

# Lenguaje Natural y Ontologías

Héctor Jiménez Salazar

`hgimenezs@gmail.com`

1. Necesidad de las ontologías.
2. Conocimiento y lenguaje.
3. Léxico.
4. Relaciones semánticas.

# Numeralia<sup>a</sup>

## *Algunos datos sobre la web (2003)*

1. Cada año se recopila medio millón de nuevas bibliotecas (del tamaño de la B. del Congreso).
2. 5 'exabytes'  $\approx$  5 millones de 'terabytes' (1 'terabyte'  $\approx$  1 millón de 'megabytes'), aproximadamente el texto que contienen un millón de libros (=5 billones de libros).
3. El 92% de toda esta información se almacenó en medios magnéticos, sobre todo discos duros.
4. 600 millones de personas tienen acceso
5. 167 'terabytes', sólo de páginas fijas (167 millones de libros).

<sup>a</sup><http://cibernauta.elcorreodigital.com>, Peter Lyman, Hal Varian. U. de Berkeley.

## Papel: lujo no sostenible<sup>a</sup>

*Se espera aún mayor digitalización.*

1. Crecimiento 36% en el último trienio.
2. Son necesarios 786 millones de árboles para generar el papel que consume el mundo en un año (según UNESCO). (Un árbol viene a producir unas 80,500 hojas)
3. Cada habitante de la tierra consume 1,510 hojas al año (en EEUU 11,916 hojas por hab./año. Europa 7,280).
4. La mitad de todo ese volumen de papel se emplea en impresoras y fotocopiadoras de oficina.

<sup>a</sup><http://cibernauta.elcorreodigital.com>, Peter Lyman, Hal Varian. U. de Berkeley.

# Problemas de Búsqueda

1. Actualmente hacemos búsquedas dando una cadena de caracteres: `banco` es una palabra ambigua y debe darse información adicional, `banco de trabajo` → 4,430,000 páginas (el buscado aparece primero en la posición 6).
2. Puede hacerse la búsqueda dentro de una página para localizar la información, o la misma página puede contar con una biblioteca digital, glosario, etc. lo cual restringe la búsqueda.
3. Es necesario agregar información semántica a cada elemento de las páginas html. Esto lleva a etiquetar los términos con categorías semánticas (p.e. `perro` → `animal`, etc.)

Se han propuesto diferentes caminos para realizarlo, pero en todos los casos es necesario tratar con el lenguaje natural.

Aplicaciones Lugares Sistema mar 10 de abr, 23:36

banco trabajo - Buscar con Google - Mozilla Firefox

Archivo Editar Ver Historial Marcadores Herramientas Ayuda

http://www.google.com.mx/search?hl=es&q=banco+trabajo&btnG=B%C3%BAscueda&meta= Google

Getting Started Latest BBC Headlines

Acceder

Google **La Web** [Imágenes](#) [Grupos](#) [Noticias](#) [Más »](#)

banco trabajo  [Búsqueda avanzada](#) [Preferencias](#)

Búsqueda: la Web páginas en español páginas de México

**La Web** Resultados 1 - 10 de aproximadamente 1,990,000 de **banco trabajo**. (0.09 segundos)

[Grupo Altas Cumbres](#)  
www.bantra.com.pe/ - 2k - [En caché](#) - [Páginas similares](#)

[Banorte | El Banco Fuerte de México](#)  
Por segundo año consecutivo, Banorte el mejor **banco** de México ... cuyo espíritu de innovación se extiende a todos los aspectos del entorno del **trabajo**. ...  
www.banorte.com/portal/banorte.portal?\_nfpb=true&\_pageLabel=pageKnow&elementId=23 - 43k - [En caché](#) - [Páginas similares](#)

[Banco de Trabajo](#)  
**Banco de Trabajo** - Una puerta al empleo. Empresa de Servicios Eventuales avalada por 19 años de permanencia en el mercado.  
www.bancondetrabajo.com.ar/ - 9k - [En caché](#) - [Páginas similares](#)

[El Banco Mundial - Empleos](#)  
El idioma de **trabajo** del **Banco Mundial** es el inglés; por tanto, si desea postular a un cargo, debe hablarlo y escribirlo con fluidez. Además del dominio del ...  
www.bancomundial.org/empleos/ - 29k - [En caché](#) - [Páginas similares](#)

[El Banco Mundial - Empleos](#)  
El **trabajo** del **Banco Mundial**. El personal del **Banco Mundial** forma parte de un sistema de redes y sociedades globales que vinculan a los mejores ...  
www.bancomundial.org/empleos/gente.htm - 20k - [En caché](#) - [Páginas similares](#)

[Mesa o banco de trabajo para taller - Negocios - Equipo ...](#)  
**Banco de trabajo** metalico para uso pesado en buenas condiciones. Mide Terminado

Terminado

banco trabajo - Busc... hjimenez@hjimenez... ConDCCD - Navegad... [hjimenez@hjimene... [Descargas] Iniciando Capturar p...

*Búsqueda de información: Concepto vs cadena.*

Aplicaciones Lugares Sistema mar 10 de abr, 23:35

"banco de trabajo" - Buscar con Google - Mozilla Firefox

Archivo Editar Ver Historial Marcadores Herramientas Ayuda

http://www.google.com.mx/search?hl=es&q=%22banco+de+trabajo%22&btnG=Buscar+con+Google&meta Google

Getting Started Latest BBC Headlines

Acceder

Google "banco de trabajo" Búsqueda Búsqueda avanzada Preferencias

Búsqueda: la Web páginas en español páginas de México

**La Web** Resultados 1 - 10 de aproximadamente 58,300 de "banco de trabajo". (0.12 segundos)

[Grupo Altas Cumbres](#)  
www.bantra.com.pe/ - 2k - En caché - Páginas similares

**Banco de Trabajo**  
**Banco de Trabajo** - Una puerta al empleo. Empresa de Servicios Eventuales avalada por 19 años de permanencia en el mercado.  
www.bancodetrabajo.com.ar/ - 9k - En caché - Páginas similares

**Mesa o banco de trabajo para taller - Negocios - Equipo ...**  
**Banco de trabajo** metalico para uso pesado en buenas condiciones. Mide 220x60x100cm (largoXanchoXaltura). Tengo otra mesa de menos altura anunciada tambien.  
www.segundamano.com.mx/.../Mesa\_o\_banco\_de/add18580434\_14004\_14005\_14009\_4\_209\_20900003\_0/index2.aspx - 28k - En caché - Páginas similares

**Mesa o banco de trabajo para taller - Negocios - Equipo ...**  
**Banco de trabajo** metalico para uso pesado en buenas condiciones. Mide 220x60x70cm (largoXanchoXaltura). Tengo otra mesa un poco mas alta anunciada tambien.  
www.segundamano.com.mx/.../Mesa\_o\_banco\_de/add18580409\_14004\_14005\_14009\_4\_209\_20900003\_0/index2.aspx - 28k - En caché - Páginas similares

**Extrusiones Metálicas**  
**BANCO DE TRABAJO Y MONTAJE B-4000.** CARACTERÍSTICAS TÉCNICAS:. Banco especialmente diseñado para el montaje de cerramientos de aluminio y PVC. ...

Terminado

"banco de trabajo" - ... hjimenez@hjimenez... ConDCCD - Navegad... [hjimenez@hjimene... [Descargas] Iniciando Capturar p...

*Búsqueda de información: Ambigüedad.*

Aplicaciones Lugares Sistema mar 10 de abr, 23:25

Columbia Newsblaster: Iran Expands Uranium Enrichment Effort - Mozilla Firefox

Archivo Editar Ver Historial Marcadores Herramientas Ayuda

http://www1.cs.columbia.edu/nlp/newsblaster/summaries/2007-04-10-04-31-03-009.html

Getting Started Latest BBC Headlines

## Columbia Newsblaster

Summarizing all the news on the Web

Tuesday, April 10, 2007  
Articles from 04/07/2007 to 04/10/2007  
Last update: 4:31 AM EST

Search for:

Offline summarizatio Go

**U.S.**  
**World**  
**Sci/Tech**  
**Sports**

[View Today's Images](#)

[View Archive](#)

[About Newsblaster](#)

[About today's run](#)

[Newsblaster in Press](#)

[Academic Papers](#)

**Article Sources:**  
[washingtonpost.com](#)  
(29 articles)  
[boston.com](#)

### Iran Expands Uranium Enrichment Effort

Summary from multiple countries, from articles in English  
[UPDATED] (see summary with new information since yesterday)

In his speech on Monday, Iranian President Mahmoud Ahmadinejad did not say how many centrifuges - the machines that spin uranium gas in order to enrich it to levels needed for fuel - were now operational at Natanz. [\(article 2\)](#) The only surprise about Iranian President Mahmoud Ahmadinejad's announcement Monday that Iran has moved to "industrial-scale" uranium enrichment is the timing. [\(article 5\)](#) NATANZ, Iran - Iran announced today that it has begun enriching uranium with 3,000 centrifuges, defiantly expanding a nuclear program that has drawn U.N. sanctions and condemnation from the West. [\(article 3\)](#) Ahmadinejad said at a ceremony at the enrichment facility at Natanz that Iran was now capable of enriching nuclear fuel " on an industrial scale. [\(article 3\)](#) Asked if Iran has begun injecting uranium gas into 3,000 centrifuges for enrichment, top nuclear negotiator Ali Larijani replied, " Yes. [\(article 3\)](#) NATANZ, Iran Iran's president and security officials announced today that their country has begun producing nuclear fuel on an "industrial scale" in defiance of United Nations demands that Tehran stop its pursuit of nuclear technology [\(article 1\)](#) " Iran continues to defy the international community and further isolate itself by expanding its nuclear program, rather than suspending uranium enrichment said Gordon Johndroe [\(article 4\)](#)

**Other summaries about this story:**

- [Summary from the United Kingdom, from articles in English](#) (1 articles) [\[compare\]](#)
- [Summary from United States, from articles in English](#) (5 articles) [\[compare\]](#)



Terminado

Columbia Newsblast... hjimenez@hjimenez... ConDCCD - Navegad... [hjimenez@hjimene... [Descargas] Iniciando Capturar p...

Acceso a la información: Sistema de acceso a noticias.

Aplicaciones Lugares Sistema mar 10 de abr, 23:26

Columbia Newsblaster: Iran Expands Uranium Enrichment Effort - Mozilla Firefox

Archivo Editar Ver Historial Marcadores Herramientas Ayuda

http://www1.cs.columbia.edu/nlp/newsblaster/summaries/2007-04-10-04-31-03-009.html

Getting Started Latest BBC Headlines

**Article Sources:**

- [washingtonpost.com](#) (29 articles)
- [boston.com](#) (18 articles)
- [baltimoresun.com](#) (16 articles)
- [seattletimes.com](#) (15 articles)
- [nwsources.com](#) (15 articles)
- [news.bbc.co.uk](#) (14 articles)
- [dallasnews.com](#) (14 articles)
- [nypost.com](#) (10 articles)
- [latimes.com](#) (9 articles)
- [nytimes.com](#) (9 articles)
- [suntimes.com](#) (7 articles)
- [msnbc.msn.com](#) (5 articles)
- [abcnews.go.com](#) (4 articles)
- [usatoday.com](#) (3 articles)
- [time.com](#) (3 articles)
- [film.com](#) (3 articles)
- [guardian.co.uk](#) (2 articles)
- [sfgate.com](#) (2 articles)

**Other summaries about this story:**

- [Summary from the United Kingdom, from articles in English](#) (1 articles) [compare]
- [Summary from United States, from articles in English](#) (5 articles) [compare]

**Event tracking:**

- [Track this story's development in time](#)

**Story keywords**  
Iran, Nuclear, centrifuges, enrichment, uranium

**Source articles**

- [Iran defies U.N., says its nuclear fuel production is on an 'industrial scale'](#) (L.A. Times, 04/09/2007, 36 words)
- [Iran nuclear 'landmark' angers US](#) (BBC News, 04/10/2007, 403 words)
- [The Seattle Times: Iran producing nuclear fuel "on an industrial scale"](#) (seattletimes.nwsources.com, 04/09/2007, 842 words)
- [Iran Expands Uranium Enrichment Effort](#) (msnbc.msn.com, 04/09/2007, 1070 words)
- [Behind Iran's Nuclear Tough Talk](#) (time.com, 04/09/2007, 811 words)
- [Iran Looks for Missing Ex-F.B.I. Agent](#) (nytimes.com, 04/09/2007, 263 words)

Terminado

Columbia Newsblast... hjimenez@hjimenez... ConDCCD - Navegad... [hjimenez@hjimene... [Descargas] Iniciando Capturar p...

*Acceso a la información: Resumen de eventos.*



**Columbia Newsblaster**  
Summarizing all the news on the Web

Tuesday, April 10, 2007  
Articles from 04/07/2007 to 04/10/2007  
Last update: 4:31 AM EST

**Search for:**

Offline summarizatio Go

**U.S.**  
**World**  
**Sci/Tech**  
**Sports**

[View Today's Images](#)

[View Archive](#)

[About Newsblaster](#)

[About today's run](#)

[Newsblaster in Press](#)

[Academic Papers](#)

**Article Sources:**  
[washingtonpost.com](#)  
(29 articles)  
[boston.com](#)

**Today's run's statistics**

[What languages were summarized today?](#)  
[What summarizers were used today?](#)

Site	Pages downloaded	Pages with articles	Articles with dates	Recent articles	Articles in clusters
<b>Total (all sites)</b>	9585	2773	2331	880	133
<a href="#">www.washingtonpost.com</a>	479	157	138	97	28
<a href="#">www.baltimoresun.com</a>	900	476	467	89	15
<a href="#">www.boston.com</a>	587	213	198	102	14
<a href="#">www.dallasnews.com</a>	1094	347	243	62	12
<a href="#">seattletimes.nwsourc.com</a>	456	180	142	90	10
<a href="#">www.nypost.com</a>	244	60	38	37	10
<a href="#">www.latimes.com</a>	310	36	25	23	8
<a href="#">news.bbc.co.uk</a>	325	224	215	61	7
<a href="#">www.nytimes.com</a>	345	141	141	54	7
<a href="#">www.abcnews.go.com</a>	331	93	37	34	4
<a href="#">www.msnbc.msn.com</a>	160	30	29	21	4
<a href="#">www.suntimes.com</a>	252	88	82	53	3
<a href="#">www.usatoday.com</a>	201	41	16	14	3
<a href="#">www.time.com</a>	42	9	9	4	2
<a href="#">www.sfrate.com</a>	195	35	26	15	2

Terminado

Columbia Newsblaster: St... hjimenez@hjimenez-lapto... ConDCCD - Navegador de... [hjimenez@hjimenez-lapt... Iniciando Capturar pantalla

*Acceso a la información: Fuentes.*

# Sistemas de Búsqueda de Respuesta

Son sistemas que pretenden atender la demanda de textual. Para ubicar la complejidad del problema y sistematizar su desarrollo, los usuarios se clasifican:

- Usuario casual. Requiere información puntual. *¿Dónde está la Bufadora?, ¿Cuánto aumenta la población mundial cada año? La información puede aparecer en un solo documento.*
- Recopilador de información. *Se requiere recorrer varios documentos para dar respuesta: ¿Cuáles son los datos biográficos más importantes de Nelson Mandela?.* Un acercamiento es utilizar una platilla de pregunta; para la última pregunta se generarían varias preguntas puntuales sobre él: fecha de nacimiento, lugar de nacimiento, estudios realizados, etc.
- Periodista. Ayuda a la redacción de artículos. A partir de la noticia del *tsunami*, habría que recabar datos como: lugar exacto, daños materiales, causas probables, otros eventos similares, etc. Para *¿Toda la información sobre tsunami?* el SBR tendría que generar con base a conocimiento sobre este tipo de preguntas, una estrategia para hacer búsquedas en diferentes fuentes de información.
- Analista profesional. Experto en un tipo de información particular; financiero, política, etc. El SBR debe ser persistente y considera el perfil del usuario, para interactuar con él. En el proceso de búsqueda deben realizarse operaciones de recopilación y síntesis, pero además proveer herramientas de navegación.

# Complejidad de las respuestas

- Diversidad de fuentes de datos. Deberá tener la habilidad para acceder a fuentes de información tanto estructurada como no estructurada, y multimedia, multilingüe o distribuida.
- Interpretación de la información. El juicio de relevancia sobre la información es crucial para la selección e integración de las respuestas, pero además no basta la interpretación de la pregunta, pues es necesario el contexto en el que se hace para tener claro los criterios de importancia de la información recuperada.
- Integración de datos individuales. A partir de las diferentes fuentes deberá integrarse, combinar y resumir la información y ayudar a procesos posteriores, como la inferencia.

## Componentes de un SBR

1. **Análisis.** Para la pregunta *¿Quién es el emperador de Japón?* se espera encontrar un pasaje: *Hiroito, emperador de Japón, dijo ....*
2. **Recuperación de pasajes.** Los documentos se fragmentan (pasajes) y sobre ellos actúa un SRI convencional. La query sería: *emperador japon.*
3. **Extracción de respuesta.** De los pasajes recuperados deben elegirse los candidatos.  $P_1$ :*Ni siquiera el emperador de Japón acudió ...*,  $P_2$ :*Akihito dio por iniciados los trabajos en la capital del Japón.*,  $P_3$ :*Los japoneses estuvieron satisfechos de Akihito, su emperador.*

Se utiliza una colección de documentos de los cuales se extraen los pasajes y sobre los pasajes se realizan operaciones de empatamiento con la información que proporciona la pregunta.

# Validación de respuestas en SBR

*¿Qué es el Hubble?*

Telescopio espacial.

En los pasajes encontrados debe inferirse la respuesta:  
Hubble es el telescopio espacial.

Pasaje: *La NASA espera mostrar este jueves las primeras imágenes tomadas por el telescopio espacial Hubble, después de que el mes pasado los astronautas del transbordador Endeavour lograsen corregir el defecto óptico de sus lentes.*

## Pasajes falsos

¿Qué requiere el abedul para vivir?  
(sin respuesta)

Uso de una ontología: vegetal→árbol→abedul. Pero pueden obtenerse pasajes falsos:

Algunas enfermedades vegetales requieren ciertas condiciones de temperatura; por ej. *Pseudomonas syringae*...

...procesos de recolección, clasificación y prensado de los distintos elementos vegetales requieren un trabajo continuo y minucioso;...

La relación semántica: vegetal  $\xrightarrow{\text{requiere}}$  sol, puede ser usado para concluir: abedul requiere de sol.

# Ontologías en el campo computacional

*Una **ontología** es un sistema compuesto de la terminología correspondiente a la conceptualización en un dominio, organizada en clases que reúnen instancias, además existen relaciones y funciones sobre clases e instancias. La sistematización de dicha información permite abstraer reglas para incorporar nuevos elementos.*

Una ontología puede verse como un **sistema formal** (constantes, variables, funciones, relaciones, axiomas, y reglas de inferencia), por ejemplo los derivados de la lógica clásica. Aunque también las hay informales donde normalmente existe una definición de cada concepto.

En términos generales, el diseño de una ontología deberá considerar aspectos semánticos del lenguaje.

(<http://babage.dia.fi.upm.es/ontoweb/wp1/OntoRoadMap/index.htm>)

# Tipos de ontologías

La variedad de ontologías existentes ha conducido a clasificarlas según su orientación.

<b>Orientación</b>	<b>Características</b>	<b>Ejemplo</b>
Representación del Conocimiento	Mantiene convenciones, traslada conceptos a otros lenguajes	<i>Frame-Ontology</i>
Representación de Alto Nivel	Universal, identificación, no redundancia de conceptos	<i>SUO</i>
Procesamiento del Lenguaje Natural	Synsets (conjunto de sinónimos)	<i>WordNet</i>
General	Independiente del Dominio	<i>CyC</i>
Aplicaciones en un dominio particular	Modelación de problemas	<i>UMLS</i>



# Volumen del Conocimiento

El tamaño de una ontología se mide con el número de conceptos, aunque en algunos casos, se pone importancia en otros parámetros; p.e. número de axiomas.

<b>Nombre</b>	<b># C</b>	<b># R</b>	<b># A</b>	<b># I</b>
<i>Frame-Ontology</i>	10-50	10-50	100-500	-
<i>Standard Upper O</i>	100-500	100-500	-	1K-5K
<i>WordNet</i>	100K	-	0	200K
<i>CyC</i>	>5K	-	-	-
<i>UMLS</i>	5K	50-100	-	-
<i>CHEMICALS</i>	10-50	10-50	10-50	100-500
<i>Onto-Qpr</i>	-	100-200	0	50-100

Número de: conceptos (#C); relaciones (#R); axiomas (#A); e instancias (#I).

## Conocimiento implícito en los textos

**Implicaturas.** Convenciones del lenguaje: *Es gringo pero me cae bien.* → [Poca simpatía por los gringos].

**Presuposiciones.** Hechos que se desprenden de una afirmación: *Rockdrigo (no) sobrevivió al maremoto.* → [Hubo un maremoto].

**Anáforas.** Relación de términos con igual correferencia: *Nace sin que para ello exista fundamento legal. Y eso es grave.* → [eso == nacer así].

**Conocimiento del mundo.** Conocimiento adquirido por experiencia: *Salió sin cerrar la puerta.* → [La puerta quedó abierta].

## Inseparabilidad Conocimiento-Lenguaje

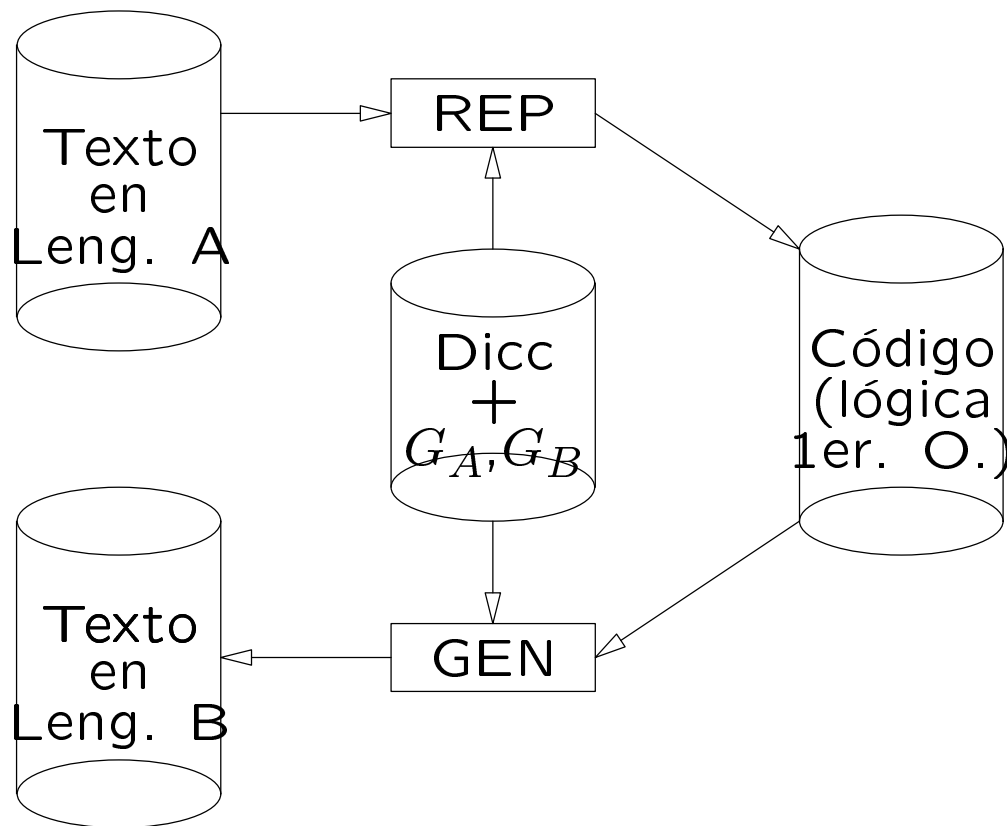
1. Información: símbolos físicos independientes de un criterio predeterminado de interpretación.
2. Conocimiento: establece relaciones múltiples con otros conocimientos, por lo cual un agente puede realizar inferencias; producir nuevo conocimiento.
3. Lenguaje: es el código para acceder al conocimiento.

# Enfoque simbólico del PLN

*Como se supone que lo hace el ser humano*

- Favorece la competencia (conocimiento del lenguaje)
- Requiere teoría lingüística:
  1. Reglas gramaticales
  2. Diccionario
  3. Lenguaje de representación
- Alta precisión vs baja cobertura.

# Traducción automática (enfoque simbólico)

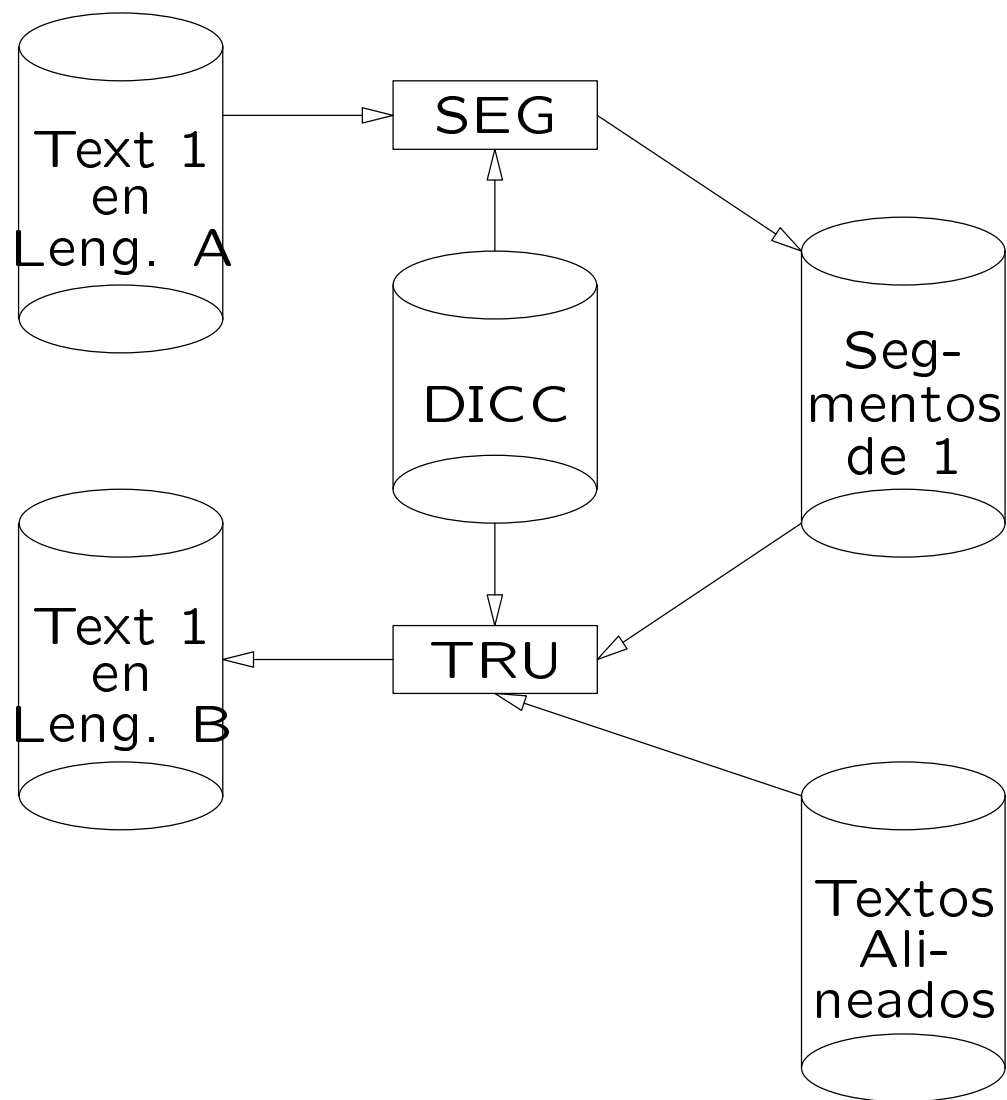


## Enfoque empírico del PLN

*Sólo es necesario que funcione aceptablemente*

- Favorece el desempeño (tareas que usan el lenguaje)
- Requiere recursos lingüísticos:
  1. Sistematización de la información (Corpus, anotación, caracterización).
  2. Uso de patrones para construir nuevos recursos lingüísticos.
  3. Clasificar textos en nuevas situaciones
- Alta cobertura vs mediana precisión.

# Traducción automática (enfoque empírico)



# Procesamiento del Lenguaje Natural

## *Taxonomía abreviada*

- Texto
  - Representación
    - \* Técnicas de selección de términos, Ponderación en la representación vectorial, *Parsing* ...
  - Generación de recursos
    - \* Corpora, Bases de Datos Léxicas (BDL), Terminología, Extracción de Información (EI), Relaciones léxico-semánticas (RL) ...
  - Clasificación
    - \* Agrupamiento de textos (AT), Desambiguación del sentido (WSD) ...
  - Búsqueda de información
    - \* Recuperación de Información (RI), Búsqueda de Respuestas (SBR) ...
  - Acceso a la información
    - \* Traducción Automática, Resumen Automático, Generación de texto ...
- Habla ...



## Características del LN

**Universalidad:** Podemos expresar cualquier cosa en el lenguaje natural, *idea, petición, saludo*, etc. (no así en otros lenguajes).

**Irregularidad:** La correspondencia no es unívoca:

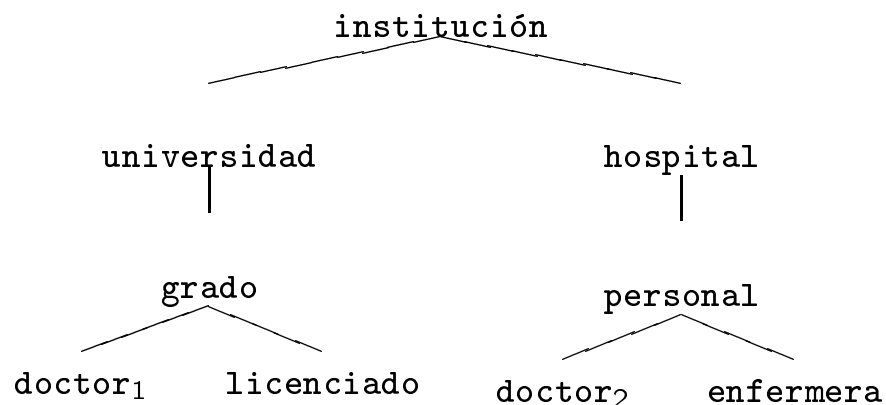
1. *Ambigüedad:* muchos significados para una expresión.
2. *Sinonimia:* muchas expresiones para un significado.

## Estado del Español

- Lengua de uso extendido: Latinoamérica, Europa, Africa y Oriente.
- 400 millones de hispanohablantes.
- Tercer lugar mundial (después del Chino e Inglés).
- Páginas *web* crecen cinco veces más rápido que en inglés.
- Cada vez más recursos disponibles (sobre todo en relación con la Comunidad Europea: traducción automática).
- No es comparable con los recursos desarrollados para lenguas como el Inglés (aunque la brecha se reduce es de más de cinco años).
- Necesidad de incentivar el uso del PLN en países en desarrollo (casi en todos los países de habla hispana hay desarrollo de PLN, pero es escaso).
- En México se empieza a formar un grupo de investigación en PLN (AMPLN).

## Uso de una ontología en la desambiguación

La desambiguación es el mayor problema del PLN. La solución a este problema facilitaría enormemente la construcción de la *web-semántica*. Considérese una parte de la jerarquía de relaciones de holonimia e hiperonimia:



La oración “El doctor instruyó a la enfermera” instancia “doctor” como personal de hospital, puesto que hay menor *distancia semántica* entre doctor<sub>2</sub> y enfermera que entre doctor<sub>1</sub> y enfermera.

## Resumen

- El problema de la informatización de la sociedad reside en la web, pero, a la vez, en ella está la solución.
- Es indispensable construir recursos y sistemas basado en el PLN.
- Los métodos de aprendizaje automático tienen importancia por la vasta cantidad de información que existe hoy en día.