[Metodika - Content - Web FIT](https://docs.google.com/document/d/1daC_IfOVNaiPzH23boFZ8jdWtHRj8-HVsUS5R2HXBsE/edit#)

Pevné mezery v češtině

Nedělitelné mezery či také &nbsp;

Tento dokument formalizuje pravidla pevných mezer pro český jazyk a definuje příslušné regulární předpisy regulárních náhrad. Opravy pevných mezer vychází z internetové jazykové příručky pro český jazyk[[1]](#footnote-0), ze slovníku pevných mezer[[2]](#footnote-1) a zkratek[[3]](#footnote-2). Regulární výrazy jsou v syntaxi Perl[[4]](#footnote-3) s využitím POSIX[[5]](#footnote-4).

Předpisy jsou koncipované, aby prováděly i nejednoznačné náhrady. Proto místo pevných mezer vkládají speciální znak „·“, tzv. interpunct nebo také middle-dot[[6]](#footnote-5) (&#183;). Před jeho náhradou za pevnou mezeru je nutné výstup projít a tzv. false-positive náhrady smazat (nahradit zpět za normální mezeru).

## Zvýraznění existujících pevných mezer

Před samotnou náhradou je doporučeno převést existující pevné mezery na jiný znak pro účely rozlišení původních pevných mezer od vygenerovaných kandidátů – např. znakem vlnovka (~). Tento krok je možné vynechat.

s/\x{00A0}/~/g

## Jednopísmenné předložky a spojky

Za jednopísmenné předložky patří pevná mezera vždy včetně velkých písmen (na začátku věty).

{ k, s, v, z, o, u, a, i }

s/\b(k|s|v|z|o|u|a|i) /$1·/gi

## Jednoslovné zkratky a značky

Za jednoslovné zkratky a značky psané bez tečky patří pevná mezera vždy včetně zkratek a značek začínajících na velké písmeno.

{ ca, cca, fa, fce, ~~viz~~ }

s/\b(Ca|Cca|Fa|Fce|ca|cca|fa|fce) /$1·/g

O něco jiná situace je u zkratek a značek psaných s tečkou. Tam je potřeba následně ověřit, zda na uvedené znaky nekončí věta. V takovém případě tam pevná mezera nepatří.

{ č., max., min., r., s., st. }

s/\b(č|max|min|r|s|st)\. /$1.·/g

## Víceslovné zkratky a značky

Víceslovné zkratky a značky mají obsahovat pevné mezery mezi jednotlivými částmi.

{ a. s., č. p., j. č., k. s., mn. č., n. l., n. m., př. n. l., s. r. o., spol. s r. o., t. č., t. r., v. r., v. o. s., v. v. i. }

s/\ba\. s\./a.·s./g  
s/\bč\. p\./č.·p./g  
s/\bj\. č\./j.·č./g  
s/\bk\. s\./k.·s./g  
s/\bmn\. č\./mn.·č./g  
s/\bn\. l\./n.·l./g  
s/\bn\. m\./n.·m./g  
s/\bpř\. n\. l\./př.·n.·l./g  
s/\bs\. r\. o\./s.·r.·o./g  
s/\bspol\. s r\. o\./spol.·s·r.·o./g  
s/\bt\. č\./t.·č./g  
s/\bt\. r\./t.·r./g  
s/\bv\. r\./v.·r./g  
s/\bv\. o\. s\./v.·o.·s./g  
s/\bv\. v\. i\./v.·v.·i./g

## Tituly a hodnosti

Mezi akademickým titulem a jménem má být pevná mezera stejně jako mezi všemi akademickými tituly před jménem. Pokud za titulem nenásleduje velké písmeno, k náhradě nedojde. To se může stát, pokud titul není uvedený před jménem nebo před jiným titulem.

{ Bc., BcA., Ing., Ing. arch., Mgr., MgA., MUDr., MDDr., MVDr., JUDr., RNDr., PhDr., PharmDr., ThDr., doc., Doc., prof., Prof., B.A., M.A., BSc., MSc., LL.B., LL.M., Mmed., M.D., Dipl.-Ing., EngD., ThLic., Dr. }

s/\b(Bc|BcA|Ing|Ing. arch|Mgr|MgA|MUDr|MDDr|MVDr|JUDr|RNDr|PhDr|PharmDr|ThDr|doc|Doc|prof|Prof|B.A|M.A|BSc|MSc|LL.B|LL.M|Mmed|M.D|Dipl.-Ing|EngD|ThLic|Dr)\. (?=[[:upper:]])/$1.·/g

Totéž platí pro vojenské hodnosti, resp. pro vojenské hodnosti začínající na velké písmeno (stejná množina na začátku věty).

{ voj., sv., svob., des., čet., rtn., rtm., nrtm., šrtm., pprap., prap., nprap., šprap., ppor., por., npor., kpt., mjr., pplk., plk., brig. gen., genmjr., genpor., arm. gen. }

s/\b(voj|sv|svob|des|čet|rtn|rtm|nrtm|šrtm|pprap|prap|nprap|šprap|ppor|por|npor|kpt|mjr|pplk|plk|brig. gen|genmjr|genpor|arm. gen)\. (?=[[:upper:]])/$1.·/g

s/\b(Voj|Sv|Svob|Des|Čet|Rtn|Rtm|Nrtm|Šrtm|Pprap|Prap|Nprap|Šprap|Ppor|Por|Npor|Kpt|Mjr|Pplk|Plk|Brig. gen|Genmjr|Genpor|Arm. gen)\. (?=[[:upper:]])/$1.·/g

Tituly bez teček (a spojka titulů) mají díky absenci tečky jednoduší předpis. Jinak se pravidlo neliší.

{ Blaw, Mlaw, ~~et~~ }

s/\b(Blaw|Mlaw) (?=[[:upper:]])/$1·/g

### Víceslovné tituly a hodnosti (pevná mezera mezi)

{ Ing. arch., brig. gen., Brig. gen. }

s/\bIng\. arch\./Ing.·arch./g  
s/\b(B|b)rig\. gen\./$1rig.·gen./g

## Číselné údaje, data, jednotky

Za číslovkou bývá typicky jednotka nebo obecně podstatné jméno, ke kterému se číslovka vztahuje. Tam patří pevná mezera. Pokud jde o číslovku řadovou, platí totéž. Zde opět pozor na potenciální konec věty.

s/(\d\.?) /$1·/g

## Pomlčka

Podle uvedené teorie patří nedělitelná mezera před každou pomlčku, aby pomlčkou nezačínal řádek.

s/ –/·–/g

# Opravy HTML

Náhrady, které vznikají konceptuálně špatně se následujícími předpisy vrátí.

s/<a·href/<a href/g

s/class="fa·/class="fa /g

# Aplikování regulárních náhrad

Uvedené výrazy je možné aplikovat obecně pomocí funkce, která podporuje náhrady regulárních výrazů v syntaxi Perlu. Např. tedy příkaz perl, sed nebo funkce preg\_replace v PHP. Na ostré nasazování textů pak stačí triviálně převést viditelné znaky na pevné mezery.

Filtrování textu tohoto dokumentu na sadu regulárních výrazů a jejich zřetězení přes středník

cat [regex\_file] | grep "^s/" | paste -sd ";" -

Náhrada pomocí příkazu perl s vyznačením zřetězení regulárních výrazů

perl -CSD -Mutf8 -pi.bak -e '[regex1];[regex2]' [data\_file]

Spojení výše uvedeného

perl -CSD -Mutf8 -pe "$(cat [regex\_file] | grep "^s/" | paste -sd ";" -)" [file]

Převod viditelného znaku middle-dot na pevnou mezeru

sed -i 's/·/\x{00A0}/g' file

Odkaz pro export dokumentu do textové podoby

https://docs.google.com/document/d/1zWGFw7E4xU4IzpqwAraRSQPbswuap8Q\_jPl5lueAzEs/export?format=txt

1. <http://prirucka.ujc.cas.cz/?id=880> [↑](#footnote-ref-0)
2. <http://www.liteera.cz/slovnik/pevna-mezera> [↑](#footnote-ref-1)
3. <http://www.liteera.cz/slovnik/zkratky> [↑](#footnote-ref-2)
4. <http://perldoc.perl.org/perlre.html> [↑](#footnote-ref-3)
5. <http://www.regular-expressions.info/posixbrackets.html> [↑](#footnote-ref-4)
6. <https://en.wikipedia.org/wiki/Interpunct> [↑](#footnote-ref-5)