

Mit CultureGraph zu einer virtuellen Systematik

Worum geht es?

- Entwicklung in der UB Dortmund:
 - Umfangreiche Anpassungen des Bestandsmanagement aufgrund der nahenden Sanierung der Bibliothek
 - "e-preferred" Strategie beim Medienerwerb
 - Folge: Festhalten an der Aufstellungssystematik der UB Dortmund unmöglich.
- Idee:
 - Entwicklung einer "virtuellen Systematik" der Bestände auf Basis der RVK

Situation

- inhaltliche Erschließung bisher eher stiefmütterlich
 - Wenn, dann nur in den Bereichsbibliotheken auf Basis veralteter sehr lokaler Systematiken (mit teilweise abenteuerlichen Systemstellen, z.B. Sammelwerke)
 - Zentralbibliothek: Numerus currens
 - RVK spielt im hbz-Verbund keine Rolle
- Idee:
 - Fremddatenübernahme durch „Clustering“ unserer Bestände mit Beständen, die eine große Zahl an inhaltserschließenden Daten (insb. RVK-Notationen) enthalten.

„Clustering“-Ansätze / Datenquellen

- Cluster
 - CultureGraph
 - WorldCat „works“
 - RDA-Werke
 - Bände versus Überordnungen
- Zusätzliche Datenquellen für inhaltserschließende Elemente
 - GND / CrissCross
 - coli-conc

Anwendung von CultureGraph

- Welche Fragen wollen wir beantworten?
 - Wie hoch wäre mit dem CultureGraph-Ansatz der mit RVK-Notationen angereicherte Anteil unseres Bestands?
 - Zunächst gedruckter Bestand inkl. Betrachtung der Standorte
 - Wie verteilt sich das Ergebnis auf die Sammlungen der UB? Gibt es Schwachstellen?
 - Wie sind die aktuellen Werke abgedeckt?
- Die Daten werden für die Analyse als Linked Data organisiert und in einer Graph-Datenbank gespeichert.
 - Vorteil: leichte Erweiterbarkeit ohne DB-Konfiguration

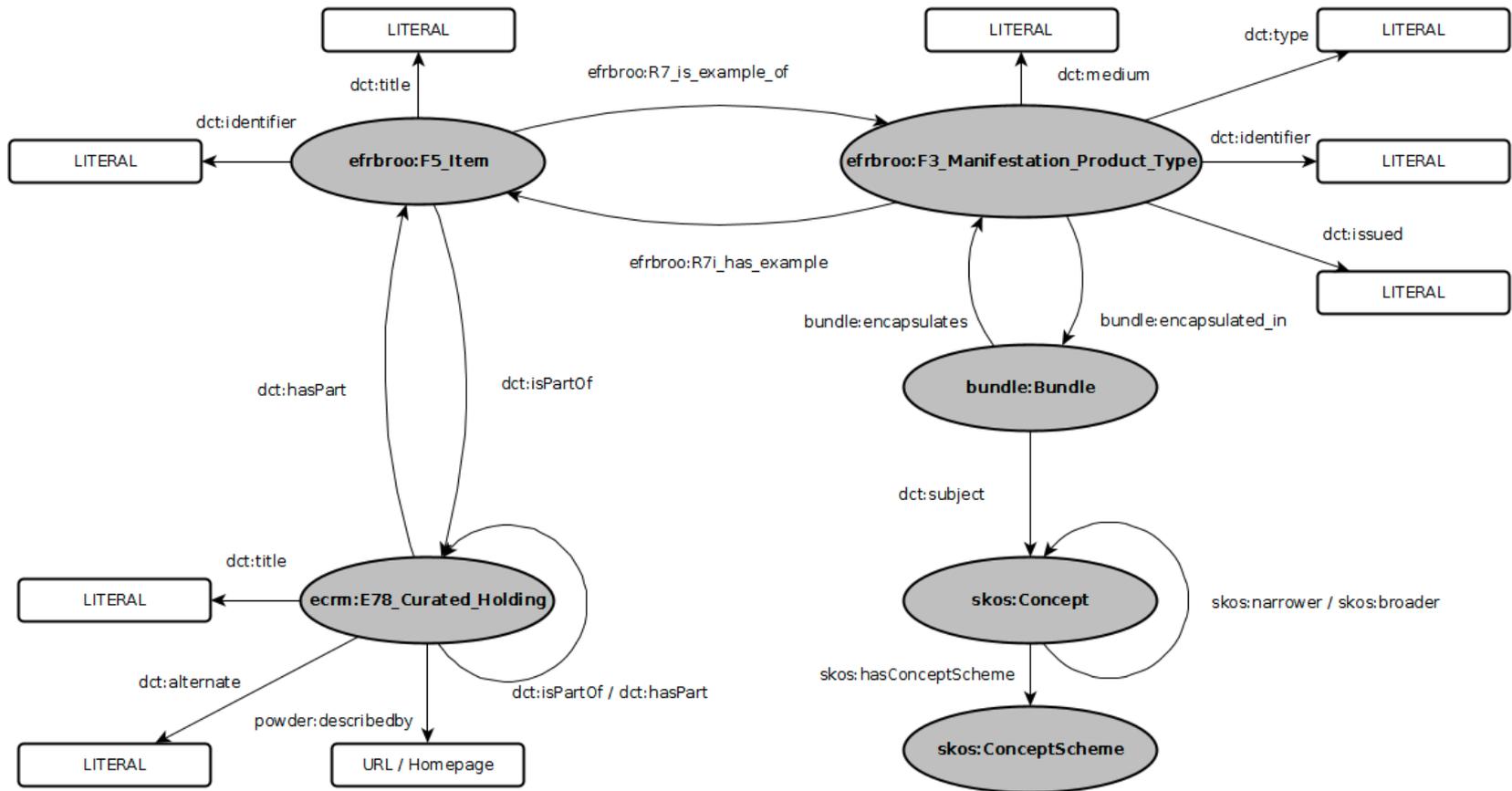
Datenmodell / „Application Profile“ (1)

- Das Datenmodell basiert auf dem Bundle- bzw. Cluster-Konzept.
 - Daher steht die Bundle-Ontologie mit ihrer zentralen Klasse "Bundle" im Mittelpunkt.
 - Die Bundle-Ontologie erlaubt es, Instanzen zu bündeln und Aussagen darüber zu machen, wie diese zustande kommen.
- Mittels FRBRoo, der objektorientierten Version der FRBR, werden die Beziehungen zwischen den im Katalog erfassten Manifestationen und deren Exemplaren dargestellt.
- Die Aufstellung der Exemplare innerhalb der UB Dortmund wird mittels Sammlungen modelliert, die hier mit Hilfe des CIDOC CRM modelliert und so an die FRBRoo-Exemplare angedockt werden.
- Es werden mittels DC Terms nur die Metadaten zu den Instanzen hinzugefügt, die für die Auswertungen benötigt werden.

Datenmodell / „Application Profile“ (2)

- Liste der verwendeten Ontologien:
 - Bundle Ontology
 - DC Terms
 - SKOS
 - Erlangen CRM (The Erlangen CRM / OWL is an OWL-DL 1.0 implementation of the CIDOC Conceptual Reference Model (CIDOC CRM).)
 - Erlangen FRBRoo (The Erlangen FRBRoo / OWL is an OWL-DL 1.0 implementation of the FRBR extension of the CIDOC Conceptual Reference Model (CIDOC CRM).)

Datenmodell / „Application Profile“ (3)



Umsetzung

- Datenbank: GraphDB von Ontotext
 - flexibles Datenmodell
 - Speichern der einzelnen Datensets als Graphs
 - Abfragespache SPARQL
 - Möglichkeit der „Inference“
- Skripte in Python
- Besondere Herausforderung:
 - RVK, CultureGraph und bibliographische Metadaten via lobid.org müssen lokal gehalten werden, da keines dieser Datensets über einen SPARQL-Endpoint nutzbar ist.

Aufbereitung der CultureGraph-Daten (1)

- CultureGraph befindet sich aktuell in einem Relaunch
 - geplant war der für Q3/2018, ist aber noch nicht erfolgt
 - Ich habe im Frühjahr 2018 bereits Daten aus dem Projekt erhalten, welche allerdings auf älteren Rohdaten beruhte.

7625070 

Source: http://data.ub.tu-dortmund.de/resource/cg/wrk_bundle/7625070

	subject	predicate	object	context
1	http://data.ub.tu-dortmund.de/resource/cg/wrk_bundle/7625070	dct:subject	rvko:AA11700	http://data.ub.tu-dortmund.de/graph/cg_bundles
2	http://data.ub.tu-dortmund.de/resource/cg/wrk_bundle/7625070	dct:subject	rvko:NR6910	http://data.ub.tu-dortmund.de/graph/cg_bundles
3	http://data.ub.tu-dortmund.de/resource/cg/wrk_bundle/7625070	http://purl.org/net/bundle#encapsulates	http://d-nb.info/967290260	http://data.ub.tu-dortmund.de/graph/cg_bundles
4	http://data.ub.tu-dortmund.de/resource/cg/wrk_bundle/7625070	http://purl.org/net/bundle#encapsulates	http://lobid.org/resources/TT001416026	http://data.ub.tu-dortmund.de/graph/cg_bundles
5	http://data.ub.tu-dortmund.de/resource/cg/wrk_bundle/7625070	http://purl.org/net/bundle#encapsulates	http://lod.b3kat.de/title/BV017078529	http://data.ub.tu-dortmund.de/graph/cg_bundles
6	http://data.ub.tu-dortmund.de/resource/cg/wrk_bundle/7625070	http://purl.org/net/bundle#encapsulates	http://swb.bsz-bw.de/DB=2.1/PRS=rdf/PPNSET?PPN=107550199	http://data.ub.tu-dortmund.de/graph/cg_bundles
7	http://data.ub.tu-dortmund.de/resource/cg/wrk_bundle/7625070	rdf:type	http://purl.org/net/bundle#Bundle	http://data.ub.tu-dortmund.de/graph/cg_bundles

Ergebnisse (1)

- Auswertung gedruckter Bestand und RVK-Notationen zu zugehörigen Manifestationen – ohne die letzten drei Jahre

Standort	Exemplare	Manifestationen	CG-Bundles	Anteil	Manifestationen mit RVK	Anteil
ZB (DE-290-0)	608.996	558.776	469.679	84,05 %	393.803	70,48 %
EFB (DE-290-12)	180.425	173.235	153.415	88,65 %	145.370	83,92 %
BR (DE-290-9)	34.841	33.420	23.907	71,54 %	20.042	59,97 %
BAB (DE-290-10)	20.914	19.899	16.418	82,51 %	13.673	68,71 %

Fazit

- Ergebnisse via CultureGraph für die inhaltliche Retro-Erschließung sind sehr gut.
- Daraufhin Entscheidung: Wenn die Aktualitätszyklen von CultureGraph angemessen sind, ist der CultureGraph-Ansatz gut.
 - Das sollte aber nicht darüber hinwegtäuschen, dass diese Zyklen nur bedingt etwas mit der Erschließung aktuellster Titel zu tun hat ;-)
- Weitere Lösungen für den Rest sollen ermittelt werden.
- Wir prüfen aktuell, auf welche Art wir etwas für den „Datenklau“ zurückgeben können.
 - Z.B. Mappen von DDC- auf RVK-Notationen via coli-conc

Vielen Dank!

Kontakt:

- Hans-Georg Becker
 - hans-georg.becker@tu-dortmund.de
 - Abteilung „Informationssysteme und Datenmanagement“
 - daten.ub@tu-dortmund.de

Universitätsbibliothek Dortmund

Vogelpothsweg 76

44227 Dortmund