

Zpravodaj Československého sdružení uživatelů TeXu

Jana Chlebíková

Štrukturované editovanie

Zpravodaj Československého sdružení uživatelů TeXu, Vol. 7 (1997), No. 4, 185–190

Persistent URL: <http://dml.cz/dmlcz/149801>

Terms of use:

© Československé sdružení uživatelů TeXu, 1997

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>

Elektronická forma dokumentu poskytuje priestor pre vytváranie rôznych aplikácií postavených nad novou formou dokumentu. Textové systémy,¹ ktoré okrem vytvárania dokumentov sú primárne určené pre výstup dokumentov na tlač (tlačiareň, osvit), sú jednou z najprirodzenejších aplikácií s elektronickými dokumentami. Inou aplikáciou nad elektronickými dokumentami môže byť použitie databázových systémov, ktoré budú spracovávať niektoré časti dokumentu. Príkladom môže byť vytváranie zoznamu autorov z nejakej kolekcie matematických článkov.

Pre vytváranie väčšiny takýchto aplikácií je dôležitá znalosť vnútornej štruktúry dokumentu a nie jeho prezentačných vlastností, ktoré sú dôležité len pre vizualizáciu dokumentu (na papieri alebo obrazovke počítača). Ak si ako príklad dokumentov zoberieme opäť matematické články, potom pre databázovú aplikáciu nie je podstatná prezentačná forma, akou je autor zobrazený (veľkosť a typ fondu, centrovanie, . . .), ale značky označujúce v dokumente miesto, kde začína a končí autorove meno.

Špeciálne a všeobecné značkovanie

Problém v súčasnosti rozšírených WYSIWYG editorov (napr. Microsoft Word) je, že používajú na uchovávanie dokumentu formát, ktorý obsahuje len prezentačné značky. Tieto prezentačné značky sú navyše zrozumiteľné len textovému systému, v ktorom bol dokument vytvorený. Z tohto dôvodu sa takéto značkovanie nazýva *špeciálnym značkováním* (special markup). Dokument v takomto formáte je prakticky nepoužiteľný pre inú aplikáciu, okrem jeho vytlačenia daným textovým systémom. Nevýhodou takéhoto prístupu je aj ťažko realizovateľný preklad do formátov používaných inými textovými systémami, resp. aplikáciami. A tak napríklad i dokumenty napísané v Microsoft Word (obľúbenom zvlášť anti \TeX istami) sú pre ďalšie aplikácie len ťažko použiteľné.

Ak chceme elektronický dokument používať aj iným spôsobom (ako len pre tlač), výhodnejšie je dokument vytvoriť v textovom systéme, ktorý využíva formát tzv. *všeobecného značkovania* (general markup). V tomto formáte si textový systém uchováva dokument spolu s označením logických prvkov, ktoré spolu vytvárajú štruktúru dokumentu. Prezentačia dokumentu je oddelená od obsahu

¹Tento pojem zahŕňa WYSIWYG interaktívne textové editory, i neinteraktívne textové formátovacie systémy typu \TeX .

dokumentu a je zabezpečovaná priamo textovým systémom (nie autorom!). Pre textový systém značky logických prvkov predstavujú volanie makier (preto niekedy aj procedurálne značkovanie), ktoré zabezpečia prezentáciu príslušného logického prvku v danom textovom systéme. Makrá na prezentáciu sú uložené nezávisle od dokumentu. Pre jeden typ dokumentu (zodpovedajúci nejakej štruktúre dokumentu, napr. ‘letter’) môže byť k dispozícii niekoľko súborov makier, ktoré určujú rôznu prezentáciu logických prvkov. Tieto podľa potreby umožňujú dokument vizualizovať v rôznych prezentáciach.

Pre autora dokumentu má oddelenie prezentácie od obsahu dokumentu veľký význam v tom, že sa nemusí starať o typografickú stránku dokumentu (tá patrí do rúk odborníkom), ale môže svoju energiu sústrediť na obsah dokumentu.

Pre takto jednotne označované dokumenty daného typu nie je problém vytvoriť žiadanú aplikáciu.

Nespornou výhodou je aj otvorenosť takéhoto prístupu, t.j. ľahšia prenositeľnosť dokumentov medzi textovými systémami, resp. aplikáciami — v ideálnom prípade ide len o zmenu názvov logických prvkov.

Ako príklad uveďme spomínaný matematický článok, ktorý sa skladá z nasledujúcich logických prvkov: názov článku, autor, autora adresa, klasifikácia, kľúčové slová, abstrakt, ... Systémom používajúcim všeobecné značkovanie je \TeX , presnejšie \LaTeX ovské štýly (resp. štýl „amsptt“ z \AmSTeX u). Dokument napísaný v \LaTeX ovskom štýle ‘article’ obsahuje jasne oddelené logické prvky pre názov článku, autora, adresu autora, kľúčové slová a ďalšie logické prvky.

Textové systémy, ktoré používajú všeobecné značkovanie, však nie vždy dôsledne „nútia“ autora dokumentu používať značky oddeľujúce logické prvky. Autor dokumentu môže dodávať vlastnú prezentáciu logickým prvkov a tým nie je nútený používať všeobecné značkovanie. Dokument teda „pekne vyzerá“ (aspoň podľa autora), ale je ďalej nepoužiteľný.

SGML

Prirodzeným vývojovým krokom vo všeobecnom značkovaní bola jeho ďalšia formalizácia. Formalizáciou sa dosiahla úplna nezávislosť všeobecného značkovania od formátu, ktorý používajú existujúce textové systémy na uchovávanie dokumentov. Táto formalizácia viedla k prijatiu ISO normy SGML (Standard Generalized Markup Language) definovanej v ISO 8879.

SGML je metajazyk poskytujúci prostriedky pre vytváranie DTD (Document Type Definition). DTD popisuje, z akých logických prvkov sa daný typ dokumentu skladá, v akom sú vzájomnom vzťahu (napríklad poradie prvkov v štruktúre) a čo je obsahom jednotlivých prvkov (iný logický prvok alebo samotná časť dokumentu). Formalizmus popisu DTD zaručuje jednoznačnosť zápisu takýchto značiek a tým ľahkú prenositeľnosť medzi rôznymi textovými systémami

alebo počítačovými architektúrami. SGML tiež poskytuje normalizované spôsoby reprezentácie špeciálnych znakov a symbolov použitím len ASCII znakov dostupných na bežných klávesniciach. Podstatné je, že SGML dokument neobsahuje žiadne² prezentačné značky.

Najrozšírenejším DTD je jazyk HTML (Hypertext Markup Language), známy z WWW (World Wide Web). Príkladom aplikácie postavenej nad HTML dokumentami sú Netscape alebo Mosaic na prezeranie takýchto dokumentov. Dôležité je uvedomiť si, že jeden a ten istý dokument nie je zobrazovaný úplne identicky v obidvoch aplikáciách preto, lebo každá zo spomínaných aplikácií priraduje odlišné prezentačné vlastnosti logickým prvkom (nadpisom, zoznamom položiek, ...).

Niektoré veľké vydavateľstvá, napr. Elsevier prechádzajú na SGML formát pre uchovávanie svojich publikácií. Tiež Americká fyzikálna spoločnosť mení svoj publikačný systém na SGML. V súčasnosti všetky informácie a vývoj okolo SGML je možné sledovať aj na Internete, ako štartovný bod možno doporučiť: <http://www.sil.org/sgml/sgml.html>.

Výhody štruktúrovaného prístupu

Na pohodlné vytváranie štruktúrovaných dokumentov slúžia (v súčasnosti čoraz populárnejšie) štruktúrované editory, ktoré by mali WYSIWYG prístupom umožňovať autorovi naeditovať štruktúrovaný dokument bez akýchkoľvek autorových znalostí o všeobecnom značkovaní. Veľmi špeciálnym prípadom takýchto editorov sú rozšírené HTML editory, ktoré ovšem nie vždy prehľadným spôsobom pracujú so štruktúrou HTML dokumentu. Navyše tieto editory majú HTML štruktúru obsiahnutú v svojom jadre (užívateľ nemá možnosť ju modifikovať) a tak pri akejkolvek oficiálne prijatej zmene v HTML jazyku sú nutné zmeny v programe.

Dôležité je poznamenať, že nie nutne každý štruktúrovaný editor musí pracovať priamo so SGML. Existujú aj iné jazyky na popis štruktúry dokumentu. Príkladom môže byť S-jazyk z projektu Opera, o ktorom je možno nájsť viac informácií na internetovskej adrese <http://opera.inrialpes.fr/OPERA/Thot.en.html>. SGML³ je ale prijatou normou a preto všetky štruktúrované editory, ktoré chcú byť otvorené pre iné aplikácie, musia s ňou spolupracovať. Z tohto dôvodu používame SGML terminológiu pri popise štruktúry dokumentu.

²Výnimku tvoria tzv. 'processing instructions', pomocou ktorých si môže textový systém vsúvať do dokumentu prezentačné informácie – napríklad miesta stránkových zlomov. Tieto sú štandardným spôsobom označené a inými aplikáciami ignorované.

³V poslednej dobe sa viac používa spojenie SGML/XML. XML je podmnožinou SGML a jeho cieľom je štandardizovaným spôsobom rozšíriť možnosti publikovania rôznych typov dokumentov na WWW, napr. matematických článkov.

Pozrime sa teraz bližšie na jeden konkrétny štrukturovaný SGML editor a výhody, ktoré vyplývajú z štrukturovaného prístupu k dokumentu. Editor GRIF bol navrhnutý Vincentom Quintom a ďalej vyvíjaný francúzskou firmou GRIF. Pre jeho flexibilitu a možnosť pridávať rôzne aplikácie nad dokument (resp. jeho logické prvky) tvorí tento editor jadro Euromath systému, o ktorom bližšie pojednáva nasledujúca kapitola.

GRIF podporuje niektoré základné DTD (odpovedajúce \LaTeX ovým štýlom ako ‘article’, ‘letter’, ‘slide’) a umožňuje užívateľovi pridávať vlastné DTD. Základné editovacie funkcie má GRIF rovnaké ako bežné WYSIWYG textové systémy (kontrola pravopisu, vkladanie obázkov, ...) Značky pre jednotlivé logické prvky sú pred užívateľom schované a tak pri jednoduchom editovaní dokumentu užívateľ ani nepostrehne, že pracuje so štrukturovaným editorom. Znalosť štruktúry dokumentu však GRIFU pridáva nový rozmer:

- *Štruktúra a prezentácia dokumentu sú dané, autor sa stará len o obsah dokumentu.* Logické prvky dokumentu spolu s ich vizuálnymi vlastnosťami sú definované a autor iba dopĺňa obsah jednotlivých prvkov. Autor nemusí mať žiadne znalosti o DTD, editor riadi pridávanie nových prvkov a nedovolí vykonať zmenu v logických prvkoch, ktorá by porušila štruktúru dokumentu (samozrejme vždy podľa odpovedajúceho DTD).
- *Pre každé DTD môže existovať niekoľko odlišných prezentácií.* Ako príklad uveďme dokument typu ‘letter’, ktorý môže byť zmenený prezentáciou na fax, súkromný list alebo memorandum. Podstatné je, že máme fyzicky len jeden dokument, v ktorom prevádzame všetky obsahové zmeny. GRIF poskytuje aj vlastný jazyk (tzv. P-jazyk) pre pridávanie vlastných prezentácií.
- *Jednotlivé logické prvky môžu byť zobrazené vo viacerých oknách.* V prípade editovania dokumentu napríklad typu ‘article’ je výhodná možnosť paralelného editovania literatúry vo vedľajšom okne.
- *‘Cross-referencie’ na jednotlivé logické prvky dokumentu rozširujú možnosti \LaTeX ovských ‘cross-referencií’ v dvoch základných smeroch: sú umožnené ‘cross-referencie’ medzi dokumentami a referencie slúžia ako hypertextové uzly v dokumente, t.j. dvojklik na hypertextový uzol zobrazí referencovanú časť dokumentu.*
- *Automatické vytváranie niektorých častí dokumentu, napríklad obsahu dokumentu.*
- *Inteligentný pohyb po dokumente.* Okrem pohybu po dokumente bežného z neštrukturovaných editorov, štruktúra dokumentu umožňuje realizovať navyše pohyb po logických prvkoch, vyhľadávanie logických prvkov, resp. referencovaných prvkov. Pohodlné je aj využitie ‘cross-referencií’ ako hypertextových uzlov.
- *Manipulácie s logickými prvkami v rámci DTD* zahŕňajú jednak zmeny atribútov, pridávanie nadštruktúry (napríklad štruktúru integrál nad štruk-

túru zlomok) logických prvkov, resp. zmeny jedného logického prvku na druhý. Sú kontextovo-závislé a editor ponúka autorovi uskutočniť len zmenu, ktorá je možná v rámci vybraného DTD.

- *Export do ostatných formátov* je dôležitý pre využitie dokumentu v iných systémoch. Príkladom môže byť export dokumentov do $\text{T}_{\text{E}}\text{X}$ u pre využitie typografických predností $\text{T}_{\text{E}}\text{X}$ u. Užívateľ má možnosť zdefinovať si export dokumentu (v tzv. T-jazyku) do ľubovoľného formátu vhodného pre príslušný typ aplikácie.

Zobecnenie predchádzajúcich výhod dáva základnú predstavu a možnostiach štrukturovaných editorov.

Euromath systém

Ako už bolo spomenuté, GRIF editor tvorí jadro Euromath systému. Vznik a vývoj Euromath systému bol koordinovaný cez European Mathematical Trust. Do jeho vývoja bolo zapojených niekoľko inštitúcií a firiem, spomeňme aspoň Fachinformationszentrum v Karlsruhe, ‘Institut National de Recherche en Informatique et en Automatique’ (INRIA) sídliacu v Grenoble a Euromath Centre v Kodani. V súčasnosti sa jeho ďalším vývojom, distribúciou a užívateľskou podporou zaoberá Euromath Support Center so sídlom v Bratislave. Posledná distribuovaná verzia Euromath 2.0 bola dostupná pod Unixom na SUNovské platformy.⁴

Cieľom Euromath systému je vytvoriť uniformné prostredie pre základné počítačové aplikácie, ktoré môže potrebovať matematik pri svojej práci a zabezpečiť komunikáciu medzi aplikáciami založenú na jednotnom dátovom modeli. To zahŕňa pohodlné vytváranie dokumentu, komunikáciu s databázami, využívanie existujúceho matematického softwaru a sieťových vecí ako napríklad elektronickej pošty, resp. zobrazovanie vzdialených SGML dokumentov. Komunikácia medzi aplikáciami musí obsahovať informácie o štruktúre prenášaných dát, preto je založená na dátovom modeli využívajúcom SGML.

Editovanie dokumentu prebieha WYSIWYG spôsobom v štrukturovanom editore GRIF, o ktorého základných vlastnostiach sme sa už zmienili. Väčšina matematikov je však zvyknutá používať na tvorbu dokumentov $\text{T}_{\text{E}}\text{X}$. $\text{T}_{\text{E}}\text{X}$ bez nejakého štandardného balíka makier (napr. $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$) neposkytuje dostatočné prostriedky pre popis štruktúry dokumentu. Výnimku tvoria matematické výrazy, štruktúra ktorých sa dá vo väčšine prípadov rozanalyzovať. Vďaka tomuto faktoru mohla byť realizovaná aplikácia umožňujúca preklad matematických výrazov z jednej štruktúry ($\text{T}_{\text{E}}\text{X}$) do druhej (SGML) a to v oboch smeroch. Dôsledkom tohto je možnosť vkladania matematických výrazov aj ako $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ ového re-

⁴Podrobnejšie informácie o Euromath systéme je možné nájsť na adrese <http://www.dcs.fmph.uniba.sk/~emt>.

ťažca. Navyše kedykoľvek je možné prepínať sa medzi L^AT_EXovskou a WYSIWYG (SGML) formou matematického výrazu a editovať výraz vo zvolenom formáte.

L^AT_EXovské štýly zodpovedajú možnostiam všeobecného značkovania a tak bolo možné realizovať preklad štruktúry L^AT_EXovského štýlu do odpovedajúceho DTD (napr. ‘article’ do ‘article’). Pre užívateľa to znamená možnosť importu L^AT_EXovských dokumentov do GRIFU, vizualizáciu matematických formuliek a možnosť pokračovať v editovaní dokumentu WYSIWYG formou. Opačný smer prekladu formátu bol zmieňovaný pri exporte do T_EXového formátu.

Databázovou externou aplikáciou je pripojenie PFS (Personal File System) k editoru, ktorý umožňuje matematikovi hľadať informácie o článkoch v databáze Zentralblattu (online pripojenie do Karlsruhe alebo z CDromov) a následne zobrazíť nájdené informácie ako dokument v GRIF editore. Výhodou takéhoto prepojenia je okamžité WYSIWYG zobrazenie formuliek vyskytujúcich sa v nájdených abstraktoch, resp. automatické vytváranie bibliografickej časti dokumentu.

Elektronická pošta je príkladom ďalšej aplikácie nad elektronickým dokumentom. Umožňuje exportovať dokument do zvoleného formátu (najčastejšie ASCII a T_EX) a odoslať. (Prijímateľ zatiaľ nebol implementovaný.)

Významné sú aj ostatné sieťové aplikácie, ktoré realizujú URL linky, t.j. umožňujú sprístupňovanie vzdialených dokumentov. Uvedenú vlastnosť majú aj WWW-prehliadače ale s tým podstatným rozdielom, že WWW-prehliadače pridávajú prezentáciu len HTML dokumentom. Euromath systém môže vizualizovať ľubovoľný dokument, ktorého DTD pozná. Špeciálne teda aj dokument plný matematických výrazov alebo tabuliek.

Komunikácia s matematickými programami bola zatiaľ implementovaná len na experimentálnej úrovni. Snahou je umožniť výpočet užívateľom označenej formuly zavolaním nejakého ‘computer algebra’ systému (napr. MuPAD alebo Maple) a automatické dosadenie výsledku do dokumentu.

Zhrnutie

Euromath systém je príkladom štruktúrovaného SGML editora, ktorý realizuje rôzne typy aplikácií postavených nad elektronickým dokumentom. Štruktúra dokumentu sa ukázala ako kľúčový pojem pre elektronické dokumenty. Štruktúrované dokumenty sú nesporne výhodnejšie aj pre dlhodobejšiu použiteľnosť dokumentu (i dnes ešte neznámou aplikáciou) a ponúkanú otvorenosť. Preto sa dá očakávať nástup štruktúrovaných textových editorov, medzi ktorými sa snáď objaví hviezdička s toľkými superlatívami, akou je T_EX medzi textovými systémami (len tri za všetky: najlacnejší, najspoľahlivejší, najinteligentnejší).

Janka Chlebíková
Katedra vyučovania informatiky, MFF UK Bratislava
chlebikj@dcs.fmph.uniba.sk