

České Wikizdroje

cs.wikisource.org

Jan Kameníček,
pro Wikikonferenci

Co patří na Wikizdroje

- Původní texty
 - Autorská práva „vypršela“
 - Svobodné licence
- Překlady původních textů
 - Zveřejněné
 - Vlastní překlady přispěvatelů Wikizdrojů

Základní způsob vkládání textu na Wikizdroje

- Podobný způsob editace, jako na Wikipedii
 - Příklad: [hesla Ottova slovníku naučného](#)
- Zdroje
 - Dostupné texty v elektronické podobě
 - OCR
 - Ruční opisování

Nástroj pro korektury

(tzv. „proofread extenze“)

Postup:

- Nahrát soubor na Commons
- Na Wikizdrojích vytvořit odpovídající stránku ve jmenném prostoru Index
- Po jednotlivých stránkách vkládat a upravovat text ve jmenném prostoru Stránka
- Text převést do hlavního jmenného prostoru

Výhody

- ❑ Snadné OCR
- ❑ Snadná spolupráce více přispěvatelů
- ❑ Jakmile jsou skeny nahrané na Commons, je k nim mnohem snazší přístup
- ❑ Snadná kontrola, zda text opravdu pochází z udaného zdroje
- ❑ Snadná kontrola překlepů a jiných chyb v přepisu
- ❑ Eliminují se některé dobře míněné špatné opravy

Výhody

- ❑ OCR může vkládat bot a člověk jen kontroluje
- ❑ Možnost označit stupeň kontroly
- ❑ Podněcuje k nezávislé kontrole jiným přispěvatelem
- ❑ Umožňuje spolupráci mezi různými jazykovými verzemi Wikisource u multilinguálních publikací

Výhody

- Veškerá příprava a kontrola nehotového textu se odehrává mimo hlavní jmenný prostor
- Snadná kontrola kvality práce přispěvatelů už během práce na konkrétní publikaci
- Pro někoho může být postupná práce po malých kouscích lepší/snadnější
- Možnost konfrontace přepsaného textu s originálním skenem odlišuje Wikizdroje od jiných webů nabízejících digitalizovaný text

Nevýhody

- Nutnost nahrávat skeny na Commons
- Transkluze do hlavního jmenného prostoru je pro nezkušené editory náročnější
- Je o něco obtížnější opravit malé chyby
- Nelze si zkopírovat zdrojový wikikód, jediné stránku po stránce

Nevýhody

- Někdy nastávají potíže při transkluzi více stránek do jednoho textu v hlavním jmenném prostoru:
 - Básně
 - Hypertextové odkazy nebo různá formátování na konci stránky přesahující do textu na následující stránku, tabulky, poznámky
 - Donedávna i rozdělená slova
- Tok textu je narušován (nezobrazenými) čísly stránek

Srovnání jazykových verzí Wikisource

No	Language	Language (local)	Wiki	Good	Total
1	English	English	en	695556	2483065
2	Polish	Polski	pl	662461	690314
3	Russian	Русский	ru	489005	803835
4	German	Deutsch	de	428997	472049
5	French	Français	fr	335454	2717321
6	Chinese	中文	zh	295667	827077
7	Hebrew	עברית	he	176793	344145
8	Italian	Italiano	it	130698	502852
9	Spanish	Español	es	112317	221192
10	Arabic	العربية	ar	79031	148162
11	Czech	Čeština	cs	44796	59795
12	Portuguese	Português	pt	30451	98226

Srovnání jazykových verzí Wikisource

		Page namespace					
	language	all pages	not proof.	problem.	w/o text	proofread	validated
1	fr	2 303 767	732 660	1 088	148 940	1 421 079	423 119
2	en	1 716 490	606 614	29 755	140 401	939 720	376 597
3	pl	468 851	147 434	2 919	30 722	287 776	215 208
4	de	268 008	17 212	126	7 945	242 725	208 161
5	it	323 338	80 792	4 321	23 212	215 013	112 256
4	es	90 041	20 022	161	6 151	63 707	34 183
7	ru	166 627	78 667	1 555	5 232	81 173	10 184
32	cs	7 780	3 470	104	987	3 219	535

Srovnání jazykových verzí Wikisource

		Main namespace				
	language	all pages	with scans	w/o scans	disamb	percent
1	fr	257 063	242 138	10 191	4 734	95,96
2	en	418 070	203 226	208 714	6 130	49,33
3	pl	188 355	180 624	4 040	3 691	97,81
4	de	159 858	41 670	117 423	765	26,19
5	it	76 546	42 653	33 763	130	55,82
4	es	92 671	10 073	81 999	599	10,94
7	ru	474 671	138 090	333 083	3 498	29,31
32	cs	41 493	974	40 235	284	13 181