

Geballte Kompetenz

Integration und verteilte Nutzung von entfernten Diensten zur Prozessmodellierung



Christian Kubczak, Tiziana Margaria,
Bernhard Steffen, Petra Grünewald

Als Teil der Lebenswissenschaftlichen Innovationsplattform Dortmund ist das 2006 errichtete Zentrum für Angewandte Proteomik (ZAP) auf die Entwicklung neuer Technologien im Bereich Quantitative Proteomics, Difference in Gel Electrophoreses (DIGE), Glyko-proteomics, Proteinbiochips, Biostatistik und Bioinformatik spezialisiert.

Autoren: Dipl.-Inform. Chr. Kubczak, Lehrstuhl 14 für Software Engineering, Universität Dortmund; Prof. Dr. T. Margaria, Lehrstuhl für Service und Software Engineering, Universität Potsdam; Prof. Dr. B. Steffen, Lehrstuhl 5 für Programmiersysteme, Universität Dortmund; Dr. P. Grünewald, BioMedizinZentrum Dortmund, Dortmund

Die mithilfe von EU- und NRW-Fördergeldern realisierte Innovationsplattform in Dortmund umfasst neben dem Zentrum für Angewandte Proteomik (ZAP) auch die Zentren für Angewandte Chemische Genomik (ZACG) und für Systembiologie (ZfS). Koordiniert wird das ZAP durch das BioMedizinZentrum Dortmund (BMZ); die wissenschaftliche Expertise liegt bei den Projektpartnern Medizinisches Proteom-Center (MPC) der Universität Bochum und der Universität Dortmund. Innerhalb des

Bausteine stellen nützliche Funktionalitäten zur Modifikation der Kontrollstruktur eines Prozesses bereit

ZAP wird zur kooperativen und interdisziplinären Arbeit zwischen den Projektpartnern eine integrative Service-Plattform eingesetzt, die den wissenschaftlichen Austausch der verschiedenen technologischen Ansätze unterstützen soll. Zu diesem Zweck findet die im Rahmen des ZAP entwickelte Bio-jETI-Plattform Anwendung. Bio-jETI (Java Electronic Tool Integration) erlaubt die Integration und verteilte Nutzung von entfernten Diensten ist und nutzt intern das jABC (Java Agent Building Center), eine Software-Umgebung zur Prozessmodellierung nach Service-orientierten Kriterien.

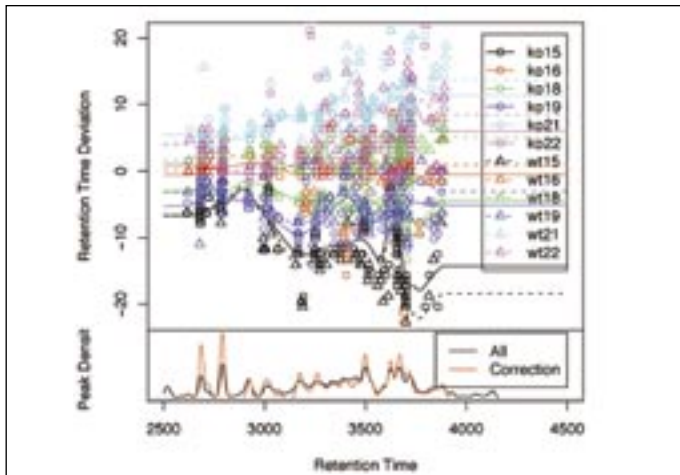
Für die auf der Biotechnica 2007 vorgestellte Forschungsarbeit wird auf eine anwendungstypische LC/MS Analyse Bezug genommen, die in einer Veröffentlichung von C. Smith im Detail erläutert wird. Das Hauptanliegen besteht dabei in der Service-orientierten Umsetzung des Analyseprozesses als ein globaler Dienst: Dieser koordiniert/orchestriert seinerseits externe Einzeldienste. Da die vorgestellte Lösung eine Fallstudie zur Service-orientierten Modellierung im Rahmen der Projektarbeit am ZAP darstellt, wird die Analyse auf einer repräsentativen Datenmenge durchgeführt, die den in Zukunft tatsächlich auszuwertenden Daten stark ähnelt. Der Analyse-Algorithmus selbst ist in der in Fachkreisen weit verbreiteten statistischen Auswertungssprache GNU R verfasst, die es in der Regel Statistikern sehr viel einfacher ermöglicht, komplexe Analysen durchzuführen.

Fallstudie: LC/MS Analyse Prozess

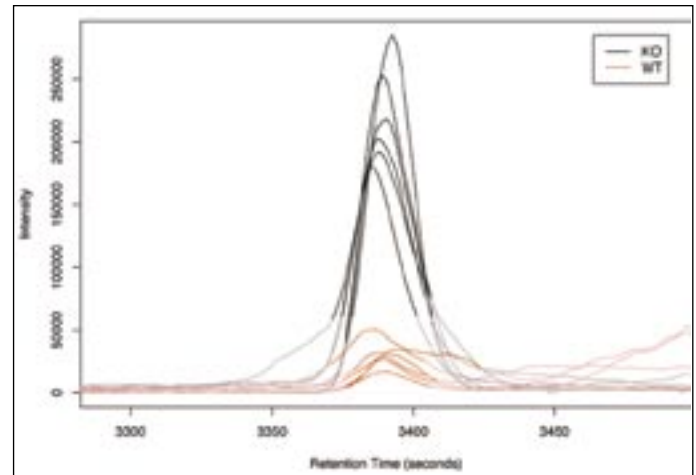
Trotz der Tatsache, dass es für die analytischen Einzelprobleme der vorgestellten Fallstudie bereits fertige Lösungen gibt, existiert kein kohärenter und automatisierter Ansatz für den gesamten Prozessablauf. Als Konsequenz verbringen Forscher substanzielle Zeit mit der manuellen Ausführung eines solchen Analyseablaufes und sind zudem gezwungen, sich Fähigkeiten, die über ihr reguläres Fachwissen hinausgehen, z. B. Programmierkenntnisse und sta-

tistische Details, anzueignen. In einer Kooperation zwischen Statistikern und Informatikern wird anhand dieser Fallstudie ein gekapselter Dienst innerhalb einer Service-orientierten Plattform für bioinformatonstechnische Systeme bereitgestellt.

Der Dienst basiert auf dem jABC, eine am Lehrstuhl für Programmiersysteme der Universität Dortmund entwickelten Modellierungs- und Entwurfsumgebung für Services. Zentrale Stärke dieser Plattform ist ihre Fähigkeit, die Lücke zwischen den abstrakten Spezifikationen der Analysen als solche, die in der Regel entsprechendes Fachpersonal mit wenig technischem Hin-



Abweichung der Retentionszeit aufgetragen gegen die Retentionszeit



Ergebnis Chromatogramm

tergrundwissen durchführt, und der eigentlichen Realisierung auf der Ebene der Techniker, zu schließen.

Im vorliegenden Beispiel wird der Analyseprozess zunächst grafisch definiert. Diese Repräsentation dient im weiteren Verlauf als abstraktes Spezifikationsmodell der zu bewältigenden Problemstellung. Der Prozess ist dabei in eine umfangreiche Phase der Datenaufbereitung und der eigentlichen Analysearbeit unterteilt.

Die LC/MS Analyse mit Bio-jETI

Da bereits alle Inhalte der Dienste zur Analyseauswertung als R-Skripte vorhanden sind, liegt der Modellierungsfokus in der Integration dieser verteilten Funktionalitäten sowie in einer korrekten Orchestrierung der Komponenten. Im Bio-jETI Ansatz wird jede Funktionalität innerhalb eines so genannten Service Independent Building Block (SIB) gekapselt. Diese atomaren „Bausteine“ werden benutzt, um den eigentlichen Prozessablauf der LC/MS Analyse in Form eines gerichteten Flussgraphen, genannt Service Logic Graph (SLG), zu modellieren.

Ein SIB kann einzelne Basisfunktionalitäten oder einen kompletten Subdienst beinhalten, was zu einer entsprechenden Makro-Darstellung führt: Dadurch werden bei Bedarf detailliertere bzw. grundlegendere Ausführungsschritte in vorhandenen Teilgraphen verdeckt. Die SLGs sind die Orchestrierungsmodelle des der Bio-jETI Umgebung unterliegenden jABC Framework und werden zur intuitiven Koordination von komplexen Funktionalitäten eingesetzt. Um auf den modellierten Graphen weitergehende Operation durchzuführen wird das jABC durch spezialisierte Komponenten auf Basis eines Plugin-Konzepts erweitert. Die Ausführung eines modellierten Prozesses wird z. B. durch eine Tracer Komponente bereitgestellt. Meist stellen SIBs in einem Modell lokale Funktionalitäten bereit, d. h. Operationen, die auf dem jewei-

gen System durchgeführt werden. Ist ein Zugriff auf entfernte und/oder verteilte Dienste (z. B. Web Services) wie in diesem Beispiel gewünscht, kommt die Bio-jETI Erweiterung zum Einsatz, die die Anbindung an Web- oder Rest-Services bereitstellt.

Bereitstellung der Basisdienste in Bio-jETI

Die auszuführenden atomaren xcms-Methoden innerhalb des globalen Kontextes der Analyse sind in Wahrheit entfernte Instanzen rechenintensiver Sub-Prozesse, die innerhalb von Bio-jETI in Form von Basisdiensten bereitgestellt werden. Diese Dienste werden auf einem entfernten Server ausgeführt und als Bio-jETI-Dienste innerhalb des jABC für Bio-jETI Nutzer zugänglich gemacht, sodass eine hohe Wiederverwendbarkeit gewährleistet wird.

Zu diesem Zweck wird erstens eine Bibliothek von R-Skripten, die die xcms-Methoden aufrufen, implementiert. Der Nutzer findet diese atomaren Dienste nach ihrer jeweiligen Aufgabe gruppiert innerhalb einer Taxonomiedarstellung. Zweitens werden die R-Skripte innerhalb von Shell-Skripten gekapselt und drittens wird den Shell-Skripten als Bio-jETI-Dienst eine zusätzliche Beschreibung zur entfernen, gleichzeitigen sowie verteilten Nutzung angefügt, woraus dann automatisiert ein lokaler Dienst für das jABC erzeugt werden kann.

Nur der dritte Schritt ist Bio-jETI spezifisch: Schritt 1 und 2 werden unabhängig vom Service-orientierten Ansatz benötigt. Konkret wird in Schritt 3 die Funktionalität eines Bio-jETI-Dienstes in Form eines XML-Dokuments beschrieben. Ein Komponentenserver übernimmt anschließend die Aufgabe der automatisierten Generierung einer lokalen SIB Bibliothek für das jABC aus eben diesen Beschreibungen. Eine entsprechende Laufzeitumgebung sorgt für den entfernten Dienstaufwurf über eine bereitgestellte Kommunikationsschnittstelle.

Modellierung des Prozesses im jABC

Nach dem klassischen Programmierparadigma müsste ein Biologe oder Statistiker ein R-Skript schreiben, das – korrekt parametrisiert – die der Anforderung entsprechend zu lösende Aufgabe adäquat erledigt. Insbesondere die Laufzeitkorrektur-Schleife in dem hier gezeigten Beispiel braucht eine Art höhere Kontrollstruktur, um den Anforderungen der im Vorhergehenden beschriebenen LC/MS Analyse gerecht zu werden. Mit dem jABC ist man in der Lage, den von den Projektpartnern spezifizierten Prozessablauf eins-zu-eins grafisch umzusetzen. Ebenso existiert bereits eine Anzahl von (Bio-jETI) SIBs, die elementare Dienste zur Verfügung stellen. Darüber hinaus existiert eine Anzahl vorhandener Zusatz-SIBs, die unabhängig von jeglicher spezifischer Anwendung ist. Sie stellen z. B. nützliche Funktionalitäten zur Modifikation der Kontrollstruktur eines Prozesses bereit.

Die neuartige Entwicklung der Service-Plattform im Rahmen des ZAP wird detailliert auf der Biotechnica vorgestellt. Hier können sich Interessenten und potenzielle Anwender direkt vor Ort über die Möglichkeit der Anwendung in Forschung und Entwicklung informieren. Außerdem werden die Schwesterzentren ZACG und ZIS ebenfalls auf der Messe vertreten sein und über aktuelle Entwicklungen informieren.

Halle 9, Stand B16

BMZ DORTMUND
000

WWW
www.vfmz.de/1500010