

# Pravidlá na kontrolu slovenskej gramatiky

Autor: Ľuboš Lehotský



# Ciele práce

Vytvorenie pravidiel na kontrolu slovenskej gramatiky

Návod na tvorbu pravidiel pre potenciálneho spoluautora

# Štruktúra práce

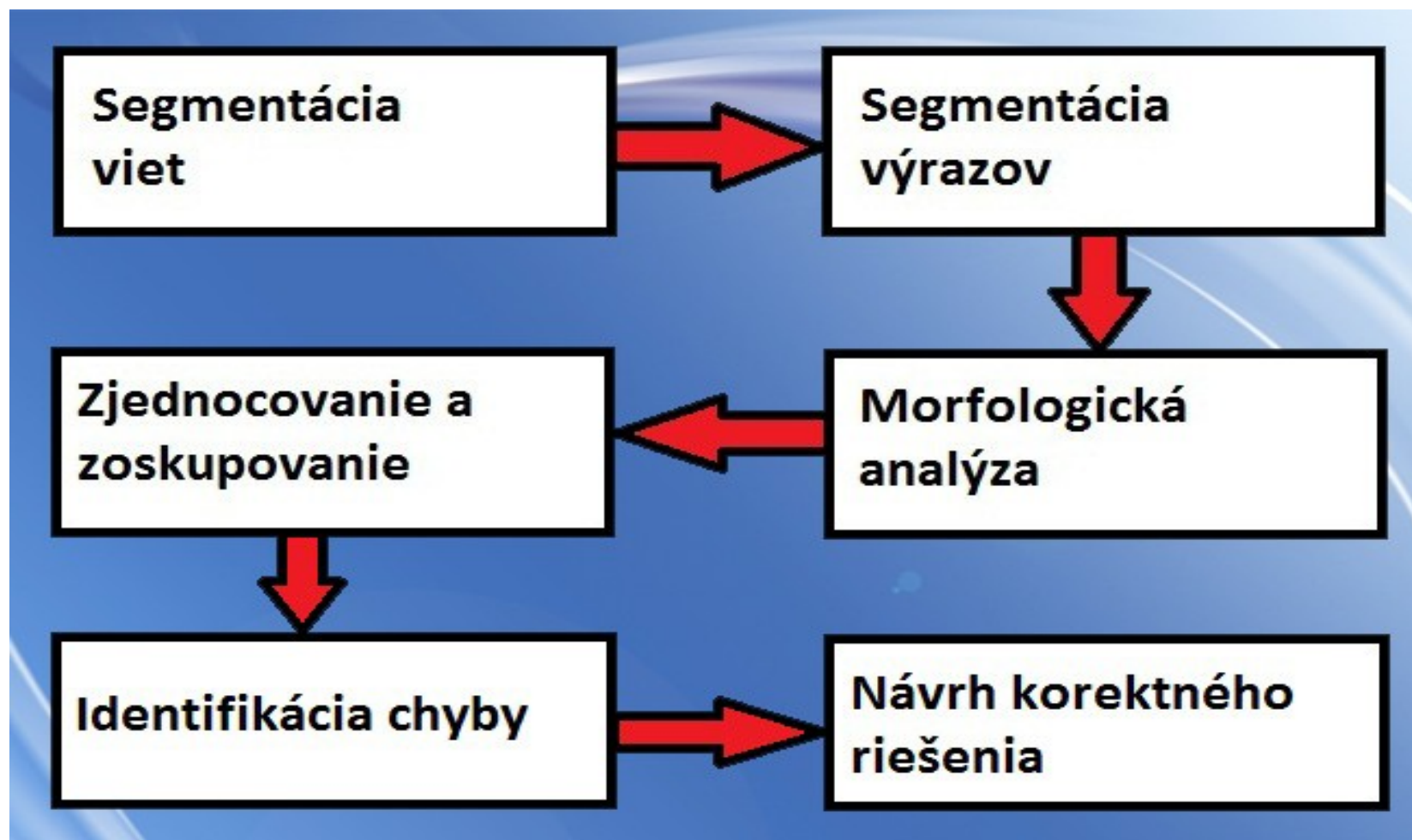
- Nástroj LanguageTool
- Vývoj slovenskej verzie kontroly gramatiky
- Výsledky



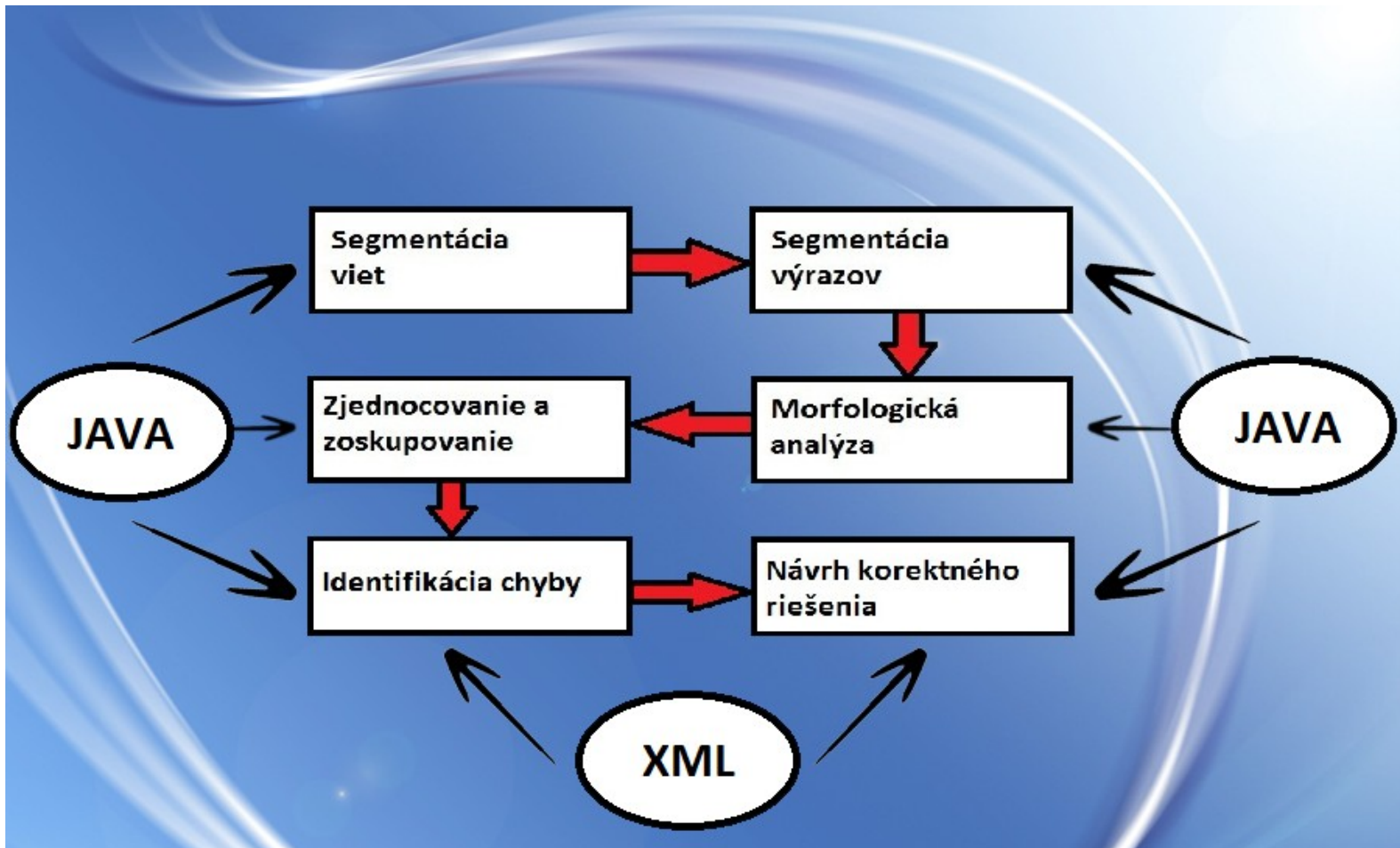
# LanguageTool (LT)

- Podpora jazykov v súčasnosti
- Architektúra a prístupy
- Použitie v praxi

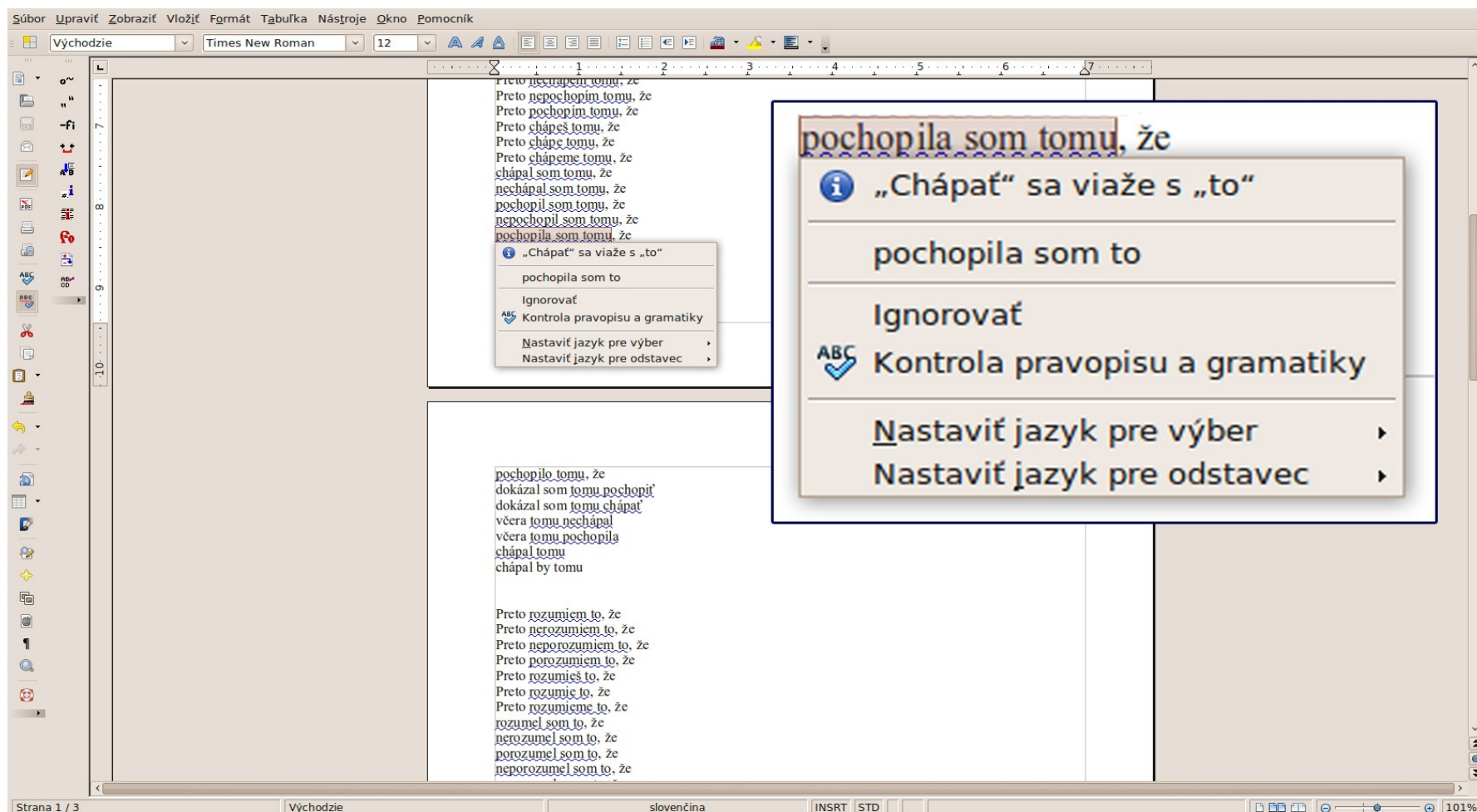
# Analýza textu v LT



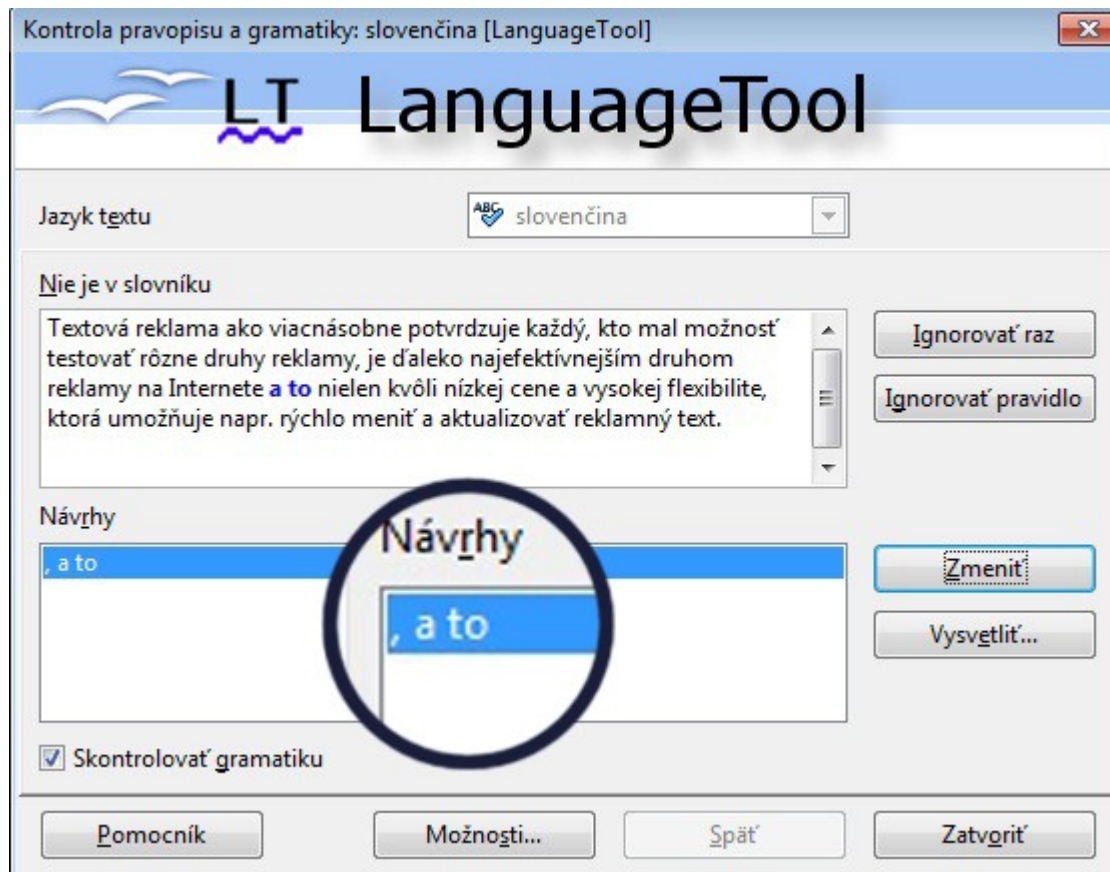
# Implementácia architektúry



# Použitie v OpenOffice.org



# Použitie; návrh riešenia





# Vývoj slovenskej verzie

- Východiská
- Zásady tvorby pravidiel
- Kompilácia
- Testovanie
- Pridanie do súčasnej verzie LT

# Tvorba pravidiel, základný postup

- Nájdí a konkretizuj neošetrenú chybu
- Použi chybné slová, prípadne celý koncept na tvorbu pravidla
- Zovšeobecni pravidlo, použi POS značky
- Testuj na rôznych prípadoch

# XML pravidlá

```
1.. grammar.xml
181 <message>Odstráňte úvodzovky po skratke tzv. <suggestion><match no="4"></match></suggestion>.</message>
182 <short>Typografická chyba.Odstráňte úvodzovky po skratke tzv.</short>
183 <example correction="testík" type="incorrect">Toto je tzv. <marker>„testík</marker> slova" pomoc.</example>
184 <example type="correct">Toto je tzv. testík slova.</example>

<rulegroup id="CITIT_SA_1_PAD" name="Cítiť sa sa spája s 1. pádom">
  <rule>
    <pattern>
      <token regexp="yes">Cítiť|Cítim|cítia|cítíme|cítiš|cíti|cítite</token>
      <token>sa</token>
      <token postag_regexp="yes" postag="A...7."/>
      <token></token>
    </pattern>
    <message>Zmeňte pád prídavných mena na nominatív: <suggestion><match no="1"></match> sa <match no="3"></match>
    <short>Typografická chyba.Cítiť sa sa spája s 1. pádom</short>
    <example correction="užitočný" type="incorrect">Cítim sa <marker>užitočným.</marker></example>
    <example type="correct">Cítim sa užitočný.</example>
  </rule>
</rulegroup>

203
204 <!-- pravidlo funguje, ale spôsobuje priveľa falosných poplachov
205 <rulegroup id="DVOJBODKA_PRIAMA_REC" name="Dajte dvojbodku pred priamu reč">
```

# XML pravidlá; kľúčové časti

- Element „pattern“
  - Atribúty „mark\_from“ a „mark\_to“
  - Atribút „skip“
- Element „suggestion“ a možnosti modifikácie návrhu

# XML pravidlá; príklady

## Pravidlo na opravu „koho/čí“

### ■ Pattern

```
<pattern>
```

```
  <token >koho</token>
```

```
  <token postag_regexp="yes" postag="S.(m|i).1"></token>
```

```
</pattern>
```

### ■ Suggestion

```
<suggestion>čí <match no="2" /></suggestion>
```

# XML pravidlá; príklady

## Miesto na

### ■ Pattern

```
<pattern mark_from="1">  
  <token >miesto</token>  
  <token regexp="yes" skip="-1" >k|ku</token>  
  <token postag_regexp="yes" postag="S...3"></token>  
</pattern>
```

### ■ Suggestion

```
<suggestion> miesto na <match no="3" postag_regexp="yes"  
  postag="S...4"/></suggestion>
```

# Tvorba pravidiel, možné problémy

- Nahradenie nesprávnym počtom slov
  - Nesprávna hodnota atribútu „mark\_from“ alebo „mark\_to“
  - Vysoká hodnota atribútu „skip“
- Veľké množstvo falošných poplachov
  - Prílišné zovšeobecnenie POS značiek
  - Vysoká hodnota atribútu skip



# JAVA pravidlá

- Vhodnosť použitia
- Implementácia a kompilácia
- Vzorové pravidlo



# JAVA pravidlá, vzor

## Equivalent pattern

```
for (int i = 1; i < tokens.length; i++) {  
  
    if (token.charAt(0)=='.' || ... {  
        bodka=true;  
    }  
  
    if (tokens[i].hasPosTag("AAfs1x") ||  
        {pady = true;  
    }  
  
    if (tag3 && !bodka) { ...  
    }  
  
    ...  
}
```

## Suggestion

```
final String msg ="Zmeňte začiatocné  
    písmeno na veľké: <suggestion> "  
    +prve + spravne + " </suggestion>";  
  
final RuleMatch ruleMatch = new  
    RuleMatch(this, pos, pos  
        + pos2, msg, "Zmeňte  
    začiatocné písmeno na veľké: ");  
  
ruleMatches.add(ruleMatch);
```

# Zhrnutie a výsledky

## ■ Osobný prínos

- Vytvorené pravidlá (asi 50 pravidiel v procese začlenenia do budúcej verzie)
- Návod pre spolupracovníkov prístupný na adrese [http://www.sk-spell.sk.cx/file\\_download/86](http://www.sk-spell.sk.cx/file_download/86)



# Výzva

- Nutnosť spolupráce jazykového odborníka
- Rozšírenie spolupracovníkov s cieľom vytvorenia komunitného projektu

# Kontakty a linky

- Zdenko Podobný – vedúci projektu  
<http://www.sk-spell.sk.cx/>  
[zdenop@gmail.com](mailto:zdenop@gmail.com)
- Luboš Lehotský:  
[lubo.lehotsky@gmail.com](mailto:lubo.lehotsky@gmail.com)
- Vývojárska verzia: [http://www.sk-spell.sk.cx/file\\_download/87](http://www.sk-spell.sk.cx/file_download/87)



Ďakujem za pozornosť