

Extending LanguageTool, a style and grammar checker

Workshop Notes

Daniel Naber, Marcin Miłkowski

11:15 – 13:00
Friday, 2009-11-06



Agenda

- Intro
- Workshop
 - Download and use LanguageTool
 - understand XML rules
 - write new XML rules
 - add support for a new language
 - the community website
 - setup Eclipse and write Java rules
- Resources

Intro



- LanguageTool = Open Source language checker for English, French, German, Polish + more
- Not (just) grammar

Intro: How to express Rules

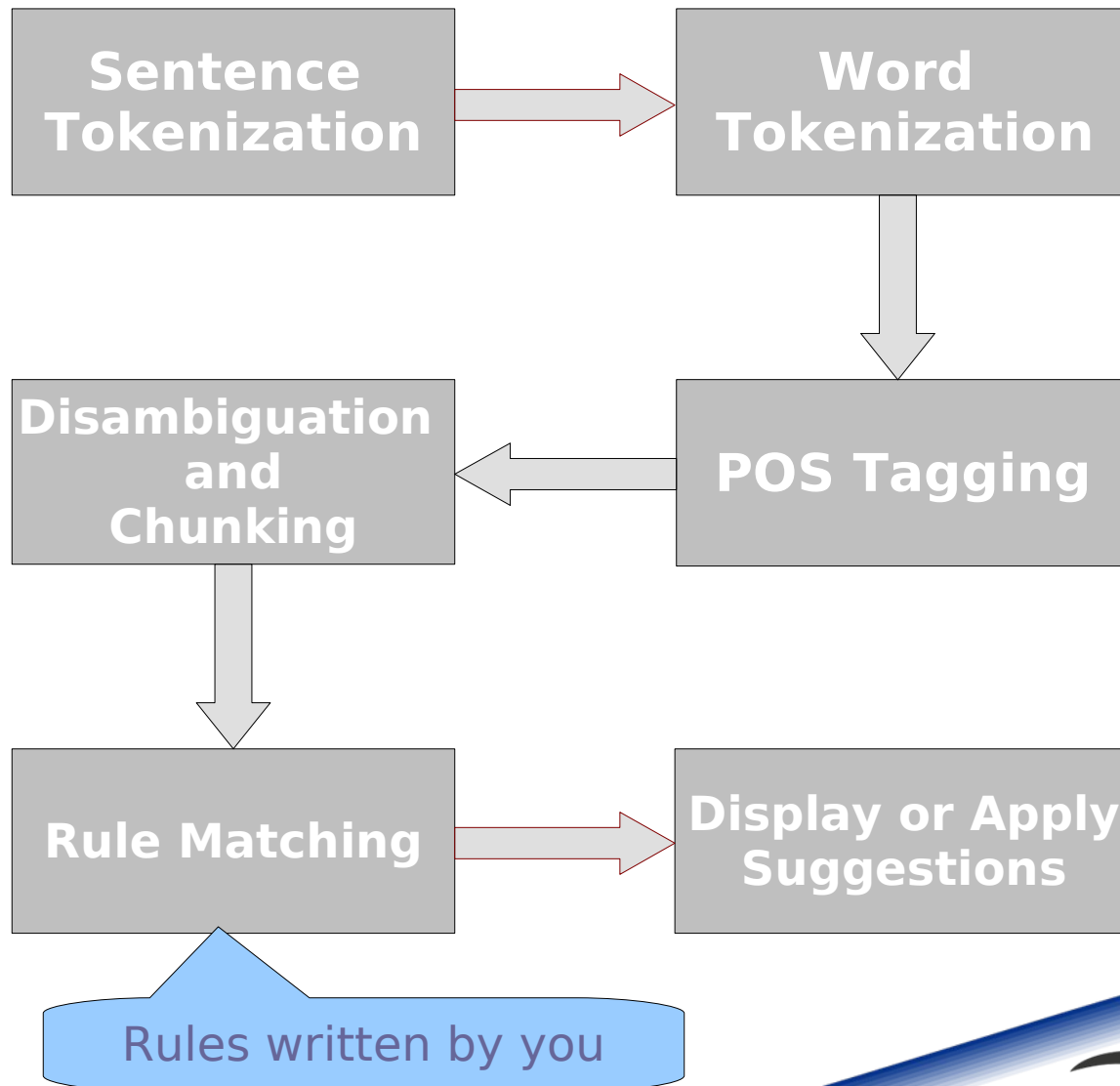


what's allowed



what's forbidden -
LanguageTool knows errors,
i.e. what's forbidden

Intro: Flow



Workshop Notes



- download the current OXT
- install it in OOo and try on English text
- try the GUI:
`java -jar LanguageToolGUI.jar`
- command line and HTTP server

Workshop Notes



- show that the English rules are in in rules/en/grammar.xml
- explain that using an XML editor is a good idea
- explain the syntax of a simple rule - one of the English redundant phrases, this is simple but useful
- create a rule to match "foo bar"

Workshop Notes



- restart LT and try it (GUI version, not OOo)
- run a test using testrules.sh/testrules.bat
- it's important to use real examples
- explain a more complex rule

Workshop Notes



- what is a POS tagger
 - mention SENT_START / SENT_END tags available to all languages
- it_is - a couple of cases, one with POS tags, and a couple others (as a rulegroup)
- “both... and” in English, as it requires skipping
- a rule for repeated phrases (can be French or Dutch)

Workshop Notes



- complex suggestions and testing corrections
- what is a synthesizer

Workshop Notes



- show how we add a new language by adapting the Java code
- check out code from CVS
 - help is at http://sourceforge.net/scm/?type=cvs&group_id=110216
 - module name is “JLanguageTool”
- call “ant”
- see <http://www.languagetool.org/development/#newlanguage>
- you can usually start working in a different language

Workshop Notes



- complex suggestions and testing corrections
- what is a synthesizer

Workshop Notes



- community.languagetool.org website
 - how to use it to find false alarms in Wikipedia
 - how to find content for rules (Wikipedia typos page, community feedback...)

Workshop Notes



- How to setup LanguageTool in Eclipse
 - check out from CVS (see above)
 - File -> Import...
 - General, Existing Projects into Workspace
 - select the "JLanguageTool" directory just checked out from CVS
- structure of a rule written in Java
- write your own simple rule

Finally... Resources



- All resources are at <http://www.languageTool.org> or are linked from there
- Don't miss the Wiki
- Discussion happens on the mailing list