



By © Raimond Spekking / CC BY-SA 4.0 (via Wikimedia Commons), CC BY-SA 4.0,
<https://commons.wikimedia.org/w/index.php?curid=74925303>

Semantic Web in Libraries

Bericht von der Linked-Open-Data-Konferenz SWIB in Bonn, die vom 26. bis 28. November 2018 zum zehnten Mal stattfand

Oliver Schöner, unter Mitarbeit von Susanne Faude, Carsten Klee und Johann Rolschewski

» In Bonn fand Ende November, inzwischen zum zehnten Mal, die Linked-Open-Data-Konferenz SWIB (<http://swib.org/swib18/>) statt. Für die Organisatoren war das Jubiläum nicht nur Anlass zurückzublicken, sondern auch, die Internationalität der Tagung hervorzuheben. Startete sie im Jahr 2009 noch als „Semantisches Web in Bibliotheken“, ist die Konferenzsprache nunmehr Englisch – ohne freilich das Akronym geändert zu haben. Quer zu dieser Internationalität – die knapp zweihundert Teilnehmer kamen aus fünfundzwanzig Ländern – und der zentralen Kongressidee, Daten und Menschen zu vernetzen, stand die traurige Tatsache, dass den aus Nigeria angekündigten Teilnehmern Babarinde Ayodeji Odewumi und Adetoun Adebisi Oyelude die Einreise nach Deutschland verweigert wurde; wohl aus Furcht, sie könnten nicht zurückkehren. Ihr Vortrag wurde als MP4-Datei wiedergegeben; erfreulich immerhin zu hören, wie pragmatisch sie bei ihrem Vorgehen, für ihre eigenen Daten die Library of Congress Subject Headings zu nutzen, angesichts solch basaler Widrig-

keiten wie instabiler Internetverbindungen umgehen. Am ersten Tag gab es wie immer keine Vorträge, sondern insgesamt sechs Workshops. In ihnen konnten die Teilnehmer nicht nur Wissenswertes erfahren, sondern das erworbene Wissen praktisch umsetzen: Themen waren Linked Open Data und wie man LOD nutzbar macht, die immer wichtiger werdenden Wiki-Familienmitglieder Wikidata und Wikibase und RDF-Software wie Fedora und ShEx.

Am Beginn der beiden eigentlichen Kongresstage standen die Keynotes von James Handler, einem langjährigen Wegbegleiter des Internetpioniers Tim Berners-Lee, und Mia Ridge von der British Library. In ihnen wurde die Vernetzung und Semantisierung von Daten aus zwei ganz unterschiedlichen Perspektiven beleuchtet. Während James Handler hervorhob, dass die Idee der automatisierten Generierung von Wissen keineswegs von Google stammt, sondern sich schon in den frühesten Artikeln zum World Wide Web wiederfindet, betonte Mia Ridge, dass die Vernetzung von Daten ohne die Vernetzung von Menschen

schlechterdings unmöglich sei: Crowdsourcing-Projekte können hier ein interessanter Ansatz sein.

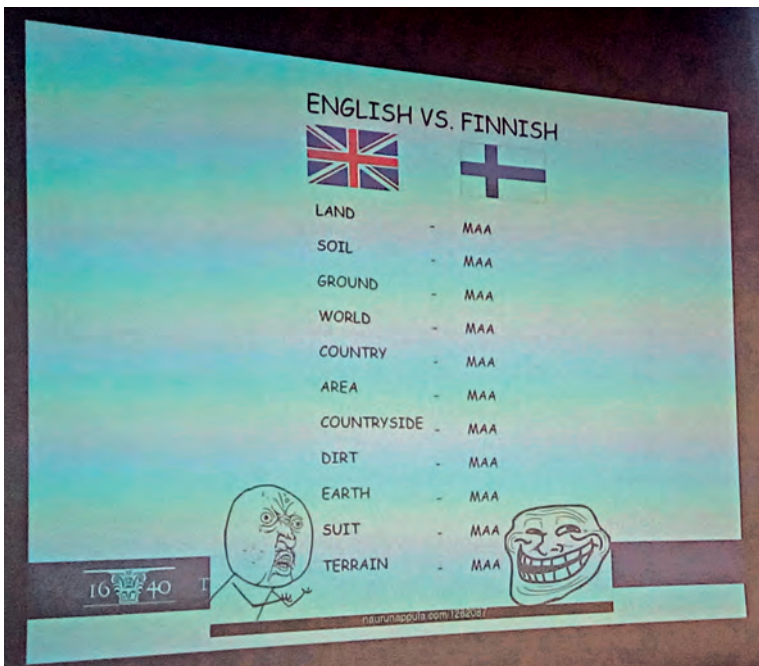
Dass Linked Data ein immer noch junges Thema im Bibliotheksumfeld ist, zeigten die zahlreichen Vorträge zu konkreten Softwarewerkzeugen – wie RDF-Daten überhaupt gespeichert (David Wilcox, Fedora) und verarbeitet (Simeon Warner und Huda Khan, SHACL und VitroLib) werden können. Und dass diese Softwarewerkzeuge noch nicht den gleichen Reifegrad

Congress Subject Headings zu verknüpfen. Hier sagt ein Bild tatsächlich mehr als tausend Worte (siehe Abbildung links).

Als Beobachter kommt man ins Grübeln und erinnert sich eines lange zurückliegenden Philosophiestudiums und der Diskussionen, ob und wie sehr die Sprache das Denken bestimmt. Die radikalste Theorie einer Determiniertheit des Denkens durch die Sprache, die Sapir-Whorf-Hypothese, wird zwar heute kaum noch verfochten. Man fragt sich aber, ob sie nicht doch zutreffend ist – was die Vernetzung von „Wissen“, das ja nur sprachlich artikuliert werden kann, sicher nicht erleichtern würde.

Aber noch einmal Finnland: Dass in der dortigen Nationalbibliothek exzellente Arbeit geleistet wird, zeigte der Vortrag von Osmo Suominen über die Machine-Learning-Software Annif, der nach meinem Dafürhalten aufregendste der gesamten Tagung. Linked Data lässt sich ja in einem ganz weiten Sinn als das Ergebnis genuin bibliothekarischer Tätigkeit fassen: Mit der Sacherschließung werden Daten vernetzt, indem Strukturen auf wenig strukturierte Daten angewendet werden. Dass eine solche Tätigkeit in keinem Widerspruch zu Machine-Learning-Verfahren stehen muss, die sich häufig mit völlig unstrukturierten Daten begnügen, zeigte Osmo Suominen auf überzeugende Weise. Er kombinierte mit der neu entwickelten Software Annif verschiedene ML-Algorithmen so, dass zu gegebenen Texten mithilfe vorgegebener Thesauri automatisch vernünftige Schlagworte erzeugt werden können. Natürlich produziert die Software gerade bei Homonymen Fehler, aber der Vergleich mit einer rein von Menschen erstellten Verschlagwortung scheint nicht ungünstig auszufallen. Gerade vor dem Hintergrund der starken Kritik, die in jüngster Zeit an einem vergleichbaren Projekt der Deutschen Nationalbibliothek geleistet wurde, verdienen solche Ansätze im Blick zu bleiben.

Für alle, die mehr wissen wollen: Das Programm der SWIB bleibt im Netz, und wer sich den einen oder anderen Vortrag anschauen will, findet die Aufzeichnung ebenfalls über die Website. **I**



Terra incognita

wie klassische, relationale Datenbanken haben, wurde in den Überlegungen von John Samuel und Katherine Thornton deutlich. Sie zeigten, wie man Software vor dem Vergessenwerden bewahrt, indem man sie in Wikidata speichert: gewissermaßen ein Repository zweiter Stufe.

Daneben gab es einige Vorträge zu konkreten LOD-Projekten – beispielhaft sei hier <https://data.bnf.fr/> genannt, die Datenschnittstelle der Bibliothèque nationale de France. Gleichermäßen erfreulich wie ernüchternd war der immer wieder geäußerte Hinweis, wie wichtig das Problem der Datenmodellierung ist. Hierzu, so die mantraartig vorgetragene und fast schon banal anmutende Erkenntnis, muss man seine Daten nämlich kennen; und dann gerät man schnell in die Verlegenheit, die Daten erst einmal bereinigen zu müssen – eine Aufgabe, die notorisch unterschätzt wird. Selbst dann, möchte man hinzufügen, wenn man sich dieses Umstands bewusst ist.

Welche praktischen Schwierigkeiten es hierbei gibt, zeigte der launige Vortrag von Satu Niininen von der Finnischen Nationalbibliothek. Sie beschrieb das Unterfangen, den finnischen Thesaurus YSO ins Englische zu übersetzen und ihn dann mit den Library of

Der Hauptautor ist

Oliver Schöner

Oliver.Schoener@sbb.spk-berlin.de,
geholffen haben

Johann.Rolschewski@sbb.spk-berlin.de,
Carsten.Klee@sbb.spk-berlin.de und
Susanne.Faude@sbb.spk-berlin.de.