

Zpravodaj Československého sdružení uživatelů TeXu

Ondřej Jakubčík

Sazba chemických vzorců v TeXu

Zpravodaj Československého sdružení uživatelů TeXu, Vol. 16 (2006), No. 1, 2–10

Persistent URL: <http://dml.cz/dmlcz/149996>

Terms of use:

© Československé sdružení uživatelů TeXu, 2006

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>

Sázecí program $\text{T}_{\text{E}}\text{X}$ je mnohokrát osvědčeným nástrojem pro sazbu všech druhů dokumentů. V poslední době jsem se zabýval sazbou chemických textů. Protože se nejedná o problematiku právě triviální, pokusil jsem se poskytnout pár základních informací širšímu okruhu uživatelů programu $\text{T}_{\text{E}}\text{X}$.

Úvod

Pro sazbu chemických vzorců mi nejprve byl doporučen program ChemSketch. I přes jeho nesporné kvality jsem jej nepoužil, neboť jednak používám jiný operační systém než Windows, a jednak se snažím používat free software.

Nejprve jsem k sazbě používal $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ a systém $\text{X}^{\text{M}}\text{T}_{\text{E}}\text{X}$, ale protože zpracování dokumentu trvalo čím dál tím déle a dokument stále nabýval na rozsahu, raději jsem přešel k plainu.

Můj zájem se tedy stočil směrem k $\text{M}\text{E}\text{TAFONT}$ u. Zde jsem zůstal u tří symbolů pro vazby, protože jsem si nedovedl představit vytváření vzorců složitých organických látek.

Nakonec mne zaujal systém $\text{P}\text{P}\text{C}\text{H}\text{T}_{\text{E}}\text{X}$, který je součástí balíku $\text{C}\text{on}\text{T}_{\text{E}}\text{X}\text{t}$. Jeho velkou předností je, že funguje i v plainu, tudíž jsem mohl zachovat stávající úpravu dokumentu.

Jaký použit formát

Při výběru, v jakém formátu budeme dokument sázet (pokud není zadán), je dobré vybírat ten, který nám nejvíce vyhovuje. Pokud je dokument velmi dlouhý, je dobré se rozhodovat i podle rozsahu dokumentu, aby např. $\text{C}\text{on}\text{T}_{\text{E}}\text{X}\text{t}$ v půlce dokumentu neskončil s nedostatkem paměti. $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ zase v základním nastavení rád „plýtvá“ papírem. Pokud si výsledný dokument nebudeme tisknout sami, nebo pokud jej nebudeme předávat k tisku v PostScriptu nebo v PDF, musíme myslet i na konfiguraci systému v tiskárně, a nejlépe použijeme formát, o kterém stoprocentně víme, že je nainstalován všude (plain).

Předávání dokumentu v PostScriptu a PDF může také vyvolat poměrně velkou nevoli, protože zvláště velký dokument se může připravovat k tisku velmi dlouhou. Sto třicet stran dlouhý dokument s dvěma sty vzorci a třiceti skenovanými obrázky v PostScriptu v normální kvalitě (600 dpi) zabíral okolo 77 MB, takže jsem jej při tisku na tiskárně s 8 MB paměti tiskl po deseti stránkách.

Celková úprava dokumentu

Chemie, zvláště pak organická, bývá velmi náročná na místo na stránce. Proto, pokud nejsme omezeni počtem stran a pokud musíme stanovit rozsah předem, je lépe počínat si velkoryse a klidně pár (desítek) stran přidat, samozřejmě podle povahy dokumentu.

Co se týče velikosti, na papír formátu A4 se toho vejde více než na A5. Pokud sázíme vzorce anorganických sloučenin nebo jen velmi jednoduchých organických látek, pak vystačíme i s formátem A5. Pokud však máme sázet anorganické rovnice o několika reaktantech nebo vzorce složitých organických látek (typicky polymerů, bílkovin, nukleových kyselin atp.), je lépe použít papír A4. P_PCHT_EX sice umožňuje zjemnit kresbu a zmenšit písmo, ale mohlo by to být na úkor čitelnosti.

Písmo

Při tvorbě dokumentu, díky kterému vlastně vznikl tento článek, jsem byl postaven před problém výběru písma. Na prvních třiceti stránkách jsem zkoušel několik desítek písem, od standardních postscriptových, přes různé volně šiřitelné fonty, vyzkoušel jsem i Lido ze Střešovické písmolijny. Přes nesporné kvality dostupných písem jsem se nakonec rozhodl pro ζ variantu Computer Modern, protože v dokumentu se vyskytovalo mnoho matematických vzorců a vzhledem k povaze práce nebylo možné použít (resp. koupit) například Math Times nebo jiný matematický font.

Jednotnost názvů

Každý dokument se čte nejlépe, pokud nejen vypadá pěkně na pohled, ale také pokud pro jednu věc používáme jeden pojem. České chemické názvosloví přímo překypuje možnostmi, ale je nejlépe držet se oficiálních názvů podle IUPAC. Pokud budeme sázet z více předloh, není špatné sjednotit označení, tzn. nepsat jednou I a podruhé J (jod). Pokud nedostaneme odlišné zadání, preferujeme raději nové názvosloví, tj. oxid místo kysličník, nebo sulfid místo siřník atp.

Systematické názvy

Systematická jména anorganických sloučenin a triviální názvy jak organických, tak i anorganických látek jsou ve své podstatě nejjednodušší z hlediska typografie. Sloučeniny jako např. síran draselný, dihydrogenfosforečnan sodný, nebo anilin sázíme většinou základním písmem dokumentu.

S organickými látkami se stává problematika podstatně zajímavější. Protože téměř vždy je nutné udávat pozice atomů v molekule, sázejí se čísla, udávající pozici substituentu, oddělená navzájem čárkou bez mezer a od názvu substituentu oddělená spojovníkem. Prakticky to vypadá asi takto: 1,2-dimethylbenzen. Měli bychom dávat pozor, aby se nám v případě rozdělení slova za spojovníkem napsal spojovník i na začátek dalšího řádku, což můžeme vyřešit použitím sekvence `\discretionary{-}{-}{-}`.

Úvod do PPCH_TE_Xu

PPCH_TE_X je soubor maker pro práci s chemickými vzorci. Ačkoli je součástí Con_TE_Xtu, je možné ho použít i v L^AT_EXu a plainu. Jediná nutnost je mít nainstalovaný P_IC_TE_X nebo P_STricks.

Pokud používáte TeX Live, pak jej možná již máte nainstalovaný, a pokud nikoliv, pak je snadné jej doinstalovat. Pokud používáte jiný systém, je Con_TE_Xt a s ním i PPCH_TE_X dostupný v CTANu (`/macros/context/current`).

Po nainstalování a aktualizaci kpathsea je systém připraven k použití.

Zavedení maker

Aby bylo možno pracovat s makry, musíme je nejprve načíst. V plainu to zařídíme přidáním následujících dvou řádků na začátek souboru:

```
\input m-pictex
\input m-ch-en
```

V L^AT_EXu vložíme toto:

```
\usepackage{m-pictex}
\usepackage{m-ch-en}
```

V Con_TE_Xtu je příprava nejsnazší:

```
\usemodule[pictex,chemic]
```

Vkládání vzorců

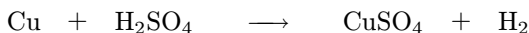
Samostatné vzorce se vkládají do bloku `\startchemical ... \stopchemical`, rovnice se vkládají do bloku uvozeného v L^AT_EXu i v plainu pomocí `$$`, v systému Con_TE_Xtu pomocí `\placeformula \startformula ... \stopformula`. V těchto blocích se jednotlivé části vzorce vkládají do příkazu `\chemical`, který se používá i pro vkládání vzorce do normálního textu.

Vzorce anorganických sloučenin

Vzorce anorganických sloučenin vytváříme pomocí příkazu `\chemical`, do kterého napíšeme atomy a indexy stejně jako v \TeX . Indexy a exponenty se nemusí uzavírat mezi dolary. Například vzorec pentahydrátu síranu měďnatého (modrá skalice) $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ vložíme do textu jako `\chemical{CuSO_4 \cdot 5 H_2O}`.

Příklad sazby vzorců anorganických sloučenin

Při výrobě elektrolytické mědi se měď rozpouští v horké kyselině sírové, přičemž vzniká síran měďnatý, chemicky CuSO_4 . Tento děj je popsán rovnicí:



Z roztoku síranu měďnatého se poté elektrolyticky vyloučí měď.

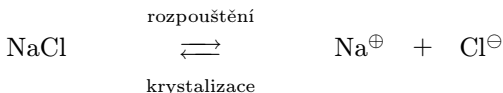
Tento odstavec byl vysázen s pomocí následujícího kódu:

Při výrobě elektrolytické mědi se měď rozpouští v horké kyselině sírové, přičemž vzniká síran měďnatý, chemicky `\chemical{CuSO_4}`. Tento děj je popsán rovnicí:

```
$$ \chemical{Cu} \chemical{PLUS} \chemical{H_2SO_4}
\chemical{GIVES} \chemical{CuSO_4} \chemical{PLUS}
\chemical{H_2} $$
```

Z roztoku síranu měďnatého se poté elektrolyticky vyloučí měď.

Jak je vidět v příkladu, $\text{PPCH}\text{\TeX}$ obsahuje i symboly jako plus, reakční šipku, a samozřejmě i jiné. Například iontovou rovnicí rozpouštění soli ve vodě



můžeme napsat takto:

```
$$ \chemical{NaCl}
\chemical{EQUILIBRIUM}{rozpuštění}{krystalizace}
\chemical{Na^{\oplus}} \chemical{PLUS} \chemical{Cl^{\ominus}} $$
```

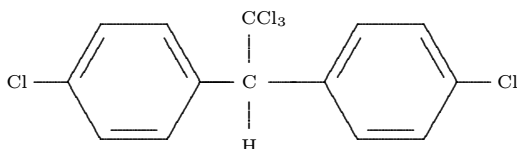
Většina symbolů má i svou zkrácenou formu, takže místo `EQUILIBRIUM` stačí psát `<->`.

Vzorce organických sloučenin

Jako u každého komplexního systému, je mnoho způsobů, jak dosáhnout kýženého výsledku. Prvním způsobem je sázení zleva doprava, tj. píšeme příkazy tak, že začneme od části vzorce nejvíce vlevo a postupně se propracováváme doprava. Tento přístup jsem používal poměrně často, avšak nevýhodu jsem poznal

až poté, co jsem musel vzorec podle zadání opravit. Proto se mi zdá s odstupem času lepší pracovat opravdu chemicky a strukturu rozložit na jednotlivé stavební bloky. Protože `PPCHTeX` umožňuje ukládat části vzorců, je asi nejlepší vytvořit si seznam stavebních látek a pak už molekuly skládáme dohromady velmi snadno. Výhoda vyvstane v okamžiku korektury, neboť pokud uděláme chybu ve zbytku, opravíme ji jen jednou a nemusíme procházet celý dokument.

Abych ukázal oba způsoby sazby, představme si, že budeme chtít vysázet následující vzorec:



Způsobem zleva doprava by vypadal asi takto:

```
\startchemical
\chemical [ONE,ZO,SIX,ROT2,B,EB135,R36] [Cl]
\chemical [MOV1,ONE,ZO37,SB37] [C,H,CCl_3]
\chemical [SIX,MOV1,ROT2,B,EB135,R63,MOV1,ONE,ZO] [Cl]
\stopchemical
```

Druhým způsobem (struktura) asi takto:

```
\definechemical [chlorfenyl1] {
\chemical [SAVE,SUB1,SIX,ROT2,B,EB135,R36,RZ6,RESTORE] [Cl]
}
```

```
\definechemical [chlorfenyl2] {
\chemical [SAVE,SUB3,SIX,ROT4,B,EB246,R36,RZ6,RESTORE] [Cl]
}
```

```
\startchemical
\chemical [ONE,ZO37,SB1357] [C,H,CCl_3]
\chemical [chlorfenyl1]
\chemical [chlorfenyl2]
\stopchemical
```

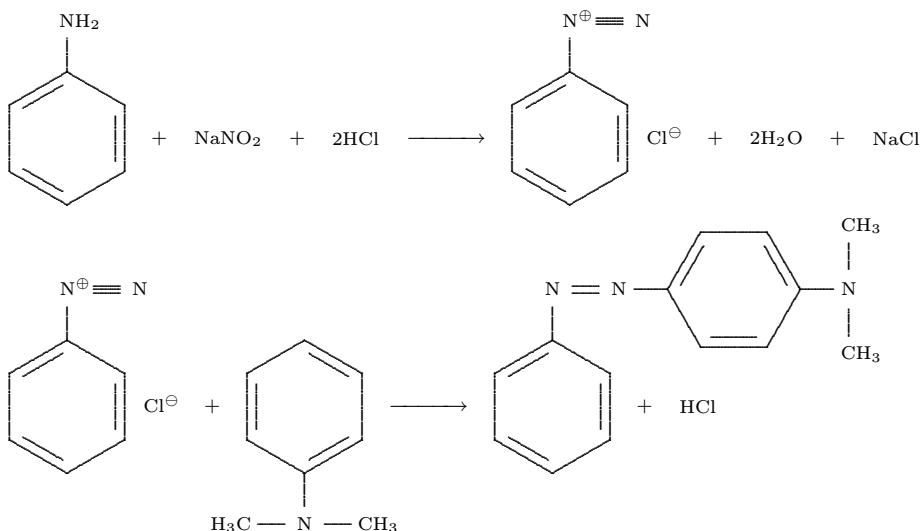
Druhý zápis je delší, nicméně struktura je jasná i z kódu. Na druhou stranu, pokud se jedná o vzorec tak jednoduchý jako $\text{H}-\text{C}\equiv\text{C}-\text{H}$, bylo by použití něčeho jiného než `\chemical{H,-,C,---,C,-,H}` čiročiré plýtvání.

Jednotlivé příkazy jsou pojmenovány poměrně intuitivně, takže např. `MOV` je posun, `ROT` otočení, `B` vazba, `SB` jednoduchá vazba, `EB` dvojná vazba uvnitř cyklu atd.

Čísla u příkazů určují směr nebo pozici. U neotočených struktur je 1 vpravo. Všechny příkazy jsou důkladně vysvětleny v manuálu [1].

Příklad sazby vzorců organických sloučenin

Diazotací anilinu pomocí NaNO_2 v kyselém prostředí vznikne benzendiazoni-
umchlorid, který kopulací s dimethylanilinem přeměníme na azobarvivo — 4-di-
methylaminoazobenzen.



Tyto vzorce vysázíme například takto:

Diazotací anilinu pomocí NaNO_2 v kyselém prostředí
vznikne benzendiazoni-
umchlorid, který kopulací s dimethylanilinem
přeměníme na azobarvivo --- 4-dimethylaminoazobenzen.

```
{\setupchemical [width=fit,height=3000,bottom=1000]
\hbox{
\startchemical
\chemical [SIX,B,EB135,R6,RZ6] [NH_2]
\stopchemical
\startchemical
\chemical [SPACE,PLUS,SPACE]
\stopchemical
\startchemical
\chemical [CHEM] [NaNO_2]
\stopchemical
\startchemical
\chemical [SPACE,PLUS,SPACE]
\stopchemical
\startchemical
```

```

\chemical [CHEM] [2HCl]
\stopchemical
\startchemical
\chemical [SPACE,GIVES,SPACE]
\stopchemical
\startchemical
\chemical [SIX,B,EB135,R6,PB:RZ6,ONE,Z0,OFF1,TB1,MOV1,Z0,
PE] [\SL{N^{\oplus}},N]
\stopchemical
\startchemical
\chemical [SPACE,SPACE,Z0] [\SL{Cl^{\ominus}}]
\stopchemical
\startchemical
\chemical [SPACE,PLUS,SPACE]
\stopchemical
\startchemical
\chemical [CHEM] [2H_2O]
\stopchemical
\startchemical
\chemical [SPACE,PLUS,SPACE]
\stopchemical
\startchemical
\chemical [CHEM] [NaCl]
\stopchemical
}}

```

```

{\setupchemical [width=fit,height=5000,bottom=2000]
\hbox{
\startchemical
\chemical [SIX,B,EB135,R6,PB:RZ6,ONE,Z0,OFF1,TB1,MOV1,Z0,
PE] [\SL{N^{\oplus}},N]
\stopchemical
\startchemical
\chemical [SPACE,SPACE,Z0] [\SL{Cl^{\ominus}}]
\stopchemical
\startchemical
\chemical [SPACE,PLUS,SPACE]
\stopchemical
\startchemical
\chemical [SIX,ROT3,B,EB135,R6,PB:RZ6,ONE,Z051,SB51,PE] [N,H_3C,
CH_3]
\stopchemical
}

```

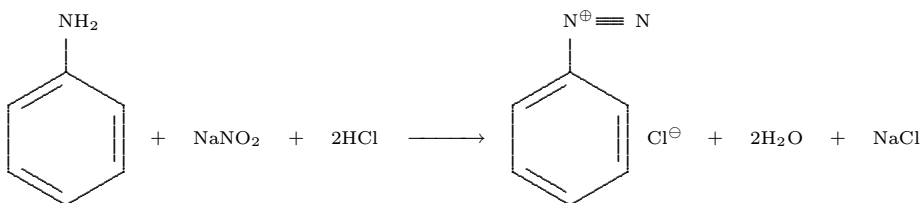


```

\startchemical
\chemical [SPACE,GIVES,SPACE]
\stopchemical
\startchemical
\chemical [SIX,B,EB135,R6,PB:RZ6,ONE,ZO,DB1,MOV1,ZO,SIX,MOV1,ROT2,
B,EB135,R36,MOV1,ONE,ZO37,SB37,PE] [N,N,N,CH_3,CH_3]
\stopchemical
\startchemical
\chemical [SPACE,PLUS,SPACE]
\stopchemical
\startchemical
\chemical [CHEM] [HCl]
\stopchemical
}}

```

Oku typografa samozřejmě neunikne, že účaří NaNO_2 a HCl nebo $2\text{H}_2\text{O}$ a NaCl jsou vůči sobě posunuta. Tento nedostatek $\text{PPCHT}_{\text{E}}\text{X}$ neřeší a ani nastavení $\backslash\text{fontdimen}16$ a $\backslash\text{fontdimen}17$ nepomůže, neboť $\text{PPCHT}_{\text{E}}\text{X}$ oba registry při sazbě přepíše vlastní hodnotou, a obnoví hodnoty, jakmile dosází vzorec. Opticky toto nepůsobí pěkně, proto pokud na vzhledu obzvláště záleží, pokud např. $[\text{HCl}]$ přepíšeme na $[\text{HCl}_{\{\vphantom{2}\}}]$, dostaneme již opticky pěkný výsledek:



Závěr

Sazba chemických vzorců v $\text{T}_{\text{E}}\text{X}$ u je poměrně náročná, ale protože $\text{PPCHT}_{\text{E}}\text{X}$ je dobře dokumentován, jde o náročnost spíše časovou. Při práci na chemickém dokumentu není na škodu znalost aspoň středoškolské chemie a je dobré spolupracovat s chemiky, neboť pokud si nejsme jisti například vaznostmi, je pravděpodobné vnesení chyb, které se zpětně špatně hledají a opravují. Periodická tabulka prvků a stručný přehled chemického názvosloví jsou nenahraditelné při jakékoli práci s chemickými vzorci. Korekturu je dobré rozdělit na korekturu faktickou, kterou provádí chemik, a typografickou, kterou provádí sazeč/typograf.

Tento článek jsem psal spíše jako ukázkou toho, co je se systémem $\text{PPCHT}_{\text{E}}\text{X}$ možné vysázet, a vzhledem k rozsahu není možné suplovat dokumentaci, o což

jsem se ani nesnažil. Přestože je dokumentace v angličtině, každý příkaz je ilustrován v několika příkladech, a tak se domnívám, že každý, kdo bude sázet chemické vzorce, bude schopen se v manuálu velmi rychle zorientovat.

Reference

- [1] Hagen, J., Otten, A. F.: *PPCH_TE_X: A macropackage for typesetting chemical structure formulas*, Pragma ADE, 2001 (<http://www.pragma-ade.com/general/manuals/mp-ch-en.pdf>)
- [2] Olšák, P.: *T_EXbook naruby*, Konvoj Brno, 2001
- [3] Vacík, J. a kol.: *Přehled středoškolské chemie*, SPN Praha, 1990
- [4] Bína, J. a kol.: *Malá encyklopédia chémie*, Obzor Bratislava, 1981

Summary: Typesetting chemical formulas in T_EX

Typesetting program T_EX is used worldwide for formatting various kinds of documents. In this article, a short presentation of its abilities on typesetting various chemical structures is given.

Ondřej Jakubčík
e-mail: ojakubcik@seznam.cz

Hrstka tipů pro T_EXovskou sazbu

PETR BŘEZINA

Rozsah vyjádřený pomlčkou

Jednou z věcí, kterými se odlišuje knižní sazba od strojopisu, je pomlčka. Zatímco se na psacím stroji používá pro spojovník a pomlčku stejný znak, ve fontech knižního písma nalezneme kromě spojovníku dokonce dvě pomlčky. Jedna je dlouhá (čtverčiková) a druhá je krátká (půlčtverčiková). V anglických textech se každá z těchto dvou pomlček objevuje v jiné funkci. V české sazbě se však ve všech funkcích používá zpravidla jen jedna pomlčka. Začátečníci bývají na rozpacích, kterou pomlčku si mají vybrat. Podíváme-li se do starších českých knih, zjistíme, že mistři sazeči dávali přednost spíše dlouhé pomlčce.