

DET SEMANTISKE INTERNET

Hvad er Linked Open Data? Og hvad betyder de for bibliotekerne?

AF CARSTEN H. ANDERSEN, CHA@DBC.DK, DBC A/S

Internetet er i udpræget grad menneske/maskine-interaktion. Vi arbejder i websider, som i det store hele er opmærkede, så browseren kan layoute siden pænt. Når vi klikker på et link, er det kun vores tolkning, der skaber vores forventning til, hvad vi bliver præsenteret for, når linket aktiveres. Der er ingen information knyttet til linket, der muliggør, at en applikation kan foretage en semantisk afkodning af, hvad linket fører til.

Men Tim Berners-Lee, direktør for World Wide Web Consortium, har formuleret en vision om et semantisk web. Hans idé er et internet, hvor websiderne er opbrudt i entydigt identificerbare og relaterede dataelementer med tilknyttet semantik. Det er så applikationernes opgave at sammensætte og præsentere relevante data for brugeren.

Berners-Lee's fire principper

Tim Berners-Lee formulerede følgende fire principper for de såkaldte *linked data* som en vej til at realisere det semantiske web (her i en lettere bearbejdet udgave):

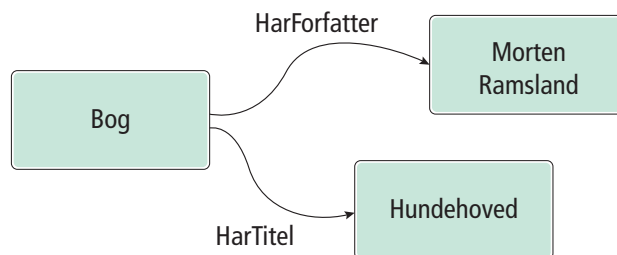
1. Brug en URI (Uniform Resource Identifier) som entydig identifikator i stedet for eksempelvis forfatterens navn.
2. Brug URL'er til de entydige identifikatorer, så de kan linkes til og opløses, for eksempel til forfatternavnet.
3. Tilbyd yderligere information om for eksempel forfatteren i et standardiseret format, når URI-linket forfølges.
4. Suppler informationen med links til relaterede URI'er, for eksempel forfatterens fødested.

Pointen er, at der anvendes helt gængse www-standarder som HTTP-protokollen og XML til at beskrive og koble dataelementer sammen på tværs af webben. Grundlæggende er der tale om linked data, men oftest taler vi om linked *open* data, fordi linked data implicit forstås som data, der bør være frit tilgængelige.

Dissekering af linked data

I linked data er informationen atomiseret i såkaldte tripler. En triple består af henholdsvis subjekt, prædikat og objekt, hvor prædikatet også kan ses som relationer (eksempel: 'Bog' – 'HarTitel' – 'Hundehoved'), og hvor subjekter og objekter kan indgå i flere tripler.

Applikationernes forståelse af triplerne opnås ved at anvende veldefinerede ontologier mv. til regulering og beskrivelse af de anvendte subjekter, prædikater og objekter.



Bibliotekerne og linked open data

Generelt har bibliotekssektoren har været fantastisk dygtig til at udvikle og anvende standarder *inden* for sektoren. Til gengæld er disse standarder vanskeligt tilgængelige for interesserede *uden* for sektoren. Fordi linked data publiceres med HTTP- og XML-standarder, åbnes op for en bredere anvendelse af data, og det betyder, at flere forskellige aktører kan lave spændende applikationer til brugerne.

Linked data giver mulighed for at dele ansvar for data, i stedet for at den enkelte skal skabe og vedligeholde alle data selv. Ultimativt kan en bibliografisk post bestå af links til dataelementer forskellige steder på webben. F.eks. vil autoritative udenlandske navneformer kunne trækkes fra VIAF (Virtual International Authority File).

Deling af data forudsætter pålidelige data og tro på, at data også er til stede næste år. Bibliotekssektoren betragtes i den forbindelse som garant for pålidelige og persistente data, og sektoren vil derfor kunne spille rollen som den datahub, der linker datasæt i et web af data.

Hot or not?

I Tim Berners-Lee's vision lover linked data os guld og grønne skove, men faktum er, at der endnu ikke er sket så meget. Internationalt set er der dog tegn på, at der er ved at komme gang i sagerne. Library of Congress, British Library, Deutsche Nationalbibliothek og Bibliothèque Nationale de France er alle stærkt engagerede i at publicere datasæt som linked open data.

Men selvom afprøvning og udvikling af linked data efterhånden fylder en del i den internationale biblioteksverden, er det overraskende, hvor lidt der tilsyneladende sker herhjemme. Vi ser eller hører ikke om mange projekter, hvor teknikkerne afprøves.

På DBC tror vi så meget på linked data, at vi har tillagt det strategisk betydning at opbygge kompetencer på området. Derfor har vi netop gennemført et analyse- og vidensopbygningsprojekt, som gik ud på at publicere en meget begrænset delmængde af Dansk Bogfortegnelse, og i 2013 gennemfører vi flere projekter. ●