



Roger Eggenberger



Armin Thommen

GeoKettle

Extending a Data Integration Tool with Geodata

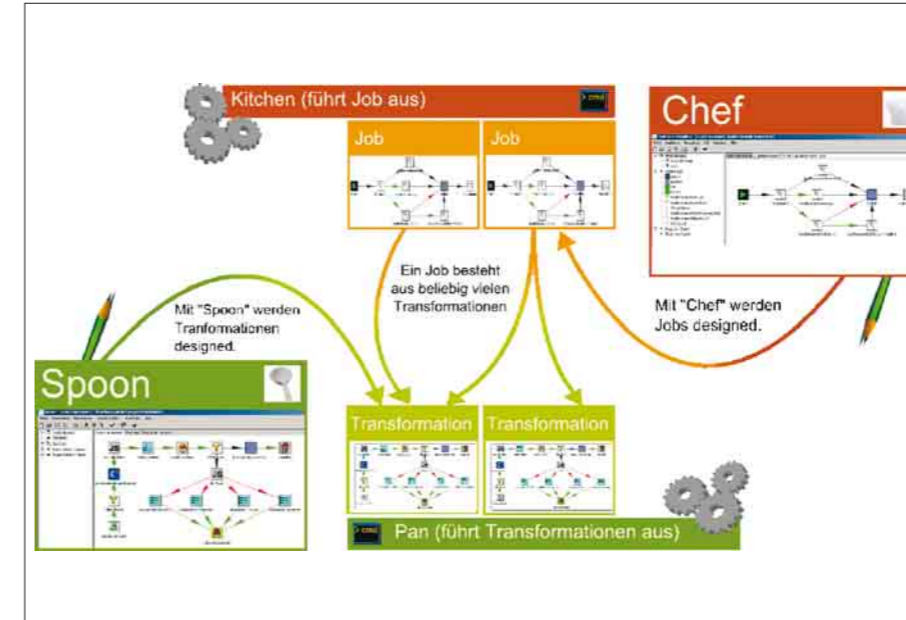
Diplomanden	Roger Eggenberger, Armin Thommen
Examinator	Prof. Stefan F. Keller
Experte	Claude Eisenhut, Eisenhut Informatik AG, Burgdorf
Themengebiet	Software
Projektpartner	Michael Germann, infoGrips GmbH, Zürich



Aufgabenstellung: Die Integration von Daten in Unternehmen, Organisationen und in die öffentliche Verwaltung ist trotz Fortschritten in der Informationstechnologie immer noch ein aufwändiges Unterfangen. In manchen Projekten kann der Aufwand zur Beschaffung von Daten das Budget teilweise bis ganz ausfüllen. Dieses Problem trifft sicher auf Arbeiten mit Geodaten zu, die typischerweise aus verschiedenen Datenquellen stammen und in Geo-Informationssystemen (GIS) verwaltet werden. Dabei wird oft das Ziel verfolgt ein Data Warehouse aufzubauen. Für diese Art der «Enterprise Information Integrati-

on» (EII) sind kommerzielle und Open Source Produkte erhältlich. «Pentaho Kettle», als Java Open Source verfügbar, ist so eine EII-Suite. Komplexe «Extract, Transform und Load»-Vorgänge können mit Kettle auf einfache Weise über eine grafische Benutzeroberfläche definiert werden.

Ziel der Arbeit: Kettle soll im Hinblick auf den produktiven Einsatz in zwei Szenarien untersucht, dokumentiert und mit anderen EII-Produkten verglichen werden. Für die Verwendung im Umfeld von Geodaten soll Kettle um zusätzliche Komponenten (Plugins) erweitert werden,



Aufbau der Datenintegrationssoftware Kettle

die den Zugriff auf ein verbreitetes GIS-Format, bestimmte Daten-Transformationen sowie die Ablage in einer Datenbank ermöglichen.

Lösung: Mit dem Szenario «Bevölkerungsstatistik» wurde ein erstes Data Warehouse realisiert und dokumentiert. Das zweite Szenario «Wanderweg-Routen» erforderte die Entwicklung von drei Kettle-Komponenten mit GIS-Funktionalität. Die erste Komponente kann ESRI Shapefile-Dateien lesen. Die zweite kann beliebige Geometrien in einzelne Punkte zerlegen sowie Linienlängen berechnen. Die dritte Komponente erlaubt es, die gelesenen Geodaten in eine PostgreSQL-Datenbank abzulegen.

Weitere Infos: <http://wiki.hsr.ch/StefanKeller/wiki.cgi?GeoKettle>