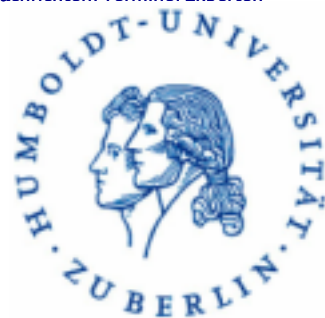




Sonderforschungsbereich FONDA - Einfachere und wiederverwendbare Datenanalyse für die Naturwissenschaften

Neue Methoden zur Analyse großer Datensätze erforscht der kürzlich eingerichtete Sonderforschungsbereich „FONDA – Foundations of Workflows for Large-Scale Scientific Data Analysis“ (Grundlagen von Workflows für die Analyse großer naturwissenschaftlicher Daten).

Solche Daten entstehen bei Experimenten in allen Bereichen der modernen Naturwissenschaften und ihre zeitnahe Analyse erfordert den Einsatz komplexer und schwer programmierbarer rechnergestützter Infrastruktur. Das Hauptziel des Sonderforschungsbereichs ist es, den Aufwand für die Entwicklung solcher Programme zu reduzieren. Daran arbeitet im Sonderforschungsbereich ein höchst interdisziplinäres Team von Forscherinnen und Forschern aus den Bereichen Informatik, Biowissenschaften, Materialwissenschaften und Fernerkundung. Es umfasst alle Berliner Universitäten und eine Reihe weiterer Forschungsinstitute aus Berlin und Brandenburg.

Ein exemplarisches Problem, das FONDA untersuchen wird, sind Workflows für die Analyse sehr großer Satellitenbildsätze. Prof. Patrick Hostert, Experte für Fernerkundung an der Humboldt-Universität zu Berlin, erklärt:

„Moderne Satelliten erfassen kontinuierlich rund um den Erdball hochauflösende Bilder. Wir analysieren umfangreiche Serien solcher Bilder, um beispielsweise Brennpunkte der Entwaldung oder Wüstenbildung auf der ganzen Welt aufzuspüren. Dazu programmieren wir unter Verwendung von Methoden des maschinellen Lernens komplexe Workflows, die aus mehreren Schritten der Bildvorverarbeitung, Registrierung, Filterung und Klassifizierung bestehen.“

Für größere geografische Regionen wie etwa Länder oder Kontinente ist die Ausführung dieser Workflows jedoch nur auf großen Rechenclustern realisierbar, was den Code zusätzlich komplexer macht. Gemeinsam mit Prof. Ulf Leser, ebenfalls von der Humboldt-Universität zu Berlin, erforscht Patrick Hostert, wie sich die Komplexität bei der Programmierung dieser Workflows reduzieren lässt. Prof. Leser, der auch Sprecher von FONDA ist:

„Ein einzigartiges Merkmal von FONDA ist dieser Fokus auf die Verkürzung von Entwicklungszeiten. Wir haben beobachtet, dass Wissenschaftler oft Wochen oder Monate benötigen, um nur einen für eine einzelne Maschine vorgesehenen Workflow so anzupassen, dass er auch auf einem Compute-Cluster ausführbar ist. Dies ist ein viel stärkeres Hemmnis für wissenschaftliche Fortschritte als die eigentliche Laufzeit der Workflows.“

In einem weiteren Projekt untersucht Prof. Kerstin Ritter von der Charité Berlin verwandte Probleme bei der biomedizinischen Bildanalyse. Sie erklärt:

„Wir arbeiten an der Vorhersage der Alzheimer-Krankheit anhand von Hirnscans, einem hochgradig explorativen Forschungsgebiet, das interaktive Methoden zur Datenanalyse erfordert – während wir gleichzeitig sehr große Bildsammlungen für die Ausbildung moderner Methoden des maschinellen Lernens berücksichtigen müssen. Dies erfordert derzeit viel zeitaufwendiges Programmieren auf niedrigem Niveau.“

Gemeinsam mit Dr. Dagmar Kainmüller vom MDC Berlin verfolgt sie innerhalb von FONDA das Ziel, eine neuartige, intuitive Programmiersprache zur Spezifikation solcher Bildanalyse-Workflows zu entwickeln. Dagmar Kainmüller:

„Unser Traum ist es, medizinischen Beratern oder Forschern, die keine Experten auf dem Gebiet der Bildverarbeitung sind, zu ermöglichen, einen Workflow einfach und interaktiv an ihre Daten und ihre Bedürfnisse anzupassen, während er bereits auf einem großen Compute-Cluster läuft.“

Solche Probleme lassen sich nur durch eine enge Zusammenarbeit von Informatikern und Informatikerinnen sowie Forschern und Forscherinnen aus den Naturwissenschaften angehen. Deshalb besteht die Hälfte der FONDA-Projekte aus solchen Teams. Sie untersuchen nicht nur Probleme der Bildanalyse, sondern auch Workflows für die Genomdatenanalyse oder für die Materialwissenschaft. Die andere Hälfte sind reine Informatikprojekte, die von theoretischen Untersuchungen über Eigenschaften von Workflow-Systemen bis hin zu verteilten Dateisystemen und neuen Scheduling-Algorithmen reichen. Prof. Matthias Weidlich, stellvertretender Sprecher des Sonderforschungsbereichs:

„Eines unserer ehrgeizigsten Forschungsprojekte befasst sich mit der Echtzeitanalyse von hochdichten Messströmen aus Elektronenmikroskopen. Gegenwärtig kann eine solche Analyse nur offline durchgeführt werden, mit einer strikten Trennung von Mess- und Analysephase. Wir zielen darauf ab, neue Methoden zur Beseitigung dieser Einschränkung zu entwickeln, die es erlauben würden, den Verlauf einer Messung auf der Grundlage laufender Beobachtungen in Echtzeit anzupassen.“

Dieses Projekt leiten der Physiker Prof. Christoph Koch und Prof. Peter Eisert, ein Informatiker. Die Lösung solcher anspruchsvoller Probleme erfordert den Input vieler weiterer Experten und Expertinnen, wie z.B. Prof. Volker Markl von der Technischen Universität Berlin, der ein Experte für Workflows über Streaming-Daten ist, oder Prof. Tilmann Rabl vom Hasso-Plattner-Institut der Universität Potsdam, Fachmann für verteilte Systeme.

Berlin ist der ideale Ort für solche Forschungen. Die hohe Dichte an Universitäten und Forschungsinstituten, die alle wissenschaftlichen Disziplinen umfassen, bildet die Grundlage, auf der Kooperationsprojekte wie FONDA möglich werden. Die Bündelung dieser Stärken steht auch im Mittelpunkt der Berlin University Alliance (BUA), deren Mitglieder sich alle an FONDA beteiligen.

„Wir betrachten FONDA mit seinem integrativen Ansatz der institutionen- und disziplinübergreifenden Forschung, der nahtlosen gemeinsamen Nutzung von Ressourcen und der Fokussierung auf hochkarätige, reproduzierbare Wissenschaft fast schon als Blaupause für die BUA“,

sagt Prof. Peter Frensch, Vizepräsident für Forschung der Humboldt-Universität zu Berlin.

„Natürlich stellen wir uns vor, dass der Erfolg von FONDA auch andere hochschulübergreifende Initiativen fördern wird.“

contact for scientific information:

Prof. Dr. Ulf Leser
Institut für Informatik
Humboldt-Universität zu Berlin
<https://informatik.hu-berlin.de/wbi>

URL for press release: <https://fonda.hu-berlin.de/>

D