

LanguageIdentifierPlugin

Identity

- plugin name: languageidentifier
- plugin version: none
- provider: [SamiSiren](#), [JeromeCharron](#)
- plugin home url: [LanguageIdentifierPlugin](#)
- plugin download url: Included with nutch source distribution
- license: Same as Nutch
- short description: Analyzer plugin that identifies the language of documents.
- long description:
- configureable parameters: lang.ngram.min.length, lang.ngram.max.length, lang.analyze.max.length
- meta data added to index: lang
- required jars:
- plugin extension points:
- plugin extension point interface:
- plugin extension point xml snippet:

Documentation

- [LanguageIdentifier](#) - some notes about the new language identifier version
- [LanguageIdentifierBenchs](#) - some benchmarks on language identifier with many parameters combinations.

Implemented Languages and their ISO 636 Codes

- da Danish
- de German
- el Greek
- en English
- es Spanish
- fi Finnish
- fr French
- hu Hungarian
- it Italian
- nl Dutch
- pl Polish
- pt Portuguese
- ru Russian
- sv Swedish