

V tomto článku formulujeme obecné principy dobrého dokumentu a klademe požadavky na jeho editaci, zpracování a vizualizaci. Na základě těchto požadavků ukazujeme, že vhodným formátem je strukturně značkový dokument. Vysvětlujeme strukturní značkování a popisujeme jeho vlastnosti. Na závěr uvádíme nástroje pro práci se strukturně značkováním dokumentem a ukazujeme způsob jeho formátování $\text{T}_{\text{E}}\text{X}$ em.

1. Cíle, definice

*Cílem dokumentu je předat informaci*¹. Protože dobře vysázený dokument ke snadnému pochopení a nezkreslenému podání informace nabádá, není důležitý jen jeho obsah, ale i forma. Proto je dobré při tvorbě dokumentu věnovat pozornost jeho formátování.

Aby k předání mohlo dojít, je třeba ji podávat *srozumitelnou formou* a přihlídnout ke způsobilostem cílového čtenáře. Takový čtenář má svá očekávání, která nejsou-li naplněna, bývá zklamán či zmaten. V takovém případě hrozí, že autorovy myšlenky nebudou pochopeny, nebo že čtenář ani dokument nedočte. Pravda, najdeme i úspěšné práce, které toto pravidlo očividně porušují, viz obr. 1, 2. I ty však splňují čtenářovo očekávání. Jde o očekávání nonkonformity, novátorství či zábavy a ne o předání obsahové informace, jak je tomu ve většině našich případů. Informace je nejlépe vstřebávána, proudí-li přirozeně a bez formálních rušivých vlivů.

Abychom se vyhnuli spekulacím o platnosti zde nabízených pravidel a doporučení, vymezíme oblast našeho zájmu. *Za dokument* budeme v tomto článku považovat:

- periodicky vydávaný/aktualizovaný/sériový text dané struktury (zákon, předpis, nařízení, podniková zpráva, memorandum, konferenční příspěvek),
- rozsáhlý text jednotné struktury (softwarová dokumentace, encyklopedie),
- text využitelný k mnoha různým účelům (podnikový adresář, informace o obcích, statistická data).

Pojem *formátování* definujeme jako proces, který vtiskne datům formu vhodnou ke vnímání člověkem. Ačkoliv tato forma může být zvuková (hlasová syntéza) či hmatová (Brailovo písmo), budeme se věnovat té nejběžnější: vizuální – tiště-

¹Informaci zde chápeme v nejobecnějším smyslu, včetně vizuální i estetické.



Obrázek 2: Celostránkové dřevorezy Josefa Váchala jsou těžko čitelné svou kostrbatou formou, archaickým jazykem i obsahem. Přesto jsou dnes velmi ceněny (a to nejen penězně). Důvodem není jen výjimečná řemeslná i umělecká kvalita a nadčasový obsah. Formální stránka dodává dílu složitě se deroucímú na svět věrohodnosti. Podtrhává osobnostní komplikovanost a existenční obtíže autora, který se pochopení svého díla nedožil. *Josef Váchal, Vidění sedmera dnů a planet. Paseka 1998.*

2. Principy dobrého dokumentu

Philip Taylor ve svém článku [26] formuloval tři principy dobrého dokumentu. Připojíme-li k nim ještě čtvrtý, pak dobrý dokument charakterizuje:

1. informace;
2. struktura;
3. jednotnost;
4. tradice, estetika.

2.1. Informace

Tvůrčí svoboda a lehkost internetového publikování (textu i obrazu) násobí tzv. informační explozi. Možnost snadno veřejně vyjadřovat svůj názor je bezesporným kladem, na druhé straně však vede k zaplevelení zajímavých, nezkraslených a komplexních zpráv. Mělo by proto být ctí autora přinášet pouze nová data prositá od těch méně kvalitních. A mít na paměti, že dokument bez přidané hodnoty nemá cenu tvořit!

Když už dokument vytvoříme, je potřeba věnovat pozornost kontrole dat. Je nutné, aby autor text po sobě několikrát přečetl. Hodně užitečné práce zastanou programy na kontrolu překlepů a pravopisu. Přesto je nezbytné, aby text četl ještě někdo jiný, nejlépe školený *korektor* znalý oboru. Korektury se dělají na papírové verzi. Odhalí se tam mnohem více chyb a označí se snadněji než na obrazovce.

Informace je téměř výhradně věcí autora, zatímco úkolem sazeče je formátování. Ačkoliv obě činnosti lze provádět nezávisle, je dobré je koordinovat. Některé úpravy dokumentu totiž sazeč již nemůže ovlivnit, pokud na ně autor při jeho vytváření nemyslel. Pokud by např. autor tohoto textu popsal výše uvedené požadavky na formu dokumentu v odstavci – namísto přehledných a stručných odrážek, použije nejen mnoho slov navíc, ale čtenáři se tyto informace budou i hůře vyhledávat. Podobným příkladem je nepoužití tabulky či obrázku. Nejčastější příčinou nevhodné struktury ‚rukopisu‘ není neznalost náležitých prvků, ale jejich špatná dostupnost. Autor by proto měl být před vytvářením dokumentu poučen o možných a vhodných elementech pro plánovaný druh textu i se způsoby jejich vytváření a vkládání.

2.2. Struktura

Struktura (pochází z latinského *struere*, skládat, sestavovat, budovat, pořádat) označuje způsob složení, vnitřního uspořádání nějakého objektu, zejména pokud vykazuje nějaké pravidelnosti a zákonitosti. Je to souhrn vztahů mezi prvky nějakého seskupení [Wikipedie].

Struktura dokumentu je souhrn jeho vizuálně upravených obsahových celků a jejich vztahů. Obsahové celky nazýváme *strukturní* (logické, vnitřní, obsahové)² *elementy* a realizujeme je pomocí *elementů grafických* (vnějších, formátovacích, formálních). Funkcí grafických elementů je vzájemně oddělovat a odlišovat elementy strukturní a umožnit tak čtenáři se v dokumentu orientovat.

2.2.1. Strukturní elementy

Každý dokument se může skládat z mnoha různých strukturních elementů závislých na typu dat. Jiné elementy bude obsahovat adresář a jiné kniha. Námí definovaný domument může obsahovat tyto obecné části a podčásti tvořící hierarchickou strukturu [21]:

- dokument \subset patitul, protitul, titulní list, vydavatelský záznam, anotace, tiráž, abstrakt, věnování, předmluva, mezititul, obsah, literatura, rejstříky, oddíly, kapitoly, podkapitoly;
- oddíl, kapitola, podkapitola, ... \subset nadpis, odstavce, obrázky, tabulky, algoritmy i s jejich popisky;
- odstavec \subset poznámky pod čarou, marginálie, písmena, symboly, matematické výrazy, zvýraznění, odrážky číslované i nečíslované, definice, tabulky, obrázky, akce, hypertextový link.

Čím obecnější je definice logické struktury dokumentu, tím je obtížněji použitelná. To platí zvláště pro autory, kteří nerozlišují mezi strukturními a grafickými

²Možná by bylo lepší používat termín *logické elementy*, abychom zdůraznili významovou složku nad grafickou. Budeme se však držet převládajícího názvosloví.

částmi. Proto je užitečné pro každý dokument vytvořit vlastní definici struktury dokumentu používající jen potřebné elementy a tyto pojmenovat přiléhavými názvy. Toto je typická práce redaktora dokumentu.

2.2.2. Grafické elementy

Grafické elementy můžeme rozdělit na viditelné a neviditelné.

Viditelné elementy dokumentu jsou vcelku intuitivní grafické prvky. Vyjme-
me ty nejvýznamnější, s jejich ⟨atributy⟩ a typickými vnořenými viditelnými
elementy:

- kniha, brožura, . . . , článek ⟨médium, stránkové zrcadlo⟩ ⊂ stránky, sloupce;
- stránka ⟨velikost, řádkový rejstřík⟩ ⊂ sloupce, záhlaví, zápatí, stránková
číslice, marginálie, plovoucí objekty včetně jejich popisek;
- sloupec ⟨pořadí, šířka⟩ ⊂ řádky, rámečky, linky;
- řádka ⟨šíře, výplň⟩ ⊂ znaky, linky, rámečky, obrázky;
- znak ⟨písmo, velikost, barva, atribut⟩;
- linka ⟨tloušťka, počátek, konec⟩;
- rámeček ⟨okraj, šířka, výška⟩;
- obrázek ⟨grafický formát, velikost⟩;
- akce a hypertextový link.

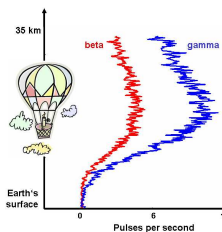
I METEOROLOGICAL EXPERIMENT

Analyzed data come from Prague-Libuš upper air meteorological station of the Czech Hydrometeorological Institute, where every month the vertical profiles of beta and gamma radioactivity are measured by the radioactivity sonde system. The radioactivity sensor consisting of two Geiger-Müller gamma and beta tubes is a part of meteorological balloons which ascends from the earth's surface up to 35 km and detects short current pulses coming from the interaction between the radiation and the tube wall material.

After several "re-calculations" the complicated measurement process results into the data pairs $(x_i, y_i)_{i=1}^n$, where x represents the altitude and y the **average number of pulses per second** in the fixed altitude x . Let us note that y is proportional to the **radiation intensity**.

II APPLIED REGRESSION MODELS

The first aim of the analysis was to suggest a parametric regression model $Y_i = m(X_i) + \varepsilon_i$, where $m(\cdot)$ represents mean amount of radiation and ε_i a random "error" term. As described by Hlubinka (2004), the models $m(\cdot)$ are based on **Richards growth curve $R(x)$ and its derivative $r(x)$** . Unfortunately, Richards curve itself does not describe properly the measurements in low atmospheric layers. That's why we propose two extended additive models. The first model consists of $r(x)$ and a **simple linear function**, meanwhile the second model of $r(x)$ and **logistic growth curve** being of the form



Obrázek 3: Příklad špatného použití mezer při formátování titulku. Velikost mezer by měla odpovídat logickému vztahu sousedních viditelných elementů. Proto před titulkem má být mezera větší než ta pod ním. Oddělíme tak kapitoly od sebe a zároveň svážeme titulek s tělem kapitoly. Výřez.

Neviditelné elementy bývají často opomíjeny [26, 25]. Jde totiž o bílá místa na stránce. Přitom tyto elementy hrají rozhodující roli v oddělení jednotlivých viditelných elementů, i v chápání správných vztahů mezi nimi. Tak např. text **Toto není nadpis!**

budeme chápat jako příkaz sumarizující předchozí odstavec a ne jako titulek další kapitoly. Rozhodující pro náš úsudek je vertikální odsazení, viz obr. 3. Mezi neviditelné elementy řadíme

- Okraje stránky – jejich funkcí je, aby prsty nepřekážely čtení při držení dokumentu, a bylo snadné vyhledat stránkovou číslici (vhodné místo je při vnějším okraji na všech stranách bez výjimky), identifikovat záhlaví. Komponování stránky a jejich okrajů je popsáno např. [2, 1].
- Mezery okolo textových elementů – oddělují elementy od sebe. Příkladem je vyznačení odstavce (obr. 4), již zmíněné mezerování titulků kapitol, obrázků ap., oddělení plovoucích elementů a jejich titulků od hlavního textu, oddělení záhlaví a zápatí.
- Řádkování – mělo by využívat hodnoty, které navrhl tvůrce použitého písma, optimalizované pro čitelnost a neslití řádků.

tomu musí být v horké sazby dodnes. Dává mnohem více volnosti v tvaru i návaznosti, než jí měl tradiční písmař. A tato tvůrčí přednost se již u některých novinek projevila.

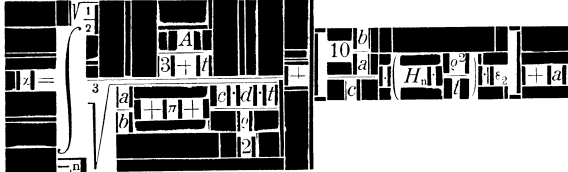
Johannes Gutenberg, když se v jeho ruce zrodil zázrak knihtisku, mohl ztěžít předpokládat směr vývoje v naší době. Věřil ve význam svého vynálezu a uvědomoval si „v báni boží i ve vši skromnosti“ jeho dosah, i když se musel trápit přízemmějšími starostmi, např. jak vyrovná zadlužení, jímž by hradil svůj objev. Ale více než sto padesát let po něm zhodnotil Gutenbergův čin jiný světoobčan, Čech, Jan Amos Komenský. Ve své „Živé tiskárně“ píše prorocky kromě jiného: „*Nejkrásnějším darem božím je vynález tiskových liter, jimiž se knihy nesmírně rychle rozmnožují. Tím nejkrásnějším bude posléze umění, jímž se budou knihy a jejich moudrost s podobnou rychlostí vtiskovat v lidskou duši, dopřeje-li Bůh, aby jednou bylo vynalezeno.*“ I když současný vynález toto Komenského předvídaté přání nesplňuje, posunuje je možná dále k jeho uskutečnění.

Písmo, typografie, ilustrace a tisk mají ještě důležitějšího spojence, který umožňuje — vedle jiných faktorů — jejich reálný život. Je to materiál, na který tiskneme a ještě dlouho budeme tisk-

Obrázek 4: Odstavcová zarážka může mít různou podobu. Nejčastěji používané je odsazení prvního řádku dovnitř textu, ale někdy i vně, jako v tomto popisku. Též vertikální odsunutí odstavců je běžné. Méně časté je odsazení prvního řádku o šířku východového (posledního) řádku předchozího odstavce tak, jako v reprodukcované ukázce z Typographie Oldřicha Hlavsy [12].

Též je možné oddělit odstavce speciálním znakem, nejčastěji ¶ (z latinského capitulum, anglicky zvaného pilcrow). Tento způsob se užíval hlavně ve středověku.

- Sloupce a buňky tabulky – mají zajistit jednoznačné oddělení sloupců a zarovnat hodnoty.
- Interpunkční mezerování – vychází z logického oddělení a tradice: např. tečka ukončuje předchozí větu, následující mezera věty odděluje.
- Mezerování matematických výrazů – odlišně mezerujeme unární, binární a relační operátory. Mnoho práce tu vykoná T_EX sám, ale i tak zbývá dost ruční sazečkové práce [31, 24], obr. 5.



Obrázek 5: Ukázka složitosti ruční matematické sazby s potřebnými výplňky. Už víte, proč se matematický text zapisuje v T_EXu mezi dolary? Ukázka z *Nauky o sazbě obyčejné, tabulkové, matematické a chemie* Jindřicha Špalka z roku 1925 [23].

2.3. Jednotnost

Dalším principem dobrého dokumentu je *jednotnost*. Jednotnost se projevuje hned v několika rovinách. Každé se budeme věnovat zvlášť.

2.3.1. Stejně elementy stejně

Formátování stejných elementů v celém dokumentu stejně je základní požadavek jednotnosti. Znamená to např. stejné odsazení odstavců, stejný způsob zvýraznění slov, stejné formátování titulků, stejnou sílu linek. Není-li dodrženo toto pravidlo, nebude struktura dokumentu jednoznačná a čtenář může být zmaten.

2.3.2. Různé elementy se sjednocující myšlenkou

Dobrý typograf vtiskne dokumentu jednotnou myšlenku tím, že opakuje různé variace jednoho nápadu pro různé strukturní elementy. Přitom je třeba zřetelně zachovat hierarchickou strukturu dat. Příkladem může být zvýraznění titulku obrázku stejným ale menším písmem jako v nadpisu kapitoly.

Úprava dokumentu by měla být ve vzájemné harmonii. Zvláštní zřetel by měl být při tvorbě knih kladen na kompozici protilehlých dvoustran. Jednostranný tisk ztratil opodstatnění z důvodů ekonomických (cena papíru), ekologických (spotřeba papíru), technických (zmizel průklep psacího stroje) i funkčních (přehlednost

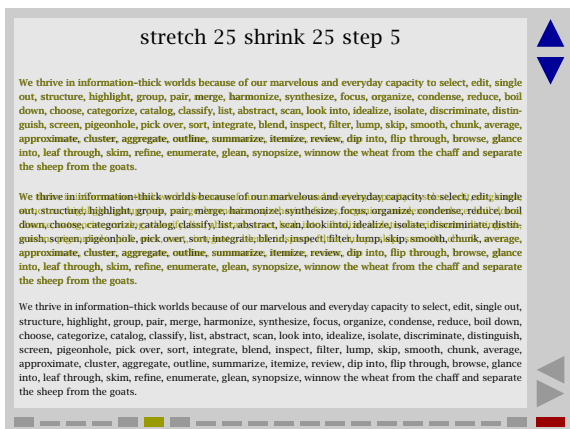
většího kusu textu bez listování). Obě stránky mají mít stejný počet řádek a pro umístění plovoucího objektu platí pravidlo, že má být vidět na dvoustraně, kde je odkazován.

2.3.3. Rovnoměrné pokrytí strany

Stránka při pohledu z větší vzdálenosti by měla působit stejnoměrnou šedou barvou, neměla by být roztráštěná. Taková působí chaoticky, jednotlivé strukturní elementy nejsou zřetelné. Naším cílem by měla být kompaktní strana. Prostředkem k jeho dosažení je citlivá práce s neviditelnými elementy.

Kontrast písma a pozadí by měl být co nejvyšší, ale na celé stránce podobný, jinak bude obojí nadměru unavovat čtenářův zrak. Prostředkem je volba písma podobné síly, jak pro hlavní text, tak zvýraznění. Z tohoto důvodu se pro zvýraznění často používá italika stejného fontu; kresba je dostatečně odlišná, aby zvýraznění bylo jednoznačné, zároveň ale stejné optické hustoty. O *prostrkávání* Karel Dyrynk píše [5, str. 18]: *Vyznačování prostrkáním hodí se pro denní list nebo časopis, na jehož úpravě příliš nezáleží. Ale v klidné stránce knihy je takové roztrhané slovo světlé a rušivé, a zvláště v krásné knize je prostrkání naprosto nepřijatelné.*

Kontrast textu též zvyšuje stará typografická finesa – sazba na *řádkový rejstřík*. Při něm si řádky na rubu i lici stran přesně výškově odpovídají. Implicitní

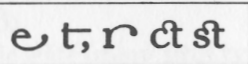


Obrázek 6: Jedním z předpokladů rovnoměrného pokrytí strany je neděravý odstavcový zlom. Při problematické sazbě (např. do úzkých odstavců) k němu napomáhá tzv. *hz-algorithmus* implementovaný do pdf \TeX u [32, 27, 34, 22]. O optické vyrovnání hran textu se stará technika *prostrkání okrajů* [33]. Ukázka z testovacího dokumentu Hanse Hageny *Does pdf \TeX make things better?* [7].

vě přikročeno k sazbě. Aby písmo bylo malebnější, nakreslil umělec od jednotlivých typů několik způsobů (e e e a a t t k k) a určil kde kterých upotřebovat, aby řádky byly světlem vyrovnané, pokusmo stanovil šířku mezer mezi slovy, úpravu titulků atd. A pro každou

(a) O Preissigově úpravě *Slezských písní*.

Úmyslně zvolen větší stupeň, tercie, aby bylo dosaženo i vnějšího souhlasu s vážným, slavnostním, evangelickým obsahem knihy. Ježto by bylo nemožno tak velké a široké písmo typograficky správně sázet na šířku 20 cicer, bylo nutno je doplnit



(b) O Komenského *Kšaftu umírající matky jednoty bratrské* vydaného Spolkem českých bibliofilů roku 1909.

Obrázek 7: K vyrovnání světlosti řádek lze použít písmenné varianty. Některé nahrazují kerning, některé vyrovnávají řádku variantní šířkou. Ukázky z knihy Karla Dyrynka *Typograf o knihách* [5].

nastavení plain $\text{T}_{\text{E}}\text{X}$ u i standardních tříd $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ u řádkový rejstřík nedodržuje. Hlavním důvodem je problematická sazba matematických vzorců – původně hlavní účel $\text{T}_{\text{E}}\text{X}$ u – do řádkového rejstříku. Zachovat řádkový rejstřík v $\text{T}_{\text{E}}\text{X}$ u jde, ale ne triviálně, viz [19], [9, CONTEXT]. U vícsloupcového textu je dodržení řádkového rejstříku estetickou nutností.

V publikacích, kde pracujeme s barvou (plakáty, časopisy, on-line dokumenty), je nutno věnovat pozornost barevnému kontrastu mezi textem a pozadím. Nápomocí mohou být programové pomůcky, např. webový Colour Contrast Visualiser [13].

Dalším prostředkem rovnoměrného pokrytí je neděravá sazba. Kromě velmi propracovaného odstavcového zlomu, který nám $\text{T}_{\text{E}}\text{X}$ nabízí, můžeme využít mikrotypografické rozšíření pdf $\text{T}_{\text{E}}\text{X}$ u [28], viz obr. 6. Jinou možností, která byla hojně používána již prvním typografem Johannesem Gutenbergem, je použití písmenných variant, obr. 7. Jimi lze potlačit jak nevhodný zlom řádku, tak i tzv. řeky – několik mezer v odstavci nad sebou. Pokud je např. z důvodů velmi úzkého sloupce nemožné dosáhnout neděravého odstavce, je lepší místo sazby do bloku použít sazbu na praporek, viz popisek obr. 8.

Domácí meta vznikne odstřížením rohů ze čtverce velikosti 43 cm, ostatní mety mají tvar čtverce o hraně 38 cm. Vnitřní hrana obdelníkového území pálkaře o rozměrech 91×213 cm je odsazena od hrany domácí mety o 15 cm a přečnává o 91 cm roh hřiště.

Hraji jej dvě mužstva (oficiálně o devíti členech). Hráči družstva v poli (polaři) mohou být kdekoli v poli, jen nadhazovač a chytač musí být ve svém území. Družstva se pravidelně střídají ve hře na pálce i v poli. Směna je období hry, kdy totéž mužstvo hrálo na pálce i v poli. Ke střídání úloh dojde, pokud družstvo v poli dosáhne třetího autu.

Účelem hráče na pálce je správně odpálit, proběhnout metovou dráhu (dotknout se v pořadí první, druhé třetí a domácí mety) dříve, než je vyautován. Za takový oběh získá jeho mužstvo bod. Úkolem polařů je zabránit běžci v získání bodu tím, že jej autují dříve, než dokončí oběh.

Nadhazování

- Nadhazovač musí zaujmut postavení oběma nohama na zemi a musí se nohama dotýkat nadhazovací mety.

4

Obrázek 8: Zcela kompaktním stranám napomáhali hlavně středověcí typografové výplní odstavcových zarážek, čímž dosahovali obdelného tvaru odstavců. Tento přístup byl použit pro sazbu drobné brožurky Pravidla softballu [6], jejíž stránku zde reprodukuje. Využívá se ‚pružného‘ obrázku pomocí `\cleaders`:

```
\def\softbat{\batA\nobreak
\cleaders\hbox{\batB}\hfil\nobreak%
\batC\nobreak
\cleaders\hbox{\batD}\hskip0pt plus1.5fil
\nobreak\batE}
```

kde `\bat?` jsou jednotlivé dílky pálky.

2.3.4. Dodržení národních, oborových a lokálních konvencí

Jednotnost by měla být zachována i v rámci vyšších celků, než je samotný dokument. Jde o národní typografické odlišnosti (pozor např. na velikost mezer okolo interpunkce, zejména pomlček [3]), dále o oborové konvence (značení bibliografie), o jednotné formátování publikací v edičních řadách (třeba diplomových prací [20]) a jednotlivých příspěvků ve sborníku. To vše splňuje očekávání čtenáře a napomáhá srozumitelnosti.

2.3.5. Chronologická jednotnost

Dobrym rysem úpravy je chronologická korespondence mezi obsahem dokumentu a použitým písmem. Sázíme-li např. starý renezanční text, působí amatérsky, použijeme-li moderní písmo. Klasifikace písem je dobře popsána v [11, 2, 4].

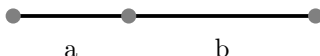
2.4. Tradice, estetika

Měli bychom mít na paměti, že úprava dokumentů se vyvíjí po několik tisíciletí a za tu dobu přirozeně konvergovala podoba i umístění jednotlivých elementů k praktické použitelnosti. Designérské experimentování při formátování by mělo být prováděno s plným vědomím tohoto dlouhého vývoje.

Příkladem tradičního estetického hlediska je potlačení vdov (stránkový zlom za prvním řádkem odstavce...) a sirotků (... před posledním řádkem odstavce), obecně parchantů. Sázíme-li do řádkového rejstříku, můžeme změnit délku textu

změnou zlomu dlouhého odstavce pomocí `\looseness`, přidat půlřádek okolo nadpisu kapitoly nebo upravit velikost obrázku.

Těžko formulovat nějaká obecná pravidla tohoto principu. Charakterizuje jej střídmost, dodržování vzájemných proporcí, kompozice (viz obr. 9) a celková harmonie. Pro náš návrh se můžeme inspirovat bezpočtem dobrých i špatných příkladů.



Obrázek 9: Při komponování se často využívá čistých geometrických tvarů a někdy i hudební harmonie [2]. Jednoduchým příkladem je formátování nadpisů na střed. Estetické vnímání však zcela nekopíruje naprostou symetrii. Tak jako lépe vnímáme, když je fotografovaný objekt vychýlený z geometrického středu, tak třeba i horizontální umístění stránkové číslice nebo vertikální poloha marginálie je lepší mimo střed či kraj. Nápomocí nám může být tzv. *zlatý řez*. Pro něj platí formule $b : a = (a + b) : b$ a poměr stran $a : b \approx 0,618$. Šířka obr. 2 je komponována právě jím.

2.5. Obecné pravidlo formátování

Při pozorném přečtení všech čtyř principů formátování zjistíme, že jednotlivá pravidla si vzájemně protirečí. Zcela jednotně pokrytá stránka nepopisuje strukturu dat a dotaženo ad absurdum, naprosto homogenně šedá stránka je sice krásně jednotná, ale nepřináší (téměř) žádnou informaci. Pro aplikování pravidel je proto třeba si uvědomit, že jejich platnost je hierarchická; přednost je v tom pořadí, v jakém jsou zde prezentovány. Koliduje-li tedy princip informace a struktura (což nebývá moc časté), je třeba dát přednost informaci. Pokud je v rozporu vyjádření struktury a jednotnosti, je třeba dát přednost struktuře ap., obr. 10. Protože tuto kolizi musí sazeč řešit každodenně, shrneme si ji následujícím tvrzením: *Povolme právě tolik vyjímek z uniformity, kolik je nezbytné pro vyznačení zvolené struktury; ne více.* Proto je žádoucí použít právě jednu odstavcovou zarážku z těch zmíněných v popisku obr. 4.

Závěrem této kapitoly poznamenejme, že výše popsaná pravidla jsou pouze zobecňujícím návodem, nikoliv dogmatem. Úprava dokumentů patří mezi umělecká řemesla a neplatí proto pro ně striktní pravidla. Formalizovali jsme je, abychom z nich vyvodili následující požadavky na editaci dat.

Výška znaku integrálu \int ve vzorcích na zvláštním řádku je 26 bodů bez ohledu na výšku vzorce; v textu, kde se nemají vyskytovat zlomky s čitatelem a jmenovatelem nad sebou, nýbrž (v nezbytných případech) se šikmým lomítkem nebo se záporným exponentem, je znaménko integrálu \int 12bodové, a to i v případě, že má meze.

$$\int_{-1}^1 \left(\frac{\partial^2 u}{\partial x \partial t} \right)^2 dx ; \int_{-1}^1 (\partial^2 u / \partial x \partial t)^2 dx .$$

Obrázek 10: Tento příklad ilustruje rozpor mezi principy informace a jednotnosti. Velký matematický symbol by při zachování jednotného řádkování způsobil slítí s textem okolních řádek, nečitelnost a proto ztrátu informace. V takovém případě je třeba oželeť pravidelné řádkování. Mnohem lepším řešením je nahradit tento symbol menší variantou či vložit jej na zvláštní řádek (display mód). Ukázka z knihy Karla Wicka *Pravidla matematické sazby* [31].

3. Požadavky na editaci dat

Naplnění výše popsaných principů není jednoduché vzhledem k jejich počtu, variabilitě, vzájemnému provázání a dělbě kompetencí mezi několik osob (autoři, redaktor, sazeč, korektor). Práci by usnadnily nástroje, které navádějí autory k přípravě strukturovaných zdrojových textů dokumentů (dále rukopisů) a systémově zjednodušují práci sazečům. Pokusme se požadavky na takové nástroje a formáty dokumentů formulovat. Učiníme tak zvlášť pro autory dokumentů a zvlášť pro sazeče a správce dat.

3.1. Z hlediska autorů dokumentů

Většinou nelze předpokládat, že autor je zároveň IT specialista. Proto nástroj pro tvorbu rukopisů musí být uživatelsky co nejjednodušší a intuitivní. Domnívám se, že jeho ovládání musí vycházet z všeobecně propagované počítačové gramotnosti ovládání toho pravého a jediného textového procesoru.

Zároveň však musí autora nabádat k *výhradnému používání strukturních elementů*. Vkládání grafických elementů by mělo být značně ztěženo nebo dokonce zakázáno. Tento požadavek je zcela v protikladu s praxí současných textových WYSIWYG procesorů.

Takto důsledně vnučená logická struktura dokumentu přináší dvě pozitiva: a) systémově umožní jednotný grafický vzhled stejných elementů a b) oddělí a odloží formátování dokumentu do dalšího kroku. Separace funkcí podporuje celkovou kvalitu dokumentu, protože o každou složku se postará specialista: o obsah autor a o formátování typograf, sazeč či designér. Hranice kompetencí je přitom jasně definovaná.

Nástroj by též měl asistovat ve vkládání náležitých elementů (tj. povolit vložit jen ty dle kontextu povolené), jejich automatické číslování (stránky, odrážky), generování (datum, obsah, křížové odkazy, citace), zpracování (řazení) a verifikaci (struktura, url). Vše centralizované do jednoho místa co nejlépe relevantní obsahové informací.

Požadavky z hlediska autorů shrneme do těchto bodů:

- snadnost a intuitivnost použití,
- výhradní používání strukturních elementů s jejich kontextovou nabídkou,
- zákaz či potlačení grafických elementů,
- hlídání udržování definované struktury textu,
- odstranění rutinních a k chybám náchylných úkonů (číslování, obsah).

3.2. Z hlediska sazeče a správce dat

Úkolem sazeče je vtisknout dokumentu pěknou a jednotnou tvář. Mezi požadavky proto patří možnost využít všechny dostupné grafické elementy. Technika by neměla omezovat návrháře.

Abyste byla zaručena jednotnost formátování, neměl by být graficky upravován každý strukturní element zvlášť, ale centrálně pro všechny instance jednoho strukturního elementu. Formátovací práce tak vyžaduje vyšší míru abstrakce, ale kromě vyšší efektivity u rozsáhlých dokumentů je vyvážena možností experimentovat s formátovacími variantami na celém dokumentu najednou.

Protože struktura dokumentu a jeho formátování se vzájemně ovlivňují, je v praxi výhodné, když sazeč může experimentovat s úpravou již od počátku tvorby dokumentu a flexibilně navrhnout autorům vhodné strukturní elementy. I zde je proto nezbytné, aby formátování bylo oddělené od dat, aby autoři úpravou textu nezrušili všechnu sazečem dosud vykonanou práci. Požadavky sazeče můžeme shrnout do těchto bodů:

- grafická obecnost,
- podpora jednotnosti,
- snadnost experimentování,
- oddělení formátování od obsahu,
- čitelnost a
- minimalizace práce.

Kromě toho musíme připočítat požadavky správce dat. Jsou jimi

- dlouhodobá platnost dat (využívání standardů),
- multiúčelovost (zpracování variabilním softwarem, různý výstup, otevřený formát),
- validace definované struktury dat a
- platformní nezávislost.

4. Strukturní značkování

Jak záhy uvidíme, strukturní značkování je způsob vytváření dokumentu, který dobře vyhovuje výše uvedeným požadavkům. Spočívá v doplnění textu o předem stanovené značky vyznačující jednotlivé strukturní elementy.

4.1. Výhody a nevýhody strukturního značkování

Strukturně označovaný dokument přináší tyto výhody:

- oddělení formátování od obsahu,
- oddělení formátování od struktury,
- expertní rozdělení kompetencí pro autora a sazeče,
- snadné udržování jednotnosti,
- vyšší přehlednost u větších projektů a pro více uživatelů,
- definování a verifikování struktury a platnosti dat,
- možnost globální změny formátování,
- snadné generování různých výstupů z jednoho zdroje (znovuvyužití textu),
- zvýšení efektivity psaní využíváním předdefinovaných elementů,
- snazší správa (jazykových) variant,
- vyšší časová stálost (snazší archivace).

A tyto nevýhody:

- požaduje větší disciplínu autora,
- menší flexibilitu u elementů, na které se nemyslelo dopředu.

Obecně lze říci, že výhody vedou k vyšší efektivitě zpracování dokumentu, ačkoliv to vyžaduje více počátečního úsilí.

4.2. Formáty

Obecným standardním formátem strukturního značkování je SGML (Standard Generalized Markup Language), ale největší popularitu získaly jeho zjednodušené aplikace HTML (HyperText Markup Language) a XML (Extensible Markup Language) [14]. Protože HTML míchá strukturní a grafické značky, je XML dnes de facto standardem strukturního značkování (obr. 11).

Pro značkování jazykem XML je nejprve třeba definovat hierarchickou strukturu dokumentu jazykem DTD (Document Type Definition).³ Naše datové XML pak bude s tímto souborem nerozlučně spjato. Pokud najdeme pro náš typ dokumentu definici s vhodnými elementy, můžeme ji použít a nemusíme si ji vymýšlet. Příkladem je běžně používaná definice DocBook [18, 30] zvláště vhodná pro odborné texty a manuály. Pro jednoduché či netypické dokumenty lze však DocBook přirovnat ke střílení kanónem na komára.

³Jinou možností je XML Schema nebo Relax NG.

```

<!-- ===== Uvod -->
<chapter>
<title>Uvod</title>
<p>
Cílem projektu <em>Analýza meteorologických
analogií</em> <footnote>z anglického názvu <em>Meteorologic
Analogy Analysis</em> byla odvozena zkratka MANA-ANA. </footnote>
bylo vytvořeno nástroje pro vyhledávání „podobných“ meteorologických
situací v databázi meteorologických měření
<a href="http://www.cdc.noaa.gov/cdc/data.ncep.reanalysis.html"
>NCEP/NCAR</a>
amerického střediska
<a href="http://www.cdc.noaa.gov/climate_diagnostics_center/<a>
Podobnost meteorologických situací je odvozena od použité podobnostní
funkce. V systému jsou implementovány dvě základní podobnostní funkce:
RMS rozdílů dvou měření a RMS úhlů mezi normálami daných měření.
</p>
<p>
Systém dále umožňuje:
<itemize>
<li>Automatické stahování (aktualizací) dat z NCEP/NCAR
databáze.
<li>Výběh meteorologických veličin, referenčního období,
časového rozsahu prohledávání, geografické oblasti a typu
podobnostní funkce.
<li>Vyhodnocení podobnostní funkce a následné seřazení situací
na základě podobnosti.
<li>Vizuální zprávy výsledků.
</itemize>
</p>
Projekt byl navržen jako serverová webová aplikace. Tato realizace
umožňuje uživateli vzdálený přístup prostřednictvím webového rozhraní

```

(a) XML

```

<ENTITY % in "in" >
<ENTITY % sum "sum" >
<ENTITY % yes "yes" >
<ENTITY % no "no" >
<ENTITY % lang "(cz|en|un)" >
<ENTITY % lang-default "un" >
<ENTITY % item-type "(0|1)" >
<ENTITY % item-type-default "0" >
<ELEMENT doc
(title|subtitle|authors|copyright|basename?,dtr-fig?,body)
<!ATTLIST doc version CDATA #REQUIRED
<!ATTLIST doc charset CDATA #IMPLIED
<ELEMENT out-HTML (#PCDATA)
<ELEMENT out-PDF (#PCDATA)
<ELEMENT titled (#PCDATA)
<ELEMENT titlet (#PCDATA)
<!ELEMENT titlet label (#PCDATA)
<ELEMENT subtitle (#PCDATA)
<ELEMENT authors (institution?,logo?,author*)
<ELEMENT basename (#PCDATA)
<ELEMENT dtr-fig (#PCDATA)
<ELEMENT institution (#PCDATA)
<ELEMENT logo (Figure)
<ELEMENT author (name,email?,address?)
<ELEMENT name (#PCDATA)
<ELEMENT email (#PCDATA)
<ELEMENT address (#PCDATA)

```

(b) DTD

```

<hr />
<h2>1. Úvod</h2>
<p>
Cílem projektu <em>Analýza meteorologických
analogií</em> <footnote>z anglického názvu <em>Meteorologic
Analogy Analysis</em> byla odvozena zkratka MANA-ANA. </footnote>
bylo vytvořeno nástroje pro vyhledávání „podobných“ meteorologických
situací v databázi meteorologických měření
<a href="http://www.cdc.noaa.gov/cdc/data.ncep.reanalysis.html"
>NCEP/NCAR</a>
amerického střediska
<a href="http://www.cdc.noaa.gov/climate_diagnostics_center/<a>
Podobnost meteorologických situací je odvozena od použité podobnostní
funkce. V systému jsou implementovány dvě základní podobnostní funkce:
RMS rozdílů dvou měření a RMS úhlů mezi normálami daných měření.
</p>
<p>
Systém dále umožňuje:
<itemize style="list-style-type: disc" type="0">
<li>Automatické stahování (aktualizací) dat z NCEP/NCAR
databáze.
<li>Výběh meteorologických veličin, referenčního období,
časového rozsahu prohledávání, geografické oblasti a typu
podobnostní funkce.
<li>Vyhodnocení podobnostní funkce a následné seřazení
situací na základě podobnosti.
<li>Vizuální zprávy výsledků.
</li>
</itemize>
</p>
Projekt byl navržen jako serverová webová aplikace. Tato
realizace umožňuje uživateli vzdálený přístup
prostřednictvím webového rozhraní bez závislosti na
konkrétním operačním systému, uživatelské rozhraní
je dvojjazyčné, české a anglické.

```

(c) HTML

```

% ===== Uvod
\chapter{Úvod}
Cílem projektu \emph{Analýza meteorologických
analogií}(\footnote{z anglického názvu \emph{Meteorologic
Analogy Analysis}}) byla odvozena zkratka MANA-ANA.
bylo vytvořeno nástroje pro vyhledávání „podobných“ meteorologických
situací v databázi meteorologických měření
\href{http://www.cdc.noaa.gov/cdc/data.ncep.reanalysis.html}
{NCEP/NCAR} amerického střediska
\href{http://www.cdc.noaa.gov/climate_diagnostics_center/}
{Climate diagnostics center}.
Podobnost meteorologických situací je odvozena od použité podobnostní
funkce. V systému jsou implementovány dvě základní podobnostní funkce:
RMS rozdílů dvou měření a RMS úhlů mezi normálami daných měření.
Systém dále umožňuje:
\begin{itemize}
\item Automatické stahování (aktualizací) dat z NCEP/NCAR
databáze.
\item Výběh meteorologických veličin, referenčního období,
časového rozsahu prohledávání, geografické oblasti a typu
podobnostní funkce.
\item Vyhodnocení podobnostní funkce a následné seřazení situací
na základě podobnosti.
\item Vizuální zprávy výsledků.
\end{itemize}
Projekt byl navržen jako serverová webová aplikace. Tato realizace
umožňuje uživateli vzdálený přístup prostřednictvím webového rozhraní
bez závislosti na konkrétním operačním systému, uživatelské rozhraní
je dvojjazyčné, české a anglické.
% ===== Spuštění programu
\section{Spuštění programu}

```

(d) L^AT_EX

Obrázek 11: Příklady strukturně označovaného textu a definice jeho struktury.

XML je bezkompromisní strukturně značovací jazyk, netvrdím však, že je vždy nutné psát dokument v něm. Pokud jsme sami sobě autory i sazeči drobného článku psaného T_EXem, byla by užvaněnost XML spíše překážkou. Poměrně dobře lze strukturně značkovat i v T_EXu. Vystačíme si se stylovým souborem, kde si makrojazykem T_EXu definujeme všechny potřebné a implicitně nenabízené strukturní značky (v plain T_EXu nám to dá více úsilí, protože implicitně téměř žádné takové značky nenabízí); tento soubor pak vložíme pomocí \input do vlastního zdrojového souboru se strukturně označovaným textem. Dbáme, abychom

v něm nepoužili žádný formátovací příkaz a centralizovali příbuznou informaci do jednoho místa, viz obr. 12.⁴ Např. namísto `\vspace` píšeme `\bigskip` nebo `\includegraphics [width=5.47cm]{ma/strasna/cesta/soubor.pdf}` nahradíme definicí

```
\graphicspath{{ma/strasna/cesta/}}
\includegraphics [width=.5\linewidth]{soubor}
```

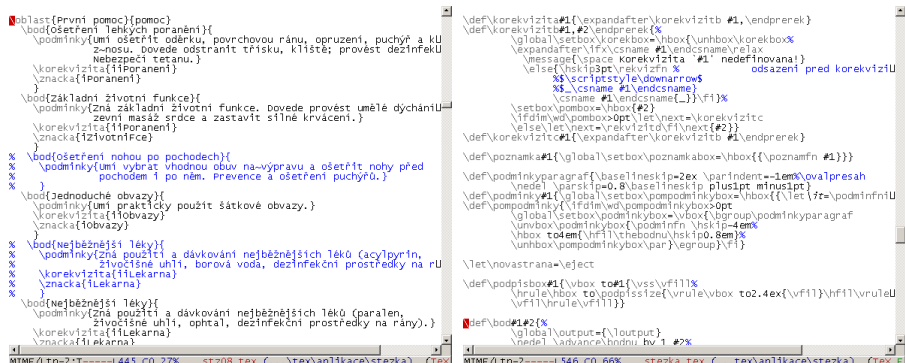
nebo ještě lépe

```
\newdimen\FigSize \FigSize=.5\linewidth
\includegraphics [width=\FigSize]{soubor}
```

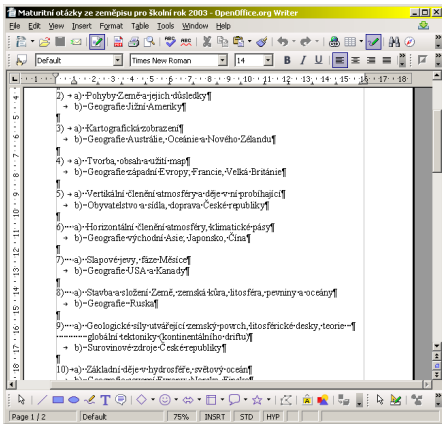
Nevýhodou tohoto přístupu je nízká formalizace a tudíž neexistence přímého validačního nástroje.

Strukturní značkování již začaly nabízet i textové procesory, ale použitelnost je mizerná. Uživatel musí projevit nebývalou disciplínu a znalost, aby nesklouzl k použití – v okenní liště snadno přístupných – formátovacích příkazů (obr. 13). I zpracovatel takto označovaných dat není v lehké pozici, neboť formát dat je dosti kryptický, ve své obecnosti značně složitý a často obsahuje ‚smetí‘ vzniklé editací (např. v podobě prázdných elementů).

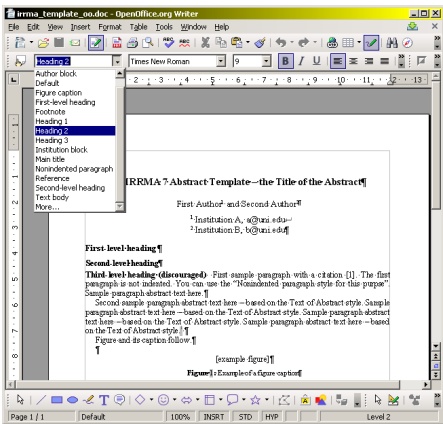
⁴V praxi zjistíme, že v závěrečné fázi formátování potřebujeme učinit drobné výjimky z obecného formátování, abychom vyhověli typografickým požadavkům. Třeba vložit `\looseness` pro zamezení vzniku parchanta. Vyplatí se i tyto příkazy strukturně označovat speciální značkou, např. `\finalTYP0a{\looseness=-1\par}` při definici `\long\def\finalTYP0a{#1}`. Jedna taková značka odpovídá jednomu způsobu formátování. V případě potřeby se pak dají tyto formátovací výjimky přepínat/zneaktivnit předefinováním `\def\finalTYP0a{}`.



Obrázek 12: Příklad důsledného značkování v plain $\text{T}_{\text{E}}\text{X}$ u. Poměrně mnoho práce s vytvořením definičního souboru je kompenzováno faktem, že datový soubor může snadno editovat kdokoliv i bez znalosti $\text{T}_{\text{E}}\text{X}$ u.



(a) Špatné



(b) Dobré

Obrázek 13: Příklady špatného a dobrého značkování v OpenOffice.org.

```

@BOOK{ byrynk93,
  AUTHOR = { Karel Byrynk },
  TITLE = { Typografie knih(\'a)ch },
  PUBLISHER = { Koutník polygrafie, Praha },
  YEAR = { 1993 },
  EDITION = { 3. },
  KEYWORDS = { sazba, typografie, oprava knih },
  NOTE = { První vyd(\'a)n(\'i) 1911. },
  WHO = { Zlka },
}

@BOOK{ Hlavsa76,
  AUTHOR = { old(\v r)jch Hlavsa and Karel Wick },
  TITLE = { Typografie 1 },
  SUBTITLE = { P(\'i)smo, ilustrace, kniha },
  PUBLISHER = { Nakladatelstv(\'i) technick(\'e) literatury, Praha },
  YEAR = { 1976 },
  EDITION = { 1. },
  KEYWORDS = { sazba, typografie },
  WHO = { Zlka },
}

@BOOK{ Wick66,
  AUTHOR = { Karel Wick },
  TITLE = { Pravidla matematick(\'e) sazby },
  PUBLISHER = { Academia, Praha },
  YEAR = { 1966 },
  EDITION = { 1. },
  KEYWORDS = { sazba, typografie, matematika },
  WHO = { Zlka },
}

@MANUAL{ Starý05,
  AUTHOR = { Richard Star(\'i) },

```

(a) BIB

Obrázek 14: I starý známý Bib_TE_Xový formát je strukturně značkový text.

4.3. Prostředky

Zde uvedený seznam prostředků pro editaci, zpracování a formátování strukturně značkových dat není a ani nemůže být úplný (obr. 14). Slouží pro nasměrování čtenáře. Největším problémem se autorovi jeví snadno dostupný strukturní open XML editor, s ovládním nepříliš vzdáleným textovým procesorům.

4.3.1. Editory

Pro vytváření strukturně značkových dokumentů lze používat různé typy editorů. Pokusíme se je stručně charakterizovat.

- a) Klasický textový:
 - + lze zapsat všechny možnosti formátu (je-li tento textový),
 - zcela nutná znalost formátu,
 - ukecanost formátu může vést až k praktické nepoužitelnosti,
 - bez přímé validace,
 - bez grafického formátování.
- b) Textový s podporou pro strukturní formát (nabídky, klávesové zkratky) [např. Emacs, Vim, jEdit]:
 - + bez omezení na možnosti formátu,
 - + kontextová nabídka,
 - + strukturní pohyb,
 - + syntaktické zvýraznění,
 - + validace,
 - nutná znalost formátu,
 - bez grafického formátování.
- c) Klasický WYSIWYG editor [Microsoft Word, OpenOffice.org Writer, LyX]:
 - + snadná editace,
 - bez přímé validace,
 - nenabádá ke strukturnímu značkování,
 - problematické získání strukturovaného obsahu.
- d) Strukturní WYSIWY(G|M) editor [Syntex Serna, oXygen, Altova, XMLmind]:
 - + vstup jen dovolené struktury (on-line validace),
 - + kontextová nabídka,
 - + strukturní pohyb,
 - + není nutná znalost formátu,
 - + grafické formátování,
 - z formátu použitelné, jen co je implementováno.

Vytvářet dokumenty pomocí editorů typu b) může být velice efektivní, je třeba se jen naučit několik nejpoužívanějších klávesových zkratk. Práce s tímto typem editorů však není v osnovách výuky počítačové gramotnosti a těžko si například představit, že bychom žádali každého autora sborníkového příspěvku, aby jej použil s dodaným DTD. Ten většinou není schopen použít ani dodanou šablonu do textového procesoru typu c).

Nadějná je skupina d). V nich na pozadí editace probíhá transformace XML → XML FO (Format Object) pomocí ručně či grafickým nástrojem připraveného XSL (eXtensible Stylesheet Language). Specifikuje se v něm, jak má obsah toho kterého elementu při editaci vypadat. Např. že `<chapter-title>` má být

na samostatném vertikálně odděleném řádku zobrazen větším tučným písmem. Uživatel běžně nemá ke změně tohoto písma přístup, může pouze volit z kontextově dostupných strukturních elementů. Grafické formátování přibližuje editaci textovému procesoru, čímž je pro běžného uživatele stravitelnější. Výstupním formátem je – po odstranění obalových FO – validní XML.

4.3.2. Zpracování

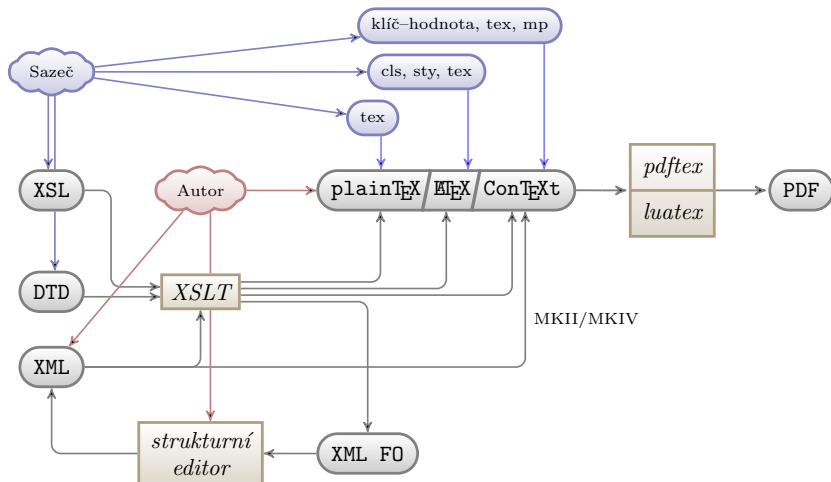
Efektivní a značně flexibilní nástroj pro konverzi a zpracování XML dokumentu je XSLT (eXtensible Stylesheet Language Transformations) procesor, např. Saxon. Snadno jím můžeme transformovat náš zdrojový text do $\text{T}_{\text{E}}\text{X}$ u/ $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ u/ $\text{C}^{\text{O}}\text{N}\text{T}_{\text{E}}\text{X}$ Tu, do (X)HTML, nebo provést výběr či řazení dat. Některé zpracování XML souboru zvládá $\text{T}_{\text{E}}\text{X}$ sám. Asi největší možnosti nabízí nová verze $\text{C}^{\text{O}}\text{N}\text{T}_{\text{E}}\text{X}$ Tu MKIV založená na $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ u [10].

Protože většina strukturně značkovacích formátů je textová, na mnohé zpracování těchto dokumentů můžeme snadno využít skriptovací jazyky (jako Perl, AWK, Ruby, Lua, ...). $\text{T}_{\text{E}}\text{X}$ istům jsou dobře známé nástroje Bib $\text{T}_{\text{E}}\text{X}$ a MakeIndex, načítající strukturovaná data vytvářená samotným $\text{T}_{\text{E}}\text{X}$ em (obr. 14). Na tomto příkladě je vidět, že mnoho práce vykoná $\text{T}_{\text{E}}\text{X}$ sám, nicméně pro mnohé úlohy, např. řazení, je efektivní využít externí nástroj a několik průchodů $\text{T}_{\text{E}}\text{X}$ u. S příchodem $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ u se toto opakované startování $\text{T}_{\text{E}}\text{X}$ u a externích programů může odstranit. Ubyde dočasných souborů, závislosti na jiných programech a celá kompilace se zrychlí.

4.3.3. Formátování

Existuje mnoho způsobů jak formátovat strukturně značkový dokument $\text{T}_{\text{E}}\text{X}$ em. O některých obecně pojednávají články [17, 15, 16, plain, $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$], [8, 10, $\text{C}^{\text{O}}\text{N}\text{T}_{\text{E}}\text{X}$ T], prakticky pak [29].

Na obr. 15 je ukázáno několik přístupů, které autor tohoto textu používá nebo je považuje za perspektivní. Dobře vidíme dělbu práce mezi autorem a sazečem i úroveň znalostí, které musí oba znát k zvládnutí své práce. Např. sazeč formátující v plain $\text{T}_{\text{E}}\text{X}$ u musí znát makrojazyk $\text{T}_{\text{E}}\text{X}$ u. Autor je plně závislý na míře abstrakce, kterou mu sazeč svými definicemi připraví. V $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ u si autor vystačí s volbou třídy, balíčků a základní definicí zkratk; sazeč však kromě makrojazyka $\text{T}_{\text{E}}\text{X}$ u musí zvládnout upravovat (složitá) makra jádra $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ u. Autor $\text{C}^{\text{O}}\text{N}\text{T}_{\text{E}}\text{X}$ Tového textu je na tom podobně jako autor využívající $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$, ale sazeč je ve výhodě. Spoustu designérské práce udělá pomocí uživatelské parametrizace klíč-hodnota bez znalosti jazyka $\text{T}_{\text{E}}\text{X}$ u. Grafickou složku návrhu realizuje v jazyce METAPOST. Pro méně běžnou sazbu je však i zde netriviální znalost $\text{T}_{\text{E}}\text{X}$ u nutností.



Obrázek 15: Datový tok přípravy a formátování strukturně značkováného dokumentu $\text{T}_{\text{E}}\text{X}$ em. Znáznorněna je dělba práce mezi autorem a sazečem. Profese redaktora, designéra a korektora jsme záměrně vypustili ze dvou důvodů: za prvé většinou komunikují jinými prostředky než přímou editací dokumentu, na kterou se zde zaměřujeme, a za druhé pro zachování únosné složitosti diagramu.

5. Závěr

V článku jsme ukázali, jak je užitečné oddělit obsah dokumentu od formátování použitím strukturního značkování. Musíme však konstatovat, že nevíme o žádném nekomerčním editoru, který by byl vhodný pro autora odmítajícího své dokumenty ručně značkovat. Do doby, než se objeví a realizuje open free projekt WYSIWYG strukturního editoru, zůstane strukturní značkování doménou IT specialistů nebo velkých koncernů a státních institucí, kteří využívají pro přípravu dokumentů komerční řešení.

Reference

- [1] Bohuslav Blažej. *Grafická úprava tiskovin*. SPN, 1. edition, 1990.
- [2] Robert Bringhurst. *The Elements of Typographic Style*. Hartley & Marks, Point Roberts, WA, USA, version 2.4 edition, 2001.
- [3] Karel Dostál. O jedné vraždě a jejím možném odčinění. *Zpravodaj Česko-slovenského sdružení uživatelů $\text{T}_{\text{E}}\text{X}$* , 4(4):163–173, 1994.
- [4] Jean-Luc Dusong and Fabienne Siegartová. *Typografie. Od olova k počítači*.

- čům. Svojtka a Vašut, 1. edition, 1997.
- [5] Karel Dyrnık. *Typograf o knihách*. Kentaur Polygrafia, Praha, 3. edition, 1993. První vydání 1911.
- [6] Jiří Gerlich and Vít Zýka. Pravidla softballu, March 1998. <http://zyka.net/soft/soft-avan.pdf>
- [7] Hans Hagen. Does pdfT_EX make things better?, 1999. <http://www.pragma-ade.com/pdftex/pre-hz.pdf>
- [8] Hans Hagen. *XML in ConT_EXt*, 2001. <http://www.pragma-ade.com/general/manuals/example.pdf>
- [9] Hans Hagen. *It's in the details*, 2003. <http://www.pragma-ade.com/general/manuals/details.pdf>
- [10] Hans Hagen. *Dealing with XML in ConT_EXt MKIV*, 2008.
- [11] Oldřich Hlavsa and František Sedláček. *Typografická písma latinková*. Státní nakladatelství technické literatury, Praha, 1. edition, 1957.
- [12] Oldřich Hlavsa and Karel Wick. *Typographia I*. Nakladatelství technické literatury, Praha, 1. edition, 1976.
- [13] Thomas Hooper. Colour contrast visualiser. <http://www.stainlessvision.com/files/ColourContrastVisualiser.swf>
- [14] Jiří Kosek. Lehký úvod do XML. *Zpravodaj Československého sdružení uživatelů T_EXu*, 10(4):122–132, 2000.
- [15] Jiří Kosek. JadeT_EX. *Zpravodaj Československého sdružení uživatelů T_EXu*, 13(1):15–26, 2003.
- [16] Jiří Kosek. PassiveT_EX. *Zpravodaj Československého sdružení uživatelů T_EXu*, 13(1):26–38, 2003.
- [17] Jiří Kosek. Použití parseru XML v T_EXu. *Zpravodaj Československého sdružení uživatelů T_EXu*, 13(1):6–14, 2003.
- [18] Jiří Kosek. *DocBook: Stručný úvod do tvorby a zpracování dokumentů*, 2007. <http://www.kosek.cz/xml/db/>
- [19] Petr Olšák. *T_EXbook naruby*. Konvoj Brno, 1. edition, 1997.
- [20] Eduard Polách. *Pravidla sazby diplomových prací*, 2000. <http://home.pf.jcu.cz/~edpo/pravidla/pravidla.pdf>
- [21] Pavel Pop, Jindřich Fléger, and Vladimír Pop. *Ruční sazba I*. Státní pedagogické nakladatelství Praha, 2. edition, 1989.
- [22] R Schlicht. *The microtype package*, 2008. <http://www.ctan.org/get/macros/latex/contrib/microtype/microtype.pdf>.
- [23] Jindřich Špalek. *Nauka o sazbě obyčejné, tabulkové, matematiky a chemie*. Spolek faktorů knihtiskáren, Praha, 3. edition, 1925.
- [24] Richard Starý. *Matematická sazba*, 2005. <http://richardstary.wz.cz/>
- [25] Philip Taylor. Knižní úprava pro uživatele T_EXu. část druhá: Praxe. In *Zpravodaj C_STUGu*, pages 38–60, 1995. Překlad Ladislav Šenkyřík.
- [26] Philip Taylor. Knižní úprava pro uživatele T_EXu. část první: Teorie. In

- Zpravodaj C_STUGu*, pages 20–37, 1995. Překlad Ladislav Šenkyřík.
- [27] Hàn Thế Thành. *Micro-typographic extensions to the T_EX typesetting system*. PhD thesis, Masarykova univerzita v Brně, fakulta informatiky, 2000. Or: TUGboat 21,4, Dec 2000.
- [28] Hàn Thế Thành, Sebastian Rahtz, and Hans Hagen. *The pdfT_EX user manual*. Pragma, 2001.
<http://sarovar.org/docman/view.php/106/66/pdfstex-s.pdf>
- [29] Zdeněk Wagner. Využití XML a L^AT_EXu při sazbě odborných knih. *Zpravodaj Československého sdružení uživatelů T_EXu*, 13(1):188–211, 2002.
- [30] Norman Walsh and Leonard Mueller. *DocBook: The Definitive Guide*, 2006.
<http://www.docbook.org/tdg/en/html/docbook.html>
- [31] Karel Wick. *Pravidla matematické sazby*. Academia, Praha, 1. edition, 1966.
- [32] Hermann Zapf. About micro-typography and the hz-program. *Electronic Publishing*, 6(3):283–288, 1993. <http://cajun.cs.nott.ac.uk/compsci/epo/papers/volume6/issue3/zapf.pdf>
- [33] Vít Zýka. Používáme pdfT_EX IV: mikrotypografické rozšíření. *Zpravodaj Československého sdružení uživatelů T_EXu*, 14(2):47–53, November 2004.
- [34] Vít Zýka. Používáme pdfT_EX IVa: hz-algoritmus jednodušeji. *Zpravodaj Československého sdružení uživatelů T_EXu*, 15(1):90–92, 2005.

Summary: Documents preparation for typesetting

In this article we express the general principles of a good document and we pose the requirements for their editing, processing and visualisation. Based on these requirements we show that an appropriate format is a structure marked document. We explain what structure marking is and describe its features. Finally we mention the tools for manipulating with structure marked documents and we sketch the ways they are formatted by T_EX.