

Frågor (och svar?)

från utbildningsdagen om Wikidata
den 13 oktober

Bakgrund

Den 13 oktober erbjöd Wikimedia Sverige i samarbete med Kungliga Biblioteket en utbildning om Wikidata. Detta som en del i vårt gemensamma projekt [Strategisk inkludering av biblioteksdata på Wikidata](#). Under dagen fick vi många frågor. Vi besvarade de allra flesta, alla hoppas vi, i chatten och muntligt.

Frågorna (och svaren?)

Egenskaper för personer, hur ser diskussionen ut i relation till dataskydd?

Svar

Wikidatas riktlinjer [\[\[:wikidata:Wikidata:Living_people|finns här\]\]](https://wikidata:Wikidata:Living_people). I korthet: när man redigerar data om levande personer måste man vara särskilt noga med att källbelägga informationen med tillförlitliga källor och tänka på huruvida redigeringen skulle kunna orsaka skada för personen. Information som inte anses vara i allmänt intresse kan tas bort av administratörer.

Eftersom Wikidatas servrar ligger i USA är det amerikansk lag som gäller vid eventuella tvister.

Hur ser du att en Wikipediaartikel är eller inte är kopplad till ett Wikidataobjekt?

Svar

Om en artikel är kopplad till ett Wikidataobjekt så syns länken *Wikidata-objekt* under rubriken *Verktyg* i vänsterkolumnen.

Mer information om Wikidata just från ett Wikipedia-perspektiv finns på `[[[:wikipedia:sv:Wikipedia:Wikidata|denna hjälpsida på svenskspråkiga Wikipedia]]`

Hur använder forskare Wikidata?

Svar

Som ett kollaborativt skapas, flerspråkig och länkad databas är Wikidata särskilt intressant för forskare inom data- och informationsvetenskap. Se följande artikel för en systematisk översikt av forskning om Wikidata:

[<https://arxiv.org/abs/1908.11153> Mariam Farda-Sarbas, and Claudia Müller-Birn. (2019). Wikidata from a Research Perspective – A Systematic Mapping Study of Wikidata].

I det jag arbetar med är PIDs (Persistent Identifiers) en stor fråga, är det så även för Wikidata?

Svar

Detta är väldigt viktigt för Wikidata. Det finns över ett tusen egenskaper av typen “extern identifierare”, alltså länkar till andra resurser. Bland de viktigaste finns ISNI, VIAF och ORCID. Att de inte ruttar bort är nyckeln till Wikidatas roll som ett nav för länkade data på webben.

Även Wikidatas egna identifierare (QID) ska vara beständiga. Om ett objekt tas bort så kan inte dess QID återanvändas, och om två objekt slås ihop (t.ex. på grund av dubletter) så blir den ena identifieraren en omdirigering till den andra.

Har Wikidata en definierad målgrupp? Alltså vem skall använda Wikidata?

Svar

Wikidata `[[[:wikidata:Wikidata:Introduction|beskriver sig självt]]` som en kollaborativt skapad sekundärdatabas som tillhandahåller strukturerad data för Wikipedia och de andra Wikimediaplattformarna men även för alla människor i världen.

Wikipediagemenskapen har nytta av Wikidata eftersom det gör det möjligt att länka samman artiklar om samma ämne på olika språk. Vidare kan information från Wikidata visas på Wikipedia, vilket gör det lättare att hålla artiklar på olika språk uppdaterade; om en uppgift förändras på Wikidata uppdateras även alla de Wikipediaartiklarna där informationen visas.

Data på Wikidata är maskinläsbar och gratis att använda, vilken gör den attraktiv för utvecklare av digitala tjänster, appar m.m.

Ser ni Wikidata som informationskälla eller aggregator för information?

Svar

I Wikidata samlas information från många olika källor, och alla uttalanden bör källbeläggas. Det är alltså ingen ny information som skapas på det sättet. Det är också så Wikipedia fungerar – som en sammanfattning av information från tillförlitliga källor.

Om jag vill ha info om Tove Jansson, vad avgör vilken jag skall välja – Wikidata eller Wikipedia?

Svar

För en vanlig läsare som vill lära sig om något ämne är Wikipedia det bästa valet. Wikidata lämpar sig bättre om man vill t.ex. komma åt stora mängder faktauppgifter för att bearbeta dem maskinellt, som att skapa en tidslinje över prisbelönade finländska författare på en webbplats.

Hur förhåller sig upphovsrätten här? Både wikidata och vi på bibliotek gillar öppen data, men producenter gör det inte nödvändigtvis – oavsett om det är kommersiella eller offentliga resurser(läs kulturinstitutioner).

Svar

Data på Wikidata omfattas av [https://creativecommons.org/share-your-work/public-domain/cc0/ licensen CC0]. Den är alltså gratis att använda för både privata och kommersiella ändamål utan krav på att man ska hänvisa till uphovsperson eller källa.

En baksida av denna mycket generösa licens är att data som importeras till Wikidata från externa källor också måste omfattas av CC0. Licenser som CC-BY eller liknande som ibland används i "öppna" datamängder är inte kompatibla med Wikidata.

Jag är oklar över skillnaden med att ladda upp grejer och länka till.

Svar

Att länka till innebär att man lägger till en extern identifierare i någon annan resurs, till exempel Libris auktoritetsnumret [<http://libris.kb.se/resource/auth/191503> 191503] för Tove Jansson. Användare kan genom att klicka på numret på vidare till motsvarande post i Libris.

Att ladda upp innebär att man lägger till faktiskt innehåll till objektet, alltså uttalanden som "födelsedatum", "födelseplats" osv.

Måste all data importeras? Kan det inte räcka med länkar till andra källor, som också använder sig av länkad data?

Svar

Wikidatas styrka ligger i att data från många olika källor finns samlad på samma plats och på det sättet berikar varandra. När man tittar på ett Wikidataobjekt kan man ta del av all information mycket enklare än genom att besöka de länkade resurserna.

Devil's advocate: med så mycket data som hänger ihop. Vad är den långsiktiga planen för förvaltning? Kan någon köpa upp wikidata om kollektivet lessnar?

Svar

Wikidata, Wikipedia och de övriga Wikimediaprojekten drivs av stiftelsen [<https://wikimediafoundation.org/> Wikimedia Foundation]. Deras arbete finansieras av donationer, både av enskilda användare och organisationer (t.ex. Alfred P. Sloan Foundation och flera andra stiftelser, Google). Tack vare detta har vi kunnat använda Wikipedia i nästan 20 år utan att behöva se reklamannonser el. dyl.

Målsättningen är att Wikimediaprojekten ska alltid vara helt gratis att använda. Inget tyder på att Wikimedia Foundation inte skulle kunna fortsätta sitt arbete i en lång tid framöver.

Hur uppdateras Wikidata när de olika källorna ändrar information?

Svar

Det finns inget automatiskt uppdateringssystem. Det är alltid upp till användaren som lagt till informationen att uppdatera den.

När det gäller storskaliga uppladdningar från externa källor så är det ofta så att datan laddas upp vid ett tillfälle, utan att någon systematisk uppföljning/avstämning är inplanerad.

De 30 miljonerna vetenskapliga artiklarna verkar sammanfalla med de i PubMed. Används även MeSH linked data för ämnesbeksrivningarna i Wikidata?

Svar

Flera MeSH-egenskaper finns i Wikidata:

- * [[[:wikidata:Property:P486|MeSH descriptor ID (P486)]] – 37 000 objekt som använder
- * [[[:wikidata:Property:P672|MeSH tree code (P672)]] – 26 000 objekt som använder
- * [[[:wikidata:Property:P6680|MeSH term ID (P6680)]] – 59 objekt som använder
- * [[[:wikidata:Property:P6694|MeSH concept ID (P6694)]] – 720 objekt som använder

Det finns [[[:wikidata:Wikidata:WikiProject_Medicine|ett medicinskt wikiprojekt]] som bland annat koordinerar arbetet [[[:wikidata:Wikidata:WikiProject_Medicine/MeSH|med MeSH]]

Vad är det som är inläst nu från KB? Auktorietsdatabas? Något mer?

Svar

Vi har bland annat importerat:

- * 60 000 identifierare från Libris auktoritetsdatabas
- * 500 bokposter för de böcker som är mest använda som källor på svenskspråkiga Wikipedia
- * 6 000 poster för Statens Offentliga Utredningar
- * 33 000 poster för tidningar och tidskrifter

På

[[Projekt:Strategisk_inkludering_av_biblioteksdata_på_Wikidata_2018/Rapportering/Stödsidor_och_material_runt_biblioteksdata_och_dess_massuppladdningar|denna rapportsida kan man ta del av aktiviteter under projektens första del (2018–2019). Det nuvarande projektet (2019–2020) avslutas i december 2020.

Går det att komma åt referenser eller andra "egenskaper" av påståenden med SPARQL?

Svar

Absolut, det finns sätt att komma åt det mesta som vi inte gått igenom. En bra referensresurs finns t.ex. på [[:wikibooks:en:SPARQL|Wikibooks]] – här är [[:wikibooks:en:SPARQL/WIKIDATA Qualifiers, References and Ranks|kapitlet om referenser och bestämningsord]].

Vem skall använda sökningarna? datorprogram eller bibliotekarier?

Svar

Ibland är det väldigt kraftfullt som användare att kunna söka genom stora mängder information, som man kan göra på Wikidata med hjälp av SPARQL. Men också komplext. Så ibland gör datorprogram det åt en för att göra det enklare.

Till exempel är [<https://scholia.toolforge.org/> Scholia], som vi tittade på under dagen, i princip bara i förväg definierade SPARQL-frågor med författare som enda parameter man som användare väljer.

Ett annan webbtjänst som använder Wikidata som datakälla är Histropedia. Där kan man skapa snygga enkelt [<http://histropedia.com/timeline/xqbkzlv0st/Explorers-from-the-Age-of-Discovery> delbara och interaktiva tidslinjer]

Många enskilda vanliga Internetanvändare kan tycka att Wikidata och Wikidata Query Service för tekniskt komplext och abstrakt. I de flesta fall är man då bättre betjänt av Wikipedia. Men om man vill kunna ställa mer komplexa frågor, av researchsnatur, kunna visualisera resultaten, och exportera svaren som data för bearbetning offline är Wikidata mycket värdefull.

De flesta Internetanvändare är omedvetna användare av Wikidata: När man googlar efter t.ex. Selma Lagerlöf, får man en faktaruta bredvid sökresultaten. Den faktarutan är till stor del fylld med data från Wikidata.

Finns det något område där ni jämfört er täckning med andra källors täckning?

Svar

Ett projekt som Wikimedia Sverige genomfört som fokuserat på just detta var [\[:meta:FindingGLAMs\]](#). Projektet syftade till att kartlägga och förbättra täckningen av kulturarvsinstitutioner – GLAMs (Galleries, Libraries, Archives, Museums) på Wikidata. Projektet gjorde det tydligt att det finns väldigt mycket kvar att göra, i synnerhet i områden utanför västvärlden / anglosfären. Just nu finns det till exempel endast 2 bibliotek i (vardera) Turkmenistan, Tadzjikistan eller Myanmar.

Ett sätt att jämföra Wikidatas täckning med andra resurser är att titta på externa identifierare som används som egenskaper. Till exempel länkar 36 tusen Wikidataobjekt till [\[:wikidata:Property_talk:P3222#Documentation|Nationalencyklopedins onlineutgåva\]](#) och 58 tusen till [\[:Property_talk:P761#Documentation|Vatteninformationssystem Sverige\]](#).

Är det relevant att t ex söka efter vetenskapliga artiklar utgivna i Sverige och sortera på lärosäten eller publikationer?

Svar

Just när det gäller vetenskaplig publicering är Sverige och Norden ett eftersatt område på Wikidata. De storskaliga uppladdningar av data om vetenskapliga artiklar som pågått inkluderar inga svenska databaser. Anledningen är att det saknas öppet licensierade datakällor. Varken <http://swepub.kb.se/> Swepub eller <https://www.diva-portal.org/> Diva tillhandahåller CC0-licensierade metadata.

En översikt av CC-0/CC-BY-data om vetenskaplig publicering kan ses i tjänsten <https://graph.openaire.eu/> Open Aire].

Om man som bibliotekarie och efter det här känner att man vill börja göra något i wikidata, var ska man engageras sig, var börjar man? Kan inte sparql som är lite svårt.

Svar

Att lägga till etiketter och beskrivningar på svenska och andra språk man kan är ett enkelt första steg. Många objekt har enbart engelska etiketter och beskrivningar.

När man läser en Wikipediaartikel kan man gå vidare till dess Wikidataobjekt och tänka på huruvida det finns något som saknas. Har objektet minst ett "instans av" (P31) uttalande? Är uttalandena källbelagda? Det finns många objekt som är helt eller nästintill tomma, och därför enkla att förbättra.

Slå upp ditt favoritmuseum, -bibliotek eller -författare på Wikidata. Vad finns där, och hur kan det förbättras? Har museet geografiska koordinater? Är författarens födelsedatum källbelagt? Är bibliotekets officiella webbplats länkad från Wikidataobjektet?

Finns det någon lista över vilka egenskaper som kan läggas till?

Svar

Alla egenskaper finns listade [[[Wikidata:List_of_properties](#)|på denna sida på Wikidata]]. Det finns även verktyg där man kan bläddra bland egenskaperna på ett mer visuellt tilltalande sätt, som [<https://prop-explorer.toolforge.org/> Prop explorer] och [Propbrowse <https://hay.toolforge.org/propbrowse/>]