

# Infopragmática: un método eficiente para la recuperación de información\*

Rafael Ibarra,\*\* Silvia Ballesteros\*\*\*

## RESUMEN

Basada en un algoritmo lingüístico, que se apoya en un teorema de la incertidumbre, la infopragmática es un nuevo método que ofrece una solución eficiente, aunque no limitada, a aquellos usuarios de habla hispana que intentan conseguir la información más útil de bases de datos académicas cuyo contenido está en idioma inglés. Presenta un análisis de búsqueda, considerando una tabla de niveles de comprensión; así como breves consideraciones sobre la ambigüedad del término “relevancia” y algunas razones estadísticas para brindar la infopragmática a la comunidad de usuarios del Sistema Bibliotecario de la Universidad Nacional.

**PALABRAS CLAVE:** Infopragmática, recuperación de información (RI), algoritmo lingüístico, teorema de incertidumbre, pragmática, lingüística, hablantes del español.

## Abstract

Based on a linguistic algorithm, supported by an uncertainty theorem, the infopragmatics is a new method that offers an efficient solution, but not limited, to those Spanish speaking users who try to get the most useful information from academic databases which contents is in English. Presents a search analysis, an application of the language of levels understanding table, brief considerations on the ambiguity of the term “relevance” and statistical reasons to put infopragmatics into action at our National University Library System.

**KEYWORDS:** Infopragmatics, information retrieval (IR), linguistic algorithm, uncertainty theorem, pragmatics, linguistics, Spanish speakers.

\* La versión original en inglés fue presentada en la Qualitative and Quantitative methos in libraries, International Conference, llevada a cabo en Chania, Creta, Grecia, del 26 al 29 de mayo de 2009; puede ser consultada en: [http://www.isast.org/proceedings/QQML2009/PAPERS\\_PDF/Ibarra\\_Ballesteros-Infopragmatics-an\\_efficient\\_method\\_for\\_information\\_retrieval\\_PAPER-QQML2009.pdf](http://www.isast.org/proceedings/QQML2009/PAPERS_PDF/Ibarra_Ballesteros-Infopragmatics-an_efficient_method_for_information_retrieval_PAPER-QQML2009.pdf).

\*\* Maestro en Lingüística Aplicada; DGB-UNAM; edificio anexo de la DGB, Circuito de la Investigación s/n, Ciudad Universitaria, 04510 México, D.F. Email: rafael@dgb.unam.mx.

\*\*\* Maestra en Ciencia e Ingeniería de la Computación; DGB-UNAM; edificio anexo de la DGB, Circuito de la Investigación s/n, Ciudad Universitaria, 04510 México, D.F. Email: silviabe@dgb.unam.mx.



## Introducción

La recuperación de información es un desafío inacabado que enfrentan los bibliotecarios de hoy para dar satisfacción a sus usuarios con la información pertinente. Hay varias razones que hacen este problema insoluble, entre otras: interfaces caóticas de los sistemas de información; categorías múltiples de usuarios y la comparación inevitable entre el vocabulario controlado y el lenguaje natural. Estos tres aspectos críticos y los cimientos subyacentes conducen a los usuarios, en general, a experimentar una fase de incertidumbre que los abruma e inhibe para obtener la información que requieren de los sistemas de información (Ssi), también conocidos como bases de datos (BD).

Aun más, aunque existen varias alternativas para la búsqueda y recuperación de información ofrecidas por los proveedores de Ssi, de incontables contribuciones teóricas y prácticas de investigadores de diversas áreas del conocimiento, las demandas de los usuarios de un método conveniente para satisfacer sus necesidades de información no han disminuido con la suficiencia deseada: todavía están buscando un método apropiado. La infoprágmatika se ofrece para satisfacer la demanda de un método eficiente que consiste en tres herramientas lingüísticas: un algoritmo lingüístico, un teorema de incertidumbre y una lluvia lingüística (lingstorm), que resultan en un acercamiento eficaz para obtener una solución no-robótica eficiente para satisfacer las demandas de información de usuarios, especialmente para, pero no limitado a, los nativos hablantes del español (NHE).

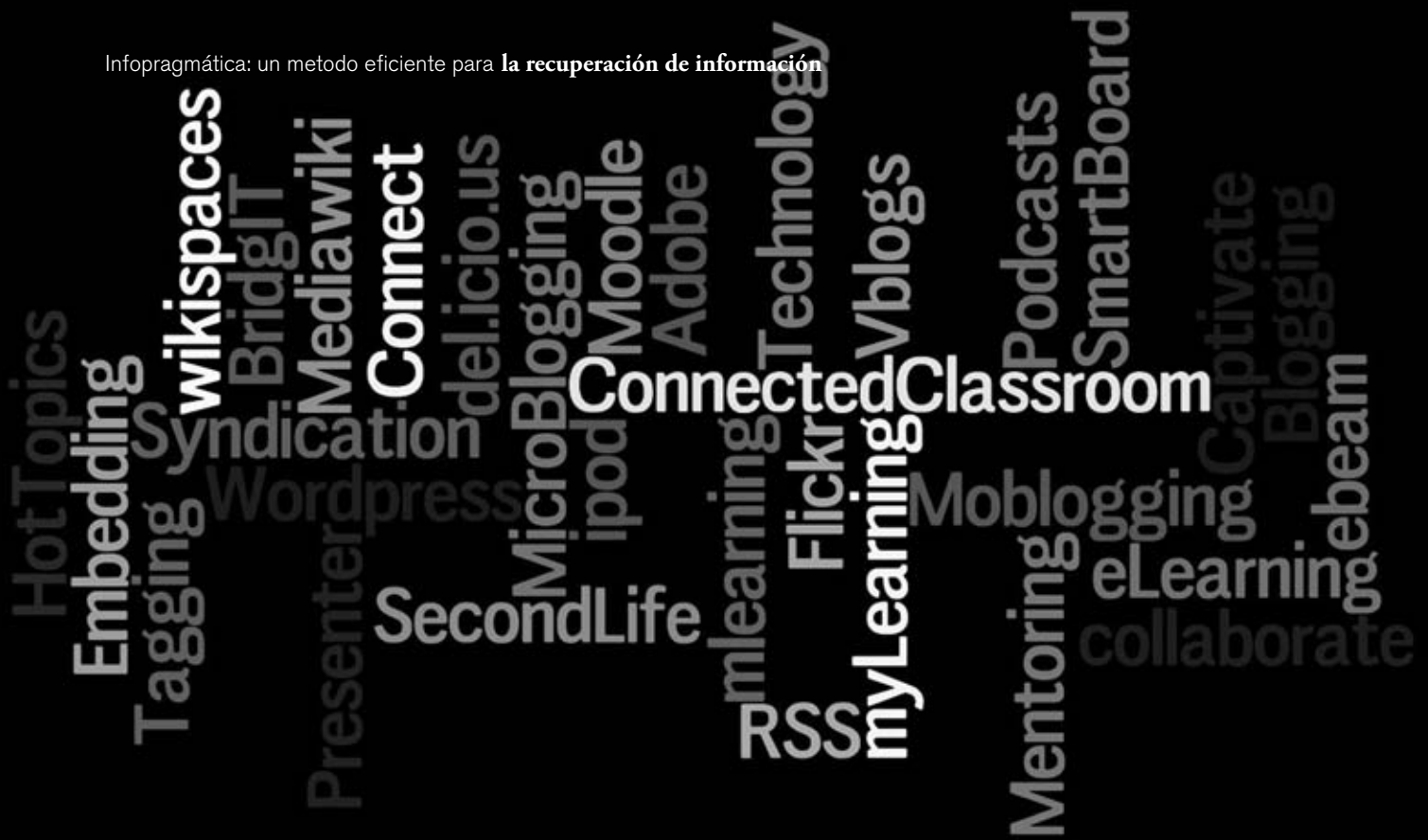
Antes de entrar a profundidad, se debe mencionar que aunque existe una variedad de Ssi que proporciona

datos en varios formatos: texto, imagen, audio, video, y sus combinaciones posibles, en este documento se considerarán solamente a las compañías comerciales de los Ssi de carácter académico, tales como las ofrecidas por ProQuest, Ebsco y Elsevier, por su elevado uso entre la comunidad involucrada.

## Sistemas de información

Los Ssi comerciales mencionados anteriormente presentan páginas de búsqueda con interfases caóticas y una tecla de ayuda de acceso rápido (F1) poco efectiva, que demuestra claramente su ineficiencia para los numerosos y diversos tipos de usuarios, pero especialmente para un usuario principiante cuya lengua nativa no es la inglesa. Por ejemplo, las interfaces de los Ssi presentan más de 25 elementos que no son esenciales para un usuario principiante. Algunos de tales elementos son: *pago por ver, alertas, journals preferidos, vínculos rápidos, vínculos externos, mis preferencias, número de páginas, noticias de portada, búsqueda inteligente, búsqueda dentro del texto completo de los artículos, nuevas características*, y varios otros elementos que están definitivamente fuera del interés del usuario principiante. El resto de los elementos - diez - es más apropiado para usuarios con experiencia intermedia o expertos.

Según el último informe publicado, las tres bases de datos más consultadas en la UNAM en 2006 fueron: Academic Search Premier, con 25,723 consultas; Elsevier Science, con 7,753 consultas; y PsychINFO, con 5,472 consultas. Después de analizar las interfaces de cada uno de estos Ssi, se descubrió que son adecuadas para un bibliotecario o un usuario certificado por



las compañías proveedoras, al igual que para un robot que lee y selecciona listas de verificación presentadas en un instante. No obstante, tales interfaces no son fáciles de utilizar para un bibliotecario o un usuario estándar cuya lengua sea española o alguna otra lengua diferente de la inglesa. La falta de uniformidad y las complejas interfaces de los Ssl inhiben, en su uso, a varios usuarios y éstos deciden realizar sus consultas en motores de búsqueda populares como Google Scholar, que ofrece una interfaz muy conveniente de no más de nueve elementos en idioma inglés y cinco en idioma español; interfaz que es un componente clave que se abordará en la siguiente sección.

### **La experiencia de NHE con los Ssl para su uso en idioma inglés**

El Sistema Bibliotecario de la Universidad Nacional Autónoma de México (SB UNAM) ofrece acceso a sus usuarios a más de 198 Ssl relacionados con sus campos de estudio y el 91% de ellos están disponibles

para su uso en lengua inglesa. Sus servicios se ofrecen en 140 bibliotecas, pero el análisis presentado y referido en este trabajo ocurrió en la Biblioteca Central (BC), escenario en el cual la interacción entre un bibliotecario y un usuario tiene dos características: la comunicación es oral y en idioma español. Sin embargo, la interacción entre los seres humanos y los Ssl debe ser escrita y en inglés, lo cual establece ya una serie de problemas probables: mala ortografía, afijación incorrecta, uso de lenguaje natural, en contraste con el vocabulario controlado y un caótico sinnúmero de herramientas de los Ssl que vienen dentro de ellas.

Cada año más de 9,000 estudiantes de maestría y más de 37,000 nuevos alumnos se inscriben como estudiantes en la UNAM y son usuarios potenciales de los Ssl. Lo anterior nos conduce a observar que, a pesar de que muchos bibliotecarios ocupan sus puestos de trabajo durante varios años, y se ofrecen cursos anuales relacionados con el uso de los Ssl, miles de usuarios son novatos en el uso de los Ssl.

## Tipos de usuarios y de su idioma

Los usuarios son únicos en sus ideas, pensamientos y necesidades, así que la manera en la cual expresan sus demandas de información es tan diferente como puedan ser. Por otro lado, los usuarios de habla española, estudiantes o bibliotecarios, no son de un sólo tipo, e internacionalmente comparten diversas características que se pueden resumir en las siguientes categorías, aunque parciales, referidas por Stubinz y Whighly<sup>1</sup>: Pip, el impaciente; Odiseo, el tenaz; Ismael, el exploratorio; Hamlet, el confuso; Ophelia, la trastornada y Don Quijote, el idealista. Aunque ésta es una variedad limitada de usuarios, se puede apreciar que, además de las características psicológicas o emocionales humanas, los usuarios representativos de esta lista hablan diversos idiomas: inglés, griego, hebreo, danés y español.

Con este panorama, no es difícil imaginarse los efectos que la fusión entre estas clases de usuarios, diversos idiomas e interfaces caóticas de los Ssl comerciales puede causar. Hoy en día es evidente que los investigadores y profesores de todos los niveles cuentan con una gran cantidad de estrategias, con modificaciones numerosas e intereses para conocer, manejar y mejorar las maneras y los canales de los servicios para el usuario. Katsirikou y Skiadas<sup>2</sup>, enumeran 23 acciones del difícil proceso de búsqueda de información desde la apertura hasta el cierre del diálogo en una petición de información, que parte desde encontrar el recurso electrónico apropiado y llega hasta el hecho en el cual el usuario proporciona, indirecta e involuntariamente, datos personales a los proveedores de información, sin importar en qué idioma se plantean las peticiones.

## Vocabulario controlado y lenguaje natural

Uno de los más antiguos antecedentes del vocabulario controlado y del lenguaje natural es el conflicto entre

<sup>1</sup> STUBINZ J., WHIGHLY S. *Information Retrieval System Design for Very High Effectiveness*. [en línea]. <<http://goanna.cs.rmit.edu.au/~jz/sci/p3.pdf>>

<sup>2</sup> KATSIRIKOU, A., SKIADAS, C. H. Chaos in the Library Environment. *Library Management...*, 2001.

dos grupos: los *anomalistas* y los *analogistas*. El primer grupo sostenía que la lengua humana se debe expresar y vivir de la manera misma en la cual se articula; y el segundo grupo afirmaba que los seres humanos debían expresar sus ideas de manera apropiada, formal. Después de un tiempo, los grupos se convirtieron en *descriptivos* y *prescriptivos*, respectivamente. Este conflicto no ha acabado todavía, tiene lugar en las bibliotecas modernas, cuando los usuarios están utilizando su lenguaje natural para buscar información descrita con vocabulario controlado para redactar su investigación. Este tipo de usuarios no está consciente de que las grietas semánticas que existen entre los Ssl y ellos son en la mayoría de los casos, la razón por la cual, después de una búsqueda de información infructuosa, ellos terminan con la sensación de frustración y de incertidumbre. Sin lugar a dudas, ellos preferirían sentirse seguros a considerarse perplejos si supieran cómo. Una opción, en este caso, sería manejar el teorema de incertidumbre presentado en este trabajo.<sup>3</sup>

## La infopragmática

La infopragmática es un método eficiente para obtener información pertinente de los Ssl. Está diseñada desde una perspectiva lingüística y dirigida a aquellos usuarios cuyo idioma no es el inglés, poniendo énfasis, aunque no de manera limitada, en los NHE. La infopragmática puede ser de gran ayuda para los académicos que no pueden conseguir información pertinente de los Ssl comerciales, tales como los proporcionados por Elsevier, Ebsco, ProQuest y demás similares; consiste en un teorema de incertidumbre, un algoritmo lingüístico y una lluvia lingüística (lingstorm).

## El teorema de incertidumbre

Después de que los usuarios no pueden obtener la información que necesitan de un sistema de información, ya sean asistidos por algún bibliotecario o mediante el

<sup>3</sup> IBARRA CONTRERAS, Rafael. Algunas grietas semánticas en la recuperación de información: una perspectiva deconstructiva para una solución pragmática..., 2007.

uso de las guías rápidas ofrecidas por los proveedores de Ssl, se preguntan el por qué, y no hay oportunidad de darles un indicio para revelar cuáles pudieron haber sido las razones, puesto que tuvieron la ocasión de leer registros *relevantes* desplegados obtenidos de un sistema de información académico especializado. Como ejemplo, tenemos que Jaffe<sup>4</sup> indica que después de que se aplicara una prueba a usuarios de Ssl sin más ayuda que la que proporciona el proveedor mismo, se sintieron frustrados con las combinaciones booleanas: *"They sought the quick "hits" possible in an index-controlled database, and they were frequently confused when non-relevant citations appeared in a Boolean search, even though all the search terms were represented in the citation. After long reading sessions users got tired and fall into a state of uncertainty."* Ibarra<sup>5</sup> presentó un trabajo de acercamiento al problema de incertidumbre considerando diversas teorías lingüísticas desde una perspectiva etnolingüística y varios análisis basados en 120 búsquedas de información reales con resultados similares.

Hasta ahora, los Ssl (SI en singular) no ofrecen ninguna ayuda directa para mantener a los usuarios lejos de la perplejidad resultante, boquete que se espera cubrir con el siguiente teorema de incertidumbre, como una orientación eficiente para imbuir confianza en los usuarios para alcanzar sus metas informativas:

1. Si no hay información en el SI es debido a:
  - El usuario carece de capacidades léxicas
  - La sintaxis es incorrecta.
  - Los términos usados no son equivalentes en significado.
  - El contexto no es el apropiado.
  - No hay nada en la base de datos usada.
2. Si la información no es adecuada es debido a:
  - El lenguaje técnico es inadecuado.
  - La ortografía es incorrecta.
  - Las palabras clave o los descriptores no son tales.

Para reforzar las consideraciones anteriores, es conveniente apoyarse en las seis posibilidades lingüísticas diversas y sintetizadas de Liddy<sup>6</sup>, que hacen que los NHE no logren alcanzar sus metas de información en los siguientes niveles de comprensión: *Morfológico*, mala ortografía; *Léxico*, expresión de sus necesidades; *Sintáctico*, reglas gramaticales; *Semántico*, oposición entre el lenguaje natural y el vocabulario controlado; *Discursivo*, la manera en la cual el bibliotecario y el usuario interactúan al ofrecer y recibir información del SI; y, *Pragmático*, el conjunto de estrategias que permiten a los usuarios/bibliotecarios satisfacer sus necesidades de información.

La siguiente figura representa una petición de información sobre *Bolsa Mexicana de Valores*:

**Looking for information on "Bolsa Mexicana de Valores":  
Mexican bag of values vs. Mexican stock exchanges**

Adj + N	MEXICAN STOCK EXCHANGE
N + Adj	INDICADOR BURSÁTIL
N	FINANZAS
N + Prep + N	
N + Prep + N	ÍNDICE DE COTIZACIONES
N + Prep + N N + Prep + N N + Prep + N	BOLSA DE VOLARES BALSA DE VALORES BOLSA DE VALORES

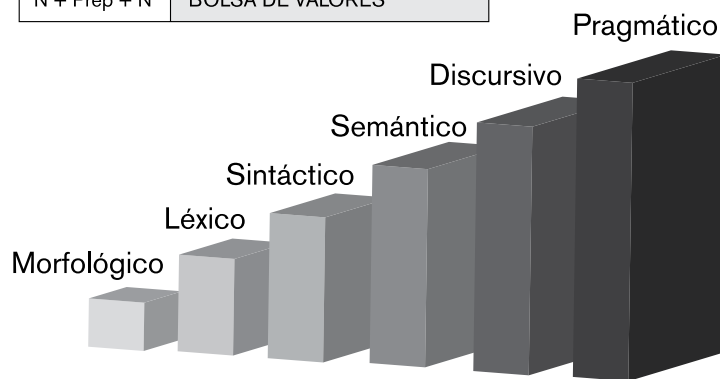


Fig. 1. Niveles de comprensión de idioma.

<sup>4</sup> JAFFE, J. G. For Undergraduates: InfoTrac MAGAZINE INDEX Plus or WILSONDISC with Reader's Guide and Humanities Index?...,1988.

<sup>5</sup> IBARRA CONTRERAS, Rafael. *Aprovechamiento y optimización de los recursos tecnológicos en la búsqueda y recuperación de información en CD-ROM basados en estrategias lingüísticas...*, 1999.

<sup>6</sup> LIDDY, Elizabeth D. Whither Come the Words? [en línea]. En: *CENDI Subject Analysis and Retrieval Working Group Conference: Controlled Vocabulary and the Internet*. (2007). <[www.cendi.gov/presentations/liddy.PPT](http://www.cendi.gov/presentations/liddy.PPT)>

En este ejemplo es evidente el mal uso del idioma inglés, pero no es posible negar que el *bag* significa *bolsa*, *mexican*, *mexicana* y *values*, *valores*; ni tampoco en esta otra posibilidad: *sac de valeurs vs bourse*.

Aunque esta solicitud de información fue tomada de varios años atrás, se ingresaron los mismos términos erróneos en diversos Ssl el 19 de abril de 2009. Se descubrió que Academic Search Premier, de Ebsco, y Science Direct, de Elsevier, presentaron como resultados relevantes: 365,540 y 1,475 registros respectivamente.

Es común que los usuarios sigan un recorrido crítico: no cuentan con disponibilidad léxica de vocabulario controlado; carecen de diccionarios técnicos, índices, encabezamientos de materia o tesauros, además de que tienen numerosas faltas de ortografía. El segundo paso es enfrentar las caóticas interfaces de los Ssl y la ayuda poco eficiente que constituye la tecla F1. El tercer paso es leer los registros desplegados y después seleccionar e imprimir o enviar por correo electrónico los registros no deseados, para justificar, en la mayoría de los casos, el uso de los Ssl.

Cada uno de los errores previamente identificados puede causar una variedad de problemas no sólo para nativos hablantes del español, sino para todos esos nativos cuya lengua no sea la inglesa.

Oportunamente, hay otra herramienta que identifica el recorrido crítico y sugiere una ruta ideal, además de indicar los pasos que se deben tomar para garantizar una solución práctica que satisfaga la demanda urgente de los usuarios para obtener registros pertinentes de los Ssl (Sl en singular).

### Relevancia

Un hecho llamativo y *oculto* que, de cierta manera, contribuye a la incertidumbre de los usuarios es el término *relevancia*. Por una parte, los resultados presentados por el

Sl se ofrecen considerando la relevancia que representan desde el Sl, lo cual significa, desde el punto de vista de la ingeniería en computación, en varios casos, el número de veces que aparece un término en un documento.

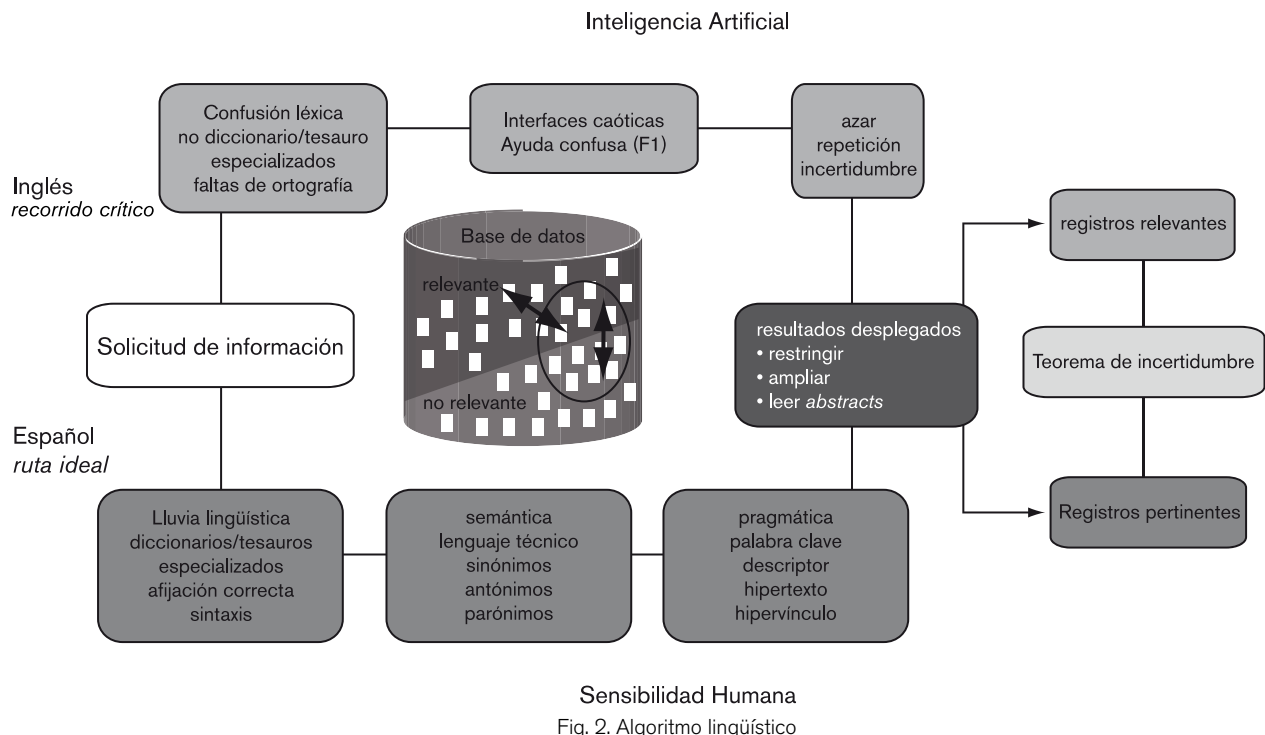
Por otra parte, al entrar a las definiciones de sentido, en el escenario de las ciencias de la información y la lingüística, se ha llegado a una serie de tareas que ya han cerrado un círculo completo en investigación, pues se vuelven a emplear métodos empíricos y análisis basados en recopilación que caracterizan algunas de las tentativas más tempranas para solucionar el problema de la ambigüedad. Entre los numerosos esfuerzos que se han hecho, Ide y Véronis<sup>7</sup> observan:

*"Al-base methods, symbolic methods, connectionist methods, knowledge based-methods, machine readable dictionaries, computational lexicons, corpus based methods (automatic sense-tagging and overcoming data sparseness)."* Sin embargo, el término relevancia usado por los proveedores de Ssl para presentar los resultados de sus sistemas de recuperación de información enmascaran, de manera involuntaria, la acepción del término relevancia como equivalente a *repetición* y *anáfora*, de modo que los usuarios tienden a seguir el recorrido crítico con resultados infructuosos, como se muestra en los análisis mostrados en el apéndice de este trabajo.

### Algoritmo lingüístico

El algoritmo lingüístico (figura 2) es la ilustración de una serie de operaciones lingüísticas para resolver los problemas intrínsecos a la RI. Puede ser utilizado como mapa guía para que los usuarios observen el par de rutas intrínsecas y, alternativamente, consideren el uso de acciones rápidas para satisfacer sus necesidades de información, en oposición al recorrido crítico cuyos resultados se discutieron previamente.

<sup>7</sup> IDE, N., VÉRONIS, J. Introduction to the Special Issue on Word Sense Disambiguation: the state of the art.



El primer elemento de la ruta ideal abarca una acción rápida que requiere el uso de fuentes lingüísticas, ya sean impresas o en versiones electrónicas, para asegurar los pasos siguientes.

El diccionario y el tesoro especializados permiten que el usuario diseñe una "lluvia lingüística" (fig. 3), a ser utilizada de manera similar a la que funciona como una lluvia de ideas en un grupo de trabajo o un ejercicio individual de creatividad, una constante a lo largo de la RI.

La lluvia lingüística es necesaria para crear un puente entre el lenguaje natural y el vocabulario controlado; la afijación correcta descifra las diferencias entre verbos, sustantivos y adverbios que causan, comúnmente, grietas semánticas cuando los usuarios cambian del idioma español al inglés; la sintaxis determina los resultados en frases sustantivas; los sinónimos amplían los horizontes semánticos e incluyen variedades de ortografía al ocuparse de caracteres no latinos; aunque los antónimos y los homófonos ocurren durante la interacción verbal, evitan una afasia temporal pro-

longada; las palabras clave y los descriptores se encuentran fácilmente a lo largo de la lectura del registro y se pueden recuperar para establecer una estrategia alternativa de búsqueda; el hipertexto y los enlaces de hipertexto son opciones que se vuelven atractivas a seguir tan pronto como aparezcan en los registros de información y continuar la armonía semántica establecida.

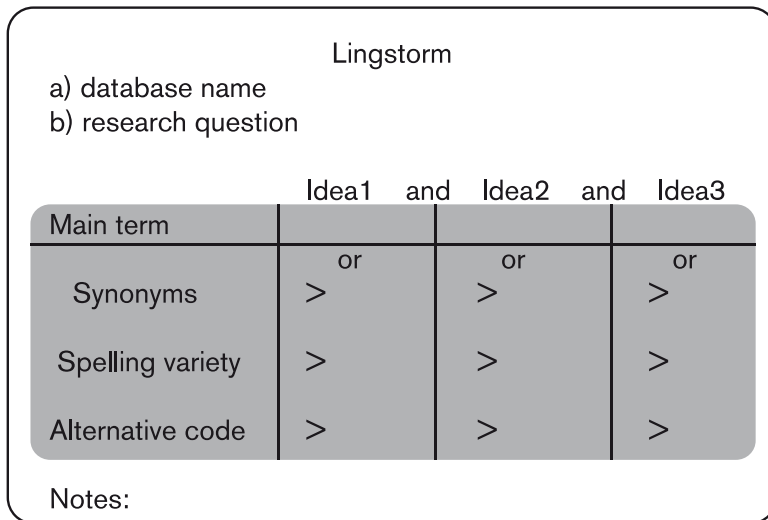


Fig. 3 Lluvia lingüística.

### Razones estadísticas e implementación técnica de la infopragmática en el Sistema Bibliotecario de la Universidad Nacional Autónoma de México

La puesta en práctica de la infopragmática en el portal o portales de las diversas bibliotecas del Sistema Bibliotecario de la UNAM considera la observación del número de usuarios, de las colecciones y de dos escenarios: a distancia y en sitio. Como se mencionó anteriormente, la UNAM recibió a 46,000 estudiantes de pregrado y de posgrado (usuarios potenciales de los Ssl); y, según los últimos datos estadísticos, en el Sistema Bibliotecario de la UNAM, hasta abril de 2009, existen cerca de nueve millones de documentos, de material bibliográfico.<sup>8</sup>

Específicamente, los 150 Ssl disponibles en idioma inglés comprenden el acceso a varios millones de registros. Por otro lado, el promedio semanal de consultas de información al sitio web de la Biblioteca Digital es de 20,000.

Aunque se desconoce el número exacto de aquellos usuarios que no reciben asesoría personal de algún bibliotecario, y los horarios regulares en los cuales ellos utilizan los Ssl, se puede asumir que una gran cantidad de usuarios estarían interesados en recibir algún curso/taller sobre la infopragmática.

En vista de los puntos anteriores, además del nivel y tipos de usuarios, se está diseñando un par de talleres:

*in situ* y a distancia. Por el momento hay una guía rápida “recupera información pertinente en inglés” en el sitio web de la Biblioteca Digital, que está disponible en <http://132.248.9.9:8084/infopragmatica/>.

Cabe mencionar que la guía rápida incluye enlaces a traductores electrónicos, a diccionarios y tesauros especializados cuya calidad fue evaluada y, posteriormente, fueron incluidos en la colección de recursos gratuitos de la Biblioteca Digital de la UNAM. Asimismo, existe un enlace que permite a los usuarios expresar sus opiniones sobre la utilización de la guía rápida en la parte inferior de la página. Con los resultados obtenidos, que se almacenan automáticamente en una base de datos, el diseño será mejorado.

### Conclusiones

La RI abarca una gran variedad de problemas que se pueden resolver con soluciones apropiadas, es decir, si hay discrepancias en el uso de la lengua las soluciones se deben delimitar de manera lingüística. En este documento se identificaron algunas dificultades lingüísticas comunes que afectan a los nativos hablantes del español así como las propuestas de solución correspondientes. La infopragmática es un método ágil que ofrece un agudo dinamismo que permite a los usuarios de Ssl conseguir información pertinente, en oposición a la información relevante que resulta de un recorrido crítico. El algoritmo lingüístico representa una herramienta cuyo diseño permite su adaptación teórica a la mayoría de las lenguas modernas, la inglesa incluida, para servir como dispositivo productivo de información. ☞

<sup>8</sup> Universidad Nacional Autónoma de México. Agenda Estadística 2009 [en línea]. <<http://www.planeacion.unam.mx/Agenda/2009/>>





## Obras Consultadas

IBARRA CONTRERAS, Rafael. Algunas grietas semánticas en la recuperación de información: una perspectiva deconstructiva para una solución pragmática. En: Simposio Internacional sobre Organización del Conocimiento Bibliotecología y Terminología (2007 : México, D. F.). *Memorias del I Simposio Internacional sobre Organización del Conocimiento: Bibliotecología y Terminología*, 27 al 29 de agosto de 2007, México: UNAM, Centro Universitario de Investigaciones Bibliotecológicas, 2007, p. 305-320.

----- *Aprovechamiento y optimización de los recursos tecnológicos en la búsqueda y recuperación de información en CD-ROM basados en estrategias lingüísticas*. México: el autor, 1999. 171 h. (Tesis Maestría: UNAM, Facultad de Filosofía y Letras)

IDE, N., VÉRONIS, J. Introduction to the Special Issue on Word Sense Disambiguation: the state of the art. *Computational Linguistics*, 1998, vol. 24, no. 1, p. 1-40.

JAFFE, J. G. For Undergraduates: InfoTrac MAGAZINE INDEX Plus or WILSONDISC with Reader's Guide and Humanities Index? *American Libraries*, 1988, vol. 19, no. 9, p. 759-61.

KATSIRIKOU, A., SKIADAS, C. H. Chaos in the Library Environment. *Library Management*, 2001, vol. 22, no. 6/7, p. 278-287.

LIDDY, Elizabeth D. Whither Come the Words? [en línea]. En: *CENDI Subject Analysis and Retrieval Working Group Conference: Controlled Vocabulary and the Internet*. (2007). <[www.cendi.gov/presentations/liddy.PPT](http://www.cendi.gov/presentations/liddy.PPT)> [Consulta: 5 julio 2007].

STUBINZ J., WHIGHLY, S. *Information Retrieval System Design for Very High Effectiveness*. [en línea]. <<http://goanna.cs.rmit.edu.au/~jz/sci/p3.pdf>> [Consulta: 7 abril 2008].

## Apéndice

Los análisis descritos aquí se dividen en dos grupos: el primero muestra la transición entre el lenguaje natural y el vocabulario controlado y, el segundo, algunos intentos empleados por los usuarios para satisfacer sus necesidades de información.

Petición Original: <i>Bolsa Mexicana de Valores</i>		PROQUEST ABI/Inform records	EBSCO Academic Research Premier records	Elsevier Science Direct records
Lenguaje Natural	Mexican Bag of Values	4	1	1,475
Vocabulario controlado	Mexican Stock Exchange	610	164	2,727

Tabla 1, grupo 1.

Petición Original: <i>Odontología</i>		PROQUEST Cambridge Medline records	EBSCO Academic Research Premier records	Elsevier Science Direct records
Lenguaje Natural	Odontology	1,893	1,394	13,563
Vocabulario controlado	Dentistry	134,846	29,404	108,143

Tabla 2, grupo 1.

Petición Original: <i>Rana verde</i>		PROQUEST Biological Sciences record	EBSCO Academic Research Premier records	Elsevier Science Direct records
Lenguaje Natural	Frog Green	11	1	19,574
Lenguaje Natural	Green Frog	309	64	19,574
Vocabulario controlado	Lithobates clamitans	1	52	13
Vocabulario controlado	Rana Clamitans	262	40	550

Tabla 3, grupo 1.

Petición Original: <i>Protección Radiológica</i>		Syntax and parts of the speech Noun + Adj.	PROQUEST Cambridge Medline records	EBSCO Health Source: Nursing/ Academic Edition records	Elsevier Science Direct records
Primer intento	Radiology Protection	Noun + Noun	1,396	1	14,645
Segundo intento	Radiologic Protection	Adj. + Noun	5	15,519	4,027
Tercer intento	Radiological Protection	Adj. + Noun	1,336	67	16,152

Tabla 4, grupo 2.