



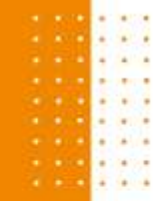
kulturgut scanner
bürogut scanner
Marken der MIK-CENTER GmbH

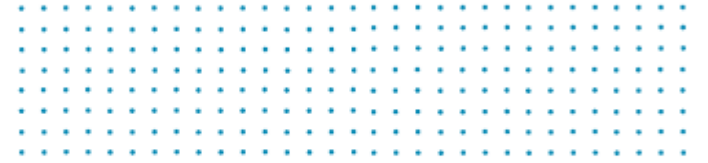


Hibernate Search

Der Weg zur Framework-Lösung

MIK-CENTER GmbH, Berlin
repräsentiert durch Matthias Ronge





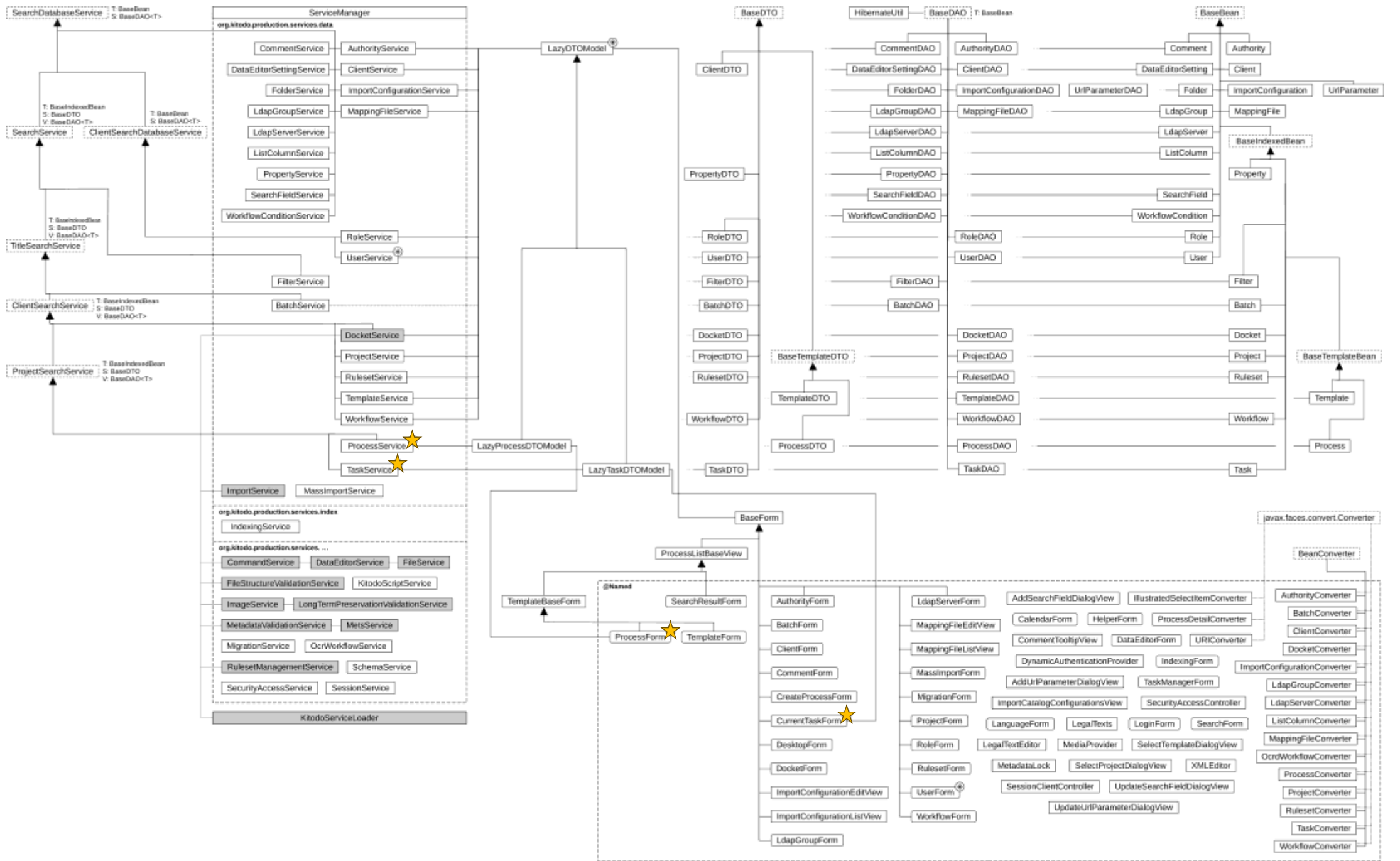
Probleme?

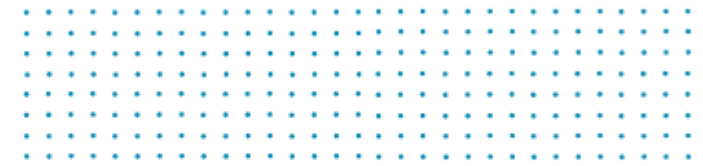
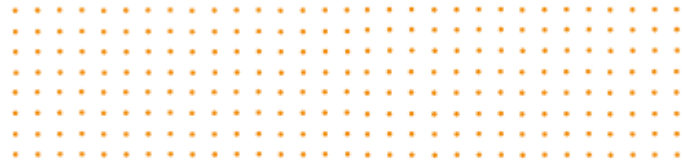
- **Wildcard search for properties can be used in the filter field but not via extended search** bug search
#6154 opened on Jul 30 by oliver-stoehr
- **Cannot filter for workflow steps whose names contain spaces in parenthesis** bug search
#5907 opened on Jan 31 by aetherfaerber
- **Extended search by "stepdone" does not work correctly** bug search
#5820 opened on Oct 24, 2023 by andre-hohmann
- **List filter broken after changing client** bug search
#5791 opened on Sep 28, 2023 by solth
- **Searching for task with project limitation is resulting in wrong results** bug search
#5730 opened on Jul 21, 2023 by henning-gerhardt
- **Search - Missing properties for processes of newspaper issues and periodical volumes** bug search
#5707 opened on Jul 13, 2023 by andre-hohmann
- **Errors while filtering are not caught and properly displayed** bug search
#5530 opened on Feb 6, 2023 by oliver-stoehr
- **Exact search for project name in task/process search** feature search
#5492 opened on Dec 15, 2022 by mnscholz
- **Exact search for project name in advanced process/task search** bug search
#5473 opened on Dec 2, 2022 by mnscholz



Ein Blick unter die Haube

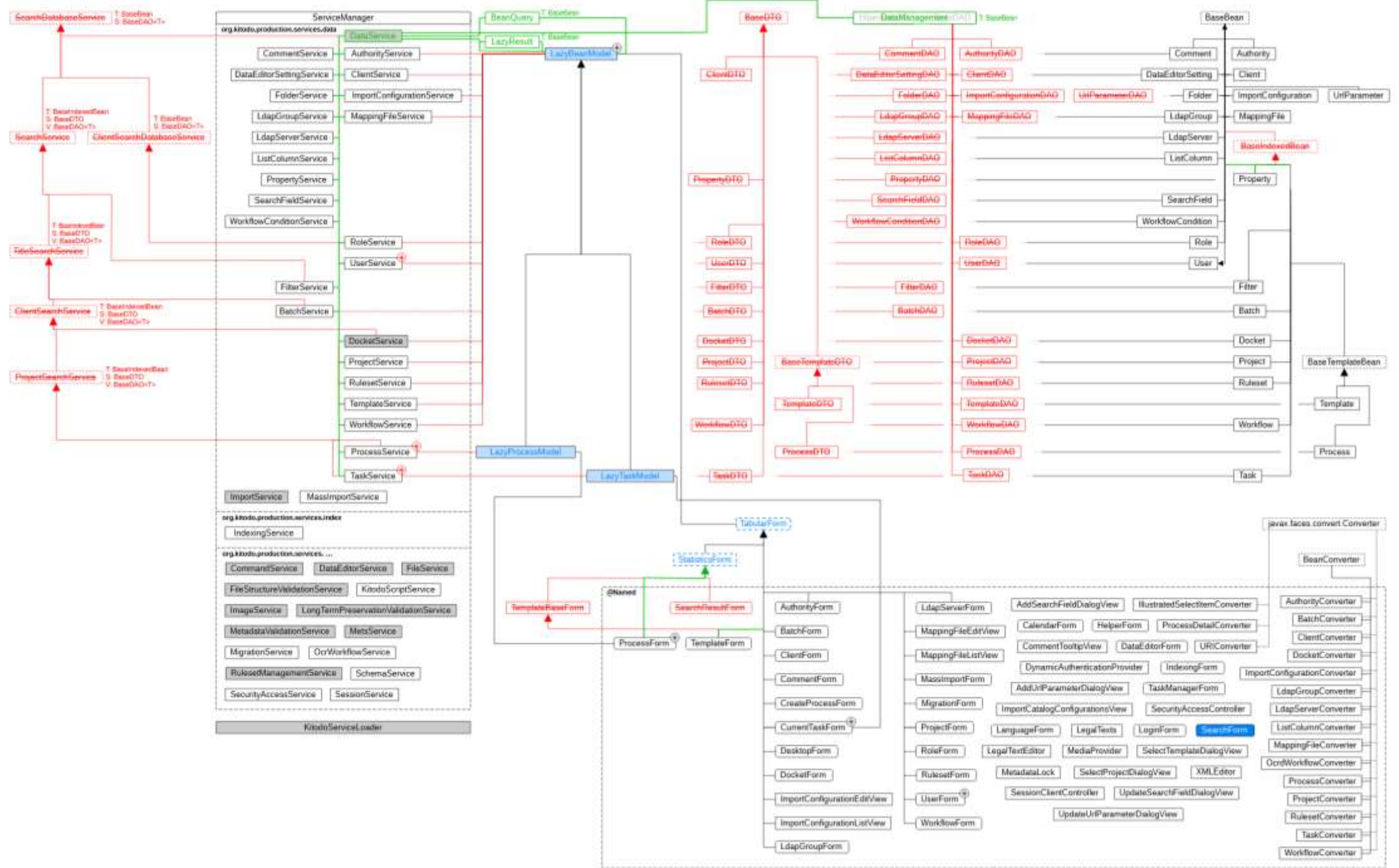


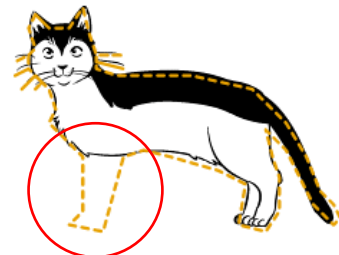
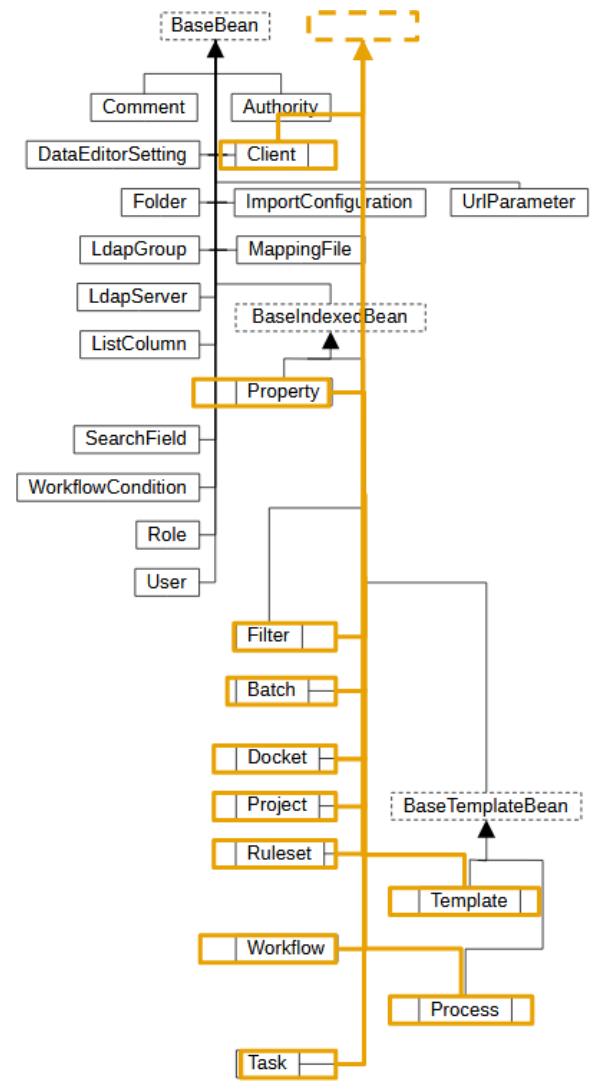
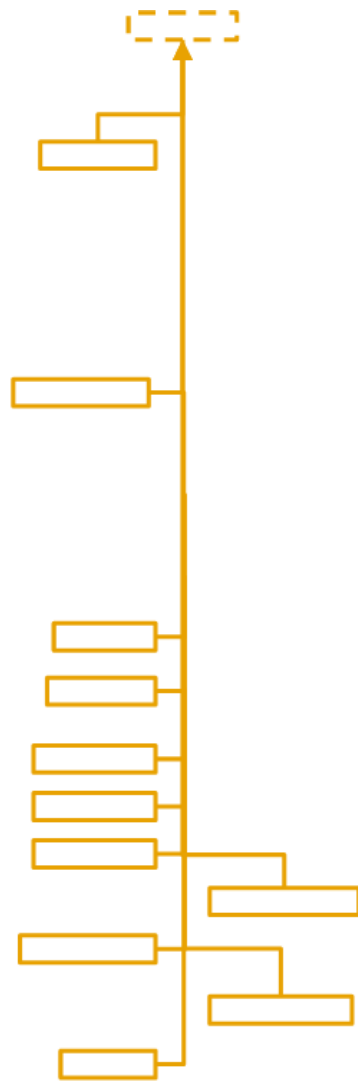
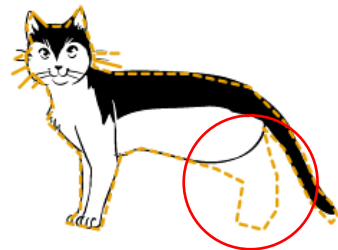
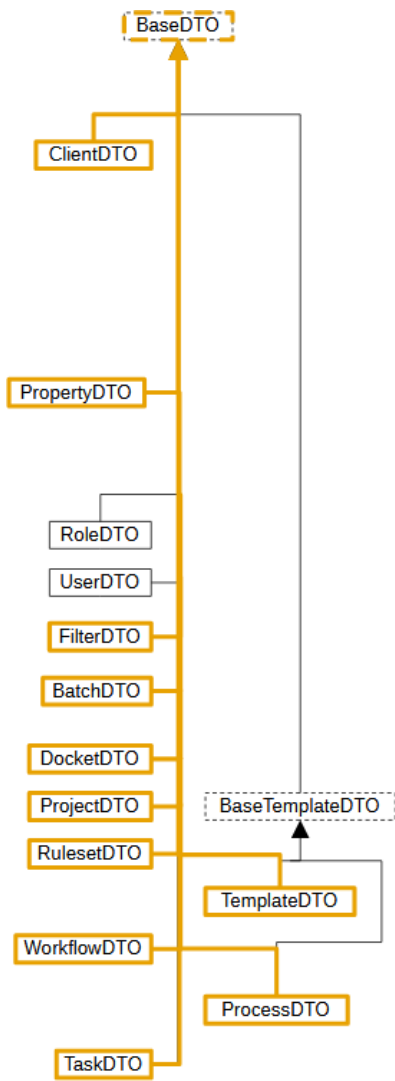


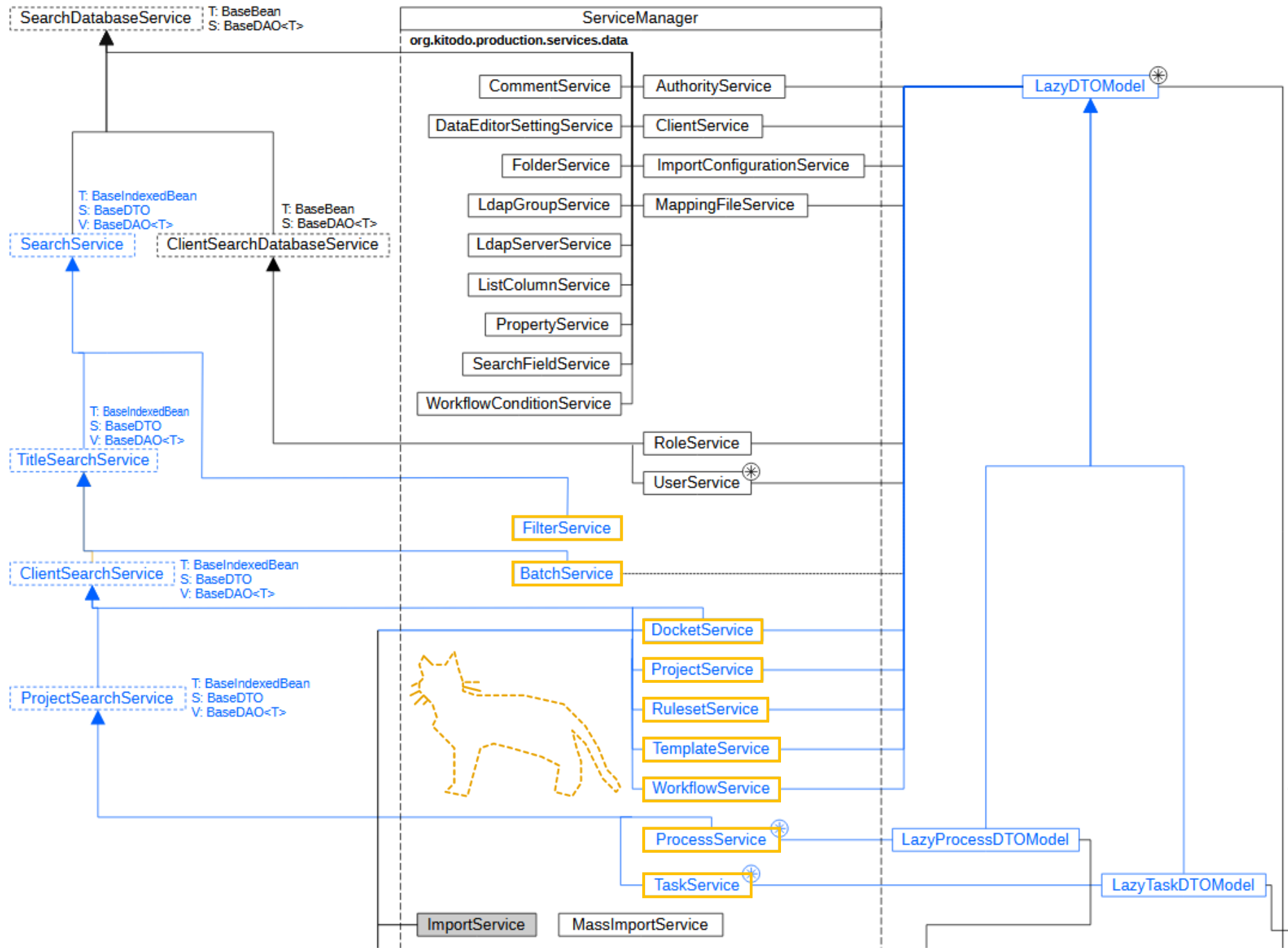


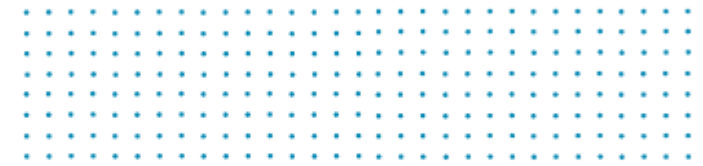
Der Umbau

- Einführung einer gemeinsamen Schnittstelle für DTO-Objekte und Data Beans
- Einführung von Schnittstellen für die Datenbank- und Suchfunktionen der Dienste
- Suchdienste auf Datenbank basieren
- Selbstentwickelte Elasticsearch-Anbindung löschen
- Einführung des Hibernate Search-Frameworks und Implementierung einer Indexierungsseite
- Implementierung von Indexierung und Suche



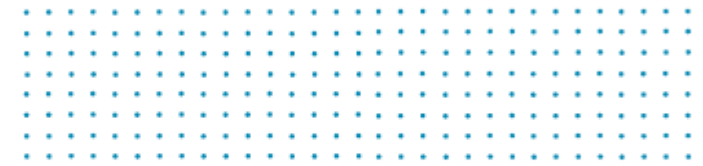
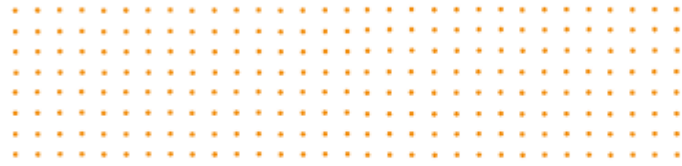






Analyse

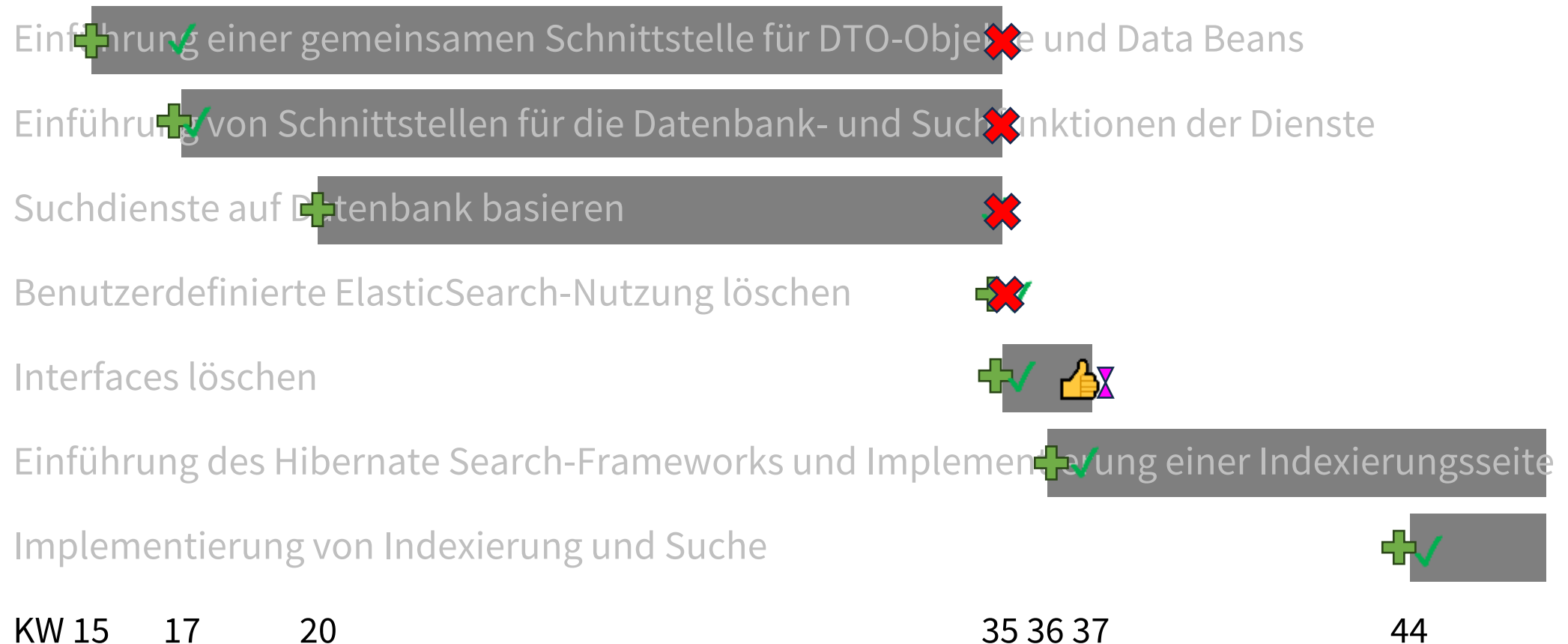
- Beispiel ProcessService: 110 Funktionen
 - 14 Funktionen finden oder bearbeiten Vorgänge *von allen Mandanten*
 - 6 Funktion finden nur Vorgänge, *die dem Mandanten zugeordnet sind*, für den der Benutzer, dessen Session dem ausführenden Thread gerade zugeordnet ist, gerade arbeitet
 - 90 Funktion haben gar nichts mit der Suche zutun

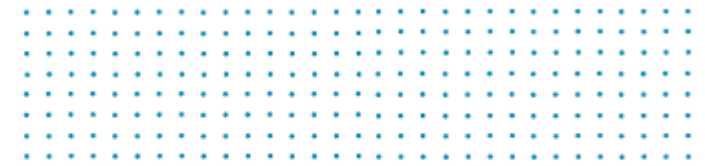
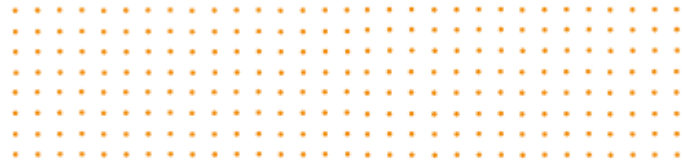


Ungewöhnliche Programmierung (Auswahl)

- Funktion getProcessDataDirectory („hole den Pfad zum Vorgangsverzeichnis“) *schreibt* die URL zum Vorgangsordner in den Vorgang **und speichert den Vorgang ab**
- Sortierung nach Erstelldatum mit dem Schlüsselwort „duration“ (Dauer)
- Ergebnis der Funktion findByRuleset (lade *alle* Vorgänge, die einen bestimmten Regelsatz haben) wird nur auf „leer/nicht leer“ geprüft

Ablauf





Was hat sich geändert?

- Anzeige von Listen direkt aus der Datenbank
- Suche nur für Titel und Metadaten
- Anwendung funktioniert weitestgehend ohne Index
- Indexprofil wird automatisch erstellt
- Keine Anzeige der Anzahl Objekte im Index möglich



Was geht

- Suche nach einfachen Metadaten
- Suche nach Metadaten in bestimmten Feldern (Key)
- Suche nach Metadaten in bestimmten Feldern (Label)
- Suche mehrere Wörter
- Suche “enthält nicht”
- Suche nach Metadatenfeldern

Kapitel

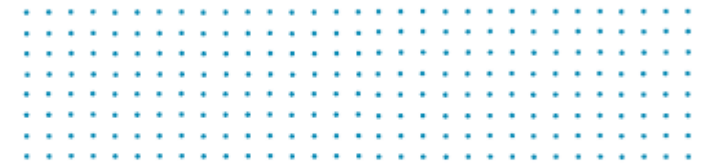
TitleDocMain:Kapitel

Hauptsachtitel:Kapitel

"Autor:Max Frisch"

-Revolutionär

isbn_falsch



Was geht

- Suche nach Werkstückmetadaten (dmdSec)
- Suche nach Vorlagenmetadaten (amdSec/sourceMD)
- Suche nach Prozessmetadaten (amdSec/techMD)
- auch in bestimmten Feldern
- auch mit deutschen Suchschlüsseln

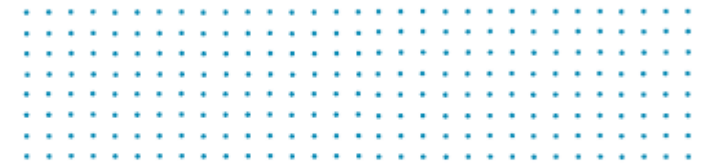
workpiece:1234567X

"template:Signatur:Ars Hydr."

processproperty:ocr:Fraktur

workpiece:Hauptsachtitel:Kapitel

werkstueck:Kapitel



Was geht

- Suche ID (datenbankbasiert)
- Suche mehrere IDs (datenbankbasiert)
- Suche nach Vorgangstitel, wortweise rechtstrunkiert
- Suche nach Projekt
- Suche nach Batch

id:42

id:122-188

process>LoremIps

project:DigiWunschbuch

"gruppe:Kiste 48"



Was geht

- Suche Aufgabe
- Suche gesperrte Aufgabe
- Suche offene Aufgabe
- Suche in Bearbeitung befindliche Aufgabe
- Suche abgeschlossene Aufgabe

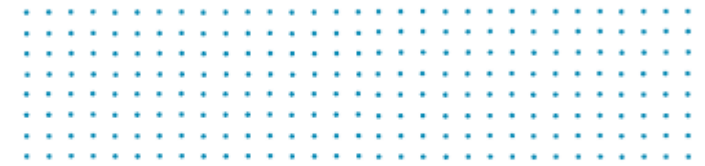
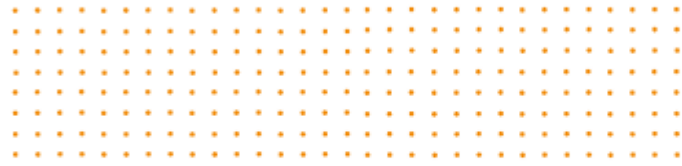
step:OCR

steplocked:Quality

stepopen:scanning

stepinwork:Scannen

stepdone:QC



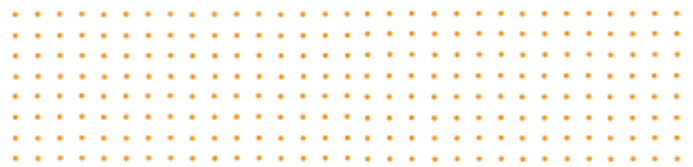
Was geht

- Suche automatische Aufgabe
- Suche irgendeine Aufgabe
- Benutzer, der die Aufgabe abgeschlossen hat

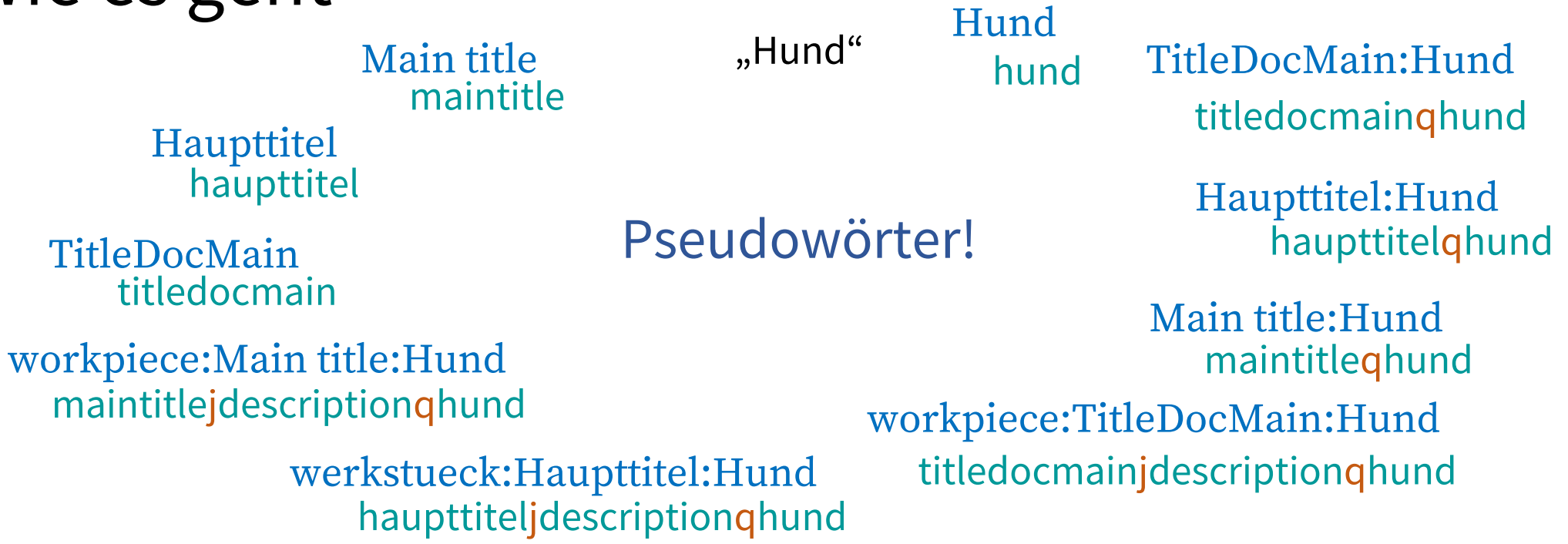
stepautomatic:OCR

stepautomatic

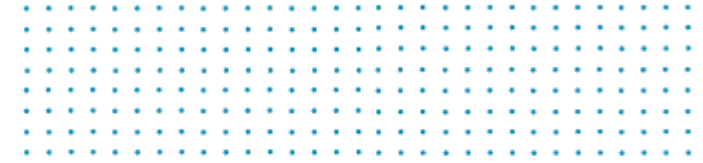
stepdoneuser:Mustermann



Wie es geht



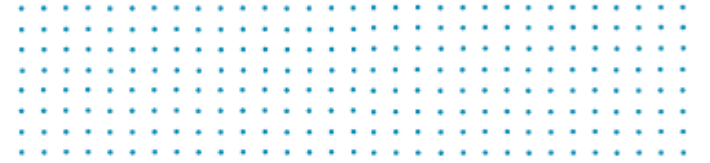
Korrekturmeldung



Wie es geht

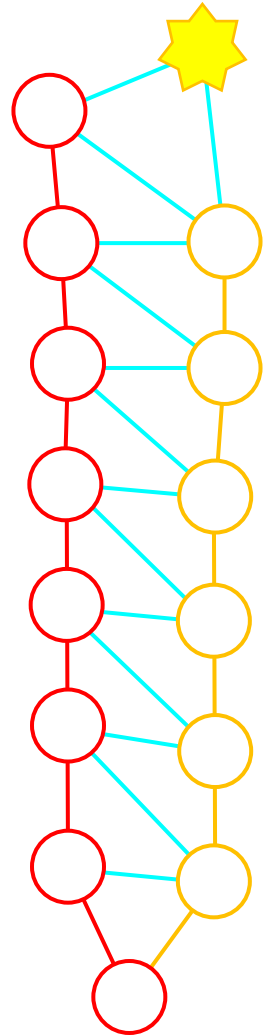
„Korrekturmeldung“

k ko kor korr korre korrek **korrekt** korrektu **korrektur** korrekturm korrekturme korrekturmel
korrekturmeld korrekturmeldu korrekturmeldun **korrekturmeldung** o or orr orre orrek orrekt orrektu
orrektur orrekturm orrekturme orrekturmel orrekturmeld orrekturmeldu orrekturmeldun
orrekturmeldung r rr rre rrek rrekt rrektu rrektur rrekturm rrekturme rrekturmel rrekturmeld
rrekturmeldu rrekturmeldun rrekturmeldung r re rek rekt rektu rektur rekturm rekturme rekturmel
rekturmeld rekturmeldu rekturmeldun rekturmeldung e **ek** ekt ektu ektur ekturm ekturme ekturmel
ekturmeld ekturmeldu ekturmeldun ekturmeldung k kt ktu ktur kturm kturme kturmel kturmeld
kturmeldu kturmeldun kturmeldung t **tu** tur **turm** turme turmel turmeld turmeldu turmeldun
turmeldung u ur urm urme **urmel** urmeld urmeldu urmeldun urmeldung r rm rme rmel rmeld rmeldu
rmeldun rmeldung m me mel meld meldu meldun **meldung** e el eld eldu eldun eldung l ld ldu ldun
ldung d du dun **dung** u un ung n ng g



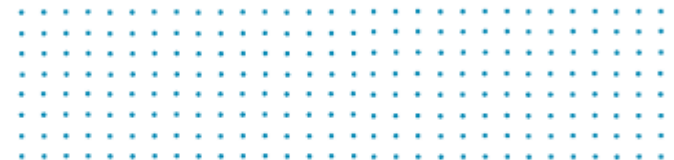
To do ...

master



hibernate-search

*Every tree
needs a
little friend.*

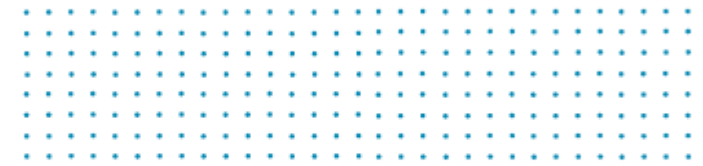
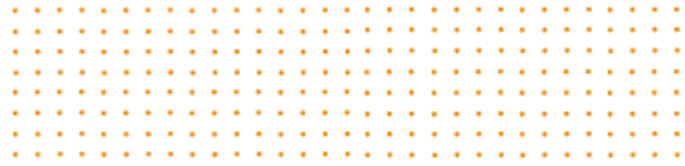


Offene Pull-Requests

Einführung des Hibernate Search-Frameworks und Implementierung einer Indexierungsseite

Implementierung von Indexierung und Suche



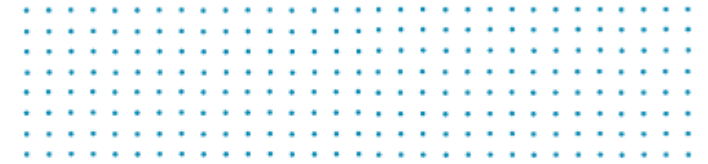


Fragen? Barcamp!

- Indexierung von Objekten ohne Metadaten?
- Mehrwort-Nicht-Suche kann zu viel nicht-finden
- Brauchen wir eine Stoppwort-Funktionalität?
- „Meldung“ beidseitig trunkiert indexieren?
- Overhead durch erneutes Einlesen der meta.xml pro Aufgabe ... ?



kulturgut | scanner
bürogut | scanner
Marken der MIK-CENTER GmbH



Vielen Dank für Ihre Aufmerksamkeit!