



etytree tool

database of lexical info + etymological relationships +
interactive tool
based on the English Wiktionary

Epantaleo
WWW.EPANTALEO.COM
ESTERPANTALEO@GMAIL.COM



the workflow

English Wiktionary XML dump



parser based on <http://kaiko.getalp.org/> by prof.Sérasset *
https://bitbucket.org/esterpantaleo/dbnary_etymology

RDF database



SPARQL endpoint

<http://etytree-virtuoso.wmflabs.org/sparql>
<http://etytree-virtuoso.wmflabs.org/dbnary/eng/pistachio>



d3.js visualisation
<https://github.com/esterpantaleo/etymology>

etytree

<http://tools.wmflabs.org/etytree>

why the English Wiktionary?

- multilingual
- etymology sections in the English Wiktionary are relatively easy to parse with a machine:
 - + follow **a clear set of standard rules**
<https://en.wiktionary.org/wiki/Wiktionary:Etymology>
 - **there are many exceptions**
e.g.: use of links: "[[fortia]]" in Galician "forza"

English [edit]

Wiktionary

Etymology [edit]

From Italian *pistacchio*, from Latin *pistacium* ("pistachio"), from Ancient Greek *πιστάκιον* (*pistákion*), from *πιστάκη* (*pistákē*, "pistachio tree"). Of Iranian origin. Compare Kurdish *pisteq*, Persian *پسته* (*pista*), Middle Persian *pstk'* (*pistag*, "pistachio nut"), Old Armenian *պիստակ* (*pistak*) (from Iranian).

wiki text

===Etymology===

From {{etyl|it|en}} {{m|it|pistacchio}}, from {{etyl|la|en}} {{m|la|pistacium|pistachio}}, from {{etyl|grc|en}} {{m|grc|πιστάκιον}}, from {{m|grc|πιστάκη|pistachio tree}}. Of {{etyl|ira|en}} origin. Compare {{cog|ku|pisteq}}, {{cog|fa|پسته|tr=pista}}, {{cog|pal|pstk'|tr=pistag|pistachio nut}}, Old Armenian {{m|xcl|պիստակ}} (from Iranian).



regex and grammar

FROM LANGUAGE LEMMA1 COMMA LANGUAGE LEMMA2 COMMA[...] DOT
from which the triples:

LEMMA etymologicallyDerivesFrom LEMMA1
LEMMA1 etymologicallyDerivesFrom LEMMA2
[...]



after substitution

eng:pistachio etymologicallyDerivesFrom ita:pistacchio .
ita:pistacchio etymologicallyDerivesFrom lat:pistachium .
[...]

RDF

the idea

- create a database of etymological relationships exportable into Wikidata (wordA etymologicallyderivesFrom wordB) for:
 - linguists
 - data scientists
- create a tool to visualise the etymological tree of any word for:
 - the general user interested in etymologies
 - people that are learning a language (“meaningful learning [...]: connected to prior learning [...] more highly retainable and generalisable”*)
 - polyglots
 - to discover new words

* Herbert D. Pierson in “Using etymology in the classroom”

future steps

- debug parser
- debug visualisation
- merge with Wikidata? after the WikibaseLexeme data model has been implemented
- find new collaborators