

Zpravodaj Československého sdružení uživatelů TeXu

Tereza Vrabcová

Digitální archivace Zpravodaje CSTUGu

Zpravodaj Československého sdružení uživatelů TeXu, Vol. 32 (2022), No. 1-4, 11–17

Persistent URL: <http://dml.cz/dmlcz/151105>

Terms of use:

© Československé sdružení uživatelů TeXu, 2022

Institute of Mathematics of the Czech Academy of Sciences provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these *Terms of use*.



This document has been digitized, optimized for electronic delivery and stamped with digital signature within the project *DML-CZ*:
The Czech Digital Mathematics Library <http://dml.cz>

Je-li člověk hrdý na svůj výtvar, chce jej sdílet s co největším množstvím lidí. A tak nemůže být divu, že ζ TUG sáhl po příležitosti digitálně archivovat Zpravodaj a zpřístupnit své články širší odborné veřejnosti v rámci České digitální matematické knihovny (DML-CZ). Tento článek představuje technické řešení a výsledky této digitalizace společně se statistickými zajímavostmi, na které se přišlo v jejím průběhu. Jedná se o interpretaci autorčiny přednášky na valném shromáždění ζ TUGu 14. května 2022 [1].

Klíčová slova: digitalizace, DML-CZ, statistika

Zpravodaj Československého sdružení uživatelů $\text{T}_{\text{E}}\text{X}$ u (ζ TUG) tímto vydáním počíná 32. ročník své publikace. Během této doby v něm bylo publikováno velké množství článků pojednávajících o složitostech a krásách $\text{T}_{\text{E}}\text{X}$ u a mnoha dalších nadstavbách a odvozeninách tohoto programu pro počítačovou sazbu, kolem kterého bylo toto sdružení utvořeno. Mnoho hodin práce bylo věnováno jejich napsání, editaci a přípravě pro jejich publikování. Je tedy pochopitelné, že během tohoto procesu vyvstala touha zachovat informace a detaily uložené v těchto stránkách, stejně tak, jako pokusit se rozšířit o nich povědomí v rámci širší odborné veřejnosti našich dvou zemí.

Česká digitální matematická knihovna (DML-CZ) je evropský projekt (2005 až 2009) zajišťující akvizici, digitalizaci, archivaci a vyhledávání nad československými matematickými texty. V roce 2019 tým DML-CZ nabídl vzhledem k významu $\text{T}_{\text{E}}\text{X}$ u pro matematiku Zpravodaj ζ TUG digitálně archivovat a zpřístupnit v rámci knihovny, která usnadňuje dosažitelnost těchto článků pro širší veřejnost. Na valném shromáždění 2020 bylo Vítém Novotným v prezentaci *Konverze Zpravodaje pro Českou digitální knihovnu* [2] navrženo řešení tohoto problému.

Průběh archivace

Nejprve bylo potřeba připravit články ve formátu přijatelném pro DML-CZ. Každé číslo bylo rozděleno na jednotlivé články, pro které pak byly připraveny následující dokumenty:

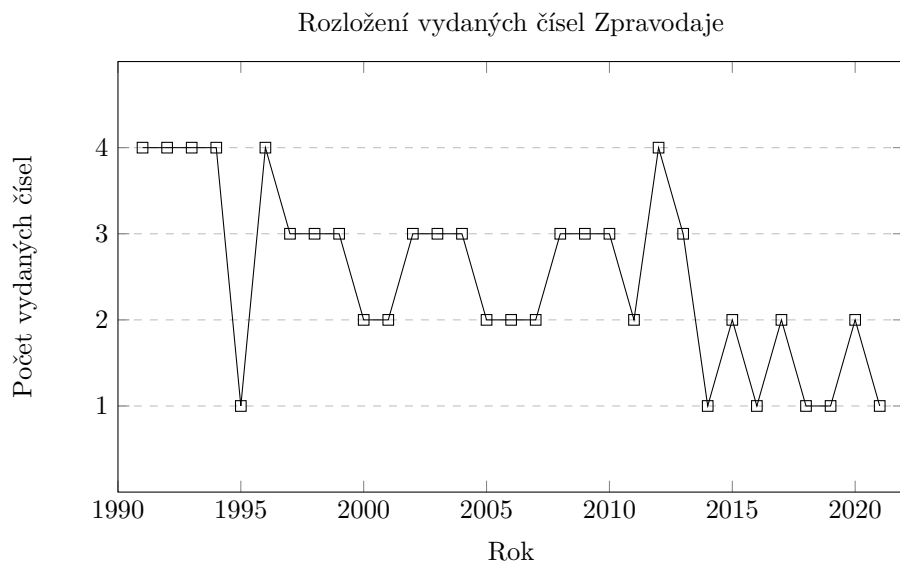
- **meta.xml:** Tento soubor obsahuje informace o daném článku: jméno článku a autora, jazyk článku, typ článku, rozsah stránek, abstrakt a mnoho dalších.
- **source.pdf:** Tento soubor obsahuje článek jako takový.
- **references.pdf:** Tento soubor obsahuje bibliografické odkazy článku. Cílem bylo vytvoření co nejbohatších metadat pro usnadnění vyhledávání a jednoznačnosti odkazů.

Ukázky XML souborů je možné si prohlédnout na čtvrté straně mé prezentace z valného shromáždění ζ TUGu 14. května 2022 [1]. Výsledky této archivace je již v tuto chvíli možno procházet na stránkách DML-CZ [3].

Statistiky Zpravodaje

Knihovna DML-CZ na svých stránkách aktuálně nezobrazuje všechna metadata připravená při archivaci tohoto časopisu. Prohlédněme si tedy tato data formou statistik a zajímavostí objevených při archivaci. Pokud nebude stanoveno jinak, grafy v tomto článku zobrazují statistiky odborných článků na rozdíl od grafů v prezentaci z valného shromáždění, které zobrazovaly statistiky všech článků. Zobrazovaná data pocházejí z let 1991–2021.

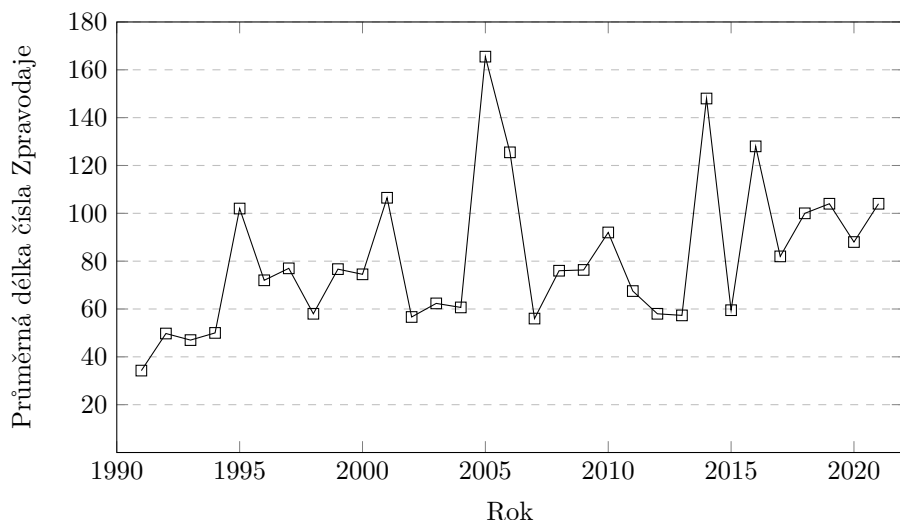
Během let 1991–2021 bylo ve 31 ročnících celkově vydáno 78 čísel Zpravodaje. Dle mediánu i modu se počet čísel pohyboval na třech číslech každý rok, průměr se blížil spíše ke dvěma a půl (2,52) číslům. V posledních letech se pak počet čísel pohyboval mezi jedním až dvěma čísly.



Průměrná délka čísla Zpravodaje naopak v průběhu let rostla. Medián i modus čísla se rovná 64 stránkám, průměrný počet stránek všech čísel pak je 72,55¹. Nejdelším vydáním je číslo 2–4 z roku 2005 [4] s úctyhodnými 239 stránkami.

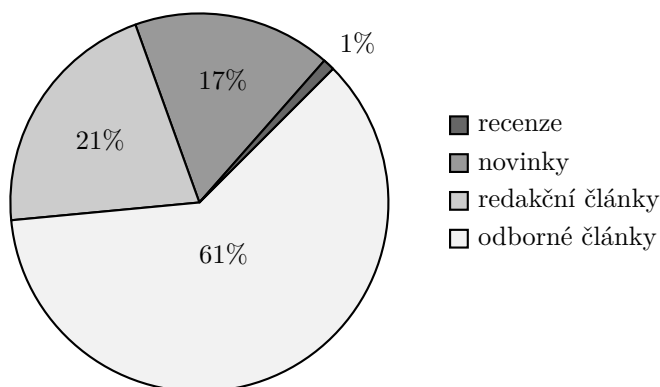
¹Budeme-li pokládat dvojčísla, trojčísla a čtyřčísla za dvě, tři a čtyři separátní čísla, průměrný počet stran je pak 59,57, medián se rovná 57 stránkám a modus zůstává na 64 stránkách.

Průměrné délky čísel Zpravodaje na rok



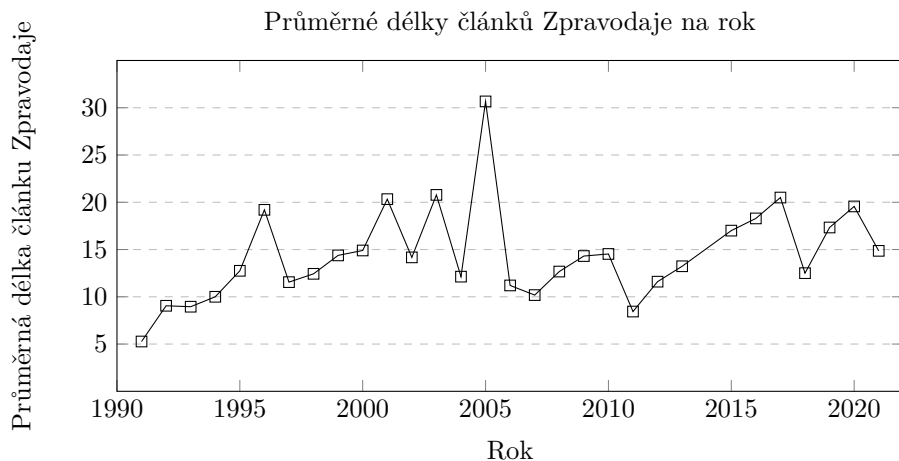
Víme tedy, kolik máme čísel a jak jsou dlouhá, ale co na nás čeká uvnitř? Každý digitalizovaný článek spadá do jedné ze čtyř kategorií – redakční článek, novinky, recenze, či odborný článek. V množství vítězí kategorie odborných článků s více než třemi pětinami z celkového počtu 645 článků (391 článků), na posledním místě se pak nachází kategorie recenze se sedmi články.

Rozložení kategorií článků Zpravodaje



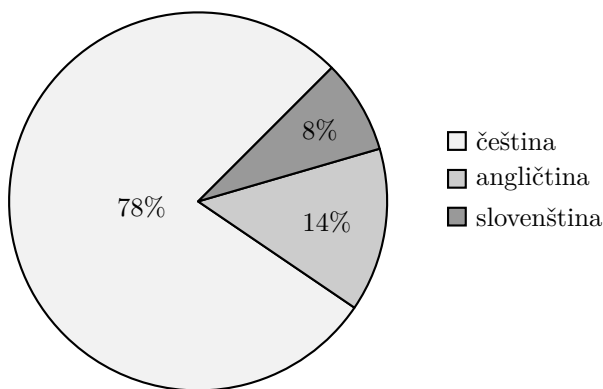
Zaměříme se nyní na kategorii odborných článků. Nejčastější délka jednoho takového článku se pohybuje na sedmi stránkách, hodnota mediánu pak na

devíti. Pro výpočet střední hodnoty radši odeberme neobvykle dlouhé články, jako například článek publikovaný v roce 2014, *TEX pro pragmatiky* [5] od Petra Olšáka se 148 stránkami, či *CS TUG FAQ* [6] z roku 2005 s 228 stránkami. Pak průměrný článek nabývá délky 9,48 stránky.



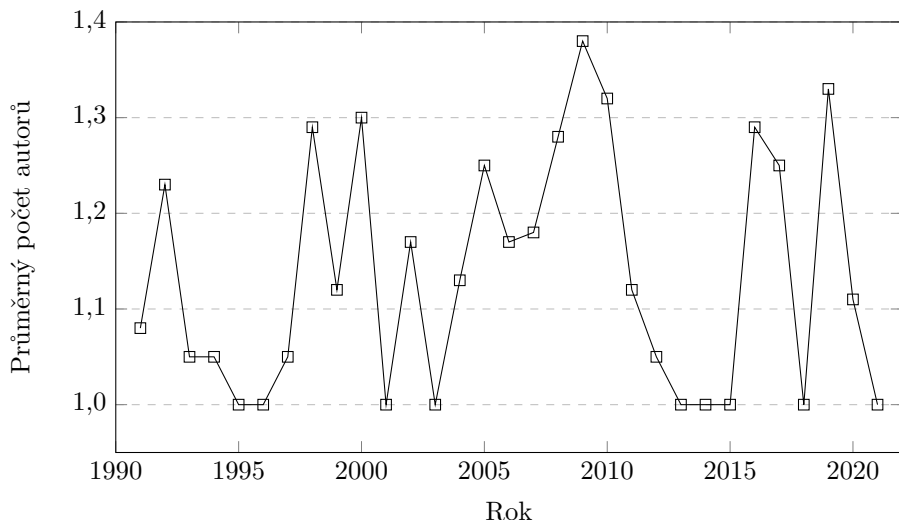
Jazykové rozložení článků může být pro některé překvapující. Na prvním místě s nejvyšší četností sedí čeština s více než třemi čtvrtinami článků (307), na druhém angličtina (53) a slovenština přichází na řadu poslední (31).

Rozložení jazyků článků Zpravodaje



Na druhou stranu překvapující snad nebude průměrný počet autorů jednoho článku, kde se medián a modus drží na jedničce, průměrná hodnota pak na 1,14.

Roční průměr počtu autorů na jeden článek Zpravodaje



Na psaní odborných článků ve Zpravodaji se podílelo celkem 180 unikátních osob: 165 mužů a 15 žen.

Bibliografické zajímavosti

Množství bibliografických odkazů využitých v článku závisí na uvážení autora, kvantifikovat je pouze průměrem přes všechny články se zdá nedostatečné. Podívejme se tedy radši na články s nejbohatšími bibliografiemi z našeho výběru:

1. *T_EX na školách? Samozřejmě ano! Příklad užití T_EXu na Fakultě informatiky Masarykovy university* (2017, 50 odkazů)
2. *Sadzba bibliografie podľa normy ISO 690 v systéme L^AT_EX* (2016, 42 odkazů)
3. *E-T_EX: Sprievodca budúcimi rozšíreniami T_EXu* (1994, 35 odkazů)

Informace ve článcích Zpravodaje nebyly čerpány pouze z těch nejnovějších zdrojů v době jejich psaní, jak je možné vidět při procházení nejstarších citovaných bibliografických odkazů:

1. *The Abacus, in Its Historic and Scientific Aspects* (1886)
2. *On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling* (1900)
3. *A History of Japanese Mathematics* (1914)
4. *Nauka o sazbe obyčejné, tabulkové, matematiky a chemie* (1925)
5. *Tibetan Yoga And Secret Doctrines* (1935)

Z důvodu tematického zaměření Zpravodaje je přirozené, že určité zdroje informací jsou citovány častěji než ostatní. Ukažme si, které odkazy byly citovány nejčastěji napříč odbornými články:

1. *The T_EXbook* (30 citací)
2. *T_EXbook naruby* (29 citací)
3. *Typografický systém T_EX* (16 citací)
4. *The pdfT_EX user manual* (14 citací)
5. *T_EX: The Program* (11 citací)

Závěr

Díky digitalizaci Zpravodaje $\mathcal{C}\mathcal{S}TUGu$ je procházení a nalézání článků jednodušší než kdy předtím. Proč tedy nezavítat na stránky DML-CZ a neobjevit polozapomenuté moudrosti z let předešlých? (Samozřejmě nejlépe až po dočtení tohoto čísla.)

Odkazy

1. VRABCOVÁ, Tereza. *Digitální archivace Zpravodaje $\mathcal{C}\mathcal{S}TUG$* [online]. 2022. [vid. 2022-09-30]. Dostupné z: <https://www.cstug.cz/informace/zpravy/2022-04-01-valna-hromada-2022/files/vrabcova-dmlcz-2022-05-14.pdf>.
2. NOVOTNÝ, Vít. *Konverze Zpravodaje $\mathcal{C}\mathcal{S}TUGu$ pro Českou digitální matematickou knihovnu (DML-CZ)* [online]. 2021. [vid. 2022-09-30]. Dostupné z: <https://www.cstug.cz/informace/zpravy/2021-02-06-valne-shromazdeni-2020/files/novotny-konverze-2021-03-06.pdf>.
3. *Zpravodaj Československého sdružení uživatelů T_EXu* [online]. Česká digitální matematická knihovna. [vid. 2022-09-30]. Dostupné z: <https://www.dml.cz/handle/10338.dmlcz/148724>.
4. *Zpravodaj Československého sdružení uživatelů T_EXu* [online]. 2005, roč. 15, č. 2–4 [vid. 2022-09-30]. ISSN 1211-6661. Dostupné z: <https://www.dml.cz/handle/10338.dmlcz/149988>.
5. OLŠÁK, Petr. T_EX pro pragmatiky. *Zpravodaj Československého sdružení uživatelů T_EXu* [online]. 2014, roč. 24, č. 1–4 [vid. 2022-09-30]. ISSN 1211-6661. Dostupné z DOI: 10.5300/2014-1-4/1.
6. $\mathcal{C}\mathcal{S}TUG$ FAQ. *Zpravodaj Československého sdružení uživatelů T_EXu* [online]. 2005, roč. 15, č. 2–4, s. 94–331 [vid. 2022-09-30]. ISSN 1211-6661. Dostupné z DOI: 10.5300/2005-2-4/94.

Summary: Digital Archival of the ζ TUG Bulletin

When a person is proud of their creation, they want to share it with the largest amount of people possible. Therefore, it cannot be a surprise that ζ TUG took the opportunity to digitally archive the ζ TUG Bulletin and make its articles accessible to the general public with the help of the Czech Digital Mathematics Library (DML-CZ). This article introduces the technical solution and the results of this digitization, along with some interesting statistics that have been discovered during the process. It is an interpretation of the author's talk at the general assembly of ζ TUG on May 14, 2022.

Keywords: digitization, DML-CZ, statistics

Tereza Vrabcová, xvrabcov@fi.muni.cz