

# Nástroje open source pro zpracování textů

## Open Source Tools for Text Processing

Jiří Rybička

**Abstrakt:** V rámci projektu Innovative Open Source Courses for Computer Science je řešen kurz zabývající se zpracováním textů pomocí počítače. Cílem příspěvku je představit hlavní motivy zvolené koncepce tohoto předmětu, stručně seznámit s nejdůležitějšími prvky a inovativním pohledem na problematiku a uvést ukázky vybraných částí návrhu předmětu.

**Klíčová slova:** open source, zpracování textů, koncepce VŠ výuky, technologie na bázi  $\text{\TeX}$

**Abstract:** It is being solved within the project Innovative Open Source Courses for Computer Science computer-based word processing course. The aim of the paper is to introduce the main motives of the chosen concept of this subject, briefly acquaint with the most important elements and an innovative view of the issue and state examples of selected parts of the course design.

**Key words:** open source, text processing, college teaching concept, technology based on  $\text{\TeX}$

## 1 Úvod

Zpracování textů pomocí programového vybavení osobních počítačů má dlouholetou tradici a vyznačuje se masivním zastoupením ve všech oblastech, kde se mohou počítače využívat. Dlouhodobě se programy pro zpracování textů umisťují mezi nepoužívanějšími aplikacemi, v posledních letech je téměř samozřejmostí, že v každém počítači je instalován některý z kancelářských balíků, jehož součástí a nepoužívanější aplikací je textový procesor.

Na tuto situaci již také zareagovala norma ČSN 01 6910 s názvem „Úprava dokumentů zpracovaných textovými procesory“ (ČSN 01 6910, 2014). Není zde přímo uvedeno, o jaký procesor se jedná, z kontextu a odkazů na některé funkce však můžeme usoudit, že jde o textový procesor Word firmy Microsoft, případně o podobně se chovající textové procesory z kancelářských balíků z oblasti open source – Open Office Writer, Libre Office Writer. Kromě toho lze v některých případech uvažovat i cloudová řešení (Google, Microsoft), která jsou používána stále častěji a představují značný benefit především při kooperaci na dokumentech a jejich sdílení.

Základní práce s textovým procesorem nepochybně patří do výuky, nikoliv však na vysokoškolské úrovni. V běžně praktikovaných rámcových vzdělávacích programech je tato úroveň zahrnuta do základního školství a je určitým způsobem přítomna i na školách středních. Vzhledem k masivnímu rozšíření těchto textových procesorů je samozřejmé, že tomu tak bude i po poměrně zásadní změně koncepce výuky informatiky na základních a středních školách. Můžeme si v této chvíli přirozeně položit zásadní otázku, proč a jak zpracování textů realizovat jako vysokoškolský předmět.

Problematika zpracování textů má několik rovin:

1. Zmíněný obsah vyučovaný na nižších stupních škol lze zařadit do roviny *základního ovládnutí* určitého programového systému – poznání nabídek, některých funkcí, editačních příkazů textu atd.
2. Druhou rovinou je účel a smysl jednotlivých nástrojů – *typografická a jazyková pravidla* vyúsťující v souhrnné požadavky na cílový tvar dokumentu.
3. Konečně třetí rovinou je *technická efektivnost* zpracování dokumentů zahrnující zejména rozšíření možností snadných úprav a znovupoužití zpracovaného materiálu.

## 2 Koncepce výuky zpracování textů

Ve vysokoškolské výuce se předpokládá, že obsah předmětu navazuje na určité předchozí znalosti a dovednosti, ale zároveň se rozvíjejí další, vyšší a komplikovanější prvky vyžadující rozšířený, zobecněný a celistvý pohled na problematiku. Proto se logicky do koncepce předmětu z oblasti zpracování textů na vysokoškolské úrovni promítá především část roviny druhé (část neobsažená v předchozích vzdělávacích stupních) a téměř celá rovina třetí.

Zároveň je stěžejní součástí také celkový a zobecněný pohled na metodiku zpracování – oddělení jednotlivých fází a použitých (použitelných) nástrojů. Specificky ve vysokoškolské variantě je v této oblasti kladen důraz na odborné a vědecké dokumenty, protože se s nimi studenti setkávají, potřebují je vytvářet a potřebují k tomu dostatek informací.

Významným prvkem konkrétně s ohledem na projekt inovace studijních předmětů za použití nástrojů z oblasti open source je dosažení volně dostupných komponent do celého procesu zpracování textů. Zde se jedná nejčastěji o použitelná písma a vhodné programové vybavení. Za přínosný lze považovat i záměr vybavit studenty informacemi o různých variantách tohoto vybavení, aby byli schopni aplikovat na určitý druh dokumentu optimální způsob jeho zpracování, tj. vybrat takový software, který co nejefektivnějším způsobem podpoří realizaci výsledného díla.

### 3 Obsah kurzu a nejdůležitější součásti

Z uvedených skutečností vychází obsah konstruovaného kurzu. Jako softwarová podpora je nejvíce diskutován systém postavený na principu  $\text{\TeX}$ . Hlavním důvodem pro tuto volbu je orientace na odborné a vědecké dokumenty, pro něž je tento systém ve většině případů dostatečně efektivní a umožňuje realizovat i velmi náročné typografické a technické požadavky. Protože se nejedná o masově rozšířený textový procesor typu Writer nebo Word, nepředpokládá se, že by se s ním studenti v předchozí výuce setkali. Nutně se zde proto musí věnovat určitý prostor i první zmíněné rovině, tj. základnímu ovládání komponent, které k systému patří.

#### 3.1 Seznam témat kurzu

Návrh sylabu kurzu počítá s 12 kapitolami, v každé z nich je prostor pro teoretickou část, na niž pak navazuje praktická část – cvičení u počítače. Počet kapitol odpovídá zhruba počtu využitelných týdnů typického jednosemestrálního vysokoškolského kurzu. Přibližný obsah je uveden v následujícím přehledu:

##### 1. Dokument a metoda jeho zpracování

*Teoretická témata:* • Dokumentní prvky – princip • Identifikace prvků v dokumentu • Typografický návrh dokumentu – reprezentace prvků • Technologie – realizace typografického návrhu • Technologický princip na bázi  $\text{\TeX}$  • Technologický princip na bázi systémů open office

*Praktické cvičení:* • Systém  $\text{\TeX}$  – základní principy • Distribuce, instalace • Editory, první dokument, překlad, protokol o překladu

##### 2. Základní parametry dokumentu

*Teoretická témata:* • Knižní písmo, volba typu písma • Základní písmo, stupeň základního písma • Elektronický/tištěný dokument, formát (rozměry) stránek • Technologie – definice makropříkazů

*Praktické cvičení:* • Zdroje písem, přehled dostupných písem, příklady • Parametry základního písma, volba základního písma • Definice maker s parametry, přístup  $\text{\LaTeX}$ , přístup  $\text{\TeX}$

##### 3. Speciální znaky, národní prostředí

*Teoretická témata:* • Kódování dokumentu • Nastavení národního prostředí (jazykově závislé texty, dělení slov) • Nastavení dělicího algoritmu • Speciální znaky a jejich řešení • Technologie – délkové jednotky, měrné typografické systémy

*Praktické cvičení:* • Kódování UTF-8, speciální znaky, vkládání kódů • Vzory dělení slov • Práce s délkovými jednotkami, výpočet délek, měření délek

##### 4. Sazba odstavců, algoritmy, parametry

*Teoretická témata:* • Základní text – parametry odstavců (zarážky × odsazení; zarovnání) • Odstavcové prvky jiné než základní text – parametry (citáty, výčty) • Technologie – délky, délkové registry, operace s délkami

*Praktické cvičení:* • Parametry odstavců, hladká sazba s různými parametry, příklady • Zvláštní odstavce – odrážkové a číslované seznamy, citáty • Délkové registry, aditivní a multiplikativní operace s registry

## 5. Smíšená sazba

*Teoretická témata:* • Vyznačování • Využití doplňkového typu písma • Využití různých řezů písma (kromě vyznačovacích) • Barva písma a její využití (technologie – barvy, modely, definice)

*Praktické cvičení:* • Písmové řezy • Různé typy písma v jednom dokumentu, volba kompatibilních fontů • Práce s barvami (definice uživatelských barev, barevné modely)

## 6. Členění dokumentu

*Teoretická témata:* • Systémy mezinadpisů • Iniciály • Tvorba obsahu • Technologie – číslování (čítače, reference)

*Praktické cvičení:* • Předdefinované mezinadpisy, uživatelské definice • Technologie iniciál • Čítače a křížové odkazy na ně

## 7. Stránky

*Teoretická témata:* • Odstavec a stránkový zlom • Stránková záhlaví a paty • Poznámky pod čarou • Marginálie • Stránkový design spec. stránek (titul, vyd. znám, tiráž)

*Praktické cvičení:* • Odstavcové parametry pro optimální stránkový zlom • Vkládání poznámek pod čarou, vkládání marginálií • Úprava stránky s různými řezy a velikostmi písma

## 8. Matematické a podobné výrazy

*Teoretická témata:* • Prvky výrazů • Textová a vysazená matematika • Začlenění výrazů do dokumentu, křížové odkazy

*Praktické cvičení:* • Přehled matematických prvků (exponenty, indexy, zlomky...) • Prostředí pro matematickou sazbu a jejich možnosti • Výrazy se sumami, limitami a maticemi

## 9. Tabulky

*Teoretická témata:* • Typy tabulek • Způsoby zarovnání tab. obsahu • Začlenění tabulky do dokumentu – plovoucí/neplovoucí objekty, popisky

*Praktické cvičení:* • Prostředí tabbing a tabular • Zarovnání číselných dat v tabulkách • Cvičení s různými typy tabulek

## 10. Obrazový materiál a grafika

*Teoretická témata:* • Typy obrazů – podle bar, hloubky, podle zdroje • Grafické prvky v dokumentu • Technologie – možnosti pořizování grafických prvků nástroji systému • Požadované vlastnosti grafických prvků importovaných zvnějšku • Popisky obrázků, vazba na popisky tabulek, plovoucí/neplovoucí objekty

*Praktické cvičení:* • Příprava grafiky – rastrový formát, vektorový formát • Možnosti vektorových formátů, vkládání souborů PDF • Prostředí picture • Prostředí pro vkládání tabulek a obrázků

## 11. Dokument

*Teoretická témata:* • Sestava stránek • Obsahy, rejstříky, křížové odkazy • Vyřazení stránek pro tisk, vazby, zpracování tištěného dokumentu

*Praktické cvičení:* • Návrh stránkových prvků: běžná záhlaví, marginálie, číslování stránek • Technologie obsahu, seznamů tabulek a obrázků • Uspořádání stránek, vyřazení stránek pro tisk více stránek na jeden list

## 12. Návrh a realizace vlastního dokumentu

*Teoretická témata:* • Procvičení typografického návrhu a technické realizace celého dokumentu

*Praktické cvičení:* • Typografický návrh • Určení prvků dokumentu • Technologické zpracování (styly, makra)

### 3.2 Komentář k první kapitole

Témata první kapitoly se věnují základnímu konceptu přístupu ke zpracování dokumentů, proto se tomuto obsahu budeme věnovat poněkud podrobněji. Východiskem zobecněné metody zpracování dokumentů je dlouhodobě funkční (až historická) dělba práce rozdělená mezi nejméně tři osoby (možná přesněji pracovní role):

1. **Autor** – dodává obsahový materiál dokumentu a určuje jeho prvky (vyznačení, nadpisy, poznámky, popisky atd.).
2. **Úpravce (designer)** – určuje vzhled dokumentu stanovením parametrů jednotlivých prvků definovaných autorem.
3. **Sazeč** – podle parametrů stanovených úpravcem realizuje dokument a jeho prvky v dané technologii.

To, že v dnešní době tyto role často splývají do jedné fyzické osoby, vůbec neznamená, že neexistují, spíše naopak – zanedbáváním některých aspektů těchto rolí přirozeně a zákonitě vznikají dokumenty s řadou zásadních nedostatků. Z rámcového popisu práce zmíněných osob (rolí) vyplývá také, jak na sebe činnosti logicky navazují. Je proto neefektivní, jsou-li tyto činnosti prováděny v nahodilém pořadí, napřeskáčku – i když jde stále případně o jednu osobu. Zde by měla existovat „pracovní schizofrenie“: soustředěním se vždy právě na jednu z těch rolí se osoba postupně „převtluje“ z autora

do designéra a nakonec v sazeče, aby každou z těch fází vyřešila pokud možno bez negativního ovlivňování jinými fázemi. Tak lze dosáhnout maximálního efektu podobně jako dělbou práce mezi nezávislé odborníky.

Na první pohled je situace ve srovnání s (nedávnou) minulostí velmi obtížná. Má být jedinec zároveň autorem, úpravcem i sazečem? Nemá to mnoho odborností najednou? Když na to dříve byli školení specialisté, lze to dnes zvládnout všechno dohromady?

Ve skutečnosti se ale ani neptáme, zda je to možné, spíše předpokládáme, že je to již nezbytná nutnost. Ve srovnání s dřívějšími technologiemi máme navíc řadu podstatných výhod – co dříve trvalo řádově dny, zvládne počítač za vteřiny. Co se dříve dělalo pracně a ručně, dnes probíhá automaticky. Místo mnoha strojů a dalších pomůcek máme jen informaci v paměti počítače.

Oproti minulosti je však dnes zcela nezbytné, aby autor zvládl více věcí – chce-li být producentem dokumentu, nutně musí ovládnout taktéž práci úpravce a práci sazeče. Cílem celé koncepce tohoto kurzu je rozvíjení a podpora zejména těchto dvou rolí, a to tak, aby student získal schopnost veškerou práci provádět efektivně, s maximální podporou síly programového vybavení.

### 3.3 Typografické zásady

Role úpravce zůstává v téměř stejném rozsahu jako dříve a je to bezesporu největší meze, kterou je potřeba patrně v každém kurzu zabývajícím se zpracováním dokumentů zaplnit. Všeobecné povědomí o některých zásadách sice pomalu roste, ale stále platí, že se jedná o oblast velmi málo pokrytou běžným vzdělávacím systémem a slabě vnímanou i běžnými uživateli dokumentů (čtenáři).

Ve většině kapitol je z těchto důvodů vždy uveden „typografický základ“ – jde o cíle, kterých chceme v dokumentu dosáhnout. Na to pak teprve navazuje technologická část, kde je řečeno, jak těchto cílů dosáhnout.

Jak je podle našich výukových zkušeností vhodné, jsou typografické zásady rozděleny přibližně na čtyři velké části – písmo a hladká sazba, smíšená a odstavcová sazba, stránky a jejich prvky, sestava dokumentu. Tyto části jsou rozmístěny do jednotlivých kapitol tak, aby na ně mohly navazovat příslušné technologické prvky. Kurz si nedělá ambice být rozsáhlou typografickou učebnicí – jednak takové učebnice existují (Beran et al., 2012; Kočíčka a Blažek, 2000), jednak by se takový objem nevešel do předpokládané časové dotace.

Osoba v roli úpravce se může alespoň částečně zorientovat v tom, jaké hlavní parametry určitých prvků se používají a jakých efektů lze dosáhnout.

### 3.4 Použitá technologie

Práci dřívějšího sazeče dnes reprezentuje schopnost využít vlastností a funkcí textových procesorů a dalších počítačových aplikací. Čím dokonaleji student ovládne některý ze systémů, tím efektivněji s ním bude umět pracovat a tím efektivněji bude vytvářet své dokumenty.

Hlavní součástí kurzu je z uvedených důvodů představení technologie a vhodných metodických postupů jejího využití. Koncepti „tří osob“ lze realizovat různými programovými systémy. Zcela obecně však platí, že těžištěm je využívání tzv. *strukturních značek* – technických nástrojů definujících význam jednotlivých prvků dokumentu. Autor jimi vkládá do dokumentu nezastupitelnou informaci, na kterou pak může navázat úpravce a sazeč.

Okrajově je zmíněna technologie strukturního značkování v textových procesorech typu Writer (podrobněji o tom pojednává Rybička, 2020), dále se však kurz věnuje využití systémů postavených na principu  $\text{\TeX}$ . Konkrétně je použit systém  $\text{\LaTeX}$ , ale analogicky lze postupovat i v jiných podobných nadstavbách.

Důraz je opět kladen na strukturní značkování. Proto je oproti jiným učebnicím a návodným textům již od začátku věnován velký prostor tvorbě vlastních příkazů, které představují technickou realizaci strukturních značek. Tento přístup k tvorbě dokumentů by měl vést k daleko efektivnějšímu zpracování než postup, kdy se pro značkování používají pouze předdefinované (instantní) rekvizity. Tím ovšem nikterak nesnižujeme důležitost důkladného seznámení s předdefinovaným repetoárem, naopak – ten je používán jako „stavební materiál“ pro vlastní příkazy a v některých případech i jako inspirace, jak mohou strukturní značky vypadat.

Široké možnosti vytváření vlastních příkazů v systémech postavených na principu  $\text{\TeX}$  umožňují plně realizovat koncepti strukturních značek. Proto je velmi efektivní v tomto systému pracovat, a to zejména v oblasti odborných a vědeckých dokumentů. Nezanebatelným cílem kurzu je pak také ukázat, že i přesto, že tzv. „běžné“ textové procesory (zejména včetně placených variant) jsou masivně rozšířeny, často zdaleka nepředstavují nejvhodnější volbu pro zpracování textů. Má pak smysl investovat určité úsilí do ovládnutí alternativy nacházející se navíc v oblasti open source softwaru.

## 4 Závěr

Celková koncepce kurzu „Nástroje open source pro zpracování textů“ vychází z dlouholetých pedagogických i praktických zkušeností se zpracováním textů. Očekávané přínosy této konstrukce jsou následující:

- sofistikovaný přístup k tvorbě dokumentů,
- zvládnutí technologie postavené na principu  $\text{\TeX}$ ,
- efektivní práce s dokumenty (při úpravách nebo znovupoužití),
- rozšíření možností práce s dokumenty oproti běžným textovým procesorům,
- zvýšení povědomí o oblasti open source.

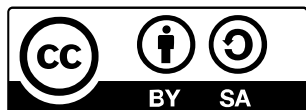
Při představení kurzu studentům v rámci Letní školy byly získány první reakce. I když se v omezené časové dotaci nebylo možné plně věnovat všem částem, byla zpětná vazba od studentů příznivá.

## Literatura

1. BERAN, V. et al. *Aktualizovaný typografický manuál*. Praha: Kafka design, 2012.
2. ČSN 016910 *Úprava dokumentů zpracovaných textovými procesory*. Praha: Ústav pro normalizaci a měření, 2014.
3. KOČIČKA, P., BLÁŽEK, F. *Praktická typografie*. Praha: CP Books, 2000.
4. RYBIČKA, J. *Zpracování textů v programu Word*. [online] [vid. 20. 11. 2021] Dostupné na <https://akela.mendelu.cz/~{}rybicka/prez/zpract/odbprace/fastword2.docx>

## Autor

Doc. Ing. Jiří Rybička, Dr., Ústav informatiky, Provozně ekonomická fakulta, Mendelova univerzita v Brně, Zemědělská 1, 613 00 Brno, e-mail: [rybicka@mendelu.cz](mailto:rybicka@mendelu.cz)



Open Access. This article is licensed under the terms of the Creative Commons Attribution-ShareAlike 4.0 International License, CC BY-SA 4.0 (<https://creativecommons.org/licenses/by-sa/4.0/>)