ART´ICULO INVITADO

Procesamiento de Lenguaje Natural y sus Aplicaciones

por Alexander Gelbukh

# Un cuento de una m´aquina parlante

En los cuentos para nin˜os, los animales y las cosas inanimadas pero m´agicas, se comportan como personas: inteligentemente. Pueden ver, o´ır, pensar, actuar. Pero

¿c´omo sabemos que un animal o una cosa son inteligen- tes? Porque son parlantes: hablan y entienden lo que les dicen. El hombre siempre ha asociado la inteligencia con el habla.

En nuestros d´ıas la ciencia convierte cada vez m´as cuentos en una realidad. Ya no nos sorprende una al- fombra voladora (aunque no parezca una alfombra) y

¿qu´e falta para que podamos conversar con Pinocho? En los nu´meros anteriores de Komputer Sapiens se ha ha- blado sobre c´omo las m´aquinas pueden ver, pensar, ac- tuar, tomar decisiones. En este nu´mero vamos a platicar sobre c´omo una m´aquina puede procesar el lenguaje, un rasgo que hasta ahora ha sido exclusivo de los humanos (y, claro, de las cosas m´agicas).

Por Procesamiento de Lenguaje Natural (PLN, deno- minado tambi´en NLP por sus siglas en ingl´es) se entiende la habilidad de la m´aquina para procesar la informaci´on comunicada, no simplemente las letras o los sonidos del lenguaje. En este sentido, un perico no es un animal par- lante; as´ı, una contestadora telef´onica comu´n, una impre- sora o un procesador de palabras como Microsoft Word tampoco son dispositivos o software de PLN, mientras que un traductor autom´atico sin duda lo es.

Diferentes programas exhiben diferente grado del procesamiento inteligente del lenguaje. Por ejemplo, un buscador de documentos puede simplemente buscar los documentos que contienen la cadena de letras especifica- da por el usuario, sin importar que esta cadena tenga o no un significado en un lenguaje (como el espan˜ol o el ingl´es). En este caso no ser´ıa una aplicaci´on del PLN. Sin embargo, el mismo buscador podr´ıa buscar los documen- tos que comuniquen la idea especificada por el usuario, sin importar con qu´e letras la comunican, y en este caso, sin duda, ser´ıa una excelente aplicaci´on de PLN, ya que entender´ıa la idea comunicada en la petici´on del usuario, la idea comunicada en cada uno de los documentos, y ser´ıa capaz de compararlas.

La ciencia que estudia el PLN se llama lingu¨´ıstica computacional. El nombre fue inventado en los tiempos cuando eso era: lingu¨´ıstica para las computadoras. Los lingu¨istas, a trav´es de la introspecci´on e intuici´on, es- crib´ıan las reglas y los diccionarios cada vez m´as exactos y detallados, acerc´andose al objetivo: dotar a la compu-

tadora con la capacidad de entender el lenguaje humano. Este camino era muy dif´ıcil y laborioso, y los avances, aunque impresionantes, eran lentos y espor´adicos.

Todo eso cambi´o con la llegada de Internet. Los in- vestigadores obtuvieron acceso a volu´menes gigantescos de textos, el objeto del estudio de nuestra ciencia, y esta u´ltima, en lugar de introspecci´on e intuici´on, se convir- ti´o en el estudio estad´ıstico directo de los datos disponi- bles. La lingu¨´ıstica computacional, en su etapa actual de desarrollo, es principalmente una rama de las tecnolog´ıas de aprendizaje autom´atico, una parte de la inteligencia artificial y la estad´ıstica.

El aprendizaje autom´atico se dedica al descubrimien- to totalmente autom´atico de las regularidades y las rela- ciones en los datos. Usualmente se aplica a datos num´eri- cos, pero la lingu¨´ıstica computacional puede ser consi- derada como el aprendizaje autom´atico sobre un tipo de datos especial, los textos en un lenguaje humano. Es as´ı como un nin˜o aprende su lenguaje natal: nadie le en- sen˜a las reglas, las gram´aticas y los diccionarios; en su lugar su cerebro analiza estad´ısticamente los sonidos del lenguaje y su relaci´on con el medio ambiente, y aprende a reaccionar adecuadamente.

El PLN tiene un gran nu´mero de aplicaciones pr´acti- cas. Aunque el gran suen˜o de los investigadores es poder algu´n d´ıa conversar en viva voz con Pinocho (de lo cual quiz´a estamos menos lejos de lo que parece), avances incluso muy pequen˜os e insignificantes en comparaci´on con este suen˜o llevan a grandes logros tecnol´ogicos en las aplicaciones de las tecnolog´ıas del PLN.

# Uso eficiente de nuestro tesoro: Bu´squeda y presentaci´on del texto

El conocimiento es el mayor tesoro que posee la hu- manidad. Durante miles de an˜os la actividad m´as im- portante del hombre ha sido el producir el conocimiento, guardarlo y pasarlo a las siguientes generaciones. Cuan- do se trata de dinero, lo guardamos de tal manera para encontrarlo r´apidamente cuando lo necesitamos y procu- ramos que no pierda valor con el tiempo. Pero cuando se trata de nuestro mayor tesoro, el conocimiento, lo ma- nejamos de manera tan negligente como nunca hacemos con el dinero.

El conocimiento se almacena y se transmite en for- ma de lenguaje humano, los textos escritos, por ejemplo, en espan˜ol o ingl´es. Sin embargo, en la actualidad usa- mos estos textos muy ineficientemente. Mencionar´e cua-

tro componentes necesarios para su uso eficiente: la digi- talizaci´on, la bu´squeda, la presentaci´on de la informaci´on y su uso directo por el software.

Primero, la digitalizaci´on de los documentos. Las bi- bliotecas tienen toneladas de libros en papel. Los archi- vos, tales como el Archivo General de la Naci´on, tienen kil´ometros de estantes llenos con documentos de gran im- portancia, muchos de los cuales est´an en tal estado f´ısico que simplemente tomarlos en la mano es problem´atico.

Por digitalizaci´on aqu´ı entiendo la obtenci´on del tex- to como una secuencia de letras, no como una fotograf´ıa digital. Esto requiere de gran fuerza de PLN. Un lec- tor humano, cuando lee un texto donde ciertas letras no son muy claras o cuando escucha una conversaci´on en un ambiente ruidoso, f´acilmente restaura las partes faltantes porque entiende su contenido. Los programas hoy en d´ıa son cada vez m´as capaces de reconocer el texto impreso o hasta escrito a mano o reconocer el habla, gracias a sus capacidades lingu¨´ısticas.

Segundo, la bu´squeda de la informaci´on relevante, llamada tambi´en recuperaci´on de informaci´on. No sirve de nada el conocimiento escrito y guardado si no pue- de encontrarse cuando se necesita. El problema de la bu´squeda es que la misma idea se puede expresar con muy diferentes palabras. Por ejemplo, el usuario expre- sa su inter´es con la frase “la derrota de Maximiliano” y el documento relevante para tal petici´on es “la vic- toria de Ju´arez”. Los dos textos no tienen ninguna pa- labra en comu´n, pero un humano, usando su experien- cia lingu¨´ıstica (derrota—victoria) y su conocimiento del mundo (Maximiliano—Ju´arez) f´acilmente detectar´ıa la relevancia del documento para la petici´on.

Progresos muy significativos se han logrado para que los programas puedan utilizar este tipo de razonamiento para satisfacer de la mejor manera las necesidades de los usuarios.

Tercero, la presentaci´on eficiente de la informaci´on contenida en los textos. El ejemplo m´as directo de esta tecnolog´ıa es la construcci´on autom´atica de resu´menes: dado un texto largo (o un mill´on de textos), un genera- dor autom´atico de resu´menes trata de detectar lo m´as importante que se comunica y presentarlo en un texto corto que se podr´a leer en un tiempo razonable. A pe- sar del mucho esfuerzo que se ha dedicado a estas tec-

nolog´ıas, los resultados obtenidos hasta ahora son au´n modestos, aunque cada vez mejores.

Otra manera de resumir la informaci´on contenida en muchos documentos y hacerlos m´as manejables es agru- parlos y clasificarlos; en lugar de tener que leer millones de archivos, el usuario s´olo necesitar´a considerar, diga- mos, cinco grupos cuyos documentos se parecen entre s´ı. O bien, diferentes personas considerar´an cada grupo de documentos. Por ejemplo, en un gobierno, alcald´ıa o en una empresa grande, las quejas y peticiones de los ciudadanos o los clientes se dirigir´an a las oficinas co- rrespondientes.

El resumen de la informaci´on relevante puede llegar a ser tan corto como una sola palabra. Es el caso de la respuesta autom´atica a preguntas. ¿Para qu´e busca los documentos el usuario de un sistema de recuperaci´on de informaci´on? Quiz´a no necesita los documentos sino tiene una duda e intenta aclararla ley´endolos. Las tec- nolog´ıas de respuesta autom´atica a preguntas lo hacen directamente: a la petici´on “¿D´onde naci´o Ju´arez?” la respuesta ser´a “¡en Guelatao!” y no la biograf´ıa comple- ta de Ju´arez. Tales sistemas se basan en un razonamiento complejo que a veces requiere de profunda comprensi´on del significado del texto: por ejemplo, pueden inferir la informaci´on requerida del texto “llegamos a Guelatao, el pueblo natal del Benem´erito de Las Am´ericas”.

Otras maneras de resumir el contenido de muchos textos incluyen la miner´ıa de texto (encontrar las opi- niones prevalecientes expresadas en los textos, las ten- dencias de cambio de estas opiniones o las relaciones inesperadas entre los eventos descritos en los textos), la extracci´on de informaci´on (llenar bases de datos so- bre un tema espec´ıfico, leyendo los textos) y sistemas de soporte a la toma de decisiones (buscar, sintetizar y pre- sentar de manera eficiente la informaci´on relevante para un directivo).

Cuarto, el uso de la informaci´on contenida en los textos por el mismo software para resolver tareas m´as complejas. La m´aquina puede encontrar el conocimiento necesario de los textos disponibles, tales como los art´ıcu- los cient´ıficos o los libros de texto. Tales aplicaciones est´an actualmente en la fase experimental, aunque en el futuro se convertir´an en la manera principal del manejo de conocimiento.

# ¿C´omo sabemos que un animal o una cosa son inteligentes? Porque son parlantes: hablan y entienden lo que les dicen. El hombre siempre ha asociado la inteligencia con el habla.

de los pueblos de la Am´erica Lati- na), as´ı como se dividen (pol´ıtica, econ´omica, social y culturalmente) por las fronteras no tanto pol´ıti- cas sino lingu¨´ısticas (como tam- bi´en se puede observar en el mapa de nuestro continente). Los indivi- duos, como las naciones, los pueblos o grupos pueden sentirse excluidos (econ´omica, social y culturalmente) por la frontera lingu¨´ıstica, la cual les dificulta el acceso a la informa- ci´on producida por la humanidad.



A los esfuerzos para combatir estos efectos negativos de la divi- si´on lingu¨´ıstica en el mundo y en nuestro pa´ıs, el PLN aporta las tec- nolog´ıas de la traducci´on autom´ati- ca. Con esta tecnolog´ıa el usuario puede leer en su propio lenguaje un texto escrito en otro lenguaje, pue- de escribir dirigi´endose a los lec- tores que hablan otros lenguajes o conversar (a trav´es de los mensajes instant´aneos o en viva voz) con un interlocutor que habla otro lengua- je.



La calidad de la traducci´on autom´atica se mejor´o dram´ati- camente en la u´ltima d´eca-

cribi´eramos as´ı. El mejorar este aspecto requiere de mucho esfuer- zo, pero es manejable y aunque a veces el texto se ve raro, no presen- ta tanta molestia en la pr´actica.

Segundo, y mucho m´as peligroso, la traducci´on incorrecta. Este pro- blema se nota mucho menos que el primero (y entre m´as necesita el usuario la ayuda del traductor, menos va a notar sus errores), pero puede tener consecuencias graves por la generaci´on de posibles ma- los entendidos e informaci´on falsa. Sin embargo, es mucho m´as dif´ıcil corregir este tipo de problemas, es decir, desarrollar un software para la traducci´on autom´atica que evite a lo m´aximo las alteraciones del sig- nificado en la traducci´on. Esta ta- rea requiere de toda la fuerza de la ciencia del PLN. En muchos casos es indispensable que el programa entienda el texto lo suficientemente bien para poder razonar sobre ´el. Con justa raz´on, la traducci´on au- tom´atica desde el mismo comienzo del PLN fue su principal motiva- ci´on, y fuente de inspiraci´on y retos.

“Mujer azteca hablando” (supe- rior) y “Malinche traduciendo” (inferior), detalles del *Co´digo Florentino*, libro 12, cap´ıtulo 18

(1580).

# Entender el lenguaje ajeno es la paz: Traducci´on autom´atica

Parafraseando la c´elebre frase: el entender el lenguaje ajeno es la paz. Los individuos, como las nacio- nes y los pueblos, se unen gracias a su lenguaje comu´n (como es el caso

da. El traductor de Google, [www.google.com.mx/language](http://www.google.com.mx/language) tools?hl=es, nos permite sin ayu- da externa leer las p´aginas de In- ternet en chino, ´arabe, ruso y mu- chas otras lenguas, sin mencionar el ingl´es. Sin embargo, mientras que el texto producido por tales traductores es muy u´til y sirve de gran ayuda, es todav´ıa muy mejora- ble. Estos sistemas son actualmente deficientes en dos aspectos princi- pales.

Primero, la calidad del texto que producen. En muchas ocasiones parece haber sido escrito por un extranjero que no habla bien el espan˜ol, y en otras de plano nos reprobar´ıan en la primaria si es-

A pesar de dichas dificultades vale la pena seguir trabajando en esta tarea, pues una vez resueltos los problemas t´ecnicos, viviremos en un mundo sin fronteras lingu¨´ısti- cas, sin limitaciones que se nos im- ponen por no hablar el ingl´es (o el chino, o el espan˜ol) y sin tanta di- visi´on cultural y social derivada de estas limitaciones.

Para hablar con un vecino del continente que no hable nuestro idioma, simplemente prenderemos el celular que se encargar´a de traducir lo que le estamos diciendo y de traducirnos tambi´en su respuesta.

# Era inform´atica para todos: Interfaces humano-computadora

Vivimos en una era inform´atica. En una era de libre acceso a la informaci´on. En una era de trabajo intelectual eficiente por ser asistido por la computadora.

Quiero decir, vivo yo, mis colegas ingenieros, mis estudiantes y seguramente usted, querido lector. Pero

¡qu´e poquitos somos quienes vivimos en la era inform´ati- ca! Cuando hablo con algunos de mis conocidos m´edicos, abogados, mu´sicos, historiadores, choferes, obreros, es- cucho “pues . . . la computadora . . . y estas cosas . . . ¡no soy bueno en esto!” No es cierto. Son buenos. La que es mala es la computadora.

Las computadoras fueron creadas para resolver nues- tros problemas y no para crearnos m´as problemas (como la necesidad de aprender inform´atica). Deben ser nues- tras ayudantes naturales, f´aciles de usar. Deben aprender nuestro lenguaje y no obligarnos a aprender el suyo.

Los robots ya son f´ısicamente capaces de ser nuestros sirvientes y ayudantes en tareas cotidianas. Segu´n el go- bierno de Corea del Sur, cada familia coreana en el an˜o 2020 tendr´a un robot ayudante en la casa [1], tal como en siglos pasados era comu´n tener sirvientes. Bill Gates, el l´ıder de Microsoft, dice tambi´en que habr´a un robot en cada hogar [2].

Pero para que un robot se convierta en un verdadero ayudante de casa tiene que entender nuestro lenguaje. Esto significar´a la era inform´atica para todos, no s´olo para los ingenieros.

Entre muchos problemas t´ecnicos en este camino mencionar´e aqu´ı cuatro. Ninguno de ellos es inherente a la tarea de las interfaces humano-computadora, pero son aqu´ı m´as evidentes y los retos que presentan son m´as dif´ıciles que en otras tareas.

El primero es el procesamiento de habla. Varias ve- ces dije que los programas de PLN procesan, clasifican, analizan el texto. Pero no debe ser estrictamente as´ı. No hablamos en texto, hablamos en voz. Para lograr una in- terfaz eficiente, las m´aquinas deben entender el lenguaje hablado (aunque internamente lo transformen a texto para analizarlo).

El segundo es la conducci´on del di´alogo, el cual pre- senta retos distintos de los de un texto normal (mon´olo- go). Por ejemplo, en el di´alogo se usan mucho las ora- ciones incompletas o hasta recortadas a una sola palabra (como “aj´a”, “pues”). Adem´as, hay ciertas reglas de con- ducta en cuanto al cambio de los turnos: ¿cu´ando dejo

de escuchar y empiezo a hablar? ¿Cu´anto puedo hablar sin ser interrumpido?

El tercer problema es la generaci´on de lenguaje: ha- blar o escribir a diferencia de escuchar o leer; componer a diferencia de analizar. ¡Cu´antas veces tenemos mucho que decir y lo queremos decir todo a la vez! Pero eso no se puede; hay que decidir cu´al parte vamos a expresar en la primera oraci´on y cu´al en la segunda (y peor au´n, dividir la idea grande en pedacitos de taman˜o de oraci´on), cu´al palabra va primero y cu´al luego; con qu´e palabra se ex- presa la misma idea en diferentes contextos. En espan˜ol, por ejemplo, dar atenci´on se dice “prestar”, dar una cla- se se dice “impartir”, dar una carta se dice “entregar”, dar una enfermedad se dice “contagiar”.

Finalmente, el cuarto problema es relacionar las pa- labras con las acciones, objetos y circunstancias en la conversaci´on. Un robot ayudante debe poder reaccionar adecuadamente a frases como “ve all´a y tr´aeme aquello”, relacionando el objeto y la direcci´on con el movimiento del dedo del usuario.

Igual como en el caso de otras aplicaciones, mientras los investigadores nos est´an acercando a lo que hoy se ve como ciencia ficci´on, existen actualmente aplicaciones pr´acticas y factibles de esta tecnolog´ıa. Una aplicaci´on pr´actica de las interfaces humano-computadora son las interfaces con las bases de datos. Normalmente las pre- guntas au´n bastante sencillas, como ¿qu´e porcentaje de los alumnos del tercer semestre reprobaron dos mate- rias?, implican programaci´on en un lenguaje especiali- zado de consulta a bases de datos llamado SQL. Mu- cho esfuerzo se ha dedicado durante d´ecadas a que las m´aquinas puedan directamente entender las preguntas en su forma natural, proporcionando as´ı el acceso a la informaci´on a los usuarios comunes sin la necesidad de un programador intermediario.

Un ejemplo de la aplicaci´on pr´actica del reconoci- miento de habla son los sistemas de dictado, los cuales permiten que se dicten textos (como este art´ıculo) con un micr´ofono en lugar de escribirlos con el teclado. La mi- niaturizaci´on de los sistemas electr´onicos aumentar´a la importancia de la comunicaci´on en voz: ser´a la u´nica (y muy natural) manera de interactuar con un reloj de pul- sera inteligente.

Como un ejemplo de los sistemas de di´alogo se pue- de mencionar los sistemas de venta de boletos de tren o avi´on por tel´efono, capaces de conducir un di´alogo simple sobre las preferencias de viaje del usuario.

# Por PNL se entiende la habilidad de una m´aquina para procesar la informaci´on comunicada, no s´olo las letras o los sonidos del lenguaje.

Y mucho, mucho m´as . . .

Adem´as de los tres grupos de aplicaciones ya men- cionados (el manejo del conocimiento, la traducci´on au- tom´atica y las interfaces humano-computadora), el PLN constituye la parte crucial de diversos tipos de sistemas relacionados con el uso de lenguaje humano. Mencione- mos aqu´ı s´olo algunos.

Los sistemas de soporte para la composici´on de tex- tos proporcionan ayuda al usuario para escribir docu- mentos: formatean el texto usando guiones; verifican la ortograf´ıa, la gram´atica y el estilo; completan las pala- bras o frases que empieza a escribir el usuario (muy u´til en los celulares); proporcionan traducciones, sin´onimos y explicaciones de las palabras o sugieren palabras segu´n su descripci´on [3]. Pueden variar en complejidad desde muy simples (tales como la divisi´on de las palabras con guiones) hasta muy complejos, por ejemplo, la verifica- ci´on l´ogica y factual del texto (en la frase “al salir de Francia, Juan visit´o su capital Londres” un buen pro- grama encontrar´ıa un error l´ogico y un error factual).

Las aplicaciones del PLN en la educaci´on incluyen la evaluaci´on automatizada de las respuestas o compo- siciones de los estudiantes en cuanto al estilo, lenguaje o exactitud. En la educaci´on asistida por computadora los m´etodos del PLN ayudan a componer los cursos y a proporcionar al estudiante la informaci´on requerida.

En la medicina, particularmente u´tiles son las apli- caciones de miner´ıa de texto y bu´squeda en las historias cl´ınicas de los pacientes, adem´as de los sistemas especia- lizados de bu´squeda y miner´ıa de texto para los m´edicos. Debido a la enorme cantidad de datos experimentales reportados, por ejemplo, en la investigaci´on de la inter- acci´on de los genes y las prote´ınas, resulta necesario el procesamiento autom´atico de tales publicaciones ya que una persona ya no puede leer ni siquiera las m´as relevan- tes para su trabajo.

La lingu¨´ıstica forense aplica los m´etodos lingu¨´ısti- cos, y sobre todo computacionales, en las investigaciones criminal´ısticas y de peritaje. Estos m´etodos incluyen la identificaci´on de la autor´ıa de los textos o bu´squeda de los fragmentos sospechosos en los mensajes o conversaciones grabadas. Dos ´areas muy afines a la lingu¨´ıstica forense son la identificaci´on de plagio (tanto en obras literarias o publicaciones cient´ıficas como en las composiciones de los estudiantes) y la esteganograf´ıa lingu¨´ıstica (los m´eto- dos para ocultar mensajes secretos en textos o habla y los m´etodos para detectar tales mensajes ocultos).

Las ideas y t´ecnicas desarrolladas originalmente pa- ra el an´alisis del lenguaje resultan aplicables en ´areas muy lejanas del lenguaje humano. Un ejemplo obvio es la teor´ıa de compiladores y los lenguajes de programaci´on, cuya creciente complejidad los aproxima cada vez m´as a los lenguajes humanos. Perl es un ejemplo de un len- guaje computacional que fue intencionalmente disen˜ado para aprovechar algunos rasgos de los lenguajes huma-

nos, tales como la ambigu¨edad, dado que su autor es lingu¨ista.

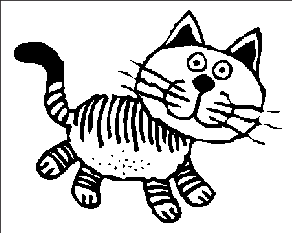
La gen´omica y la biolog´ıa molecular comparten mu- chas ideas y m´etodos con el PLN, ya que en ambos casos se trata de la codificaci´on de la informaci´on compleja en una cadena de s´ımbolos, la cual en el caso de la gen´omi- ca es la mol´ecula de DNA, RNA o las mol´eculas de las prote´ınas. Por razones similares, los m´etodos de PLN se emplean en el an´alisis y la generaci´on autom´atica de mu´sica: las estructuras repetitivas musicales se describen bien con las as´ı llamadas gram´aticas formales desarrolla- das originalmente para la descripci´on de los fen´omenos lingu¨´ısticos.

. . . . . . . . . . . . . . . . . . . . . . . .

# ¿Qu´e gato tiene Juan?

Juan usa un gato para reparar su coche.

¿Qu´e gato?



Textos de entrenamiento

|  |  |
| --- | --- |
| *Pedro usa un martillo para*  *Ana usa un desarmador para*  *el obrero usa una gru´a para* | *el gato come ratones*  *el perro come la carne*  *el ha´mster come avena* |
| alguien usa ´estos para  algo | ´estos comen algo |

El gato de Juan ha de ser m´as parecido a un martillo, un desarmador o una gru´a que a un perro o un h´amster.

Diccionario monolingu¨e

Martillo: una herramienta que ... Desarmador: una herramienta que... Gru´a: una herramienta que...

Gato 1: un animal dom´estico peludo. Gato 2: una herramienta que...

De las dos acepciones de gato, la segunda es la que m´as se parece a martillo, desarmador o gru´a. ¡Ya sabemos cu´al gato!

Diccionario bilingu¨e

Gato (1) cat

(2) jack

Ahora podemos traducir: John uses a jack to repair his car.

. . . . . . . . . . . . . . . . . . . . . . . .

# Por d´onde continuar . . .

No es el prop´osito de esta introducci´on corta el ex- plicar al lector los pormenores t´ecnicos, sino m´as bien despertar su inter´es por las tecnolog´ıas del PLN. Ahora bien, suponiendo que logr´e este prop´osito, s´olo me que- da decir d´onde el lector podr´a encontrar a los expertos del ´area. Si es usted un directivo o un empresario y en- contr´o en este art´ıculo algo que le puede servir, quiz´as

se pregunte qui´en le puede dar el servicio, y si es usted estudiante (tal vez potencial, pues nunca es tarde pa- ra estudiar), quiz´as se pregunte en d´onde puede obtener m´as informaci´on.

Al final de esta contribuci´on se proporcionan enla- ces a portales de asociaciones profesionales y a material disponible en l´ınea. ¡Espero que les sean de utilidad!✵

INFORMACIO´N ADICIONAL

El lector interesado puede encontrar m´as informaci´on y v´ınculos a las fuentes y los eventos relevantes en la p´agina de la AMPLN, la Asociaci´on Mexicana para el Procesamiento de Lenguaje Natural: [www.AMPLN.org.](http://www.AMPLN.org/) La comunidad nacional del PLN organiza anualmente dos congresos con ponencias en espan˜ol: el Coloquio de Lingu¨´ıstica Computacional en la UNAM y el Taller de Tecnolog´ıas del Lenguaje Humano organizado por el INAOE. Adem´as, el IPN organiza anualmente el congreso internacional CICLing: [www.CICLing.org,](http://www.CICLing.org/) aunque no siempre en M´exico.

Para la lectura inicial se recomiendan los libros [4-6] disponibles desde la p´agina [www.Gelbukh.com,](http://www.Gelbukh.com/) donde se puede tambi´en encontrar muchos art´ıculos cient´ıficos sobre el tema y otros materiales relevantes.

REFERENCIAS

1. “A Robot in Every Home by 2020, South Korea Says”, *National Geographic*, news.nationalgeographic.com/news/ 2006/09/060906-robots.html, visitado el 11 de febrero de 2010.
2. Gates B. (2007) “A Robot in Every Home”, *Scientific American*, [www.scientificamerican.com/article.cfm?id=a-](http://www.scientificamerican.com/article.cfm?id=a-) robot-in-every-home, visitado el 11 de febrero de 2010.
3. Sierra G. (2001) “Bu´squeda de palabras a partir de las definiciones en los diccionarios de lengua automatizados”,

*Actas de* 70 *Simposio Internacional de Comunicacio´n Social*, 2, Santiago de Cuba.

1. Bolshakov I.A., Gelbukh A. (2004) *Computational linguistics: models, resources, applications*, IPN–UNAM-Fondo de Cultura Econ´omica.
2. Gelbukh A., Sidorov G. (2010) *Procesamiento automa´tico del espan˜ol con enfoque en recursos l´exicos grandes*, Segunda edici´on, ampliada y revisada. IPN.
3. Galicia Haro S.N., Gelbukh A. (2007) *Investigaciones en ana´lisis sinta´ctico para el espan˜ol*, IPN.

SOBRE EL AUTOR

Alexander Gelbukh es maestro en ciencias con especialidad en matem´aticas y doctor en ciencias de la computaci´on. Desde 1997 es jefe del Laboratorio de Procesamiento de Len- guaje Natural del Centro de Investigaci´on en Computaci´on (CIC) del Instituto Polit´ecnico Nacional (IPN). Es miembro de la Academia Mexicana de Ciencias, Investigador Nacional de M´exico con nivel II, y secretario de la Mesa Directiva de la Sociedad Mexicana de Inte- ligencia Artificial (SMIA).



Es autor, coautor o editor de m´as de 400 publicaciones, y coautor de tres libros en las ´areas del Procesamiento de Lenguaje Natural e Inteligencia Artificial.