

ANVISSERVICE – EINE SERVICEORIENTIERTE ANALYSE- UND VISUALISIERUNGSPLATTFORM ZUR KOPPLUNG AN EINE MEDIZINISCHE DATENZENTRALE

Adl C^{1,2}, Drobics M¹, Rattay F², Schreier G¹

Kurzfassung

Um komplexe Analysen oder Visualisierungen von umfangreichen Datenbeständen, wie sie beispielsweise für medizinische Auswertungen benötigt werden, flexibel durchführen zu können, sind bestehende Standardprogrammiersprachen (z.B. Java, Python) nicht optimal geeignet. Als mögliche Alternative wird in diesem Beitrag das AnVisService beschrieben, welches über eine Webservice-Schnittstelle den großen Funktionsumfang der Statistik-Umgebung „GNU R“ bereitstellt. Das AnVisService kann durch seine serviceorientierte Konzeption einfach an bestehende Datenzentralen gekoppelt werden.

1. Einleitung

Die Datenzentralen webbasierter medizinischer Systeme, in welchen medizinische Daten abgelegt werden, enthalten oft umfangreiche aber unaufbereitete Datenbestände. Um diese Daten für die medizinische Diagnose oder Therapieverbesserung nutzen zu können, ist es notwendig, die relevante Information herauszufiltern. Diese Daten können jedoch oft nicht, oder nur unzureichend, mit jenen Programmiersprachen, in welchen die Systeme selbst entworfen wurden, analysiert oder visualisiert werden. Der Grund dafür ist, dass Standardprogrammiersprachen wie etwa Java oder Python nicht zur Erstellung komplexer Diagramme oder statistischer Analysen entworfen wurden. Es existieren zwar einige erweiternde Programmbibliotheken für Standardprogrammiersprachen, diese decken jedoch vielfach nur enge Anwendungsbereiche (z.B. nur Visualisierung oder nur Data Mining) ab. Die Kombination mehrere solcher Bibliotheken erhöht die Komplexität sehr stark, da meist keine standardisierten Schnittstellen zwischen solchen Erweiterungen verfügbar sind. Komplexere Analysen und Visualisierungen können vielfach nur durchgeführt werden, wenn die Daten in spezielle, für Statistiksoftware (SPSS, SAS, ...) kompatible Formate exportiert werden. Der eigentliche Analyse- bzw. Visualisierungsprozess findet dann *offline* in der Statistiksoftware statt.

Mit dem im Folgenden vorgestellten *AnVisService* können wiederkehrende, komplexe Analyse- und Visualisierungsaufgaben *online* über standardisierte Webserviceschnittstellen durchgeführt werden. Das *AnVisService* spielt dabei die Rolle des Service-Providers im Hintergrund (Service im Sinne der serviceorientierten Architektur (SOA)). Der Endbenutzer bekommt nur das Ergebnis des Analy-

¹ Information Management & eHealth, Austrian Research Centers GmbH – ARC, Graz, Wien

² Institut für Analysis und Scientific Computing, Technische Universität Wien

se- bzw. Visualisierungsprozesses zu Gesicht (z.B. die Visualisierung als Bilddatei oder das Analyseergebnis als Text).

Als Statistiksoftware kommt das quelloffene GNU-Projekt „The R Project for Statistical Computing“ [4] (kurz GNU R¹) zum Einsatz. GNU R als *die* neue, offen und freie „Lingua Franca“ der Datenanalyse bietet durch das eingebaute Paket-System eine einfache Erweiterungsmöglichkeit. Viele derartige Pakete existieren bereits oder können aufgrund der Offenheit des Systems für eigene Zwecke angepasst werden.

Die Nutzung des AnVisService wird in der Praxis meist in zwei Schritten ablaufen:

1. Analysten erstellen GNU R-Skripte mit Analysen und Visualisierungen in ihrer gewohnten GNU R-Umgebung
2. Diese Skripte können mit geringen Adaptionen über das AnVisService in produktiven Systemen online eingesetzt werden.

2. Material und Methoden

Das AnVisService ist ein Webservice im Sinne der serviceorientierten Architektur und kann über folgende standardisierte Webservice-Schnittstellen genutzt werden:

- SOAP²
- Extensible Markup Language Remote Procedure Call (XML-RPC³)
- Hessian⁴

Das in Java implementierte AnVisService dient als Schnittstelle zu GNU R. Die eigentliche Analyse und Visualisierung erfolgt in GNU R, welches durch seinen großen Funktionsumfang und die Erweiterbarkeit durch Eigen- und Fremdpakete dafür prädestiniert ist. Die Kommunikation zwischen AnVisService und GNU R basiert auf Rserve⁵ - einem TCP/IP Server mit zugehöriger Java-Client Bibliothek.

Für die Analyse und Visualisierung kann jede AnVisService-Instanz auf mehrere GNU R-Instanzen (GNU R-Cluster) im lokalen Netzwerk zugreifen. Dadurch kann die Last vieler paralleler Anfragen rasch auf mehrere Rechner verteilt werden.

Die Schnittstelle zum AnVisService ist sehr generisch. Neben den auszuführenden Analyse- oder Visualisierungsbefehlen (GNU R-Skript als GNU R-Funktion) müssen die Daten in einem GNU R kompatiblen Format übergeben werden. Da GNU R sehr flexibel ist, kann unter diversen Formaten gewählt (z.B. Comma Separated Values (CSV)) oder ein eigenes Format definieren werden. Dies ist möglich, da die Daten durch das auszuführende Skript interpretiert werden, welches die ganzen Möglichkeiten von GNU R zur Datenaufbereitung nutzen kann. Das AnVisService gibt die Ergeb-

¹ <http://www.r-project.org>

² <http://www.w3.org/TR/soap12-part1/>

³ <http://www.xml-rpc.com/>

⁴ <http://hessian.caucho.com/>

⁵ <http://www.rforge.net/Rserve/>

nisse der ausgeführten GNU R-Funktion als Byte-Datenstrom an den Aufrufer zurück. Dadurch können beliebige Ergebnisdaten (Bild oder Ergebnisse statistischer Tests) gleich behandelt werden.

Abbildung 1 zeigt, wie das AnVisService in ein konkretes Datenzentralenumfeld eingebunden werden kann. In der Abbildung sieht man, wie ein Benutzer im Webinterface auf einen Link klickt, der zu einem Diagramm (bzw. sonstiger grafischer Visualisierung von Daten) oder einer Seite mit textuellen Analyseergebnissen führt, die im Hintergrund vom AnVisService geliefert werden.

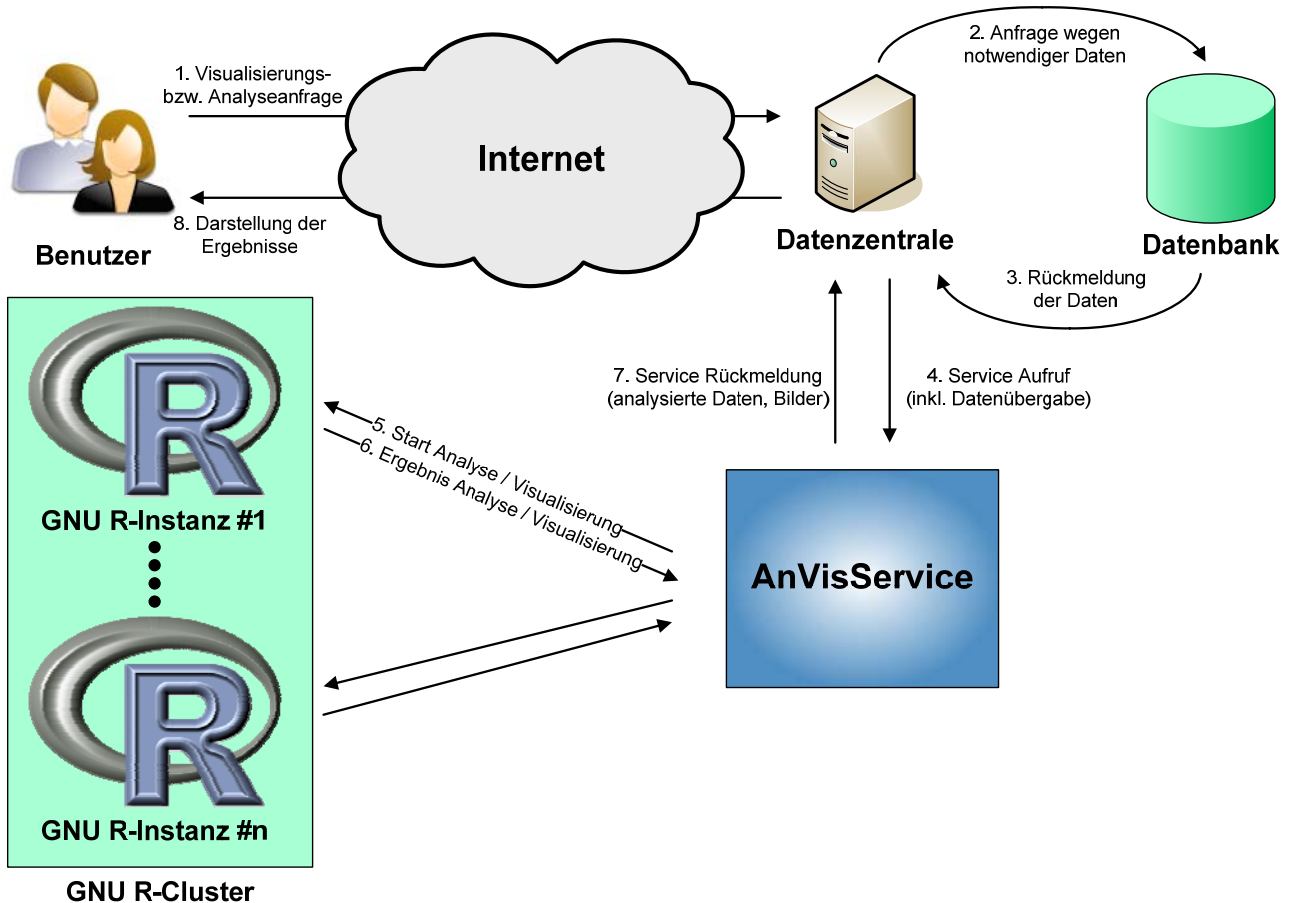


Abbildung 1: Darstellung der Abarbeitung einer über das Internet gestellten Anfrage an die Datenzentrale eines Studien-, Therapiemanagement- oder ähnlichen Systems mit Involvement des gekoppelten AnVisService.

3. Ergebnisse

Das AnVisService ist als prototypische Implementierung fertig gestellt. Für den Test wurde eine zweidimensionale Visualisierung von Blutdruckwerten mit Konfidenzintervallen umgesetzt (siehe Abbildung 2).

Der Testkontext beinhaltet ein Websystem zur Überwachung von Bluthochdruckdaten. Das GNU R-Visualisierungsskript, das beispielweise Abbildung generiert, wurde gemeinsam mit Ärzten entwickelt. Beim Aufruf einer Webseite wird im Hintergrund das AnVisService vom Websystem mit dem fertigen GNU R-Skript und den entsprechenden Patientendaten aufgerufen. Das Ergebnis dieses Aufrufs – das Bild als Byte-Datenstrom – wird vom Webserver an den Browser, der das Bild darstellt, zurückgegeben.

Das prototypische AnVisService-Interface erhält 3 Text-Parameter:

1. GNU R-Skript:
Ein auszuführendes GNU R-Skript (enthält eine oder mehrere Funktionen)
2. Daten:
Daten, auf welchen das GNU R-Skript operiert (Alternativ kann das GNU R-Skript direkt auf Daten über GNU R-Schnittstellen [z.B. Datenbankinterface] zugreifen.)
3. Aufruf:
Aufruf der Hauptfunktion des GNU R-Skripts mit den Daten als Parameter

Im AnVisService werden das GNU R-Skript und die Daten über die „source“ Funktion von GNU R geladen und danach der Aufruf ausgeführt. Das GNU R-Skript muss einen Dateinamen zurückgeben. Der Inhalt der Datei (z.B. eine Visualisierung in einem beliebigen GNU R bekannten Bilddateiformat wie png, pdf, bmp, ...) wird danach an den Aufrufer als Byte-Datenstrom zurückgeliefert. Ebenso ist es möglich Textinformation zurückzuliefern (z.B. Ergebnisse eines komplexen Analyseprozesses).

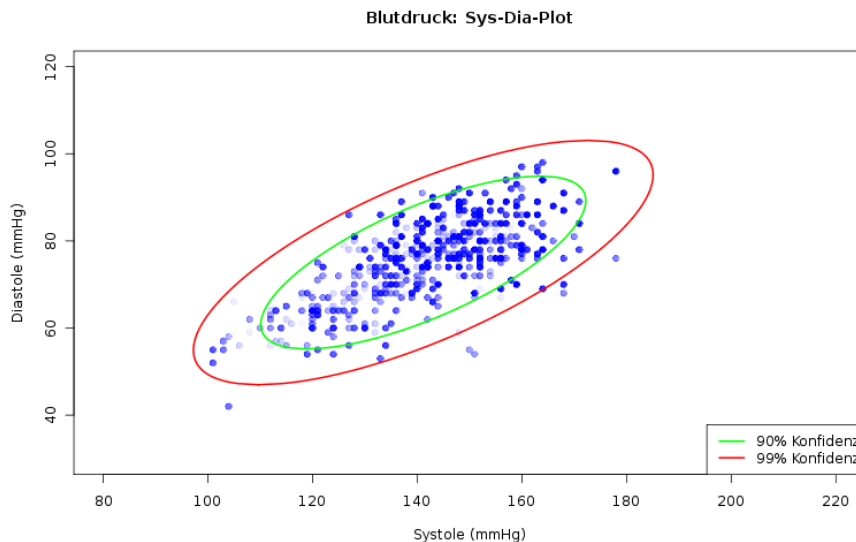


Abbildung 2: Zweidimensionale Darstellung des Blutdrucks mit Visualisierung eines 90% und 99% Konfidenzbereichs.

4. Diskussion

Das AnVisService kann in vielfältiger Weise in medizinische Studiensysteme (wie z.B. SIOPEN-R-NET [6]) oder Therapiemanagementsysteme (wie z.B. H.ELGA [3], DiabMemory 1 [2], Telemedical Compliance Management System (TCMS) [5]) integriert werden. Durch die serviceorientierte Architektur ist eine derartige Kopplung sehr einfach möglich. Komplexe Visualisierungen und Auswertungen werden bei Bedarf direkt im Webinterface eines Systems angestoßen. Die Ergebnisse liefert das AnVisService danach an den Aufrufer zurück, der sie darstellt oder die Darstellung einem Client wie z.B. einem Webbrowser überlässt. Da Anfragen an eine AnVisService-Instanz zur Bearbeitung an viele GNU R-Instanzen verteilt werden, können günstige Standardhardwarekomponenten statt eines teuren Servers mit spezialisierter Serverhardware eingesetzt werden.

Die bisherige prototypische Implementierung im Rahmen einer Masterarbeit [1] soll demnächst in den ersten Projekten produktiv eingesetzt werden. Zusätzlich zu den beschriebenen Features ist auch ein Batch-Modus für länger dauernde Analyse- oder Visualisierungsaufgaben geplant.

5. Danksagungen

Ich möchte mich bei meinen Kollegen Roman Fiedler und Gerhard Aigner für gute Ratschläge und Unterstützung bei der Konzeption, dem Design und der Umsetzung des AnVisService bedanken.

Diese Arbeit wurde teilweise vom Kompetenzzentrum HITT – health information technologies tirol finanziert.

6. Literaturhinweise

[1] ADL C., Konzeption, Design und Entwicklung einer serviceorientierten Analyse- und Visualisierungsplattform zur Integration in ein webbasiertes medizinisches Forschungsnetzwerk, Masterarbeit, in Publikation, Wien, 2009.

[2] KOLLMANN, A., RIED, M., KASTNER, P., SCHREIER, G., LUDVIK, B., Feasibility of a mobile phone based data service for functional insulin treatment of type 1 diabetes mellitus patients. JMIR, 9(5):36, 2007. Impact 2006=2.9.

[3] KOLLMANN, A., HAYN, D., GARCIA, J., TRIGO, J.D., KASTNER, P., ROTMAN, B., TSCHELIESSNIGG, K., SCHREIER G., Feasibility of a telemedicine framework for collaborative pacemaker follow-up. J Telemed Telecare, 13(7):341–347, 2007. Impact 2005 = 0.749.

[4] R DEVELOPMENT CORE TEAM, R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2008. ISBN 3-900051-07-0.

[5] SCHREIER, G., HAYN, D., KASTNER, P., KOLLER, S., SALMHOFER, W., HOFMANN-WELLENHOF, R., A mobile-phone based teledermatology system to support self-management of patients suffering from psoriasis. In Proceedings of the IEEE EMBC; 2008 Aug 20-24; Vancouver, 2008.

[6] SCHREIER, G., MESSMER, J., RAUCHEGGER, G, MODRE-OSPRIAN, R., LADENSTEIN, R., A web-based platform for interdisciplinary biomedical research. Frontiers in Bioscience, 2008. impact 2005 = 2,623.