Introduction

Scientific discoveries in the natural sciences are increasingly data-driven and computationally intensive, providing unprecedented data analysis and scientific computing challenges and opportunities. Scientific workflow systems and tools aim at simplifying the process of designing, executing, and maintaining complex "code aggregates", i.e., consisting of, e.g., a mix of simulation, data management, analysis, and visualization steps. Thus, scientific workflows target similar use cases as do scripting languages such as Perl or Python and can be seen as the "glue" for assembling and executing complex computational science experiments from pre-existing tools and programs. In addition, scientific workflow systems can provide further functionality, e.g., to capture the "provenance" (i.e., processing history and data lineage) of workflow outputs, thus allowing scientists to interpret, debug, and reproduce their computational experiments, or to provide fault-tolerance and scalability for well-designed scientific workflows.

In this talk, I will give an overview of this new and active research area, using real-world examples to illustrate the potential as well as current limitations of scientific workflow technology. At the end, I will also provide an overview of a course on scientific workflow management that will start one week after this talk. The course is aimed at both practitioners, i.e., computational scientists, bioinformaticians, etc. who would like to learn more about new (and possibly different) ways to think about their workflow automation tasks, and at computer scientists who are looking for application-oriented research problems that can make a difference for their colleagues in the natural sciences.

Kontakt:
Prof. Dr. Michael Gertz
Im Neuenheimer Feld 348
69120 HEIDELBERG
Tel.: (06221) 54 5710
E-Mail: michael.gertz@informatik.uni-heidelberg.de