



# **Universal Knowledge Core**

Phase: 2.Tools and Tutorials

W8.L15.T15.2.2

## 1 Motivation

### 2 Multilingual lexical resource

## 3 Design of UKC

## 1 Motivation



#### **3** Design of UKC

## **Motivation**

Let us take a look at this task of translation from Italian to English,

#### Problem

Translating the following Italian text to English to convey the meaning to a non-speaker of the Italian language. "Mi hanno appena comunicato di una caso di positività covid-19 di un afferente CIBIO che occupa l'ufficio 123 povo 2 corridoio lato ovest."

#### Solution

A bi-lingual dictionary can be used to translate each word to english. But with Google translate it is a very simple task.

Wordnets from various languages are used for the task for translating from source language to target language.

## **Example : Google translate**

Google	italian to english			×Q
	🔍 Tutti 🕩 Video 🛷 Shopping 🛽	🗄 Notizie  🖿	Immagini 🚦 Altro	Impostazioni
	Qualsiasi lingua 👻 Qualsiasi data 👻	Tutti i risultati 🕶	Circa 1.590.000.000 risu	ıltati (0,39 secondi)
	Italiano 👻	$\stackrel{\rightarrow}{\leftarrow}$	Inglese 👻	
	Mi hanno appena comunicato di una caso di positività covid-19 di un afferente CIBIO che occupa l'ufficio 123 povo 2 corridoio lato ovest.	× I h inf po CI oc po co	ave just been formed of a cov ositive case of a BIO afferent w ccupies office 1 wo 2 west side rridor.	vid-19 a ho 23
		C		
		<b>■</b> D		

### **1** Motivation

## 2 Multilingual lexical resource

#### **3** Design of UKC

# What is UKC ?

Large scale multilingual lexical resource composed of two layers which maintains a clear distinction between the two layers, one of which is language(s) that used to describe the world as it is perceived and the other what is being described, i.e., the world itself.

The following is the current status,

- number of concepts > 120,000
- number of languages : 1067
- number of words > 1.9m
- number of senses > 2.9m

# Example of the UKC : Language independent

The following is the representation of the language independent part of UKC, which is a graph of language-independent notions



# Example of UKC : Language dependent

This is an example of a vocabulary in various languages to represent the language dependent component of the UKC,

Lang	Synset	Gloss
eng	Canal	long and narrow strip of water made for boats or for irrigation
ita	canale; naviglio	corso d'acqua artificiale, costruito per l'irrigazione o la navigazione
mon	суваг	усжуулалт эсвэл завинд зориулсан барьсан усны урт нарийн гудамж
ben	খাল	পানির দীর্ঘ এবং সরু ধারা যা সেচ বা নাব্যতার জন্য তৈরি করা হয়েছে
zho	沟渠; 运河	人工水道或人工修缮的河流,用于旅 行、航运或灌溉
hin	नहर; कुलिया	सिंचाई, यात् <b>रा आदि के लिए छोटी नदी के</b> रूप में तैयार किया हआ जलमार्ग

### 1 Motivation



## 3 Design of UKC

## **Representation of structure of UKC**

Concept Core (CC) : consists of a semantic network where the nodes are language independent concepts.

Language Core (LC) : corresponds to the PWN, namely the set of words, senses, synsets, glosses and examples supported by the UKC.



## Wordnet view of UKC



## **Concept core**

- Concepts are language-independent notions.
- Semantic Relations connect the concepts and can be Hierarchical (e.g. is-a) or Associative (e.g.has-member)



## **Concept core : semantic relations**

Name	Description	Example
is-a	the relation between concepts denoting classes, between relations, between a value of a descriptive attribute and corresponding name and between any value (descriptive or qualitative) and a more general one.	chicken is-a animal (two classes), listen is-a perceive (two relations), autocracy is-a form of government (value and corresponding name), monarchy is-a autocracy (two descriptive values) and sanguine is-a red (two qualitative values).
has-aspect	a contingent specialization, i.e. in C has-aspect P an entity of the child class C is not essentially an entity of the parent class P, i.e. is a P in some circumstances.	chicken has-aspect food
is-agent	similar to has-aspect with the difference that in any C is-agent P the parent P of C must be the concept agent or a more specific concept.	president is-agent person
part-of	the relation between concepts denoting classes connecting a part with its whole.	subcontinent part-of continent
value-of	the relation between concepts where the source denotes a value and the target its attribute name.	red value-of colour
substance-of	a relation between concepts where the target represents a compound and the source one of its constituent substances.	hydrogen substance-of water
member-of	a relation between concepts where the target denotes a set and the source one of its members.	football player member-of football team member-of football league
metaphor-of	a relation between two concepts where the source is said to be the literal meaning and the target is said to be the metaphoric meaning.	leg (part of the human body) metaphor-of leg (a piece of furniture) corresponds to gamba (parte del corpo) metaphor-of i (sostegno di mobile) in Italian.

## Language core

- It provides support to multiple languages (e.g. English, Italian)
- Within a language, words with same sense are grouped into synsets.
- Lexical Relations are used to connect senses (e.g. antonym).



## Language core : lexical relations

Lexical Relations are relations between word Senses within a language:

Name	Description	Example
Synonymy	relation connecting two senses of different words with the same POS and the same meaning.	intelligent synonym clever
Antonym	a symmetric relation connecting two senses of different words with same POS and the opposite meaning.	black antonym white.
Pertains-to	a relation between a relational adjective (i.e. an adjective expressing a relation with a noun or another adjective) and the corresponding noun or adjective.	fraternal pertains-to brother and chemical pertains-to chemistry.
Participle-of-verb	a relation between an adjective and a verb such that the adjective is the participle of the verb.	breaking participle-of-verb break.
Also-see	a generic relation between adjectives or between verbs which are similar in meaning.	asleep also-see unaware also-see unconscious.
Related form (a.k.a. derivationally related form)	a generic relation between two words which are one a variation of the other.	personhood related-to person.
Homograph-of	is a symmetric relation used to explicitly mark two senses of the same word and POS having unrelated meanings	bank as an institution and bank as sloping land.

### 1 Motivation



Multilingual lexical resource

#### **3** Design of UKC

## Uses of UKC as a resource

The following are the downstream uses,

- computational linguistics (such as word sense disambiguation, word translation)
- industrial and research projects on semantic interoperability

It is used for the study of language diversity in the following ways,

- Ianguage-specific concepts, for instance, malga from the previous example is the Italian language specific term and another concept in the UKC.
- under-resourced languages, for instance, Scottish gaelic is a minority language spoken in Scotland.

# Use of UKC in iTelos development methodology

Recall the lecture on iTelos : Formal Modelling,

- Teleology is constituted of three particular levels of knowledge, L1 - concept space, L2- Lexical-Semantic Space, and L4 schema knowledge graph.
- the formal L4 schema knowledge graph will be annotated through the L1 and L2 concept definitions.
- The main purpose of the Annotation activity is to contribute into finding most, if not all, the possible concepts in the UKC that can help to describe the entities and relations found in the finalized ontology.

# **UKC Application UI**



# **UKC Application UI - 2**



# **UKC Application UI - 3**

SHIB KOS EM	=	)😂 🚀 admin
Entity Base	Knowledge Explorer - 3011	English ~
🕼 Etype Modeler	a, 3011	Search <
<ul> <li>Etype Explorer</li> </ul>		
Knowledge Base	Glossary Relations Provenance	
🕼 Knowledge Importer	K Filter relations	
👕 UserBase Management	😂 📧 🛪 position, post, berth, office, spot, billet, place, situation	2929
	🗁 📧 🗸 occupation, business, job, line of work, line	
		2006
		56
	event	
	😂 📭 🛤 abstraction	
	e les abstract entity	20
	a rs.A entity	1

# **UKC Application UI - 4**

SHIB KOS EM	≡		)🟳 🚀 admin
Entity Base	Knowledge Explorer - 3011		English ~
C Etype Modeler	at 3011		Search <
Etype Explorer			
E Knowledge Base	Glossary Relations Provenance		
C Knowledge Importer	Senses	professorship () , chair ()	
📽 UserBase Management	Global Id	3099	
	Reference Languages		
	Bengali		
	català		
	español		
	Basque		
	Gallegan		

## References

Links to the websites,

- UKC : http://ukc.disi.unitn.it/
- Linguarena : http://linguarena.eu/
- iTelos : https://unitn-kdi-2020.github.io/unitn-kdi-2020/materials/slides/W7.L14.M5.T14.AllPDF.pdf

Useful papers,

- Understanding and Exploiting Language Diversity F Giunchiglia, K Batsuren, G Bella International Joint Conference on Artificial Intelligence (IJCAI-17),
- One World-Seven Thousand LanguagesF Giunchiglia, K Batsuren, AA Freihat, 19th International Conference on Computational Linguistics and Intelligent Text Processing, Hanoi, Vietnam.



# W8.L15.T15.2.2

## Universal Knowledge Core Phase: 2.Tools and Tutorials