

[Workshop Wikidata pro knihovníky](#) [1]

Autor:

- [JANSOVÁ, Linda](#) [2]

Číslo:

- [2020, ročník 29, číslo 1](#) [3]

Rubrika:

- [Z regionů](#) [4]

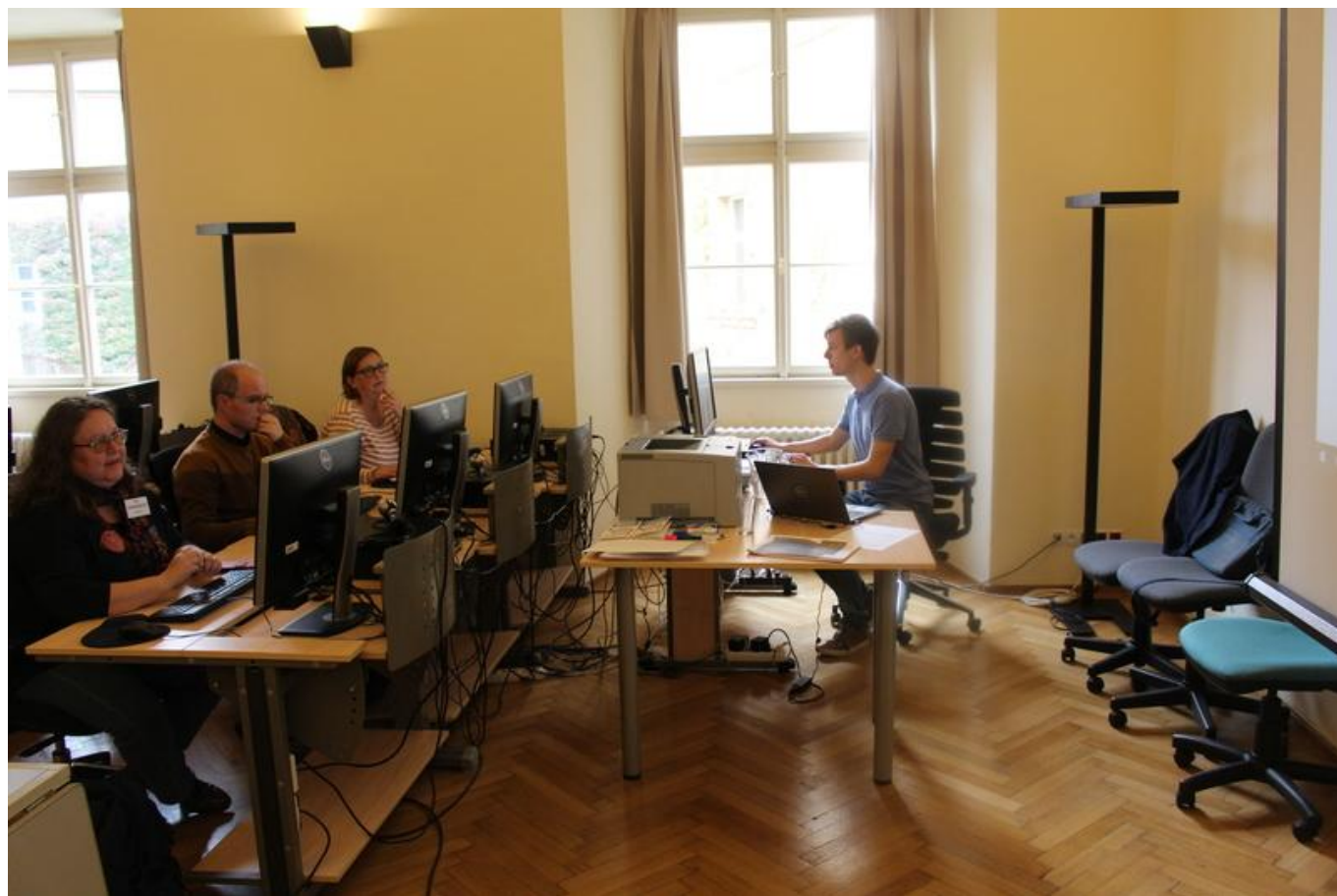
Klíčová slova:

- [propojená data](#) [5], [Wikidata](#) [6], [Wikimedia ČR](#) [7]

Region:

- [01 Praha](#) [8]

Dne 1. listopadu 2019 uspořádala [pražská organizace SKIP](#) [9] ve spolupráci se spolkem [Wikimedia ČR](#) [10] a [Národní knihovnou ČR](#) [11] v Klementinu workshop nazvaný [Wikidata pro knihovníky](#) [12]. Akce, kterou po lektorské stránce zajistil Mgr. Vojtěch Dostál ze spolku Wikimedia ČR, přilákala celkem 15 účastníků.



[13]

Vojtěch Dostál při výkladu (foto: Linda Jansová, Národní knihovna ČR)

Proč vznikla Wikidata

V. Dostál nejprve na příkladu Klementina objasnil rozdíl mezi Wikipedií (viz [heslo Klementinum](#) [14]) a Wikidaty (viz [položku Klementinum](#) [15]). Ačkoliv Wikipedie vznikla relativně nedávno – v roce 2001 –, dnes je všeobecně používaným zdrojem informací a je pokládána za stejně samozřejmou věc jako např. vyhledávač Google nebo sociální síť Facebook. Charakteristickým rysem Wikipedie je mnohojazyčnost. Lze v ní najít poměrně podrobné články na řadu témat. Vpravo v hesle se nachází tzv. infobox, který stručně shrnuje základní údaje z hesla (jde často o údaje, které se v čase mění). V levé části hesla jsou k dispozici odkazy na další jazykové verze. Obojí se nejprve vkládalo pouze ručně.

Postupně se začalo ukazovat, že v datové oblasti začíná Wikipedie silně zaostávat. Denny Vrandečič a [Wikimedia Německo](#) [16] proto vytvořili projekt nazvaný Wikidata. Ten vychází ze stejného softwaru jako Wikipedia, tj. z [MediaWiki](#) [17], ale přidává rozšíření [Wikibase](#) [18]. Hlavním stavebním kamenem Wikidat jsou tzv. položky. Jejich podstatnou součástí od počátku tvoří odkazy na seznam hesel z různých jazykových hesel Wikipedie; později začaly být přidávány další údaje, mj. ty, které jsou využívány jako obsah infoboxů v heslech Wikipedie. Obecně lze říci, že položky, které jsou součástí Wikidat, nejsou primárně určeny pro člověka, ale pro další zpracování stroji.

V. Dostál také připomněl, že Wikipedie a Wikidata nejsou jedinými projekty zastřešující organizace [Nadace Wikimedia](#) [19]. Z dalších projektů můžeme jmenovat např. [Wikimedia Commons](#) [20], [Wikizdroje](#) [21] nebo [Wikislovník](#) [22], ale také již zmíněný software MediaWiki. Vše je provozováno na nekomerční bázi, pod svobodnou licenci.

Co jsou Wikidata

Pokud jde o to, co vlastně Wikidata jsou, lze říci, že jde o svobodnou kolaborativní mnohojazyčnou strukturovanou databázi všeho. Umožňuje propojování napříč databázemi. Údaje jsou obvykle zadávány nikoliv v podobě dlouhých řetězců, kterým rozumí pouze člověk, ale ve formě výroků, jimž rozumí i program. Stejně jako Wikipedie jsou i Wikidata mnohojazyčná. Vznikají společnou prací komunity, neexistuje jedna centrální autorita, která by určovala pravidla. Přesto se časem vždy dospěje ke konsensu. Nové náměty jsou konzultovány s komunitou – takto jsou např. do Wikidat přidávány nové vlastnosti (příkladem může být [proces schválení identifikátoru pro hesla z \[23\] knihovnicko-informační terminologické báze TDKIV \[23\]](#)).

O softwarové a hardwarové zabezpečení Wikidat se starají vývojáři z německé Wikimédie ve spolupráci s Nadací Wikimedia. S daty následně může pracovat kdokoliv téměř bez jakýchkoliv dalších podmínek, tj. může je využívat např. i ke komerčním účelům. Wikidata jsou dostupná pod licencí [Creative Commons CC0 \[24\]](#), tedy pod ještě volnější licencí než Wikipedie, jejíž obsah podléhá licenci [Creative Commons Uvedte původ-Zachovejte licenci \[25\]](#). U Wikidat se totiž očekává použití např. v mobilních aplikacích, ve vyhledávacích nebo v počítačových hrách – v těchto případech by uvádění zdroje bylo pochopitelně značně nepraktické.

Jak již bylo částečně naznačeno, Wikidata vznikla kvůli potřebě rychle (prakticky v reálném čase) aktualizovat data, která se často mění (např. počty obyvatel obcí nebo sportovní výsledky), ale také získávat odpovědi na dotazy, které překrývají více vrstev vědění (příkladem je často uváděný dotaz, jaké je největší město řízené ženou, tj. starostkou/primátorkou) a v tradičních encyklopedických heslech se špatně hledají. Naopak Wikidata na tyto otázky mohou přinášet odpovědi díky přehledné datové struktuře využívající propojení jednotlivých položek.

Jako příklad konkrétní položky Wikidat vybral přednášející [položku Jaromír Jágr \[26\]](#). Ukázal na ní identifikátor položky (tvořený písmenem Q, za nímž následují číslice), vlastnosti (jejich identifikátor je opět tvořen číslicemi, ale na rozdíl od identifikátoru položky začíná písmenem P), jejich hodnoty a také zdroje. Je totiž žádoucí, aby byl pokud možno u všech výroků uveden zdroj (dá se tak přehledně sdělit např. i ta skutečnost, že různé zdroje uvádějí různá data narození jednoho člověka). Vše je tedy zaznamenáno ve strukturované podobě v podobě trojic (výroků) položka – vlastnost – hodnota (v řeči propojených dat jde o subjekt – predikát – objekt).

Wikidata zároveň kladou důraz na vizualizace dat a na jejich analýzy. Ty mohou často být hlubší než ve zdrojových databázích (tj. v databázích, ze kterých byly údaje získány) a mohou nabízet širší souvislosti.

Pojetí Wikidat vychází i z toho, jakým způsobem se v poslední době proměnil internet. Wikipedie sice nabízí encyklopedická hesla, ta jsou však dlouhá a málokdo je čte (převažuje práce s textem, kterou můžeme označit výrazem „skenování“). Je to způsobeno i tím, že začátek hesla mají uživatelé k dispozici už ve výsledcích vyhledávání. Navíc se informace stále častěji šíří v podobě videí; na tento trend Wikipedia také reaguje velmi obtížně. Čím dál více dat vzniká na základě umělé inteligence, rozeznávání hlasu a zpracovávání dat na míru. Uživatelé proto stále více očekávají odpovědi na míru.

Spolupráce Wikidat s knihovnami

Tyto trendy si uvědomují i knihovny, proto jich stále více spolupracuje s Wikidaty. Svědčí o tom i [setkání zástupců národních knihoven, které proběhlo na konferenci Wikimania 2019 \[27\]](#). Výrazně se zapojují národní knihovny Německa a Francie, do spolupráce je ale zapojena i Národní knihovna ČR. Lze se domnívat, že tyto aktivity povedou k tomu, že se budou postupně stírat rozdíly mezi různými typy databází. Na druhou stranu nelze říci, že veškeré údaje jsou vhodné k začlenění do Wikidat. Konkrétním příkladem údajů, které není žádoucí zařazovat, jsou např. časové řady s vývojem teplot a obdobné datové sady.

Pokud jde o vlastní spolupráci, vhodným prvním krokem je propojení databází (Wikidat a databáze instituce) pomocí identifikátorů. Propojování lze realizovat různými postupy, přičemž obvyklé je i určité penzum manuální práce (nejčastěji při kontrole a opravách chyb). Poté může následovat import dat; instituce může také začít provozovat vlastní instalaci Wikibase a do ní ukládat svá data.

Spolupráce instituce s Wikidaty není prospěšná jenom pro samotná Wikidata. Spolupracující instituci umožní kontrolu kvality databáze, pomůže jí najít duplicity, umožní rovněž analýzu a vizualizaci dat, kombinaci dat s daty z Wikidat nebo možnost prostřednictvím Wikidat odkazovat ze své databáze do databáze jiné. V tomto duchu probíhá i spolupráce s Národní knihovnou ČR. V roce 2018 byly přidány identifikátory hesel z báze TDKIV, o rok později byla zahájena systematická spolupráce na importu identifikátorů jmenných autoritních záznamů. Tempo importu je poměrně značné – zatímco na podzim 2018 existovalo cca 50 tisíc propojení, na podzim 2019 už těchto propojení bylo přibližně 260 tisíc.

Uvažuje-li instituce o zahájení spolupráce s Wikidaty, je vhodné se nejprve seznámit s [jejími rámcovými pravidly ve vztahu k paměťovým institucím](#) [28].

Praktická část workshopu

Po přednáškové části následovala praktická část workshopu. V ní účastníci upravovali [položku věnovanou spisovateli Borisí Steinbauerovi](#) [29]. Ten v době workshopu měl své [heslo ve Wikipedii](#) [30], ale odpovídající položka ve Wikidatech dosud neexistovala. Účastníci měli také možnost se seznámit s [Wikidata Query Service](#) [31], což je služba umožňující pokládání jednoduchých i velmi sofistikovaných dotazů do Wikidat. U příliš složitých dotazů však může dojít k tomu, že výsledkem je chyba, neboť běžný uživatel má limit na vyhledávání cca jednu minutu. Takové dotazy je třeba přeformulovat tak, aby byly výpočetně méně náročné. Výsledky je pak možné zobrazit v různých formátech; samozřejmě jsou také různé způsoby vizualizace získaných dat. Zajímavým nástrojem pro zadávání dotazů je také [VizQuery](#) [32].

Zpětná vazba

Účastníci byli s workshopem spokojeni, jak naznačily i vyplněné evaluační dotazníky. 60 % respondentů pokládalo akci za velmi přínosnou a 40 % za spíše přínosnou, 73 % z nich bezprostředně po akci uvažovalo o uplatnění získaných poznatků v praxi a přes 90 % respondentů vyjádřilo zájem o účast na další obdobné akci.

URL zdroje: <https://bulletinskip.osvobozena-knihovna.cz/vsechna-cisla/prohlizet-cisla/2020-rocnik-29-cislo-1/workshop-wikidata-pro-knihovniky>

Odkazy

- [1] <https://bulletinskip.osvobozena-knihovna.cz/vsechna-cisla/prohlizet-cisla/2020-rocnik-29-cislo-1/workshop-wikidata-pro-knihovniky>
- [2] <https://bulletinskip.osvobozena-knihovna.cz/vsechna-cisla/autori/jansova-linda>
- [3] <https://bulletinskip.osvobozena-knihovna.cz/vsechna-cisla/prohlizet-cisla/2020-rocnik-29-cislo-1>
- [4] <https://bulletinskip.osvobozena-knihovna.cz/vsechna-cisla/rubriky/z-regionu>
- [5] <https://bulletinskip.osvobozena-knihovna.cz/vsechna-cisla/klicova-slova/propojena-data>
- [6] <https://bulletinskip.osvobozena-knihovna.cz/vsechna-cisla/klicova-slova/wikidata>
- [7] <https://bulletinskip.osvobozena-knihovna.cz/vsechna-cisla/klicova-slova/wikimedia-cr>
- [8] <https://bulletinskip.osvobozena-knihovna.cz/vsechna-cisla/regiony/01-praha>
- [9] <https://www.skipcr.cz/regiony/01-praha>
- [10] <https://www.wikimedia.cz/>
- [11] <https://www.nkp.cz/>
- [12] <https://www.skipcr.cz/regiony/01-praha/akce/skip-pro-praxi/workshop-wikidata-pro-knihovniky/>
- [13] <https://bulletinskip.skipcr.cz/sites/default/files/images/661/jansova1.jpg>
- [14] <https://cs.wikipedia.org/wiki/Klementinum>
- [15] <https://www.wikidata.org/wiki/Q1100429>
- [16] <https://www.wikimedia.de/>
- [17] <https://www.mediawiki.org/wiki/MediaWiki>

- [18] <https://www.mediawiki.org/wiki/Wikibase>
- [19] <https://wikimediafoundation.org/>
- [20] https://commons.wikimedia.org/wiki/Main_Page
- [21] https://en.wikisource.org/wiki/Main_Page
- [22] https://en.wiktionary.org/wiki/Wiktionary:Main_Page
- [23] https://www.wikidata.org/wiki/Wikidata:Property_proposal/TDKIV_term_ID
- [24] <https://www.wikidata.org/wiki/Wikidata:Licensing>
- [25] https://en.wikipedia.org/wiki/Wikipedia:Text_of_Creative_Commons_Attribution-ShareAlike_3.0_Unported_License
- [26] <https://www.wikidata.org/wiki/Q295345>
- [27] https://en.wikipedia.org/wiki/Wikipedia:Wikipedia_Signpost/2019-09-30/In_focus
- [28] <https://databaze.wikimedia.cz/>
- [29] <https://www.wikidata.org/wiki/Q73119194>
- [30] https://cs.wikipedia.org/wiki/Boris_Steinbauer
- [31] <https://query.wikidata.org/>
- [32] <https://hay.toolforge.org/vizquery/>