

# Lexeme Forge

a web application for creating inflection dictionaries

## Overview

Lexeme Forge (*Kuźnia*) is a web application allowing collaborative creating of new linguistic resources using various imported sources as well as data added directly. It is developed in Python with the Django web framework and uses JavaScript with jQuery and jQueryUI on the client side.

The Forge users have different roles: Lexicographers edit existing lexeme descriptions and create new ones; Publishers define dictionaries (lexeme sets) and manage them, as well as administer the work of Lexicographers; Patternmakers correct and define inflection patterns, etc. (a user can have more than one role).

Lexeme descriptions in the system have different statuses, depending on their reliability: 'candidates' are incomplete or doubtful, 'described' are complete, but need manual checking and 'confirmed' are checked by a Lexicographer.

## Jan Szejko

Institute of Computer Science  
Polish Academy of Sciences

## Inflection tables

For every described lexeme there's a generated nice inflection table, which can be seen in the respective tab in the lexeme view.

	l. p.	l. m.
M.	acetylocholinoesteraza	acetylocholinoesterazy
D.	acetylocholinoesterazy	acetylocholinoesteraz
C.	acetylocholinoesterazie	acetylocholinoesterazom
B.	acetylocholinoesterazę	acetylocholinoesterazy
N.	acetylocholinoesterazą	acetylocholinoesterazami
Ms.	acetylocholinoesterazie	acetylocholinoesterazach
W.	acetylocholinoesterazo	acetylocholinoesterazy

## Lexeme history

All changes made by lexicographers (as well as automated corrections) are stored and can be viewed – both as a global history and as a history of changes to a single lexeme. The global history interface allows filtering, so one can see all changes made by a given lexicographer in a given period.

## Filtering

Lexeme can be filtered with many criteria, including part of speech, inflection pattern, regular expressions for entry, status, comment, qualifiers and many others. Complex filters can be saved for future use.

Wyszukiwanie...

oraz +

Hasło  kończy się na  aza

Część mowy  równe  subst

## Export

Every Publisher can produce an exported data. They can specify which dictionaries they import. Some lexemes and forms have special qualifiers like 'archaism' or 'slang' and it's possible to exclude those forms in the export.

The tags used in the export are defined in variants and there may be many variants to choose from.

A fragment of an example export:

fantazja	fantazja	subst:sg:nom:f
fantazji	fantazja	subst:sg:dat:f
fantazję	fantazja	subst:sg:acc:f
fantazje	fantazja	subst:pl:voc:f
fantazjami	fantazja	subst:pl:inst:f
fantazyj	fantazja	subst:pl:gen:f
fantazji	fantazja	subst:pl:gen:f

## List of lexemes

There are currently over 400,000 lexemes in the system, so the list has to be managed dynamically. This is achieved with a JavaScript library jqGrid and the AJAX technology.

## Cross-references

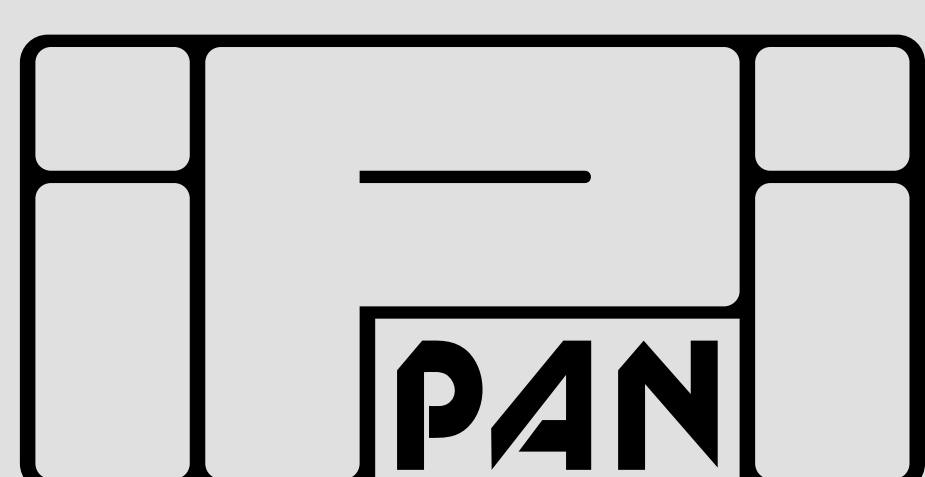
Some lexemes have defined links, or cross-references, to other lexemes, like negations (legal - illegal), names of properties (red - redness) etc. Those links may also connect different parts of the same lexeme; eg. participles and gerunds are defined as separate lexemes. During exports such connections are resolved and the basic form of the main lexeme is given for linked sublexemes.

## Pattern prompter

This tool helps finding the correct inflection pattern for a given lexeme. It looks at existing lexemes in the system and finds examples having common suffix and some other characteristics with the examined one and suggests the patterns they use. It is possible to preview forms generated using suggested patterns.

Odsyłacze: +

przymiotnik „zanegowany”	209202	niepiękny	0-	✖
przysłówek stopnia wyższego	170339	piękniej		✖
przysłówek stopnia równego	164988	pięknie		✖
przymiotnik stopnia wyższego	157244	piękniejszy		✖
nazwa cechy	130307	piękność	f	✖



Podpowiadacz wzorów

geomorfologia	f	pospolita	0127
agrofenologia	f	pospolita	0135g
absorpcjometr-la	f	pospolita	0126
acetylenowni-a	f	pospolita	0138
biżni-a	f	pospolita	A700Z
cacca-a	f	pospolita	0135g2
licentia-a	f	pospolita	0000
abatys-a	f	pospolita	0092
abbreviatura-a	f	pospolita	0098
abitell-a	f	pospolita	0135

Uwzględniaj charakterystykę fleksyjną  
 Uwzględniaj pospolitość  
 Pomijaj wzory nietypowe

sg:nom	morfologi-a
sg:gen	morfologi-i
sg:dat	morfologi-i
sg:acc	morfologi-ę
sg:inst	morfologi-ą
sg:voc	morfologi-o
pl:nom:m2	morfologi-e
pl:gen:fneut	morfologi-i hom.
pl:gen:fchar	morfologi-j arch. char.
pl:dat	morfologi-om
pl:inst	morfologi-ami
pl:loc	morfologi-ach

Anuluj Wybierz