

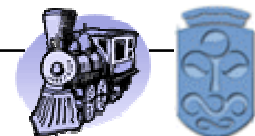
# Galnet

---

Wordnet 3.0 do galego

Grupo TALG

<http://webs.uvigo.es/talga>





# Presentación

---

- Tecnoloxías e Aplicacións da Lingua Galega (Grupo TALG da Universidade de Vigo)
  - Seminario de Lingüística Informática
  - Observatorio de Neoloxía
- Actividades de I+D
  - Creación de ferramentas, aplicacións e recursos informáticos de uso libre para a lingua galega



# Presentación

---

- Grupo de investigación interdisciplinario
  - Área de Lingüística Xeral
  - Área de Tradución e Interpretación
  - Área de Filoloxía Galega e Portuguesa
- ... e interdepartamental
  - Departamento de Tradución e Lingüística
  - Departamento de Filoloxía Galega e Latina



# Recursos

---

- Córpora textuais de lingua galega
  - Paralelo (CLUVI)
  - Técnico (CTG)
  
- Recursos léxicos de galego
  - Terminoloxía (Termoteca)
  - Neoloxía (Neoteca)
  - Dicionarios bilingües (Dicionarios CLUVI)
  - WordNet (Galnet)



# Recursos

---

- Outras aplicacións lingüísticas
  - Tradución automática (Apertium)
  - Corrección ortográfica (OpenOffice, OrtoGal)
  - Análise lingüística (FreeLing)
  - Integración de recursos (RILG)

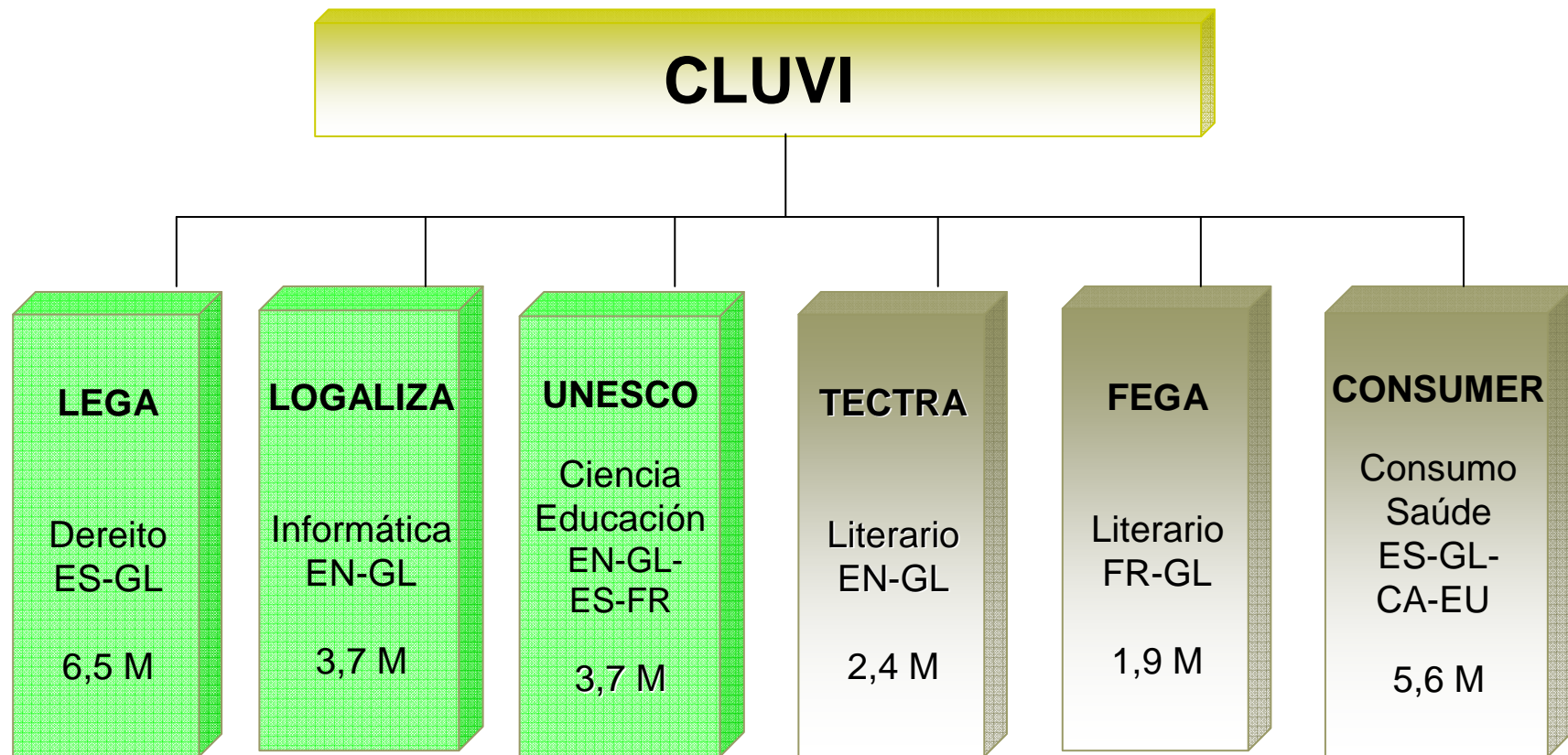
# Corpus CLUVI

```
...
<tu>
<tuv xml:lang="en"><seg>The Time
Traveller (for so it will be convenient
to speak of him) was expounding a
recondite matter to us.</seg></tuv>
<tuv xml:lang="gl"><seg>O Viaxeiro do
Tempo (este será o xeito máis
conveniente de referírmonos a el) estaba
expoñéndonos un problema cargado de
misterio para nós.</seg></tuv>
</tu>
<tu>
<tuv xml:lang="en"><seg>His grey eyes
shone and twinkled, and his usually pale
face was flushed and
animated.</seg></tuv>
<tuv xml:lang="gl"><seg>Os seus ollos
grises brillaban e escintilaban e
víaselle a cara, normalmente pálida,
acorada e animada.</seg></tuv>
</tu>
...
```

- Corpus paralelo aliñado, de traducións ao galego ou do galego
- 23 millóns de palabras en formato TMX
- Literatura, dereito, informática, *Consumer Eroski*, *The UNESCO Courier*...
- De libre consulta en <http://sli.uvigo.es/CLUVI>

# Corpus CLUVI

---





# Resultados da pescuda no Corpus Paralelo CLUVI



## (Corpus Lingüístico da Universidade de Vigo)

Os resultados das buscas efectuadas no CLUVI poden ser usados con fins educativos e de investigación, sempre que se mencione a fonte. Se desexa citar exemplos tirados do corpus, calque na ligazón que hai na esquerda de cada equivalencia para obter a súa referencia completa. A cela da esquerda da táboa contén o número de orde do exemplo, seguido do código da obra mais o número da unidade de tradución da equivalencia no texto aliñado. Se desexa consultar o contexto (frase anterior e posterior) dunha equivalencia, calque na frecha da dereita. Para se referir ao corpus como un todo, cite: *Corpus Paralelo CLUVI 2.1 - http://sli.uvigo.es/CLUVI/*. Se desexa realizar outra pescuda no Corpus Paralelo CLUVI, pode calcar [aquí](#).

Está a procurar equivalencias de tradución **inglés - galego** no Corpus literario TECTRA inglés-galego (1.476.020 palabras).

Equivalencia de tradución buscada: [en] \bboys?\b = [gl] .

1-PER (105)	Out of their brush houses they poured _ Kino's brother Juan Tomás and his fat wife Apolonia and their four children crowded in the door and blocked the entrance, while behind them others tried to look in, and one small <b>boy</b> crawled among legs to have a look.	Xurdían das súas casopas de ramaxe. Juan Tomás, o irmán de Kino, a súa gorda muller, Apolonia, e mailos seus catro fillos, amontoáronse na porta, trancando a entrada, mentres que detrás deles, outros tentaban mirar dentro, e un cativo furaba por entre as pernas dos grandes para botar unha ollada.	
2-PER (342)	News seems to move faster than small <b>boys</b> can scramble and dart to tell it, faster than women can call it over the fences.	As novas semellan correr máis de présa do que os nenos poden saír precipitados a contalas, máis de présa do que as mulleres as poden vocear por riba dos valados.	
3-PER (344)	Before panting little <b>boys</b> could strangle out the words, their mothers knew it.	Antes de que os rapaces folguexando puidesen dicir unha palabra, as súas nais xa o sabían todo.	
4-PER (450)	Some day, his mind said, that <b>boy</b> would know what things were in the books and what things were not.	Algún día, dáballo a alma, ese neno sabería as cousas que están nos libros e as que non o están.	
5-PER (685)	It was known among the neighbors in the brush huts, among the pearl fishermen; it was known among the Chinese grocery-store owners; it was known in the church, for the altar <b>boys</b> whispered about it.	Sabiase entre os veciños nas cabanas de xestas e entre os pescadores de perlas. Sabiase entre os propietarios chineses das tendas de comestibles. Sabiase na igrexa pois esa era a leria dos sancristáns.	
6-PER (687)	The little <b>boys</b> knew about it with excitement, but most of all the pearl buyers knew about it, and when the day had come, in the offices of the pearl buyers, each man sat alone with his little black velvet tray, and each man rolled the pearls about with his fingertips and considered his part in the	Os rapaciños oían entusiasmados falar do caso, pero máis ó tanto ca todos eles estábano os compradores de perlas. Cando o día chegou todos eles sentaron sós nas súas oficinas, cada un coa súa pequena bandexiña de veludo negro diante, e cada un facendo xira-las perlas coas puntas dos	



# Dicionarios CLUVI



The screenshot shows a web browser window displaying the CLUVI dictionary website. The page title is "Dicionario CLUVI inglés-galego, 2ª edición revisada" and it is based on the CLUVI corpus from the University of Vigo. The website features a search bar and navigation buttons. The entry for "pillow" is highlighted, showing its classification as a noun and its Galician translation "almofada". The entry also includes an example sentence in English and its Galician translation.

Dicionario CLUVI inglés-galego, 2ª edición revisada  
(Baseado no Corpus CLUVI da Universidade de Vigo)

Portada Limiar Nova consulta

pilfer, pilgrim, pilgrimage, pill, pillage, pillaging, pillar, pillion, pillory, pillow, pilot, pimento, pimp, pimple, pimply, pin, PIN, pinch, pincushion

**pillow**

noun

✓ almofada

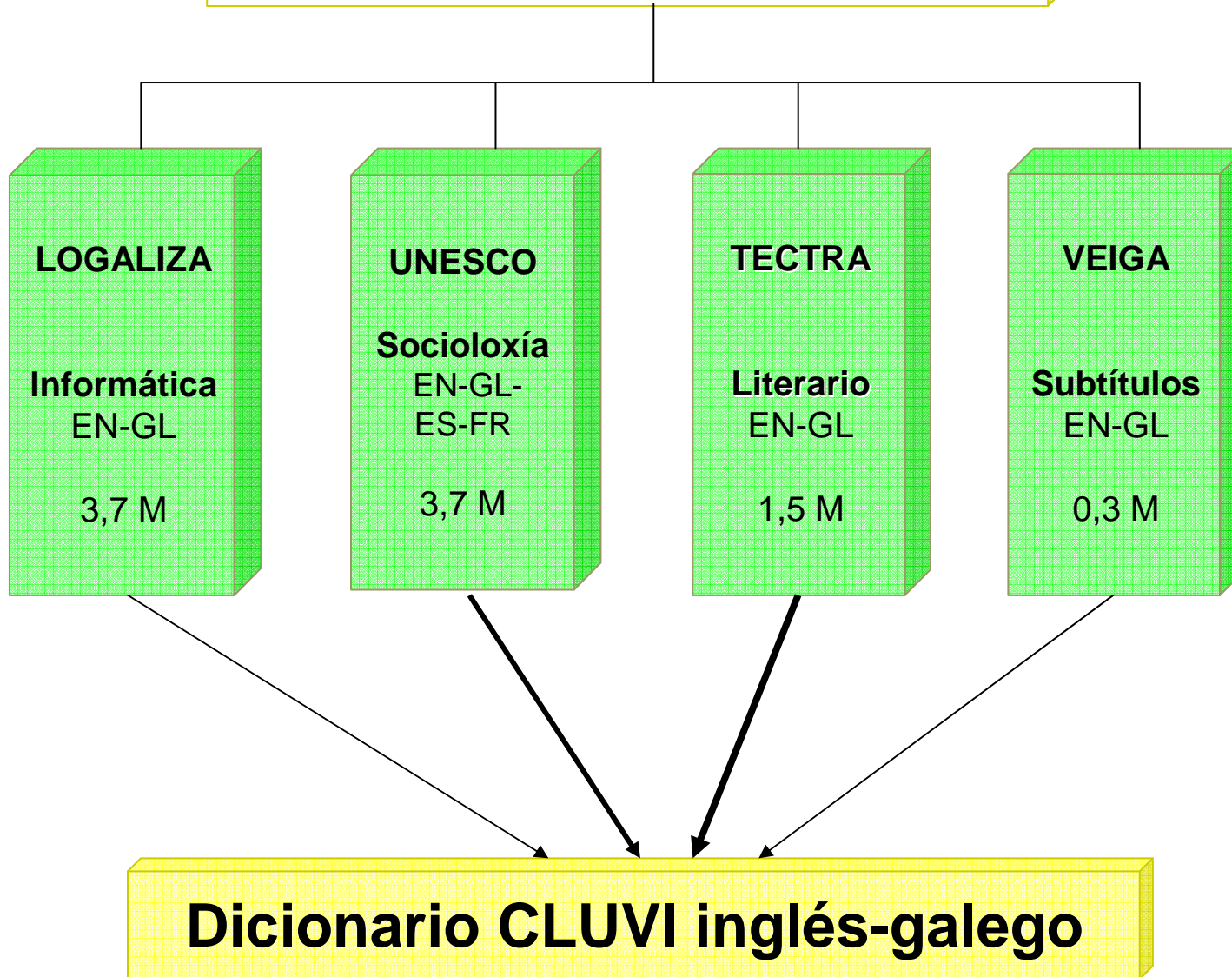
EN She soon came back and sat on the edge of the bed, leaning her head down on the **pillow**.

GL Voiveu axiña e sentou no bordo do leito, recostando a cabeza contra a **almofada**.

Fonte: ESP (131)

- Dicionarios bilingües elaborados a partir do Corpus CLUVI de traducións. O Dicionario CLUVI inglés-galego contén máis de 20.000 entradas, posúe versión en formato libro, e pode consultarse na web en [sli.uvigo.es/diccionario](http://sli.uvigo.es/diccionario)

# Corpus CLUVI



# Corpus Técnico do Galego

- Corpus textual de rexistros especializados do galego contemporáneo, de 15 millóns de palabras
- Dereito, informática, socioloxía, economía, ecoloxía e medicina
- De libre consulta en <http://sli.uvigo.es/CTG>

...

<frase>As medidas que se deben adoptar para o proceso de descarga, moedura e preparación do alimento son: </frase>

<frase>Os materiais pulverulentos ou potencialmente pulverulentos débense almacenar en colectores, en silos deseñados para este propósito ou baixo cuberta.</frase>

<frase>A distribución do alimento dende o lugar de almacenamento ata as instalacións onde comen os animais debe levarse a cabo de tal maneira que se prevexan ou minimicen as emisións ao aire de partículas, é dicir, preferentemente con sistemas pechados.</frase>

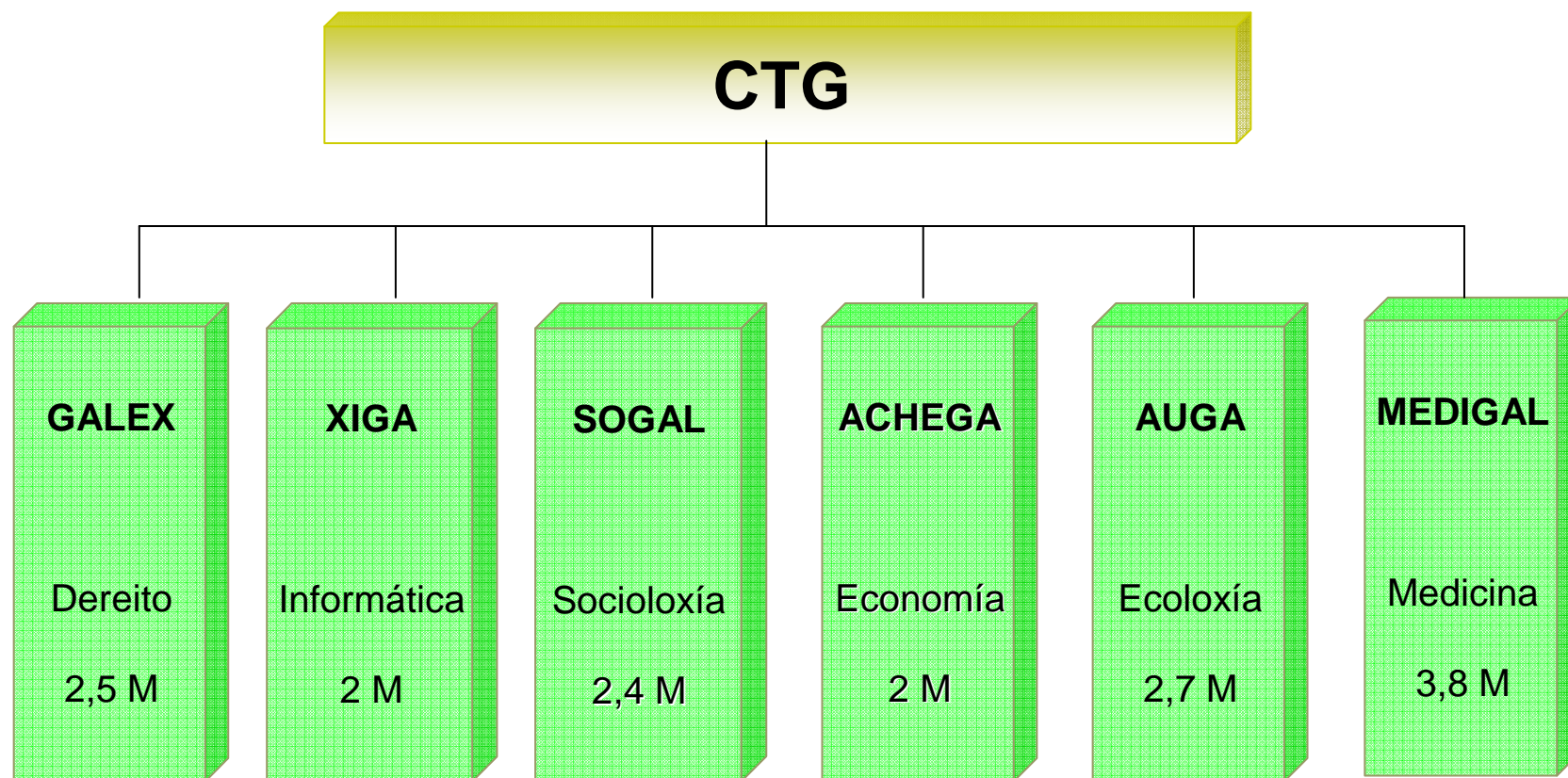
<frase>A moedura e a mestura do alimento haberá que realizala de tal forma que se prevexan ou minimicen as emisións de partículas ao aire.</frase>

<frase>Os tanques de almacenamento deberán estar protexidos de posibles danos por colisión.</frase>

...

# Corpus Técnico do Galego

---





# Corpus Técnico do Galego



## Resultados da pesquisa

Os resultados das buscas efectuadas no CTG poden ser usados con fins educativos e de investigación, sempre que se mencione a fonte. Se desexa consultar a referencia e o contexto dun exemplo, calque no símbolo [+] na cela da dereita. Para se referir ao corpus como un todo, cite: *Corpus Técnico do Galego (CTG) - http://sli.uvigo.es/CTG/*. Se desexa realizar outra pesquisa no Corpus CTG, pode calcar [aquí](#).

Está a procurar contextos do uso de **[[[:<:]]efecto invernadoiro[[:>:]]** nos textos de ecoloxía e ciencias ambientais do Corpus Técnico do Galego (CTG).

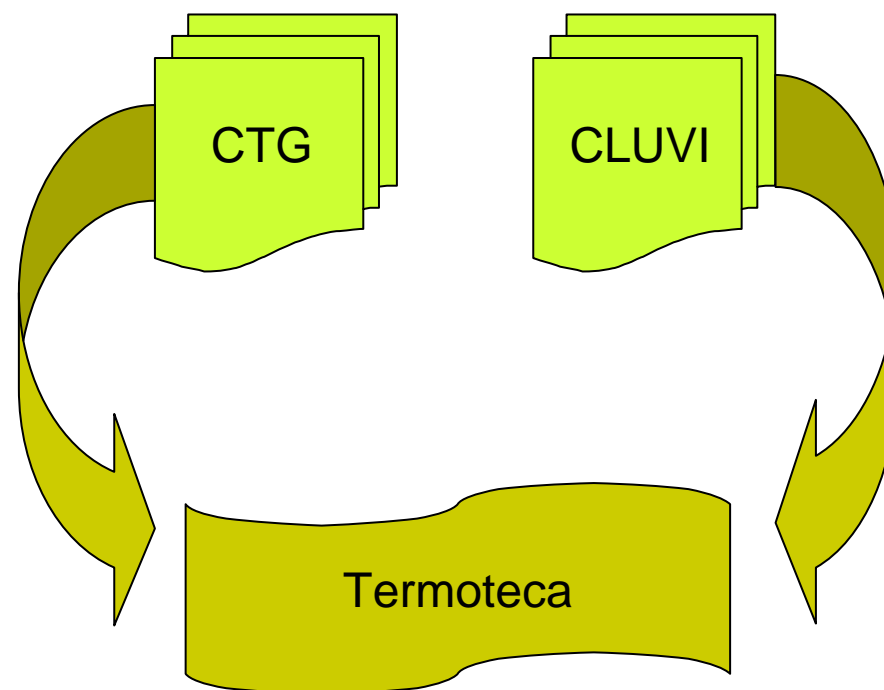
Número de contextos atopados: 320

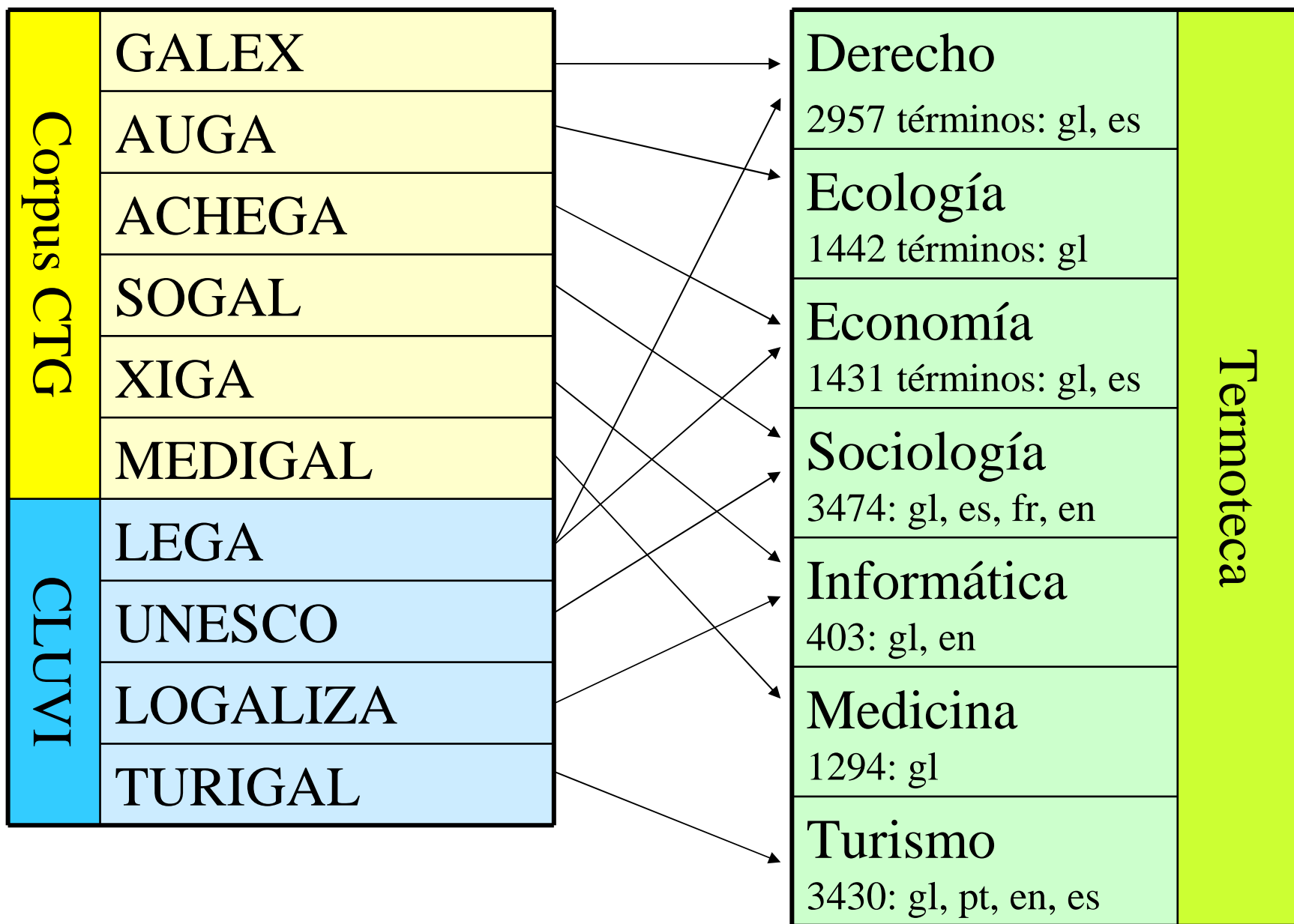
1 - [018/2]	O transporte é responsable de múltiples e graves danos ambientais, como o aumento do efecto invernadoiro (o principal problema ambiental global, con profundas implicacións máis aínda pouco presente nas preocupacións da nosa sociedade), a destrución e a fragmentación de hábitats naturais, a contaminación atmosférica, o ruído, a degradación da paisaxe ou a xeración de residuos tóxicos.	[+]
2 - [031/7]	A maioría da comunidade científica sostén que o cambio climático está causado polas emisións humanas de determinados gases (dióxido de carbono, metano, CFCs, ...), que aumentan a retención de calor na atmosfera, intensificando o chamado "efecto invernadoiro".	[+]
3 - [099/3]	A guía informa sobre as principais características e problemas do medio ambiente do noso país (sen esquecer os problemas ambientais globais, especialmente o aumento do efecto invernadoiro) e sobre o asociacionismo ecoloxista.	[+]
4 - [1000/3]	O transporte de contaminantes a longas distancias, a chuvia ácida, a destrución da capa de ozono estratosférico e a acumulación de gases con «efecto invernadoiro» son as súas manifestacións máis significativas.	[+]
5 - [1000/281]	1. A Administración pública galega elaborará anualmente un informe no que se recolla a información ambiental relevante sobre a situación da calidade do aire na Comunidade Autónoma e sobre as emisións dos distintos contaminantes, incluídas as dos gases con efecto invernadoiro, e as súas fontes, así como sobre as medidas de prevención adoptadas e os progresos acadados.	[+]
6 - [1000/303]	e) A posta en marcha de medidas e actuacións para a aplicación de técnicas de redución en orixe das emisións, co fin de lograr un alto nivel de protección do ambiente atmosférico, con especial atención ós gases de efecto invernadoiro e ás substancias destructoras da capa de ozono.	[+]
7 - [1003/30]	Os combustibles fósiles -responsables do efecto invernadoiro- seguirán representando do 75 ó 80% do consumo total de enerxía, a enerxía nuclear do 4 ó 7% e as enerxías renovables (eólica, solar, hidroeléctrica), no mellor dos casos o 20%.	[+]
8 - [1018/1]	LEI 1/2005, do 9 de marzo, pola que se regula o réxime do comercio de dereitos de emisión de gases de efecto invernadoiro.	[+]
	I A Directiva 2003/87/CE do Parlamento Europeo e do Consello, do 13 de outubro de 2003, pola que se establece un réxime para o comercio de dereitos	

# Termoteca

---

- Banco de datos terminolóxico da Universidade de Vigo
- 7.500 rexistros con información sobre 15.000 termos tirada dos corpóra CLUVI e CTG
- De libre consulta en <http://sli.uvigo.es/termoteca> e con licenza GPL en <http://metashare.upf.edu/>







<i>Ref.</i> <a href="#">8880101</a>	
<i>Campo temático:</i>	< <a href="#">UNESCO.3.1.2.6/3.1.2.6. Desviación social</a> >
<i>Definición:</i>	Crime committed using a computer and the internet to steal a person's identity or sell contraband or stalk victims or disrupt operations with malevolent programs - <a href="#">[wordref]</a>
<i>Relacións semánticas:</i>	[hipo => <a href="#">8880177</a> ]
<i>Termo GL:</i>	<b>cibercriminalidade</b>
<i>Categoría:</i>	f
<i>Morfosintaxe:</i>	cibercriminalidade/cibercriminalidade_NCFS0
<i>Variante:</i>	com
<i>Frecuencia relativa:</i>	5.54181186214
<i>Contexto de uso:</i>	En 1997, un grupo de traballo do Consello de Europa, constituído tamén por funcionarios da policía e do ministerio do Interior dalgúns países membros (Estados Unidos, Canadá e Xapón), acometeu a redacción dunha convención internacional sobre a <i>cibercriminalidade</i> . [C07]
<i>Máis exemplos:</i>	<a href="#">No CTG</a> / <a href="#">no CLUVI</a> (algún dos exemplos do termo no corpus poden non corresponder ao concepto da ficha)
<i>Termo EN:</i>	<b>cybercrime</b>
<i>Categoría:</i>	com
<i>Variante:</i>	com
<i>Frecuencia relativa:</i>	5.38968500525
<i>Contexto de uso:</i>	In 1997, a Council of Europe working group formed by law enforcement officials and interior ministers from select COE member countries, the U.S., Canada and Japan began drafting an international convention on <i>cybercrime</i> . [C07]
<i>Máis exemplos:</i>	<a href="#">No CLUVI</a> (algún dos exemplos do termo no corpus poden non corresponder ao concepto da ficha)



# Wordnet

---

- Rede léxico-semántica representada como unha base de datos léxica organizada mediante relacións semánticas, cunha orientación inicial psicolingüística (lexicón mental) (George A. Miller, Universidade de Princeton)
- Synsets (nós da rede): conxunto de sinónimos que son representacións léxicas de conceptos. Intercambiables polo menos nalgún contexto
- Relacións básicas entre synsets: hiperonimia/hiponimia, holonimia/meronimia (substantivos), antonimia (adxectivos), implicación, hiperonimia/hiponimia, causativa, oposición (verbos)



# Wordnet

---

- *Wordnet* (Universidad de Princeton)

<http://wordnetweb.princeton.edu/perl/webwn>

- *Visual Thesaurus*

Aplicación gráfica do Wordnet

<http://www.visualthesaurus.com>



# Galnet

---

- Galnet é un proxecto para crear un Wordnet do galego aliñado co ILI (índice interlingüístico) do Wordnet 3.0 do inglés
- O mesmo se está a facer co catalán, co español e co euskera de xeito coordinado co galego: EHU/UPV, UPC, UAB, UdL, UB, UOC, UPF e UVIGO



# Galnet

---

- *Multilingual Central Repository 2.0: TALG* (FFI2009-08317-E/FILO). Ministerio de Ciencia e Innovación, 2010-2011
- *Adquisición de escenarios de conocimiento a través de la lectura de textos: Desarrollo y aplicación de recursos para el procesamiento lingüístico del gallego (SKATeR-UVIGO)* (TIN2012-38584-C06-04). Ministerio de Economía y Competitividad, 2013-2015

# Galnet



The screenshot shows the Galnet web interface. At the top, there is a search bar with the word 'ouro' entered. Below the search bar, there are several dropdown menus for selecting the word and its hyponym, and checkboxes for selecting the gloss, score, relations, and full view. The search results are displayed in a table with columns for the word, its hyponym, and its gloss. The word 'ouro' is listed with its hyponym 'ouro\_1' and its gloss 'a soft yellow malleable ductile (trivalent and univalent) metallic element; occurs mainly as nuggets in rocks and alluvial deposits; does not react with most chemicals but is attacked by chlorine and aqua regia'. The gloss is provided in English, Basque, Spanish, and Galician.

Word	Nouns	Galician_3.0	English_3.0
ouro	ouro_1	ouro_1	ouro_1

Multilingual Central Repository (ILI 3.0) - [WikiMCR](#)

eng-30-14638799-n 5 gold\_3  
ili-30-14638799-n Au\_1 atomic\_number\_79\_1  
eus-30-14638799-n 5 urre\_2  
spa-30-14638799-n 5 oro\_2  
glg-30-14638799-n 5 Au\_1  
número\_atómico\_79\_1 ouro\_1  
cat-30-14638799-n 5  
número\_atòmic\_79\_1 or\_3

a soft yellow malleable ductile (trivalent and univalent) metallic element; occurs mainly as nuggets in rocks and alluvial deposits; does not react with most chemicals but is attacked by chlorine and aqua regia  
gai bakuna, kolore hori distiratsuko metal bikaina, urak, aireak eta garratzek erasotzen ez diotena, metaletan harikor eta xaflakorrena da: 1849an aurkitu zen urrea Kalifornian eta bertara ziren askoren artean ditugu euskaldunak;  
elemento metálico, (trivalente e univalente), dúctil, maleable, amarelo e suave; atópase principalmente como pebidas nas pedras e depósito de aluvial; non reacciona coa maioría dos produtos químicos, pero é atacado polo cloro e a auga rexia  
Metall preciós

- <http://sli.uvigo.es/galnet>
- no RILG  
<http://sli.uvigo.es/RILG>
- en EuroWordNet  
[adimen.si.ehu.es/web/MCR](http://adimen.si.ehu.es/web/MCR)
- no Open Multilingual WN  
<http://casta-net.jp/~kuribayashi/multi/>



# Galnet: primeira fase

---

- Creación dos synsets nominais e verbais pertencentes aos Basic Level Concepts (conceptos de Wordnet que representan o maior número de conceptos abstractos sen deixar de ser conceptos concretos)
- Tradución e adaptación dos ficheiros de nomes de partes do corpo e das substancias; tradución e adaptación parcial dos adxectivos



# Galnet: primeira fase

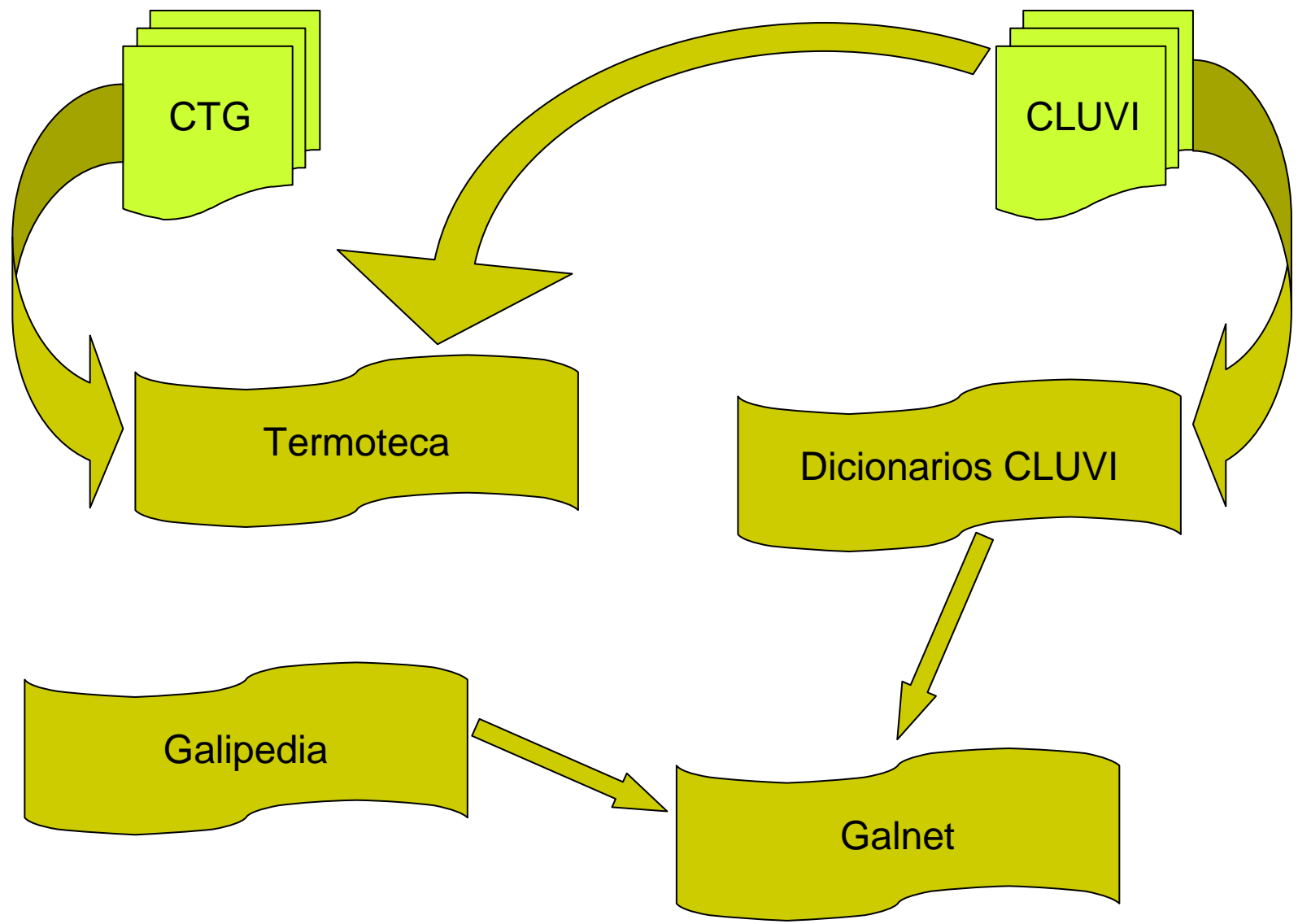
	<b>Variantes WN30</b>	<b>Synsets VN30</b>	<b>Variantes Galnet</b>	<b>Synsets Galnet</b>
<b>Nomes</b>	117798	82115	9183	5646
<b>Verbos</b>	11529	13767	1414	616
<b>Adxectivos</b>	21479	18156	4864	3114
<b>Adverbios</b>	4481	3621	0	0
<b>TOTAIS</b>	155287	117659	15461	9376



## Galnet: segunda fase

---

- Ampliación de Galnet coa ferramenta WN-Toolkit a partir de dous recursos léxicos bilingües inglés-galego xa existentes: Dicionario CLUVI e Wikipedia (Galipedia)
- O resultado da extracción automática, revisado manualmente, serviu para ampliar o Galnet con 11.677 novas variantes e 9.936 novos synsets (isto é, ao duplo do existente)



# Galnet: estado actual

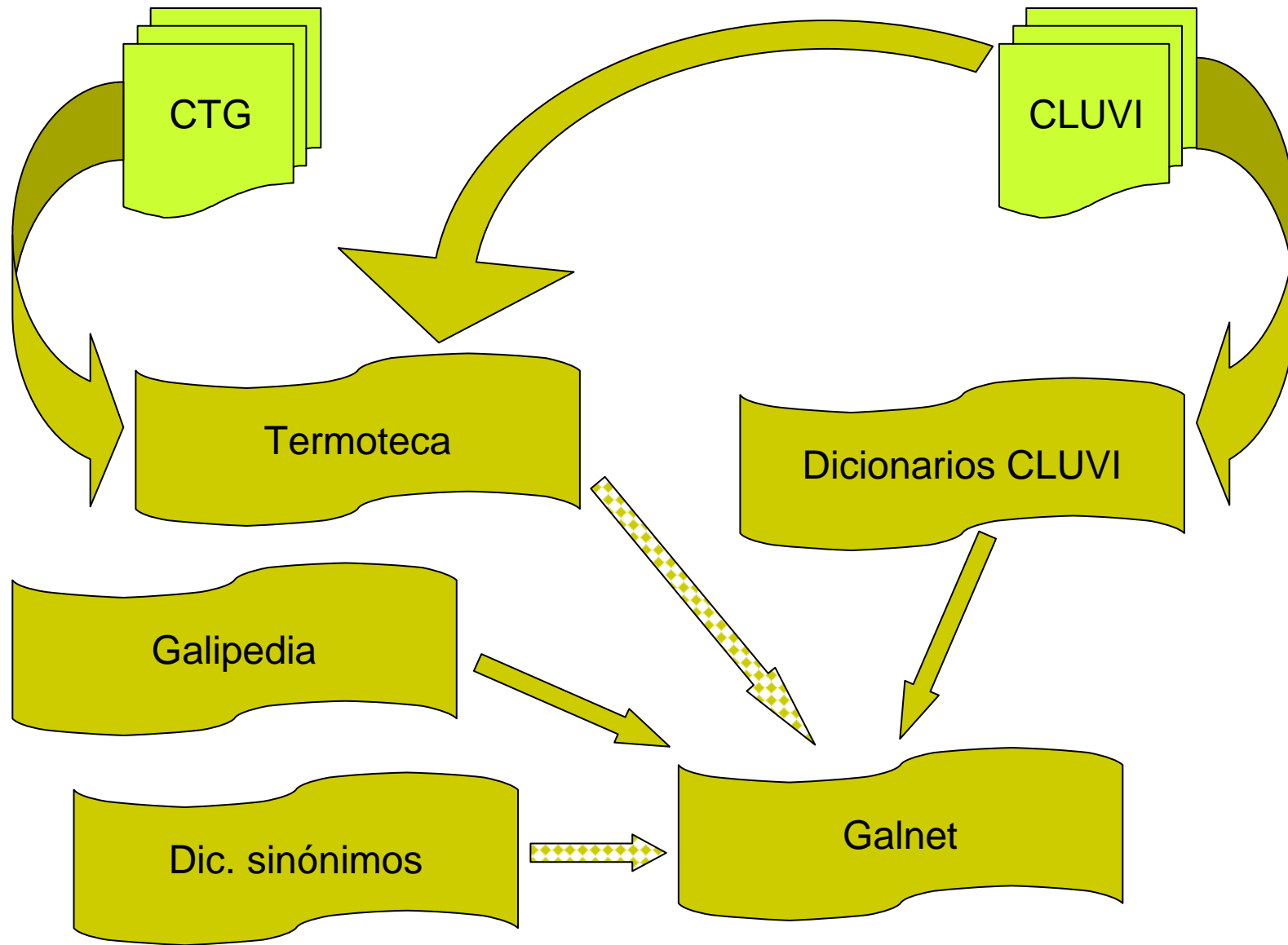
	<b>Variantes WN30</b>	<b>Synsets VN30</b>	<b>Variantes Galnet</b>	<b>Synsets Galnet</b>
<b>Nomes</b>	117798	82115	18949	14285
<b>Verbos</b>	11529	13767	1416	612
<b>Adxectivos</b>	21479	18156	6773	4415
<b>Adverbios</b>	4481	3621	0	0
<b>TOTAIS</b>	155287	117659	27138	19312



## Galnet: terceira fase (en curso)

---

- Ampliación de Galnet, no ámbito do léxico especializado, a partir das entradas terminolóxicas recollidas na Termoteca
- Ampliación de Galnet, no ámbito do léxico xeral, a partir das entradas do Dicionario de Sinónimos da Lingua Galega (Galaxia, 1997)
- Ampliación de Galnet no ámbito da fraseoloxía



# Galnet

## WordNet 3.0 do galego

---

Agardamos que fose do seu interese.  
Moitas grazas!

Grupo TALG

<http://webs.uvigo.es/talga>

