

ARCHIVACE WEBOVÝCH ZDROJŮ V CITACÍCH VĚDECKÝCH A ODBORNÝCH TEXTŮ

MGR. MARIE HAŠKOVCOVÁ, MGR. MARKÉTA HRDLIČKOVÁ, BCA. LUBOŠ SVOBODA

✉ marie.haskovcova@nkp.cz
marketa.hrdlickova@nkp.cz
lubos.svoboda@nkp.cz
Národní knihovna ČR –
Oddělení Webarchivu
www.webarchiv.cz

Purpose: Authors discuss issues connected with archiving cited web sources in scientific and professional works and propose possible solutions.

Methodology: The study is based on an analysis of foreign and Czech sources and a survey of practice in the field of web archives. It also includes a comparison and evaluation of the most important tools for the preservation of web sources, including synthesis of the acquired knowledge.

Results: The analysis of sources and the comparison of tools resulted in the proposal of procedures for users and web archivists regarding the protection of cited sources from extinction and methods of citing archival links to sources in web archives.

Conclusions: The research showed that the content of cited web sources quickly disappears, which is a problem, especially in terms of preserving the integrity of professional works. Although there is no uniform solution for their storage, the study summarises current knowledge and experience from abroad and offers approaches for protecting cited sources, including the inclusion of archival references in professional citation practice.

Keywords: web archives, citing web sources, dead web sources, link rot, citation standards, comparative analyses

ÚVOD

Citování informačních zdrojů je standardní součástí akademické kultury, opírá se o metody citování a pravidla popisu, jejichž cílem je identifikace citovaného pramene. Samozřejmou součástí citační praxe se staly webové zdroje. Některé citované zdroje mají své tištěné ekvivalenty, jiné ne. Zatímco tištěné dokumenty jsou často vnímány jako stabilnější a trvalejší, elektronické zdroje mohou nabízet snazší přístup a aktuálnější informace. Vzhledem k nestabilitě webového prostředí, v němž často dochází k zániku zdrojů, k proměně obsahu nebo jeho přesunu na jinou URL adresu, jevu označovaného jako **link rot**, může ztráta citovaného webového zdroje významně ohrožovat důvěryhodnost a autenticitu práce, která na takovýto zdroj odkazuje – ať už se jedná o akademickou práci, výzkumný projekt, oborovou databázi nebo třeba osobní archiv. Uchování citovaných webových odkazů je klíčové pro zajištění integrity uchovávání znalostí, bez něj dochází k degradaci akademických informací, která znemožňuje další šíření výsledků vědeckého poznání. Webové archivy proto hledají řešení, jak integritu informací zachovat. Pro tyto účely již vznikla řada nástrojů, jejichž cílem je webové zdroje ochránit, respektive uchovat a získat trvale archivované odkazy. Web není jen zdrojem vědeckých informací nebo prostorem pro publikování akademických prací, součástí výzkumu je mnohdy práce s online nástroji a technologiemi, k nimž patří i postupy strojového učení a umělé inteligence. Softvér může být v oborech, jako je třeba biomedicína, součástí výzkumné metodologie, je součástí citačního aparátu a bez jeho dochování práce ztrácí svou integritu (Teixeira da Silva a Nazarovets 2023). Tento typ vědeckých výstupů tak přináší nové výzvy pro zachování jejich integrity.

V textu navazujeme na naše dosavadní poznatky o zanikajícím webovém obsahu shrnuté v článku *O mrtvých webových zdrojích. Jak identifikovat a sledovat zaniklý webový obsah?* (Svoboda a Haškovcová 2023), jenž zahrnuje průzkum relevantních zahraničních zdrojů reflektující problematiku link rotu. Ty zde rozšiřujeme o zdroje a poznatky v oblasti citování webových zdrojů a jejich uchovávání. Zabýváme se způsoby, jak archivní kopie webového zdroje citovat a jaká jsou stanoviska a doporučení citační normy. Dle vybraných kritérií porovnáváme existující citačně-archivační služby – jak z hlediska uživatele, tedy toho, kdo cituje, tak z pohledu webového archivu,

který citovaný obsah uchovává a může promýšlet design takovéto služby pro vlastní využití. Jak ukážeme dále v textu, jednotné řešení pro uchovávání webových zdrojů v akademických citacích neexistuje a každý navrhovaný způsob má své limity i potenciální rizika. Součástí odborné debaty se stává také zvažování možností a navrhování postupů, jak lze odpovědnost za integritu informací v online prostředí sdílet napříč akademickou obcí. Na základě zjištěných poznatků proto navrhuje doporučení, jak k této problematice přistupovat.

ZANIKAJÍCÍ WEBOVÝ OBSAH A MOŽNOSTI JEHO UCHOVÁNÍ

Rychle rostoucí obsah na webu patří vzhledem ke své značné proměnlivosti z pohledu jeho uchování k nejhroženějšímu. Jeho archivaci se věnují webové archivy od druhé poloviny 90. let. Vzhledem k tomu, že webové prostředí je široce využívanou a masivně se rozrůstající platformou pro veřejnou i soukromou komunikaci, webové archivy se stávají zásadními prameny pro výzkum soudobé společnosti. Kromě akvizice, uchovávání a zpřístupňování dat v rámci národních legislativ, jež přináší mnoho koncepčních a technických výzev, se v posledních letech archivy snaží i o badatelské zhodnocení svého obsahu například s využitím pokročilých postupů strojového učení (Vozár, Haškovcová a Prokopová 2022).

Na možnosti nahlížení mizejícího webového obsahu nejen v oblasti zanikajících citovaných zdrojů jsme se detailně zaměřili ve zmiňovaném článku o mrtvých webových zdrojích (Svoboda a Haškovcová 2023). Představili jsme problematiku link rotu v kontextu zahraničních výzkumů a studií, pokusili jsme se o vymezení pojmu mrtvý web a nastínění možných scénářů, jak ho automatizovaně identifikovat, aby bylo možné sledovat proměny českého webu.

Mizející webový obsah začal být zkoumán právě v souvislosti se zanikáním nebo změnami obsahu citovaných webových zdrojů v akademických pracích, na něž se v tomto textu zaměříme, či v oblasti žurnalistiky. Předmětem výzkumu se stal jev označovaný mezinárodní komunitou jako link rot, který sleduje vztah mezi URL adresou a obsahem. Na

rozdíl od specializovaných průzkumů v rámci určitého definovaného celku, jako je třeba monitorování výskytu mizení odkazů v rámci nějakého vědeckého periodika nebo zpravodajského portálu, konkrétně například analýza odkazovaných zdrojů v *New York Times* (Bowers, Stanton a Zittrain 2021), nebo průzkum webových zdrojů ze sbírek *Library of Congress* (Meehan 2022), jsme se v článku o mrtvých webových zdrojích (Svoboda a Haškovcová 2023) zaměřili na komplexnější klasifikaci. Jejím smyslem je umožnit na základě dlouhodobého sledování a vyhodnocování pojmenovat proměny webu v širším kontextu, jako je například rozmach blogů nebo diskusních fór, které nahradila éra sociálních sítí a podobně. Na rozdíl od link rotu definovaného tím, že odkaz nevede tam, kam má, a je tedy nedostupný, námi definovaný mrtvý webový zdroj není z hlediska obsahu vázaný k doméně. Z tohoto komplexnějšího pohledu webových archivářů lze za mrtvý webový zdroj označit zdroj tehdy, pokud je dostupný pouze z webového archivu. Navrhli jsme aplikaci automatizovaně

detekující zanikající webový obsah *Extinct Websites* (*Extinct Websites* [bez datumu]) na základě porovnávání metadat a stavových kódů v průběhu času, přičemž první výstupy této v českém prostředí unikátní analýzy na vzorku dat jsou dostupné v *Databázi mrtvých webových zdrojů* (*Zpráva* [bez datumu]), jejímž cílem je podávat zprávu o zanikajících webových zdrojích.

Z hlediska zachování integrity informačních zdrojů je však zásadní hledání strategií, jak se vypořádat s jevem link rot, jak ze strany uživatele, tak i ze strany webových archivů. S ohledem na záměr autora může být v některých případech žádoucí kombinovat odkaz na živý web, s tím, že si je vědom rizika zániku nebo změny jeho obsahu, s archivní kopií, která představuje „zmrazený“ archivní materiál tak, aby ukazoval to, co viděl autor v době, kdy s ním pracoval. V případě zdroje kontinuálně uchovávaného ve webovém archivu může případně doložit i jeho proměny v čase. Jak si ukážeme dále, se zánikem odkazů se potýkají webové archivy již řadu let, vznikají, ale i zanikají ná-



WWA



stroje pro ochranu a uchovávání citovaných zdrojů. Problém zachování důvěryhodnosti pramenů, které se opírají o webové zdroje, se týká přirozeně i různých vědomostních databází a encyklopedií, k nimž patří Wikipedie. Zatímco lze předpokládat, že záznamy na této transparentní platformě budou dlouhodobě uchovány a zpřístupněny, nelze to samé říci o odkazovaných zdrojích. Wikipedie se proto snaží tyto webové zdroje uchovávat v americkém *Internet Archive* (*Internet* [bez datumu]). Ten je patrně největším webovým archivem s mezinárodním záběrem, umožňuje komukoli vložit zdroj k archivaci a oproti mnoha jiným institucím, podléhajícím restriktivnějším legislativám, jsou jeho data volně dostupná, což z něj momentálně činí vhodné místo pro uchovávání tohoto typu obsahu. Zůstane však *Internet Archive* navždy otevřeným zdrojem pro demokratické šíření a sdílení informací jako doposud?

Odkazy mohou linkovat do webových archivů automatizovaně softvéroví roboti (v případě Wikipedie je to *InternetArchiveBot*) nebo je mohou zajišťovat autoři hesel manuálně. Ochranou národního obsahu a externích odkazů Wikipedie se začaly iniciativně zabývat i některé národní archivy. Například archivaci externích odkazů portugalské části Wikipedie se věnoval v loňském roce archiv *Arquivo.pt* (*Arquivo.pt* [bez datumu]). Z průzkumu, který realizoval v roce 2023, vyplynulo, že 25 % externích odkazů již nebylo funkčních – ať už z důvodu zániku domény nebo přesunu obsahu (tzv. content drift) (*Arquivo.pt* 2023). Hlavním výstupem bylo vytvoření automatického procesu pro extrakci a sběr externích odkazů, které jsou ukládány do *Arquivo.pt*. Z jiného úhlu pohledu se archivaci svého národního obsahu věnuje švýcarský webový archiv, který se primárně zaměřuje na archivaci samotných hesel, se snahou o dokumentaci jejich verzí (*Wikipedia Articles* 2023). Součástí sklizených dat je i seznam přispěvatelů a změnová metadata.

Mezi výše uvedená rizika patří to, že mnohé volně dostupné služby nebo archivy dat mohou být do budoucna omezeny, zrušeny či zpoplatněny. Některé nástroje, které vznikly na konci 90. let v dobách rozvoje vědecké komunikace na webu, už nejsou aktuální – například dnes již nevyužívaný nástroj *WebCite* (*WebCite* [bez datumu]). Z novějších příkladů omezení dostupnosti určitého nástroje či postupu můžeme zmínit výrazné omezení využití API pro získání dat z někdejší platformy Twitter, nyní X, které archi-

vy mohly využívat pro sklizení tweetů a příspěvků a vytváření jejich archivních verzí. Jako strategické řešení se jeví archivace ve více archivech. Vhodné mohou být díky svému mandátu národní paměťové instituce zaměřené na uchovávání obsahu z hlediska kulturního dědictví, k nimž patří i český webový archiv Národní knihovny – *Webarchiv* (*Webarchiv* [bez datumu]).

V akademickém prostředí sílí tendence k publikování pod otevřenými licencemi. V případě obsahu za heslem nebo za paywallem se sice nejedná o jeho ztrátu, ale výsledek může být pro uživatele velmi podobný. Kromě toho, že je dostupnost obsahu limitována, může heslo bránit samotné archivaci – sklízecí roboti se k obsahu bez umožnění přístupu nedostanou. Další příčinou, proč obsah nemusí být zachycen, je záměrná technická blokáce sklizení webu ze strany provozovatele. Jak lze tedy přispět k lepší ochraně odkazovaných online informací?

Přestože má archivace webu své limity, zastává v této oblasti zásadní roli. K omezením patří například to, že v rámci běžné agendy webových archivů jsou archivovány jen určité domény, obsah není zachycován nepřetržitě a tedy kompletně, nebo že některé formáty lze sklídit jen obtížně anebo vůbec (dynamický obsah, obsah sociálních médií). Dále je potřeba vést v patrnosti, že u archivovaného obsahu se nejedná o kopii, ale o archivní verzi. Při sklizení obsahu robot stahuje jednotlivé elementy webu, které následně znovu skládá – archivní verze tak vzniká v čase a může být složena z více různých verzí původní stránky. Možnosti zpřístupňování archivovaného obsahu definuje legislativa či její národní implementace. V případě českého *Webarchivu* umožňuje všechny archivní kopie zpřístupnit na vyhrazeném terminálu v budově Národní knihovny, přičemž volně dostupné jsou pouze licencované zdroje.

V odborných kruzích se začíná diskutovat o sdílené odpovědnosti (Teixeira da Silva a Nazarovets 2023) – v případě akademického publikování jak ze strany autorů vědeckých textů, tak ze strany redakcí. Linkování zdrojů do webových archivů by se mohlo stát součástí publikační etiky na základě revidovaných zásad a doporučení pro archivaci citovaných digitálních informací. Na situaci, kdy v průběhu recenzního řízení může docházet ke změnám zdrojů, reagoval v minulosti například projekt *Hiberlink* (*Hiberlink* [bez datumu]), který se snažil o zachycení zdrojů

hned v několika fázích (před odesláním, v průběhu řízení a po publikování textu) z důvodu uchování vědecké práce včetně všech jejích změn a aktualizací. K dalším praktickým doporučením patří používání nebo vytváření nástrojů pro automatizované linkování zdrojů do webových archivů nebo třeba snaha o vytvoření co nejkratších URL adres zdrojů. Dlouhé adresy citací mohou být například z hlediska grafické úpravy (sazby) při publikování v odborném periodiku nepraktické. To může vést k jejich krácení prostřednictvím různých nástrojů, což nežádoucím způsobem zvyšuje riziko link rotu (Teixeira da Silva a Nazarovets 2023). Jedná se o jedno z kritérií, které zkoumáme dále v kapitole *Jak fungují služby pro archivaci citovaných webových zdrojů*. Pro ověření existence archivní verze zaniklého webu může badatel využít nástroj Time Travel (*Time Travel* [bez datumu]), který umí vyhledávat napříč několika spojenými webovými archivy. V případě těžko archivovatelného obsahu je doporučováno pořídit alespoň screenshot webové stránky.

CITOVÁNÍ ARCHIVNÍHO WEBOVÉHO ZDROJE

Digitální technologie otevřely široké možnosti pro práci s informacemi a jejich publikování. Webové zdroje však vyžadují jiný přístup ze strany vydavatelů, uživatelů i paměťových institucí, než tomu je u zdrojů tištěných. The Library of Congress na svých stránkách prezentuje metodiku zpracování odborné práce, kde reflektuje skutečnost, že většina studentů dnes přistupuje k primárním zdrojům prostřednictvím digitálních aplikací, což předpokládá využívání online zdrojů pro výzkumné a citační účely (*Getting* [bez datumu]). Také s rostoucím množstvím akademických online vyhledávačů (ResearchGate, Academia.edu, GoogleScholar apod.) výrazně vzrostla citovanost elektronických zdrojů v odborné literatuře.

Publikování v online časopisech s otevřeným přístupem se postupně stává běžnou součástí vědecké publikační praxe. Vznikají nejrůznější iniciativy, které usilují o to, aby výsledky vědecké práce byly publikovány v časopisech s otevřeným přístupem. Jedná se například o Budapeštskou iniciativu z roku 2002, Berlínskou deklaraci o volném přístupu k vědeckým informacím z roku 2003 či Pravidla otevřeného přístupu v progra-

mu Evropské unie pro program *Horizont 2020* (*Budapest* [bez datumu]), která stanoví příjemcům grantů povinnost zajistit otevřený přístup k recenzovaným vědeckým publikacím (u vybraných projektů také k výzkumným údajům) souvisejícím s projektovými výsledky. Vědecká komunikace se posouvá směrem k modelu publikování, který klade důraz na otevřený přístup nejen k online článkům, ale i monografiím.

Citace pramenů je klíčovým prostředkem pro odbornou komunikaci, který podporuje transparentnost, integritu a kontinuitu vědecké práce. Odkazované elektronické zdroje mohou být i hodnotným akvizičním zdrojem webových archivů. Vzájemná interakce mezi vědci probíhá skrze citační aparát. Uvedení použité literatury umožňuje ověřit původnost vědeckých východisek a uvedených tvrzení, situuje výzkum do širšího kontextu a použití relevantních zdrojů zesiluje důvěryhodnost a autoritu článku. Kvantifikace citací primárních pramenů se stala stavebním kamenem pro měření kvality vědecké práce, a tedy i hodnocení dané vědecké oblasti. Na předpokladu četnosti citovanosti konkrétní publikace je postavena metoda scientometrie, jejíž metody kvantitativních indikátorů slouží jako podklad pro hodnocení vědy. V neposlední řadě jde o dodržování autorských práv, a s tím úzce související etiku vědecké práce. Korektní citování zdrojů je projevem respektu k duševnímu vlastnictví autorů.

Citace elektronických zdrojů mají pevně daná pravidla pro uvedení obsahu, formy a struktury údajů. Tato pravidla jsou zakotvena v normě ČSN ISO 690 (010197) *Informace a dokumentace – Pravidla pro bibliografické odkazy a citace informačních zdrojů*. Citační normu lze považovat za obecnější koncept zahrnující soubor základních pravidel a doporučení pro citování zdrojů. Z této normy pak vycházejí zásady pro tvorbu citačních stylů, které jsou aplikovány v různých vědních oblastech nebo vědeckých institucích. Citační styly jsou implementací normy definované jednoznačnými a individuálními pravidly pro citování konkrétních zdrojů, k nimž patří i webové stránky. Přináší detailně rozpracované pokyny, jak prameny popsat.

Jednotlivé citační styly mají svá specifika. Mezi nejběžnější citační styly patří: APA (American Psychological Association) v oblasti společenských věd, dále citační styly Chicago nebo MLA (Modern Language Association) hojně využívané v humanitních a společenských vědách, citační styl IEEE v technických oborech nebo Vancouver či Harvard aplikovaný často v medicíně. Struktura záznamu odkazovaného

zdroje obsahuje vesměs stejné základní prvky. Jsou jimi autor (uvádí se v konvertovaném stavu), název díla, datum, název a URL adresa webové stránky, kde je článek publikován. Některé z citačních stylů požadují uvádět datum, kdy byl zdroj naposledy přístupný. Důvodem je proměnlivost virtuálního prostředí, kdy, jak jsme zmiňovali výše, může obsah zaniknout, proměnit se nebo přesunout.

Důležitým faktorem pro citování elektronických zdrojů je stabilita URL adresy. Ve vědecké publikační praxi bývá často uváděn mezinárodní unikátní a trvalý identifikátor objektu – DOI (*Přidělování* [bez datumu]), prostředek k jednoznačné identifikaci digitálního objektu na internetu. Vzhledem k proměnlivé povaze webových zdrojů se v odborných kruzích diskutuje o možnosti doplnění standardního identifikátoru ještě o jeden dynamický, který by zaznamenával průběh změn elektronického zdroje v čase. Zdroj by měl tedy identifikátory dva (Teixeira da Silva a Nazarovets 2023). Jedním z řešení pro uchování URL adres jsou webové archivy, kde zůstávají uchovány i v případě, že na živém webu již neexistují. S ohledem na link rot je proto žádoucí, aby se součástí citace stal kromě odkazu na živý web i odkaz na archivní verzi do některého z důvěryhodných webových archivů.

Existuje několik nástrojů, které proces generování a organizace citací značně usnadní. V českém prostředí je známý online nástroj pro generování citací, jejich správu a sdílení Citace.com. Ze zahraničních jsou to pak např. softvérový nástroj pro správu citací a vědeckých článků Mendeley, citační manažer Zotero, komerční softvér pro správu bibliografických citací EndNote nebo RefWorks vyvinuté společnostmi Ex Libris a ProQuest company.

V mezinárodním měřítku vznikají různá doporučení (Mills 2018) a aktivity, které se možnosti citace z webového archivu věnují. Takovým příkladem je Internet Archive, který oslovil pro spolupráci na tvorbě citací z webových archivů americkou společnost Modern Language Association (*Using* [bez datumu]). Ta stojí za vznikem MLA citačního stylu, typického pro americké prostředí. MLA style sice nemá zavedenou formu, pravidlo pro citaci zdrojů z WaybackMachine, doporučuje však citovat webový zdroj jako obvykle a závěrem citaci opatřit informací o webovém archivu, například:

McDonald, R. C. „Basic Canary Care.“ *_Robirda Online_*. 12 Sept. 2004. 18 Dec. 2006 [<http://www.robirda.com/cancare.html>]. *_Internet Archive_*. [<http://web.archive.org/web/20041009202820/http://www.robirda.com/cancare.html>].

Shodně tak Chicagský styl doporučuje na závěr připsat informaci o webovém archivu a URL adresu archivované stránky.

Přestože česká norma ISO 690 se způsobům zápisu archivní verze webových stránek podrobně věnuje, v praxi zatím není široce využíván (Farkašová et al. 2023). Obecně bývá doporučováno citovat archivovanou webovou stránku podle zvoleného citačního stylu a záznam doplnit o název webového archivu (např. Webarchiv, Internet Archive) a URL adresu archivní kopie. Česká norma ISO 69019 doporučuje uvádět archivní kopii, datum archivace, tedy datum, kdy byla kopie pořízena, a archivní službu. Datum, kdy byl webový zdroj uložen, je důležitou doplňující informací k tomu, kdy byl citován/zobrazen, a vypovídá o možných diskrepancích v obsahu zdroje, pokud je časový rozdíl větší. Norma se zabývá i možnostmi, jak přistupovat k zastaralým informacím o umístění. V případě, že citovaný informační zdroj v době vytváření citace z webu již zmizel nebo je jeho aktuální verze odlišná, měla by citace obsahovat URL a původní datum zobrazení společně s URL archivní verze, je-li k dispozici. Navrhuje uvádět dvě URL adresy – první odkazuje na primární uložení zdroje (Původně dostupné z: nebo Dříve dostupné z:), druhá na odkaz do webového archivu (Dostupné z:). Pokud nejsou archivní verze identické s citovaným informačním zdrojem, doporučuje uvést datum citace archivní verze. Zároveň upozorňuje na to, že ověření pravosti kopie může být obtížné či nemožné. Přítomnost obou URL adres je pro uživatele důležitá i proto, aby si mohl zobrazit webový zdroj v živé podobě – ten je možné lépe a rychleji procházet než archivní verzi. Ta potom slouží pro případ, kdy je zdroj nedostupný. Jak jsme již zmínili, ne všechny archivy mají své zdroje přístupné online, navíc ne vždy jsou zdroje sklizeny úplně. Nabídnout čtenáři obě verze pomůže lepší dostupnosti citovaných zdrojů i komfortu práce s nimi.

Důležitým aspektem citace je vymezení času, kdy byla archivace provedena – datum záznamu webového zdroje by se tedy měl blížit, v ideálním případě být totožný, s datem citace. Proto je doporučeno, pokud zdroj znovu nearchivujete, vyhledat archivní kopii co nejbližší citovanému datu, viz vyjádření MLA (*Using* [bez datumu]). Archivní verze mají tu výhodu, že obsahují nejen datum, ale i přesný čas (time stamp) uchování. Pracují tedy z větší granularitou, než je vyžadováno u citací živých webových zdrojů.

Z pozice webových archivářů navrhujeme uvádět, pokud je to možné, vždy obě URL adresy – jak ci-

tovaný zdroj, tak jeho archivní kopii. Pokud kopie neexistuje, doporučujeme pokusit se využít některou z níže popsaných dostupných archivačních služeb a kopii vytvořit, ideálně ve více archivech. Trvalé archivní odkazy bývají v praxi citovány jako standardní webové zdroje (Dostupné z:). Není pak zjevné, zda je to proto, že zdroj třeba již není aktivní, kvůli neznalosti normy, jejíž volně dostupný výklad citování archivních verzí webových zdrojů nezmiňuje, anebo vůli autora, který preferuje trvalejší archivovaný zdroj. I z tohoto důvodu je výhodné uvádět obě URL – čtenář se tak může přesvědčit o stavu zdroje. Další výhodou uvedení obou adres je kontinuita s pracemi, které obsahují citaci webového zdroje s URL, když byl aktivní – zkrácené URL z webových archivů totiž původní URL v citaci nemusí obsahovat. Na rizika zkracování URL adres prostřednictvím dostupných služeb – jak pro používání, tak pro práci s archivními verzemi zdrojů – poukazuje i norma ČSN ISO 690.

Norma tedy v rámci záznamu informace o sklize ní webu, který by mohl vypadat: [skliz. 2024-05-29 14:32:09Z] nebo [arch. 2024-05-29 14:32:09Z], již počítá s uvedením přesného časového údaje archivace, pokud je známý. V případě užití trvalých odkazů, např. služby perma.cc, přesný časový údaj známý není. Záznam trvalého archivního odkazu do citace webového zdroje by mohl tedy vypadat takto:

WEBARCHIV. *Webarchiv*. Online. Dostupné z: <https://webarchiv.cz/>. [cit. 2024-05-29]. Archivní kopie: Webarchiv, <https://wayback.webarchiv.cz/wayback/20240529143209/https://www.webarchiv.cz/cs/> [skliz. 2024-05-29 14:32:09Z].

V případě již neexistujícího zdroje takto:

ASOCIACE FILM & SOCIOLOGIE. Cena Pavla Kouteckého. Online. Původně dostupné z: <http://cenakoutecky.cz>. Dostupné z: Webarchiv, <https://wayback.webarchiv.cz/secure/20140607191124/http://www.cenakoutecky.cz/>. [skliz. 2014-06-07 19:11:24Z] [cit. 2024-09-27].

SLUŽBY PRO ARCHIVACI CITOVANÝCH WEBOVÝCH ZDROJŮ

Snahy o uchování webových zdrojů citovaných v akademických pracích se datují už do konce

90. let, kdy vznikl nástroj WebCite, který však pozastavil přijímání odkazů v roce 2019¹ (služba stále poskytuje dřívější archivní linky). V současnosti existuje několik způsobů, jak získat trvalý odkaz. Jednou z možností je využívat citačně-archivační službu, jichž není moc, anebo odkazovat do webových archivů.

V této kapitole se pokusíme představit přehled dostupných zahraničních nástrojů, které byly navrženy (či mohou být využívány) za účelem archivace citovaných webových zdrojů. Dále navrhneme kritéria, abychom mohli jednotlivé služby porovnat z hlediska uživatele – tedy toho, kdo archivní trvalé odkazy cituje, a také v krátkosti i na parametry, které by mohly být zajímavé pro webové archivy, které by službu mohly využít.

Zvolená kritéria pro zhodnocení stavu citačně-archivačních služeb zahrnují následující aspekty: forma služby, verze zdarma a placená verze, možnost generování citací, zkrácený odkaz, správa zdrojů v účtu, kapacitní limit u verze zdarma, rychlost získání linku (nízká – absence získání odkazů, anebo v řádu dní, střední – odkazy mailem stejný den, vysoká – ihned po dokončení sklizeň zdroje, která začíná okamžitě po vložení zdroje), hromadné nahrávání URL adres, extrakce URL ze souboru, možnost vložení metadat, možnost stažení dat, export URL adres archivovaného webu/ů v textovém formátu a uživatelská přívětivost (nízká, střední, vysoká).

Z pohledu webového archiváře, tedy toho, kdo chce archivovat již existující publikované citace, jsou klíčovými kritérii: volná dostupnost (ve smyslu, zda je softvér open source), možnost self hostingu (tzn. zda archiv může provozovat nástroj na vlastní infrastrukturu), API (pro integraci mezi stávajícími softvérovými řešeními), možnost exportu dat (pro účel ukládání na vlastním úložišti), technologie sklizení (jak se web ukládá) a technologie zobrazení (jak se web zobrazuje, což hraje roli při tom, jak moc je web interaktivní, či se jedná o screenshot). Začneme s představením existujících služeb, jimž se budeme věnovat. V této části detailně popíšeme různé nástroje a platformy, které se používají k archivaci webových stránek a akademických citací. Mezi zmiňované nástroje jsme zařadili Perma.cc, Save Page Now, Conifer, Archive.today a CitationSaver a aktuálně zamrzlou službu WebCite. Záměrně jsme si vybrali hlavní archivační služby z důvodů jejich prouživatelské orientace, přestože existují menší nástroje/utility, například na službě GitHub. Do tohoto srovnání jsme je nezařadili právě z důvodů technické obtížnosti užití. Nakonec představíme návrh na parametry a funkcionalitu služby, která by mohla být užitečná v českém kontextu, zejména pro Webarchiv.



PERMA.CC

Perma.cc je nástroj vyvinut a spravován týmem z Harvard Library Innovation Lab působícím na Harvard Law School Library. Umožňuje tvůrcům odborných článků, studentům, akademikům a dalším, archivovat své citované elektronické zdroje při vzniku práce a odkazovat se v ní přímo na archivní kopii. Služba nabízí několik možných způsobů, jak ji používat z pohledu uživatele – individuální uživatel má na výběr poměrně limitovaný účet zdarma (10 zdrojů) a dále je potřeba hradit měsíční předplatné. Perma.cc nabízí pro knihovny, fakulty a akademické časopisy archivaci odkazů zdarma, nicméně instituce musí být schválena a dle webu nejsou příliš jasná kritéria (*Perma.cc user guide/Accounts and Usage Plans* [bez datumu]). Perma.cc také slibuje uložení zdrojů i po skončení předplatného – avšak je důležité si uvědomit, že se situace může změnit (*Perma.cc user guide/FAQ* [bez datumu]). Perma.cc nabízí s hyperlinkem možnost vložit metadata – Název, Popis a Poznámku. Také je možné Perma.cc nainstalovat dle dokumentace na GitHubu na svou vlastní instanci. Rychlost získání odkazu s archivovaným zdrojem je dostatečná, děje se tak, jakmile se dokončí sklizeň. Uživatel může své odkazy spravovat v účtu

a také nahrávat větší množství URL (limit neznámý), což je funkcionalita pro autory akademických prací výhodná. Nemůže však nahrávat soubory, například PDF s akademickou prací tak, aby se z ní citace načetly automaticky a archivovaly se – podobně jako funkcionalita v případě CitationSaveru ad. níže. Nelze také generovat automaticky podobu citací. Výhodou je krátkost citačního odkazu.²

SAVE PAGE NOW

Save Page Now je služba Internet Archive, která neslouží primárně jako nástroj pro archivaci citací, ale je spíše jednoduchým nástrojem, aby běžný uživatel mohl archivovat vybrané weby, či po vytvoření účtu budovat jednoduše kolekce. Přestože získá odkaz hned po skončení sklizně, který probíhá na obrazovce před ním, a je mu dokonce poslán výsledek na mail, nevýhodou se jeví absence zkrácené URL či nemožnost generovat přímo podobu citace.³ Je taktéž možné nahrávat URL hromadně, po založení účtu, avšak pro běžného uživatele je to proces technicky složitější (*Batch* [bez datumu]). Pro hromadná nahrávání lze použít i příkazovou řádku (*Internet* [bez datumu]), což je výhodné pro webové archivy. Na webu *Internet*

Archive se dá dohledat i návod, jak zdroje z Wayback Machine citovat (*Using* [bez datumu]).

CONIFER

Conifer je archivační platforma designovaná pro potřebu běžných uživatelů, která se zaměřuje především na uchovávání webového obsahu ve vysokém rozlišení. Umožňuje archivaci dynamického webu, 3D grafiky a dalšího komplexního webového obsahu (*About* [bez datumu]). Její výhodou je, že uživatelé mohou snadno archivovat webové zdroje, které byly dříve problematické – například profily na sociálních sítích, YouTube videa atd. Platforma se oddělila v roce 2020 ze služby Webrecorder (*Webrecorder* [bez datumu]), která vznikla v roce 2015. Platformu spravuje neziskové sdružení Rhizome. Jedná se o uživatelsky přístupnou webovou archivovací službu, jež k archivaci používá jádro softvér Browsertrix (*Browsertrix* [bez datumu]). Služba Conifer není přizpůsobena pro uchovávání citací – nelze generovat citace, není možné použít zkrácený link, ani nahrávat více zdrojů najednou.⁴

ARCHIVE.TODAY

Archive.today je jednou z nejstarších fungujících nezávislých služeb – funguje od roku 2012. Není veřejně známo, kdo ji vede, či vyvíjí (*Archive.today* [bez datumu]). Služba není uzpůsobena pro použití jako citační nástroj – nelze generovat citace, nahrávat více zdrojů najednou, avšak uživatel získá zkrácený link.⁵

CITATIONSAVER

Služba, kterou provozuje portugalský webový archiv Arquivo.pt, funguje zatím spíše jako sběr citací k archivaci, poněvadž nevrací uživateli archivní linky. S touto funkcí však Arquivo počítá do budoucnosti, proto se nám zdálo vhodné zde službu zmínit. CitationSaver funguje však pouze pro portugalský webový kontext. Služba byla navržena pro zachování integrity vědeckých článků. Jako jediná služba extrahuje citované zdroje z vloženého dokumentu ve formátu PDF, TXT a HTML (uživatel může odeslat URL adresu článku) (*CitationSaver* 2023). Obsah je

archivován pro pozdější přístup. Služba nenabízí generování citací.

WEBCITE (1997 – 2019)

WebCite umožňoval ve své době archivaci většiny typů webového obsahu – HTML stránky, PDF soubory, CSS, JavaScript, formáty digitálních obrázků. Také archivoval metadata o shromážděných zdrojích, jako je čas přístupu, MIME typ a délka obsahu. Byl jednou z prvních služeb, které nabízely archivaci stránek na vyžádání, což je funkce později převzatá mnoha jinými archivačními službami, jako jsou nejstarší Archive.today a Wayback Machine. Ačkoliv WebCite je již nefunkční službou, resp. zamrzlou, neboť pozastavila přijímání zpracování odkazů, je i dnes podnětné se podívat na její funkcionalitu. Jako jediná služba totiž nabízí pokročilejší práci s metadatami, jak můžeme vidět na archivní verzi z Internet Archive (*WebCite Archive* [bez datumu]). Nabízí příklady citací (podobně jako Perma.cc) a pokouší se je definovat (*How does a WebCite-enhanced* [bez datumu]). Poskytovala také zkrácený link⁶ a umožňovala vložit URL či soubor, z něž automatizovaně vyextrahovala citační odkazy (*WebCite page* [bez datumu]).

SHRNUTÍ

Z kritérií, která jsou nejdůležitější, vybíráme pět, z nichž benefituje především uživatel: automatické generování citací (žádná existující služba), možnost nahrát více URL adres (Perma, Save Page Now, CitationSaver), vysoká rychlost archivace on demand (na vyžádání) umožňující získat archivní link okamžitě (Perma, Save Page Now, Conifer, Archive Today), možnost vložení souboru s automatickou extrakcí citovaných zdrojů (CitationSaver) a krátkost URL adres (Perma, Archive Today). Více viz Tabulku 1.

Z pohledu archivu zdůrazňujeme, zda je softvér volně dostupný a nemalou roli hraje také datová politika – možnost selfhostingu dat (Perma, CitationSaver), protože existuje národní zájem na uchovávání archivních dat na domácích uložistiších, aby se zajistil přístup k nim i do budoucna, což je v případě citací webových zdrojů klíčové. Pro bližší porovnání přikládáme srovnávací tabulku kritérií (viz Tabulka 2).

Tabulka 1 Srovnávací tabulka kritérií – uživatel

Kritéria	Citačně-archivační služby					
	Perma.cc	Save Page Now	Conifer	Archive.today	CitationSaver	WebCite (pozastavená služba)
Forma služby	Webová aplikace + rozšíření do prohlížeče	Webová aplikace + rozšíření do prohlížeče	Webová aplikace	Webová aplikace	Webová aplikace	Webová aplikace
Verze zdarma	Ano	Ano	Ano	Ano	Ano	Ano
Placená verze	Ano	Ne	Ano	Ne	Ne	Ano
Možnost generování citací	Ne	Ne	Ne	Ne	Ne	Ano
Zkrácený odkaz	Ano	Ne	Ne	Ano	Ne	Ne
Správa zdrojů v účtu	Ano	Ano	Ano	Ne	Ne	Ne
Kapacitní limit u verze zdarma	10 odkazů	Ne, či není znám	5 GB	50 MB / webový zdroj	Ne, či není znám	Ne, či není znám
Rychlost získání linku	Vysoká	Vysoká	Vysoká	Vysoká	Nízká	?
Hromadné nahrávání URL	Ano	Ano	Ne	Ne	Ano	Ano
Extrakce URL citací ze souboru	Ne	Ne	Ne	Ne	Ano (PDF)	Ano (HTML)
Možnost vložení metadat	Ano	Ne	Ne	Ne	Ne	Ano
Možnost stáhnutí dat – formát	Warc/PNG	Warc/Screenshot	Warc/Wacz	Ano, ale nefunkční – ZIP	Ne	?
Export URL archivovaných webů v textovém formátu	Ano	Ano	Ano	Ne	Ne	Ne
Uživatelská přívětivost	Vysoká	Střední	Vysoká	Nízká	Vysoká	Nízká

Další část srovnávací tabulky obsahuje kritéria z pohledu webového archivu.

Tabulka 2 Srovnávací tabulka kritérií – webový archiv

Kritéria	Citačně-archivační služby					
	Perma.cc	Save Page Now	Conifer	Archive.today	CitationSaver	WebCite (pozastavená služba)
Volně dostupné	Ano	Ano	Ne	Ne	Ano	Ne
Možnost selfhostingu dat	Ano	Ne	Ne	Ne	Ano	Ne
API	Ano	Ano	Ne	Ne	Ne	Ne
Export dat – formát	Warc/Screenshot	Warc/Screenshot	Warc/Screenshot	Komprimovaná data – aktuálně nefunkční	Ne	Ne
Technologie sklizení	Scoop	Heritrix/Brozzler	Heritrix	Chromium ⁷	Heritrix	?
Technologie zobrazení	Replay Web	Wayback Machine	Replay Web	Webové, interaktivní	Solr Wayback	?
GitHub	https://github.com/harvard-lil/perma	Ne	https://github.com/Rhizome-Conifer/conifer	https://github.com/wabarc/archive.is	https://github.com/arquivo/CitationSaver	Ne

JAK BY MOHLA VYPADAT ČESKÁ CITAČNĚ-ARCHIVAČNÍ SLUŽBA

Na základě průzkumu a zhodnocení parametrů zahraničních služeb jsme se pokusili nadizajnovat scénář, jak by vypadal ideální případ užití služby poskytující jak citační, tak archivační zázemí pro uživatele. Mohl by vypadat následovně:

- ▶ uživatel, který píše odbornou práci, si online vyhledá citačně-archivační službu;
- ▶ do webového formuláře zadá buď jednotlivý webový zdroj, anebo více webových zdrojů, anebo nahraje text hotové práce v souboru;
- ▶ služba umožní uživateli vyplnit metadata potřebná pro citaci, anebo upravit automaticky vyplněná metadata následně během procesu;
- ▶ služba vyhodnotí, zda webové zdroje již má a ty nabídne uživateli, anebo je *on demand* zaarchivuje;
- ▶ služba dle domény rozpozná základní topologii webového obsahu (akademický článek, sociální síť, či jiný webový zdroj) a dle toho citace připraví;
- ▶ uvědomí uživatele o finalizaci procesu a nabídne mu citace obsahující archivní trvalé odkazy dle běžných citačních stylů (z nichž si uživatel vybere) stáhnutelné např. ve formátu csv. v podobě načrtnuté v předchozí kapitole pro příslušnou normu.

V rámci doplňkových funkcionalit by uživatel mohl mít účet na správu zdrojů – *fine tuning* citací – a mohl by tak citace přizpůsobovat a ukládat, jako v případě citačních manažerů, o nichž se zmiňujeme výše.

Z pohledu zavádění citačně-archivačního nástroje ve Webarchivu jsme identifikovali několik problémů. Sklizeň **on demand**, tedy způsob, jak uživateli doručit dostatečně rychle archivní link, naráží na nedostatečné technické zázemí, které je uzpůsobeno především velkým a dlouhodobým sklízním. Situaci by mohl změnit přechod na sklízecí službu Browsertrix. Dále by bylo nutné řešit délku stávajícího URL.⁸ Musel by také vzniknout softvér pro různé podoby generování a rozpoznávání citací – zde se nabízí případná spolupráce s dalšími institucemi, které by jimi již disponovaly. Pro zpracovávání souborů (např.

PDF) pro extrakci citací webových zdrojů z prací by bylo možné využít stávající nástroj (CitationSaver) a integrovat ho.

K podnětům vedoucím k vhodné citační praxi archivačních linků v českém kontextu bychom zařadili nutnost uvedení pole pro archivní link u webových zdrojů na generátorech citací, jako je například citace.com. Dále seznámit odbornou veřejnost s touto problematikou a nabídnout jim vhodný nástroj – vyvinout svůj či doporučit vhodnou praxi. Oslovit odborníky, kteří seznamují studenty se základy psaní akademických prací a vybavit je nástroji či českým manuálem, a vybídnout odborné časopisy, aby požadovaly citování trvalých archivních odkazů ve svých článcích v recenzních řízeních.

ZÁVĚR

Na základě průzkumu zanikajícího obsahu, zahraniční praxe v oblasti archivace zdrojů citovaných v akademických pracích a porovnání nejzásadnějších nástrojů jsme zjistili, že momentálně neexistuje vhodná služba, která by splňovala citační i archivační kritéria. Pokusili jsme se o analýzu a výčet těchto kritérií i o návrh, jak by mohla citačně-archivační služba webových zdrojů a její uvedení do citační praxe v českém prostředí vypadat.

Pokud by se Webarchiv pokusil sehrát významnější roli v archivování citací a poskytování trvalých odkazů, mělo by to několik pozitivních dopadů. Národní zájem na umenšování a předcházení link rotu, česky také vyhívání odkazů, lépe možná úhynu odkazů či vysychání odkazů v akademických pracích je zřejmý, tedy zabránit umírání zdrojů potřebných pro akademiky a udržet kontinuitu bádání.

Dalším aspektem je budování bohemikálních odborných kolekcí Webarchivu v souladu s jeho náplní, navíc obohacených o cenná metadata – ze stejného důvodu se Webarchiv soustředí na uchovávání citací českých článků Wikipedie. V neposlední řadě by služba mohla díky integraci do akademické praxe zasáhnout širší vrstvu uživatelů než doposud, či oslovit uživatele z jiných odvětví, protože vysychání odkazů, jak jsme již zmiňovali, se týká třeba i zdrojů v mediálních výstupech.

Článek byl recenzovaný

DEDIKACE

Recenzovaný článek vznikl na základě institucionální podpory dlouhodobého koncepčního rozvoje výzkumné organizace Národní knihovna České republiky poskytované Ministerstvem kultury ČR.

POZNÁMKY

- ¹ Dle archivní kopie se oznámení o nepřijímání odkazů objevilo červenci 2019, lze však předpokládat že služba měla výpadky již dříve.
- ² Ukázka citačního linku Perma.cc: <https://perma.cc/X4E2-548U>.
- ³ Ukázka archivního záznamu Save Page Now: <https://web.archive.org/web/20240624055701/https://webarchiv.cz/en/>.
- ⁴ Ukázka archivního záznamu Conifer: <https://conifer.rhizome.org/zlodej/test0/20240626154507/https://www.webarchiv.cz/cs/>.
- ⁵ Ukázka archivního záznamu Archive.today: <https://archive.ph/z8zQY>.
- ⁶ Ukázka archivního záznamu WebCite: <http://www.webcitation.org/5Kt3PxfFl>.
- ⁷ *Archive.is blog*. [bez datumu]. Archive.today. Online. Dostupné na: <https://blog.archive.today/post/618635148292964352/what-scraper-or-headless-browser-are-you-using-it>. [zobrazené 2024-07-02].
- ⁸ Ukázka archivního záznamu Webarchiv: <https://wayback.webarchiv.cz/wayback/20240417131458/http://webarchiv.cz/cs/>.

SEZNAM POUŽITÉ LITERATURY

- About*. [bez datumu]. Conifer. Online. Dostupné na: https://conifer.rhizome.org/_faq. [zobrazené 2024-06-29].
- Archive.today: On the trail of the mysterious guerrilla archivist of the Internet*. [bez datumu]. Gyrovague. Online. Dostupné na: <https://gyrovague.com/2023/08/05/archive-today-on-the-trail-of-the-mysterious-guerrilla-archivist-of-the-internet/>. [zobrazené 2024-06-29].
- Arquivo.pt*. [bez datumu]. Online. Dostupné na: <https://arquivo.pt/>. [zobrazené 2024-06-29].
- Arquivo.pt preserves Wikipedia citations*. 2023. Arquivo.pt. Online. 2023-08-18. Dostupné na: <https://sobre.arquivo.pt/en/tag/wikipedia-en/>. [zobrazené 2024-06-29].
- Batch process Google Sheets using archive.org services*. [bez datumu]. Archive.org. Online. Dostupné na: <https://archive.org/services/wayback-gsheets/>. [zobrazené 2024-06-29].
- BOWERS, John; CLARE STANTON a JONATHAN ZITTRAIN. 2021. *What the ephemerality of the Web means for your hyperlinks*. Columbia Journalism Review. Online. 2021-05-21. Dostupné na: <https://www.cjr.org/analysis/linkrot-content-drift-new-york-times.php>. [zobrazené 2024-06-24].
- Browsertrix Crawler*. [bez datumu]. Webrecorder. Online. Dostupné na: <https://webrecorder.net/tools#browsertrix-crawler>. [zobrazené 2024-06-29].
- Budapest Open Access Initiative*. [bez datumu]. Online. Dostupné na: <https://www.budapestopenaccessinitiative.org/>. [zobrazené 2024-06-29].
- CitationSaver preserves citations to web resources*. 2023. Arquivo.pt. Online. 2023-04-07. Dostupné na: <https://sobre.arquivo.pt/en/new-citationsaver-service-preserves-cited-web-content/>. [zobrazené 2024-06-29].
- Extinct Websites*. [bez datumu]. WebarchiveCZ. Online. Dostupné na: <https://github.com/WebarchivCZ/extinct-websites>. [zobrazené 2024-06-29].
- FARKAŠOVÁ, Blanka et al. 2023. *Výklad normy ČSN ISO 690*. Online. Brno: Citace.com. Dostupné na: <https://www.citace.com/Vyklad-CSN-ISO-690-2022.pdf>. [zobrazené 2024-06-29].
- Getting Started with Primary Sources*. [bez datumu]. Library of Congress. Online. Dostupné na: <https://www.loc.gov/programs/teachers/getting-started-with-primary-sources/>. [zobrazené 2024-06-29].
- Hiberlink*. [bez datumu]. Online. Dostupné na: <https://github.com/hiberlink>. [zobrazené 2024-06-29].

- How does a WebCite-enhanced reference look like?.* [bez datumu]. WebCite. Online. Dostupné na: https://webcitation.org/index#How_look. [zobrazené 2024-06-29].
- Internet Archive Developer Portal.* [bez datumu]. Online. Dostupné na: <https://archive.org/developers/index.html>. [zobrazené 2024-06-29].
- MEEHAN, Olivia. 2022. *Diving into Digital Ephemera: Identifying Defunct URLs in the Web Archives.* Library of Congress Blogs. Online. 2022-08-03. Dostupné na: <https://blogs.loc.gov/thesignal/2022/08/diving-into-digital-ephemera-identifying-defunct-urls-in-the-web-archives/>. [zobrazené 2024-06-24].
- MILLS, Elizabeth Shown. 2018. *Citing the Wayback Machine.* Evidence Explained: Historical Analysis, Citation & Source Usage. Online. 2018-07-05. Dostupné na: <https://www.evidenceexplained.com/quicktips/citing-wayback-machine>. [zobrazené 2024-06-29].
- Nařízení Evropského parlamentu.* Online. Dostupné na: https://ec.europa.eu/research/participants/data/ref/h2020/legal_basis/rules_participation/h2020-rules-participation_cs.pdf. [zobrazené: 2024-06-29].
- Perma.cc user guide / Accounts and Usage Plans.* [bez datumu]. Perma.cc. Online. Dostupné na: <https://perma.cc/docs/accounts#usage-plans>. [zobrazené: 2024-06-29].
- Perma.cc user guide / FAQ.* [bez datumu]. Perma.cc. Online. Dostupné na: <https://perma.cc/docs/faq>. [zobrazené: 2024-06-29].
- Přidělování DOI (CrossRef).* [bez datumu]. Online. Dostupné na: <https://www.lib.cas.cz/pro-redakce/pridelovani-doi-crossref/>. [zobrazené. 2024-06-29].
- SVOBODA, Luboš a Marie HAŠKOVCOVÁ. 2023. O mrtvých webových zdrojích: Jak identifikovat a sledovat zaniklý webový obsah?. *ITlib*. Online. roč. 23, č. 3 – 4, s. 37 – 41. Dostupné na: <https://doi.org/http://doi.org/10.52036/1335793X.2023.3-4.37-41>. [zobrazené 2024-06-26].
- TEIXEIRA DA SILVA, Jaime A. a Maryna NAZAROVETS. 2023. Archiving website-based references in academic papers: Problems caused by reference rot, potential solutions and limitations. *Learned Publishing*. Online. roč. 36, s. 477 – 487. Dostupné na: <https://onlinelibrary.wiley.com/doi/full/10.1002/leap.1560>. [zobrazené 2024-06-24].
- Time Travel.* [bez datumu]. Online. Dostupné na: <https://timetravel.mementoweb.org/>. [zobrazené 2024-06-30].
- Using the wayback machine.* [bez datumu]. Internet Archive. Online. Dostupné na: <https://help.archive.org/help/using-the-wayback-machine/>. [zobrazené 2024-06-29].
- VOZÁR, Zdenko; Marie HAŠKOVCOVÁ a Andrea PROKOPOVÁ. 2022. Internet jako pramen výzkumu: Přístup k archivovaným webovým zdrojům a možnosti jejich zpracování. *Teorie vědy*. Online. roč. 44, č. 1, s. 59 – 87. Dostupné na: <https://cejsh.icm.edu.pl/cejsh/element/bwmeta1.element.48227f3a-a0de-4be9-9d09-934bce4f5c09>. [zobrazené 2024-09-29].
- Webarchiv.* [bez datumu]. Online. Dostupné na: <https://www.webarchiv.cz/>. [zobrazené 2024-06-29].
- WebCite.* [bez datumu]. Online. Dostupné na: <https://webcitation.org/>. [zobrazené 2024-06-29].
- WebCite Archive Form.* [bez datumu]. WebCite. Online. Dostupné na: <https://web.archive.org/web/20170201091742/https://webcitation.org/archive.php>. [zobrazené 2024-06-29].
- WebCite page combing form.* [bez datumu]. WebCite. Online. Dostupné na: <https://webcitation.org/comb.php>. [zobrazené 2024-06-29].
- Webrecorder.* [bez datumu]. Online. Dostupné na: <https://webrecorder.net/>. [zobrazené 2024-06-29].
- Wikipedia articles.* 2023. Online. Dostupné na: <https://www.nb.admin.ch/snl/en/home/collections/library-collections/wikipedia.html>. [zobrazené 2024-06-29].
- Zpráva o mrtvých webech: Databáze mrtvých webových zdrojů.* [bez datumu]. Online. Dostupné na: <https://www.webarchiv.cz/cs/mrtve-weby>. [zobrazené 2024-06-29].