, and Matej Vitásek. *jInfer AutoEditor automaton visualization and editor module*.
- [KMS⁺c] Michal Klempa, Mário Mikula, Robert Smetana, Michal Švirec, and Matej Vitásek. *jInfer Base Module Description*.
- [KMS⁺d] Michal Klempa, Mário Mikula, Robert Smetana, Michal Švirec, and Matej Vitásek. *jInfer BasicDTDExporter Module Description*.
- [KMS⁺e] Michal Klempa, Mário Mikula, Robert Smetana, Michal Švirec, and Matej Vitásek. *jInfer BasicIGG Module Description*.
- [KMS⁺f] Michal Klempa, Mário Mikula, Robert Smetana, Michal Švirec, and Matej Vitásek. *jInfer BasicRuleDisplayer Module Description*.
- [KMS⁺g] Michal Klempa, Mário Mikula, Robert Smetana, Michal Švirec, and Matej Vitásek. *jinfer javadoc*. <http://jinfer.sourceforge.net/javadoc>.
- [KMS⁺h] Michal Klempa, Mário Mikula, Robert Smetana, Michal Švirec, and Matej Vitásek. *jInfer TwoStep simplifier design and implementation*.
- [log] Apache log4jTM. <http://logging.apache.org/log4j/>.
- [loo] org.openide.util.class lookup. <http://bits.netbeans.org/dev/javadoc/org-openide-modules/org-openide/modules/doc-files/api.html>.
- [mod] Module system api. <http://bits.netbeans.org/dev/javadoc/org-openide-modules/org-openide/modules/doc-files/api.html>.
- [net] Todo. <http://netbeans.org>.
- [Nor] Theodore Norvell. A short introduction to regular expressions and context free grammars. <http://www.engr.mun.ca/~theo/Courses/fm/pub/context-free.pdf>.
- [pro] Project sample tutorial. <http://platform.netbeans.org/tutorials/nbm-projectsamples.html>.

- [VMP08] Ondřej Vošta, Irena Mlýnková, and Jaroslav Pokorný. Even an ant can create an xsd. In *DASFAA'08: Proceedings of the 13th international conference on Database systems for advanced applications*, pages 35–50, Berlin, Heidelberg, 2008. Springer-Verlag.
- [Vyh] Julie Vyhnanovská. Automatic construction of an xml schema for a given set of xml documents.
- [wik] Regular expression. http://en.wikipedia.org/wiki/Regular_expression.
- [xml] Xml validation api. http://download.oracle.com/docs/cd/E17802_01/webservices/webservices/docs/1.6/api/javax/xml/validation/package-summary.html.