

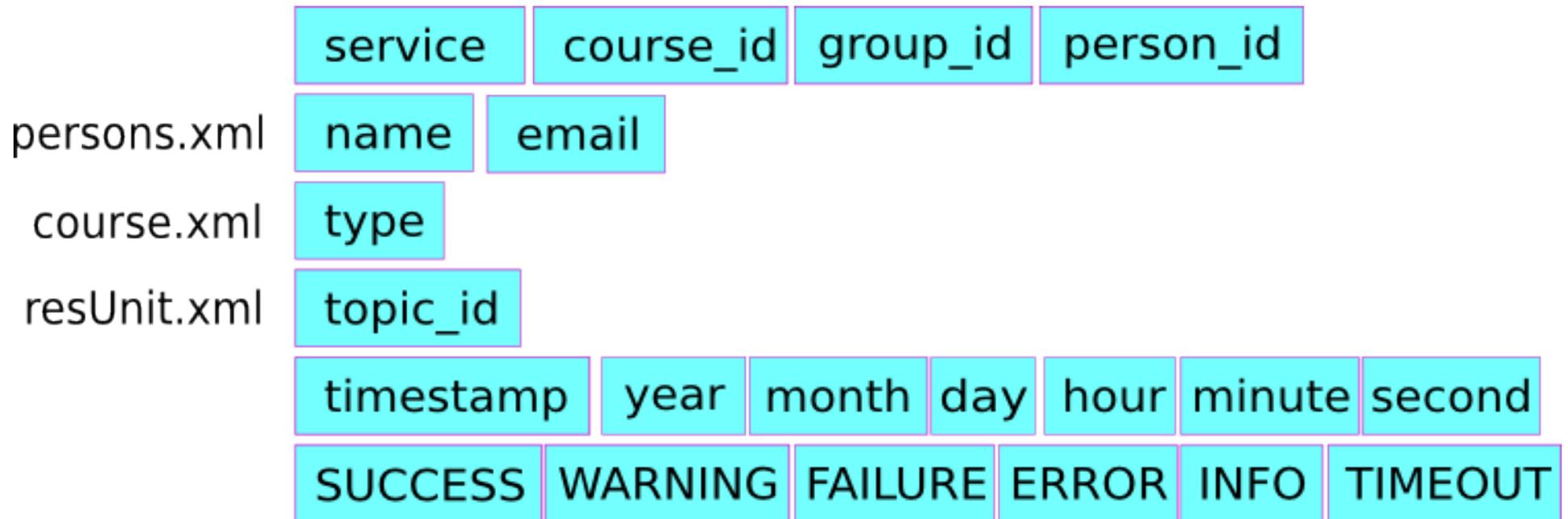
Business Intelligence

Conrad Indiono
Hans-Günter Vetter
Georg Kaes

Table of Contents

1. Service Extraction and Transformation
2. Fragestellungen
3. Foreneinträge schreiben
4. Foreneinträge lesen
5. Auswirkungen auf Performance
6. Weitere Schritte

Code Service Extract/Transform



- unassigned `group_id` (NULL values) in `persons.xml` that occur in `course.xml`
- incomplete `person_id` to `topic_id` assignment in `topics.xml`
- Verify that all `person_id`'s in `resUnit.xml` occur in `persons.xml`.
 - Result: `a710000` and `a712675` NOT in `persons.xml`
 - OK because
 - both not a valid `matrikelnr`
 - both in non-existent group "0"

Forum Service Extract/Transform

forum_persons.csv

service	course_id	group_id	person_id	name	email
---------	-----------	----------	-----------	------	-------

forum_entries.csv

service	course_id	name	fnid	fid	parent_id
date	subject_length	text_length			

forum_unique_users.csv

service	course_id	username
---------	-----------	----------

Users in Forum Entries die nicht registriert sind (persons.xml):

["a0215937","a0704803","a0726599","a0808804","voonly"]

Abgabe Service Extract/Transform

abgabe_person_lecturers.csv

service course_id group_id person_id name email type

abgabe_assessments_results.csv

service course_id user_id result_id result_value

abgabe_assessments_plusses.csv

service course_id user_id plus_date

abgabe_feedback.csv

service course_id user_id task_id subtask_id author

comment_length

Fragestellungen

bestimmter Kurs (ID: 99)

Wie hoch ist der Anteil an Kursteilnehmern (nicht Kursleiter), die etwas ins Forum geschrieben haben?

für alle Kurse, zu denen Foreneinträge existieren:

--> wie viel Prozent der Teilnehmer schreiben bzw. lesen im Forum?

Auswirkungen auf Performance?

Code: unit-test results (success, error, ...)

Länge der Forum-Einträge: Auswirkung auf irgendwas?

Durchschnittliche Länge der Forum-Einträge

Foreneinträge schreiben

- K = Liste aller Kurs-IDs, zu denen es Forumseinträge gibt

Für jedes k in K :

- A_k = alle unique user, die an diesem Kurs teilnehmen (! Kursleiter)
- B_k = alle unique user, die ins entsprechende Forum geschrieben haben (nicht Kursleiter)
- $PWE_k = (B / A) * 100$

Output:

`Kurs-ID; count(Ak); count(Bk); PWEk`

Foreneinträge lesen (1 / 2)

- K = alle Kurse, zu denen Foreneinträge existieren

Für jedes k in K :

- E_all_k = Liste aller Tupel (`kurs, nid, id, text.length`) für k in K
 - (alle Foreneinträge für Kurs k)
- $E_user_read_k$ =
[(`kurs1, nid1, id1, user1`), ..., (`kursn, nidn, idn, usern`)]
 - welcher User hat welchen Eintrag gelesen
- PRE = Prozentsatz der gelesenen Einträge pro Kurs über alle User eines Kurses
 - $PRE_k = \text{count}(E_user_read_k) / (A_k * \text{count}(E_all_k))$

Foreneinträge lesen (2 / 2)

- K = alle Kurse, zu denen Foreneinträge existieren

Für jedes k in K :

- E_all_k = Liste aller Foreneinträge für den Kurs k

AL_k = Durchschnittslänge aller Foreneinträge pro Kurs

$$\circ AL_k = \frac{\text{sum}(E_all_k.\text{text.length})}{\text{count}(E_all_k)}$$

Output:

Kurs-ID; count(A_k); count(B_k); PWE_k ; PRE_k ; AL_k

Auswirkungen auf Performance (1/2)

Code: unit-test results (success, error, failure, ...)

Kurs-ID: 99 (einziger Kurs mit Code + Forum)

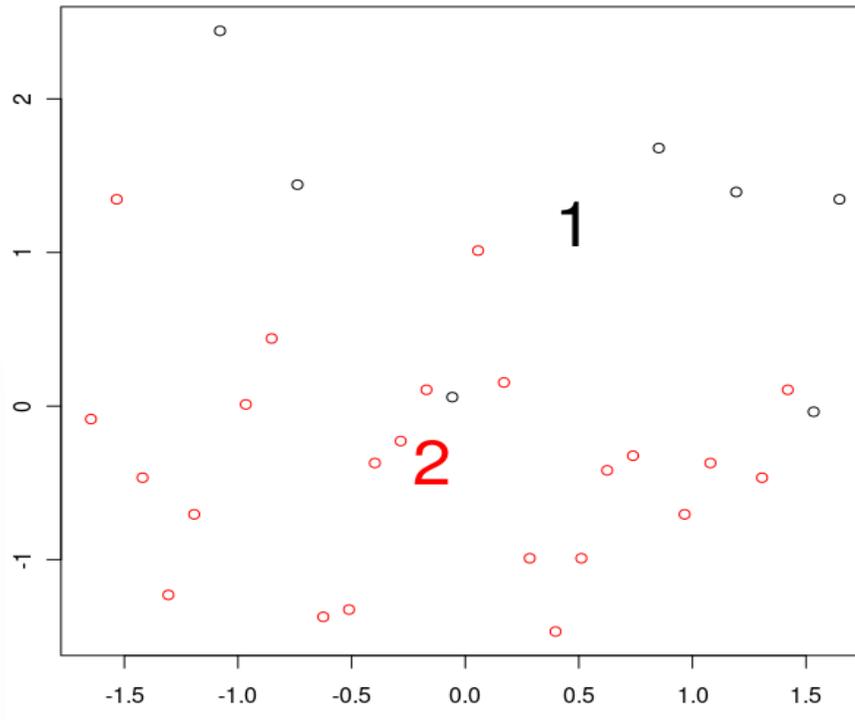
U = alle User im Code-Service die Unit-Tests ausgeführt haben
für alle u in U:

- $\text{SuccessRate}_u = \text{count}(\text{success}_u) / (\text{count}(\text{failure}_u) + \text{count}(\text{error}_u))$
- $\text{ReadByUser}_u = \text{count}(E_user_read_{k=99})$ group by (user)
- $E_written_u = [(kurs_{k=99}, fnid_1, fid_1, user_u), \dots]$
- $\text{WrittenByUser}_u = \text{count}(E_written_u)$

Auswirkungen auf Performance (2/2)

Join von Code und Forum auf User-Ebene:

```
Kurs-ID; User; SuccessRateUseru; ReadByUseru;  
WrittenByUseru
```



Weitere Schritte

Ähnliche Abfragen mit dem Abgabeservice

- Abfragen auch auf Kursebene möglich
- Abgabedatum vs Feedback (=Qualität / Score) / Plusse
mit Datum

Abstand Abgabedatum zu Deadline als int

0 = in letzter Sekunde

-2 = 2 Stunden vor Deadline

+2 = 2 Stunden zu spät