

The POLYTROPON Dictionary: towards developing a web service for querying and linking linguistic data

Voula Giouli & Nikos Sidiropoulos
Institute for Language & Speech Processing, RC Athena

outline

- the POLYTROPON dictionary
- the dictionary API
- querying POLYTROPON
- conclusions & future work

objective

- ongoing work aimed at the development of a web service that will enable robust “search & retrieve” case scenarios over LRs developed in ILSP for MG
- the web service and the dictionary jointly form an infrastructure that can be exploited
 - in application scenarios involving deep lexical information
 - by lexicon end-users (researchers interested in studying the lexicon of the Greek language, general public)

the POLYTROPON dictionary

- computational conceptual lexicon of MG
- developed within the KRIPIS project
- builds on existing models for semantic representation (i.e., Frame Semantics)
- extends the existing LR infrastructure by integrating, harmonizing & extending already existing LRs
 - architecture proposed by EKFRASIS project (Markantonatou & Fotopoulou, 2007)
 - bilingual EL-EN MITIS dictionary (Yannoutsou et al. 2005)
 - the EL counterpart of the SIMPLE semantic lexicon (Mantzari et al, 2000)
- conformant to widely accepted standards (reusability, interoperability resources and tools)

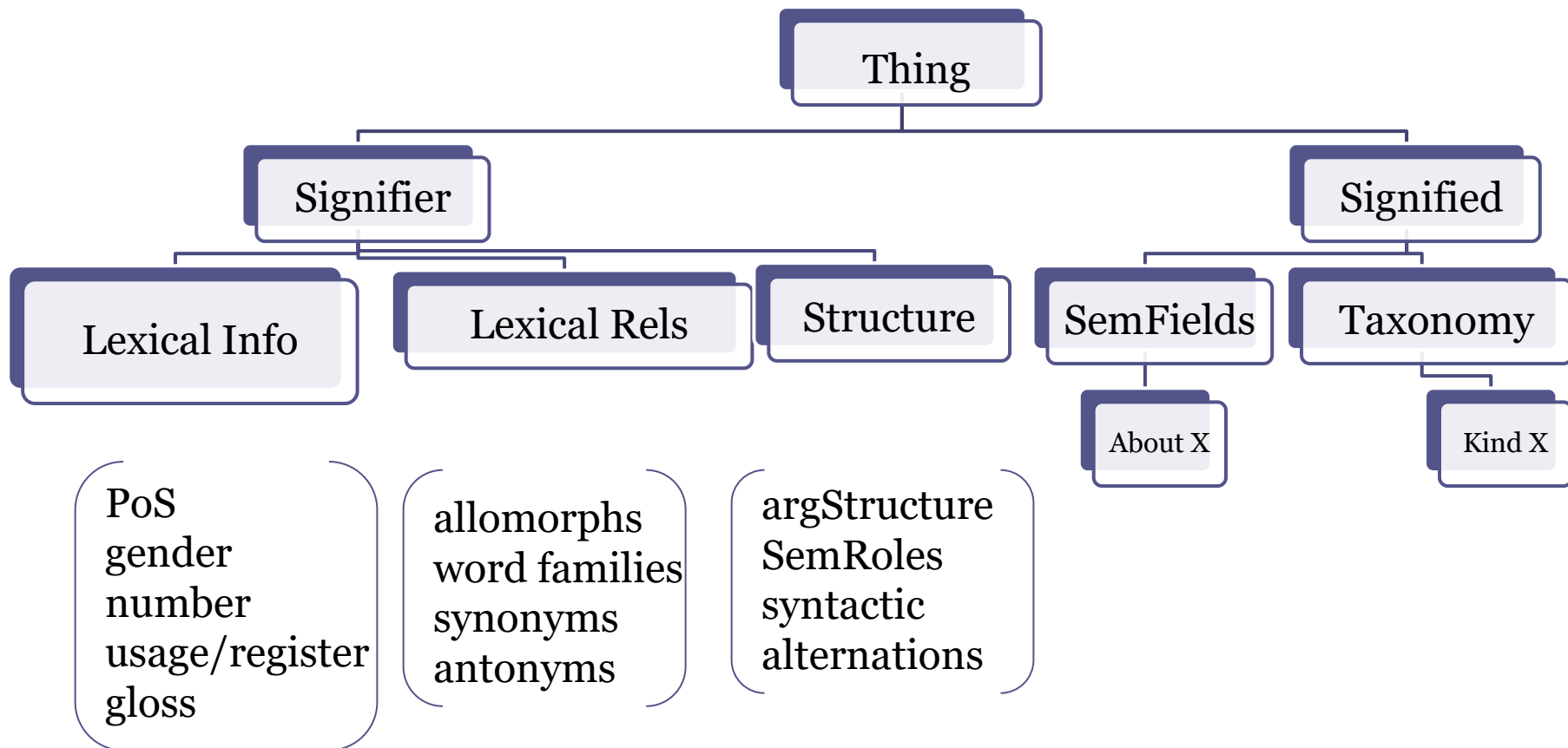
possible use scenarios

- Word sense disambiguation (WSD)
- Concept Linking
- Textual Entailment
- Coreference Resolution (i.e., bridging references, coreferring mentions, etc.)
- Semantics and Concept-based MT

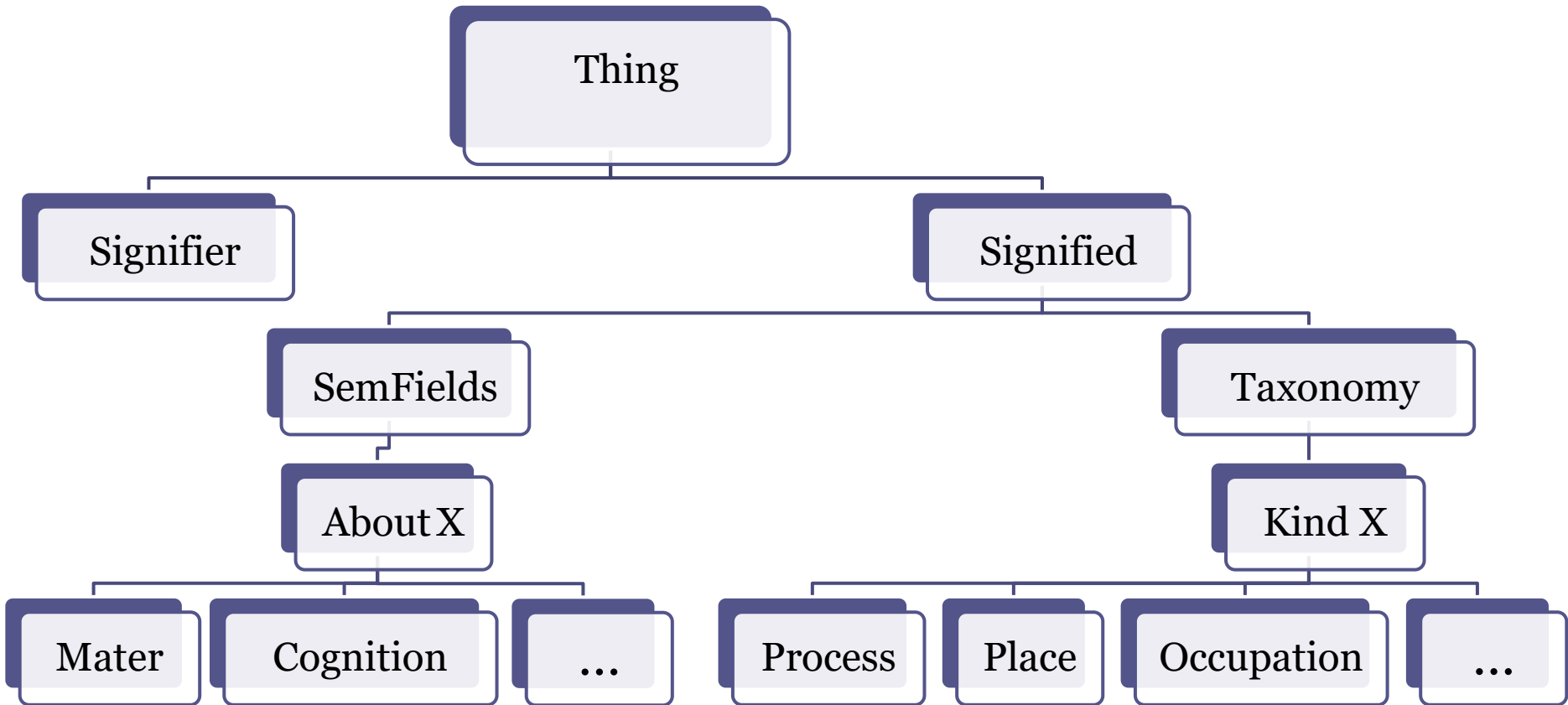
lexicon architecture

- builds on two basic notions:
 - the Saussurian notion of *sign* and its facets
 - the *SIGNIFIER* and
 - *the SIGNIFIED*
 - the notion of semantic fields
- a linguistic ontology
 - words are instances in the *SIGNIFIER* class
 - concepts are instances in the *SIGNIFIED* class
 - each word is mapped onto a concept

the linguistic ontology



the linguistic ontology



concepts are instances of classes
 one concept may belong to more than one class
 relations link concepts within or across classes

specifications

- lexicographic
- linguistic
- functional
- ontological
- coverage in terms of lemmas & semantic fields
 - words and concepts linked via rels (lexical, semantic, other) in a systematic way
 - knowledge base
 - representation (dense, extensible, flexible)

dictionary in numbers

		singleWEs	MWEs	Total
AjBa	3445	3445		
AjMWE	153		153	3598
AdBa	922	922		
AdMWE	765		765	1687
NoCm	9071	9071		
NoMWE	1899		1899	10970
VbMn	2330	2330		
VbMWE	2436		2436	4766
Total	21021	15768	5253	21021

dictionary format

- encoded in Protégé
 - pros:
 - creation & management of ontologies
 - built-in reasoners
 - efficient manipulation & control of the data
 - xml/rdf export
 - cons:
 - user un-friendly
 - Mongo-DB uses its own query language

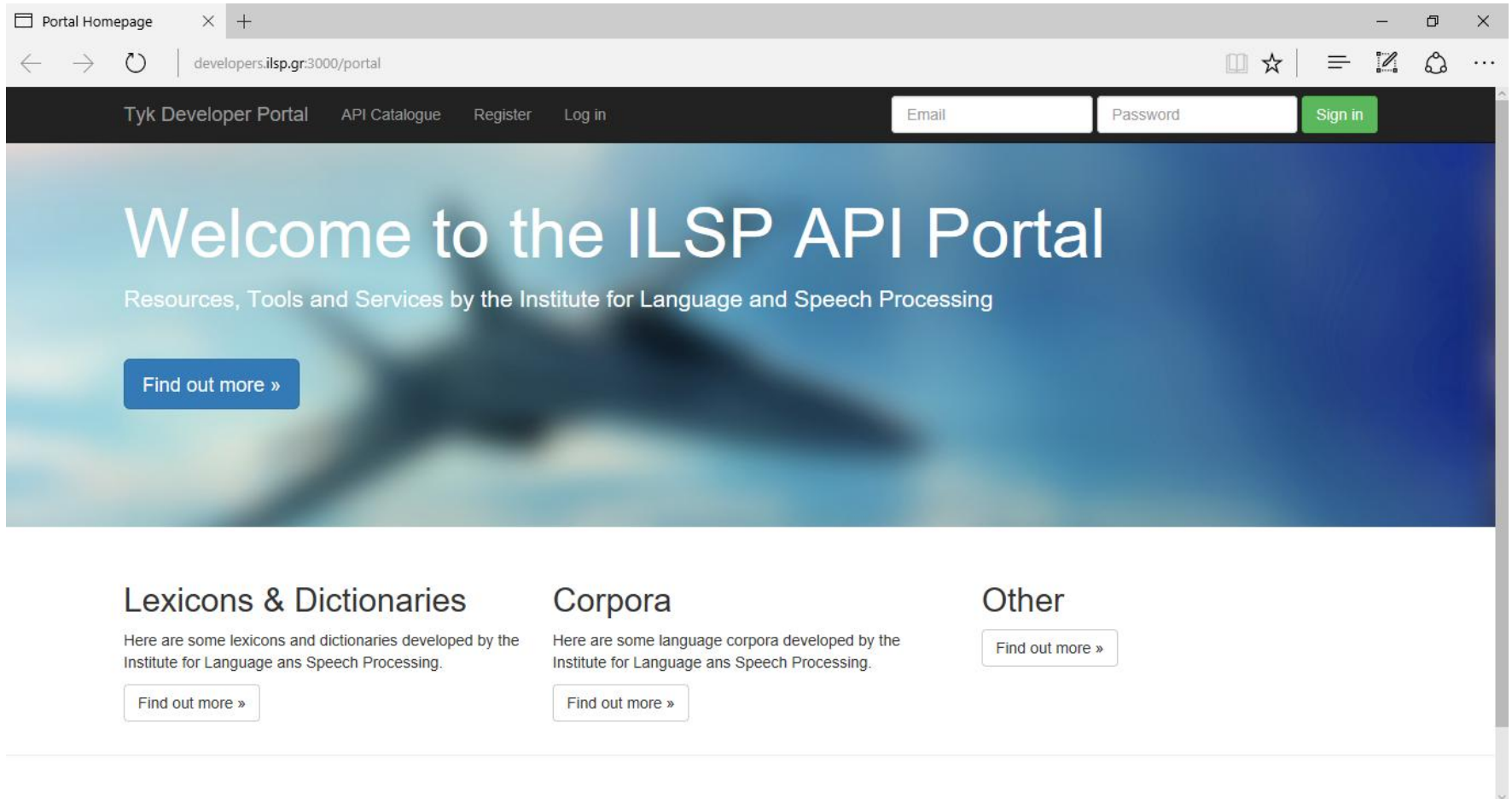
data preparation

- RESTFul API architecture
- running under an API Gateway Platform
- Protégé / Web Protégé
- rdf/xml export
- MySQL DB
- imported into the ILSP portal

ILSP web service

- access to individual dictionaries
POLYTROPON
bilingual dictionaries (EL-Xlanguage)
- query the Hellenic National Corpus (HNC)
(<http://hnc.ilsp.gr/>)
- access to NLP pipeline (<http://nlp.ilsp.gr/soaplab2-axis/>)
 - combined retrieval from the set of dictionaries and/or HNC
 - dictionary aggregator: meta-machine that aggregates the content of various dictionaries

the ILSP API



The screenshot shows a web browser window with the URL `developers.ilsp.gr:3000/portal`. The page features a dark navigation bar with links for 'Tyk Developer Portal', 'API Catalogue', 'Register', and 'Log in'. There are also input fields for 'Email' and 'Password', and a green 'Sign in' button. The main content area has a blue background with the heading 'Welcome to the ILSP API Portal' and the subtitle 'Resources, Tools and Services by the Institute for Language and Speech Processing'. A blue button labeled 'Find out more »' is positioned below the subtitle. At the bottom, there are three sections: 'Lexicons & Dictionaries', 'Corpora', and 'Other', each with a brief description and a 'Find out more »' button.

Portal Homepage × +

← → ↻ | developers.ilsp.gr:3000/portal

Tyk Developer Portal API Catalogue Register Log in

Email Password Sign in

Welcome to the ILSP API Portal

Resources, Tools and Services by the Institute for Language and Speech Processing

Find out more »

Lexicons & Dictionaries

Here are some lexicons and dictionaries developed by the Institute for Language and Speech Processing.

Find out more »

Corpora

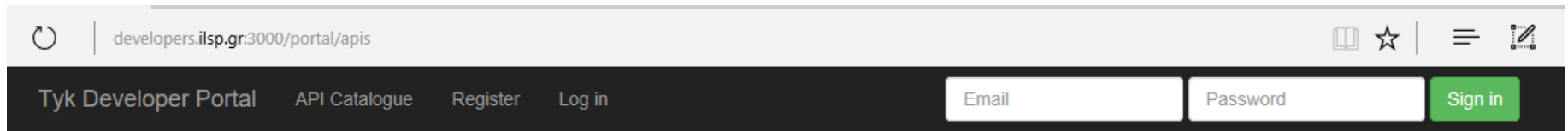
Here are some language corpora developed by the Institute for Language and Speech Processing.

Find out more »

Other

Find out more »

the ILSP API Catalogue



API Catalogue

Dictionaries API

[View documentation](#)[Request an API key](#)

HNC API

[View documentation](#)[Request an API key](#)

Polytropon API

[View documentation](#)[Request an API key](#)

Powered by [Tyk - The Open Source API Gateway](#)

access to the LRs via the API catalogue

users need to register/sign in

querying POLYTROPON

general query

Γενική αναζήτηση λέξης

Αναζήτηση λέξης:

Αναζήτηση

Αναζήτηση συνωνύμων, αντωνύμων και μορφολογικά σχετικών λέξεων

Αναζήτηση υπωνύμων-υπερωνύμων

Αναζήτηση εννοιολογικά σχετιζόμενων λέξεων

system retrieves matching sense-discriminated entries

Λέξη: στάση (1)

- word_type: single
- PoS: Nn
- wordstring: στάση
- gloss: σταμάτημα
- number: Sg
- gender: Fe
- morphologically_related: στάση (2), στάση (3), στάση (4), στάση (5), σταματάω (1)

Λέξη: στάση (2)

- word_type: single
- PoS: No
- wordstring: στάση
- number: Sg
- gender: Fe
- morphologically_related: επανάσταση (1), στάση (1), στάση (3), στάση (4), στάση (5), στασιάζω, στασιάζτρια, στασιαστής, στασιαστικός
- instantiates: επανάσταση (1), κίνημα

Λέξη: στάση (3)

- word_type: single
- PoS: No
- wordstring: στάση
- number: Sg
- gender: Fe
- morphologically_related: στάση (1), στάση (2), στάση (4), στάση (5), στέκω (1)

Λέξη: στάση (4)

- word_type: single
- PoS: No
- wordstring: στάση
- number: Sg
- gender: Fe
- morphologically_related: στάση (1), στάση (2), στάση (3), στάση (5)
- instantiates: αγωγή (1)

Λέξη: στάση (5)

- word_type: single
- PoS: Nn
- wordstring: στάση
- number: Sg

querying POLYTROPON

mwes encoded are also retrieved

Γενική αναζήτηση λέξης

Αναζήτηση λέξης:

Αναζήτηση

Αναζήτηση συνωνύμων, αντωνύμων και μορφολογικά σχετικών λέξεων

Αναζήτηση υπωνύμων-υπερωνύμων

Αναζήτηση εννοιολογικά σχετιζόμενων λέξεων

Αποτελέσματα αναζήτησης:

Λέξη: κάνω στάση

- word_type: mwe
- PoS: Vb
- wordstring: κάνω στάση

Λέξη: κρατώ στάση αναμονής

- word_type: mwe
- PoS: Vb
- wordstring: κρατώ στάση αναμονής

querying POLYTROPON

focused search (i)

Γενική αναζήτηση λέξης

Αναζήτηση συνωνύμων, αντωνύμων και μορφολογικά σχετικών λέξεων

Αναζήτηση λέξης:

Επιλογή συσχέτισης:

Συνώνυμες λέξεις ←

Αντώνυμες λέξεις ←

Μορφολογικά σχετιζόμενες λέξεις ←

Σημασιολογικά συναφείς λέξεις ←

Αναζήτηση υπωνύμων-υπερωνύμων

Αναζήτηση εννοιολογικά σχετιζόμενων λέξεων

Αποτελέσματα αναζήτησης:

Λέξη: ψηλός

Αντώνυμες λέξεις

- κοντός
- χαμηλός (1)

synonyms (strict)

antonyms

morphologically related words

semantically related words (near synonyms)

querying POLYTROPON

focused search (ii)

Γενική αναζήτηση λέξης

Αναζήτηση συνωνύμων, αντωνύμων και μορφολογικά σχετικών λέξεων

Αναζήτηση υπωνύμων-υπερωνύμων

Αναζήτηση λέξης:

Επιλογή συσθένισης:

- Υπώνυμα
- Υπερώνυμα

Επιλογή επιπέδου αναζήτησης:

Αναζήτηση

Αποτελέσματα αναζήτησης:

Υπερώνυμες λέξεις για την λέξη: μηλιά

μηλιά

δέντρο

φυτό

hyponyms
hypernyms

querying POLYTROPON

focused search (iii)

Γενική αναζήτηση λέξης

Αναζήτηση συνωνύμων, αντωνύμων και μορφολογικά σχετικών λέξεων

Αναζήτηση υπωνύμων-υπερωνύμων

Αναζήτηση εννοιολογικά σχετιζόμενων λέξεων

Αναζήτηση λέξης: καρδιά

Είδος σχέσης: Afflicts BodyPart

Αναζήτηση

Αποτελέσματα αναζήτησης:

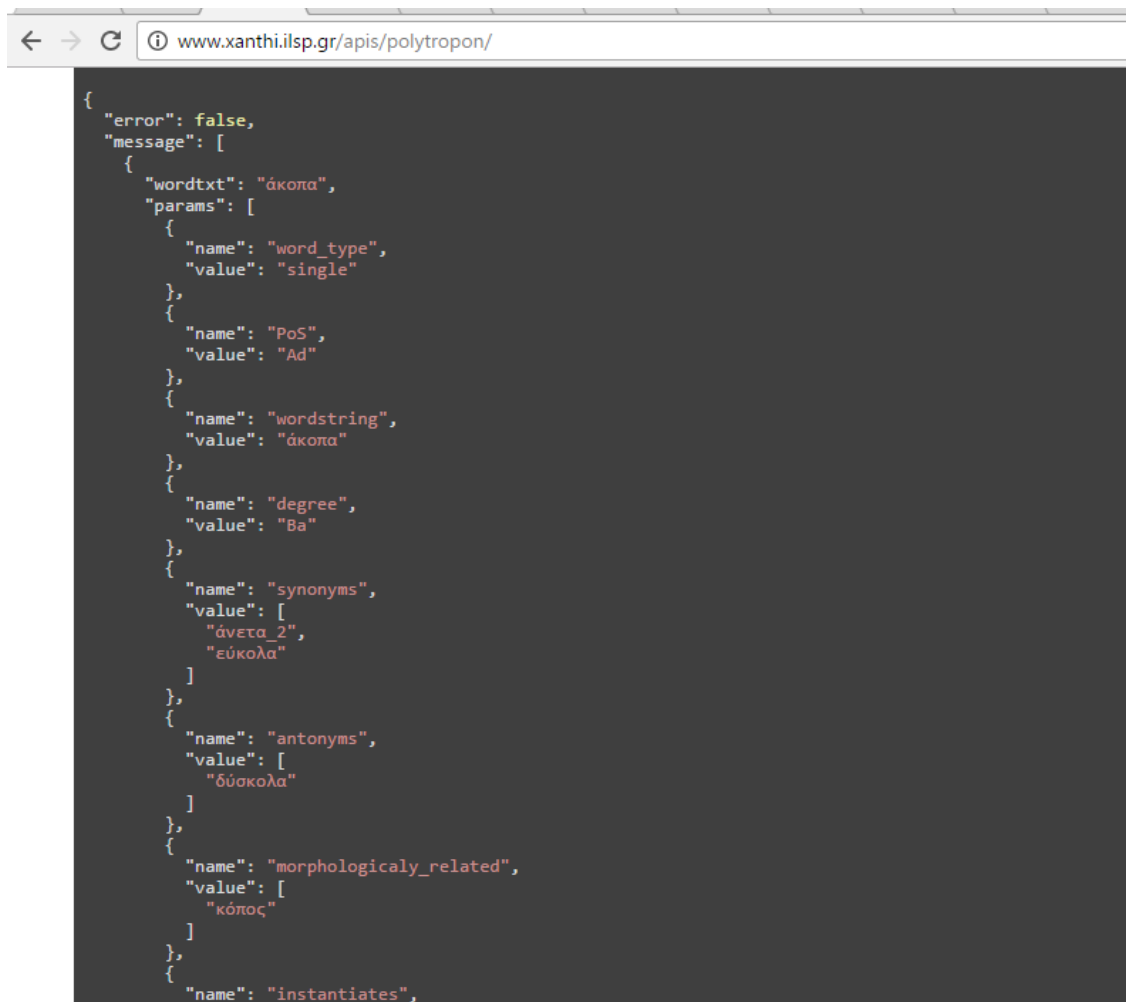
Λέξη: καρδιά (1)

Ασθένεια που προσβάλλει;

- έμφραγμα
- στηθάγχη
- στενοκαρδία
- μυοκαρδίτιδα
- ταχυκαρδία
- ταχυπαλμία
- ταχυσφυγμία
- βραδυκαρδία
- βραδυσφυγμία

retrieving semantically related words/concepts via a set of relations that link pairs of word/concepts, i.e. Afflicts_BodyPart(Disease, BodyPart)

results:JSON format



```
{
  "error": false,
  "message": [
    {
      "wordtxt": "άκοπα",
      "params": [
        {
          "name": "word_type",
          "value": "single"
        },
        {
          "name": "PoS",
          "value": "Ad"
        },
        {
          "name": "wordstring",
          "value": "άκοπα"
        },
        {
          "name": "degree",
          "value": "Ba"
        },
        {
          "name": "synonyms",
          "value": [
            "άνετα_2",
            "εύκολα"
          ]
        },
        {
          "name": "antonyms",
          "value": [
            "δύσκολα"
          ]
        },
        {
          "name": "morphologically_related",
          "value": [
            "κόπος"
          ]
        },
        {
          "name": "instantiates",
```

to sum-up

- infrastructure for exploiting dictionaries & other LRs already available in ILSP
- currently, the infrastructure facilitates access to POLYTROPON conceptual dictionary
- it is targeted to researchers of MG and to application developers in need of linguistically-aware lexical resources.

future work

- integration of other LRs (underway)
- providing combined search functionalities over the LRs available in ILSP
- extending the web service so as to enable crowd-sourcing → extend the LRs
- creation of an interface for end-users
 - visualization of the results
 - interactive access to the data
 - games-with-a-purpose

Thank you for your attention!
for queries

voula@ilsp.gr

nsidir@xanthi.ilsp.gr

