

Maven Dependency Tool

Philipp Grulich und Fabian Reiber

Idee

- ▶ Software Projekte haben Abhängigkeiten untereinander
- ▶ Im Java Umfeld wird Maven verwendet
 - ▶ In „Pom“ Dateien werden Abhängigkeiten definiert
- ▶ maven.org als Datenquelle

Konzeption

Web Crawler

- Wget
- Crawler4J
 - 7,2 GB Poms = 1.162.230 Dateien

Mongo DB

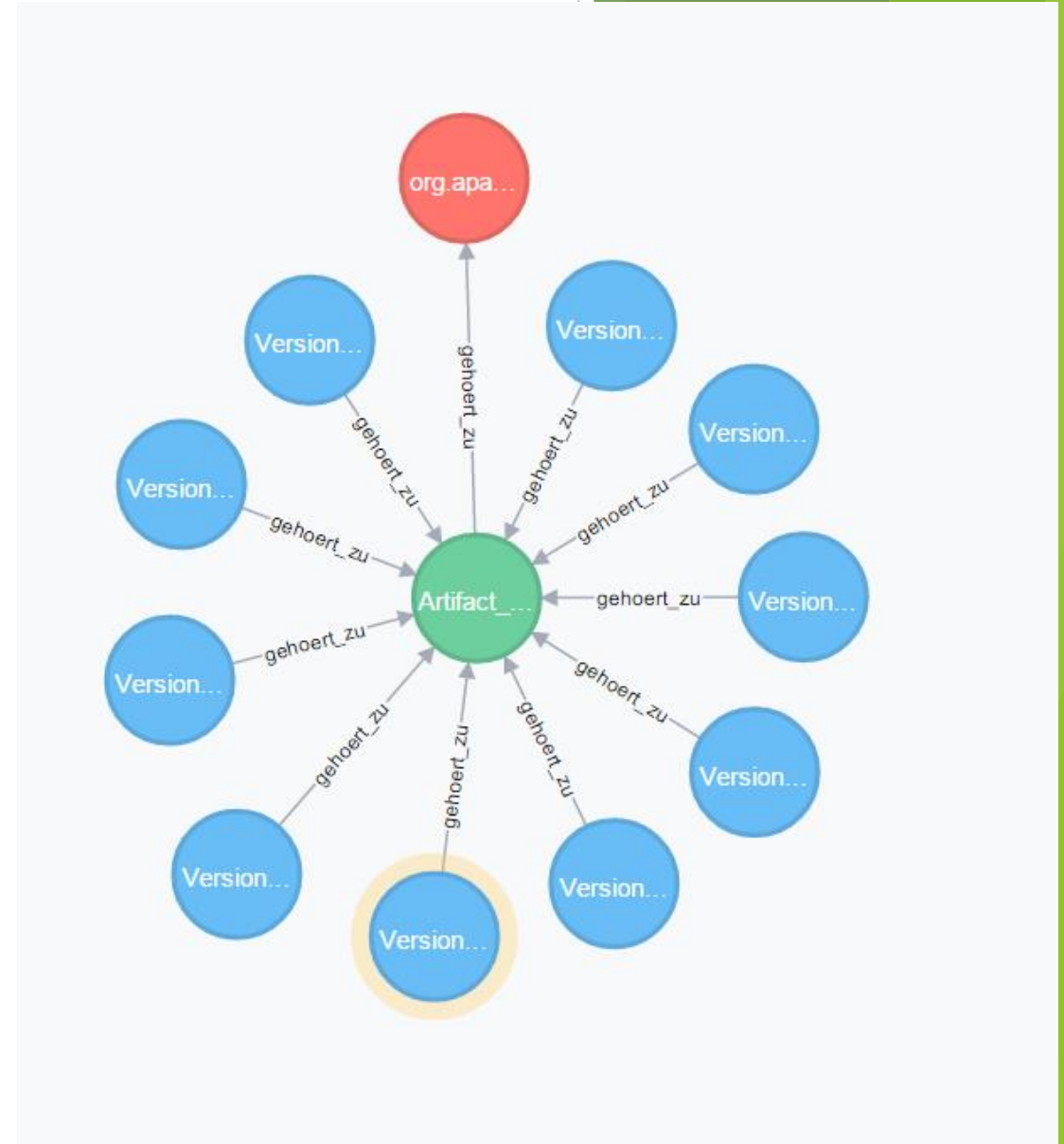
- XMLToJson
- Java Mongo-Driver
 - 1.162.086 Documents
 - 144 konnten nicht Importiert werden -> Fehlerhaftes Format
 - Mongo DB kann keine Punkte in Property-Bezeichnung

Neo4J

- Java Cypher zu langsam
- CSV Import
- Viele Probleme durch Datenqualität

Graph

- ▶ Artifact Knoten: 125.066
- ▶ Group Knoten: 20.887
- ▶ Version Knoten: 1.869.375
- ▶ Anzahl “gehört_zu” Beziehungen: 2.573.136
- ▶ Anzahl “nutzt” Beziehungen: 11.439.449



Get all dependencies of 'notification-amazon 3.6.0'

► `match (n:group) -- (m:artifact {name:'notification-amazon'}) -- (l:version {name:'3.6.0'}) -[:nutzt*]-> (o) --> (b:artifact) -- (h:group) return n,m,l,o,b,h`

► DEMOTIME 😊

Who uses junit?

- ▶ `match (n:group{name:'junit'}) -- (m:artifact {name:'junit'}) -- (l:version)`
Optional Match `l <-[:nutzt]- (o) --> (b:artifact) -- (h:group) return n,m,l,o,b,h`
Limit 100
- ▶ DEMOTIME 😊

Top 10 most used groups

- `match (g:group)--(a:artifact)--(n:version) <-[r:nutzt]- (m:version) return g.id, count(r) as counted order by counted desc limit 10`

Projekt	Count
junit	139701
org.scala-lang	98819
org.slf4j	97279
org.springframework	90930
org.apache.cxf	65443
javax.servlet	49398
org.apache.maven	49190
org.mockito	46280
org.eclipse.jetty	44536
org.apache.geronimo.specs	43020

Top 10 most used artifacts

- `match (a:artifact)--(n:version) <-[r:nutzt]- (m:version) return a.id, count(r) as counted order by counted desc limit 10`

Projekt	Count
Artifact_junit_junit	136213
Artifact_org.scala-lang_scala-library	83457
Artifact_org.slf4j_slf4j-api	49994
Artifact_commons-io_commons-io	30080
Artifact_javax.servlet_servlet-api	29945
Artifact_log4j_log4j	29436
Artifact_org.mockito_mockito-all	29411
Artifact_com.google.api-client_google-api-client	26675
Artifact_ch.qos.logback_logback-classic	25064
Artifact_com.google.guava_guava	24074

Top 10 most used versions

- `match (g:group)--(a:artifact)--(n:version) <-[r:nutzt]- (m:version) return n.id, count(r) as counted order by counted desc limit 10`

Projekt	Count
Version_junit_junit_4.11	36968
Version_junit_junit_4.10	19006
Version_javax.servlet_servlet-api_2.5	18744
Version_javax.inject_javax.inject_1	15395
Version_junit_junit_3.8.1	15278
Version_commons-io_commons-io_2.4	15119
Version_org.mockito_mockito-all_1.9.5	14098
Version_junit_junit_4.12	13638
Version_org.scala-lang_scala-library_2.10.4	13591
Version_org.springframework_spring-context_1	13271

Top 10 worst projects ;)

- `match (g:group)--(a:artifact)--(n:version) -[r:nutzt*..1]-> (m:version) return n.id, count(r) as counted order by counted desc limit 10`

Projekt	Count
Version_io.fabric8_fabric8-project-bom-with-deps_2.2.52	1003
Version_io.fabric8_fabric8-project-bom-full_2.2.73	975
Version_io.fabric8_fabric8-project-bom-full_2.2.72	973
Version_io.fabric8_fabric8-project-bom-full_2.2.70	973
Version_io.fabric8_fabric8-project-bom-full_2.2.71	973
Version_io.fabric8_fabric8-project-bom-full_2.2.69	973
Version_io.fabric8_fabric8-project-bom-full_2.2.68	972
Version_io.fabric8_fabric8-project-bom-full_2.2.67	972
Version_io.fabric8_fabric8-project-bom-with-deps_2.2.54	971
Version_io.fabric8_fabric8-project-bom-with-deps_2.2.64	971

Ausblick

- ▶ Pom-Dateien von GitHub crawlen
- ▶ Verknüpfen mit CVE-Datenbanken

Habt ihr Fragen?