

COMUNICACIONES

1 - PROCESAMIENTO DEL LENGUAJE Y GESTIÓN DOCUMENTAL

ANÁLISIS Y REPRESENTACIÓN DEL TEXTO EN DOCUMENTACIÓN: APUNTES PARA EL CAMBIO

Dr. Antonio García Gutiérrez

Profesor titular
Facultad de Ciencias de la Información
Universidad Complutense de Madrid

Resumen

El objetivo de la presente comunicación no es tanto presentar los frentes prioritarios de la investigación documentológica en los campos de análisis y lenguaje como hacerlos conectar con la problemática y avances constatables en áreas como la traducción automática, la lexicografía o la creación y gestión de bases de conocimientos en sistemas expertos.

Comencemos a exponer los temas fundamentales de investigación experimental a partir del último eslabón que debe culminar el proceso informativo: el usuario.

La necesidad de información. Modelos de demanda

En cualquier sistema de información debe modelizarse el usuario, claro es que la manera de hacerlo, especulativa o empírica, ya tiene equilibrados sus defensores y detractores. Si nos situamos en el enfoque más pragmático de la Documentología, creación de métodos para generar procesos informativos mediante análisis y lenguajes documentales, la figura del usuario modelo es insoslayable. Ahora bien, el usuario nos interesa en cuanto conjunto de elementos reales probables.

De un concepto tan difuso como "necesidad de información" se parte en investigación documentológica estando ligada la credibilidad y universalidad de los resultados al nivel de claridad que se ha dado o forzado al definir los conceptos de demanda, usuario, satisfacción, utilidad etc. Parece evidente que estas definiciones generales no son muy útiles en la investigación práctica y, por tanto, es necesario usar condiciones y parametrizar los corpus observados para poder operar lo que, de un lado, beneficia al investigador y, de otro, perjudica la conexión, fiabilidad y generalidad de investigaciones efectuadas sobre el mismo objeto.

No obstante, es posible trabajar bajo unos principios generales que dan consistencia a la disciplina. En cuanto al usuario modelo o modelo de demandas parece claro que en los sistemas convencionales se tienen presentes en función de bases temáticas solamente ignorándose las bases modales.

En efecto, el usuario se somete al sistema y no viceversa, es decir, los sistemas son planificados en función del campo de conocimiento y, hasta ahora, con excesivos recortes impuestos por la tecnología. Vayamos al planteamiento opuesto. Un usuario no desea la información que le suministra una base de datos convencional asentada sobre el método de la coincidencia formal de significantes o no quiere encorsetar su demanda en el reduccionismo de respuesta que presta un sistema experto obsesionado en los enunciados simples y declarativos. Abramos, pues, el horizonte de sus necesidades, y ya que no es posible universalizar, ciñámonos a una aplicación concreta como el discurso periodístico.

El usuario de una base de datos de actualidad desea, básicamente, información de actualidad en todas sus formas de manifestación, por lo tanto, es necesario averiguar cómo es posible solicitar información de esta naturaleza. Si un ser humano es la base de datos hipotética a la que nos referimos, seguramente la información convencional que ofrecen las existentes sería la menos procurada. Sin

embargo, análisis comparativos, verificaciones, inferencias, todo o partes de textos o datos factuales, entre otros, parecen ser focos de esas preguntas. En consecuencia, un inventario riguroso de demandas es esencial para investigar en documentación aplicada.

Documentólogos y documentalistas

El documentólogo es el agente que observa, describe y modeliza procesos de información. El documentalista quien los construye y mediatiza. En el estado actual de la Documentación, en general, y periodística en particular, la propuesta de nuevos métodos documentológicos no tendría interlocutores en la práctica debido a dos razones: los profesionales desconocen el nuevo rol que deben asumir y la tecnología, por el momento, sólo permite cristalizar avances parciales, por lo que las empresas no arriesgan inversiones en equipos efímeros.

El documentalista no solamente va a ser un transmisor que lee y sintetiza información sino un reestructurador de la misma. La tecnología debe ir haciendo admisible nuevas estructuras de información para facilitar nuevos modos de respuesta.

Entre ambos elementos del proceso situamos al documentólogo como un epistemólogo práctico, en palabras de Gardin, que se debate entre las estructuras de conocimiento y sus posibles representaciones. El desarrollo en este aspecto, es muy poco fructífero y escaso actualmente (los mayores avances documentológicos se siguen imponiendo desde los propios avances tecnológicos) pero, en cualquier caso, ya están promoviendo una conciencia diferente primero entre estudiantes de Documentación, doctorandos e investigadores y, a continuación, entre los mismos documentalistas para poder vislumbrar un cambio trascendental en la disciplina dentro de este decenio.

La aproximación pluridisciplinar

El equipo, con el que trabajo, se ha empeñado en la aproximación pluridisciplinar y, de ahí, transdisciplinar de la investigación documentológica. Estamos convencidos de que para abordar problemas cognitivos lo más sensato es situarse en las disciplinas que los estudian de manera estable y, realizando rigurosos recortes en función de nuestros fines, aprovechar el aparato teórico-conceptual y metodológico que proponen para la resolución de los casos.

La experiencia primera ha sido fructífera pero agotadora. Son demasiadas las disciplinas y muchos los autores que tratan el conocimiento y su representación y, por ello, es difícil extraer y aplicar todos los enfoques y métodos encontrados. Sin embargo, algunos de los logros obtenidos en las investigaciones que hemos realizado ya están siendo utilizados como principios investigadores y docentes. Sin extenderme más en este sentido, citaré algunas de las disciplinas que más han enriquecido, por el momento, nuestras contribuciones: Lingüística en general y Semántica y Sintaxis en particular, Lógica formal, Análisis de Contenido, Análisis del Discurso o Teoría de la Comunicación. Señalo otras prometedoras: Inteligencia Artificial, Traductología, Psicología cognitiva y Ciencias cognitivas en general.

Un nuevo modelo documentológico

Tras servirnos durante un largo periodo de modelos de investigación prestados con adaptaciones (como el estructuralista) o pocos convincentes (como el operativo propuesto por la propia empresa) hemos conseguido centrar las investigaciones en torno a unos principios generales de trabajo que son los siguientes:

*** Lectura-Transformación-Representación-Recuperación**

Estos cuatro elementos básicos constituyen los pilares del nuevo modelo permitiendo una labor coherente con el discurso más general de nuestra disciplina, integradora en cuanto a los resultados y a la previsión de aplicaciones y extrapolaciones.

* Experimentalidad y verificación

El núcleo de nuestro trabajo es experimental. Partimos de corpus contruídos, los observamos, proponemos metodologías bien de análisis, bien de construcción de vocabulario y las llevamos a la práctica. Ahora bien, no puede haber resultados fiables sin verificación.

En consecuencia, una de las prioridades de nuestras investigaciones es la de obtener métodos de validación, por el momento demasiado miopes, demasiado pobres, que den cuenta de los resultados y permitan algunas generalizaciones. También, para ello, debemos acudir a otras disciplinas que ya han hecho propuestas sobre sus respectivos corpus, en concreto, el análisis de contenido y del discurso.

La verificación afecta a todo el proceso investigador en nuestra disciplina: los corpus no están bien definidos, no disponemos de una formulación clara de sus componentes y de los elementos mínimos que las constituyen, no sabemos cómo alcanzar la representatividad de las muestras ni cómo testar los resultados satisfactoriamente. Esta es una de las mayores lagunas que tenemos presente.

Algunas experiencias emprendidas

Bajo el mencionado paraguas multidisciplinar que se vuelve transdisciplinar, conforme avanzamos, en el sentido de atravesar y verse atravesado por la Documentación y en función de nuestro modelo genérico, hemos realizado algunas experiencias sobre dos campos de observación: la generación de métodos de lectura pragmática y la propuesta de nuevos métodos de construcción de vocabulario. Veamos algún caso:

Partimos de la base de que la lectura en Documentación es rutinaria y masiva. Además, el documentalista no lee por placer sino atendiendo a unos fines prefijados y a unos criterios no explicitados. El objetivo, en este aspecto, ha sido localizar algunos métodos de análisis textual que nos permitan esquematizar el raciocinio del autor/productor y nos ofrezcan resultados similares de lectura, es decir, se persigue una mayor objetivación de la comprensión del texto mediante la lectura simulada. Este objetivo se fundamenta en las siguientes conjeturas e hipótesis: la ideología y la cultura del lector influyen excesivamente en los resultados de la comprensión lo que perjudica el sistema de recuperación. El texto propone una información que puede ser detectada por encima de esos filtros y que responde a la denominación de macroestructura, por lo tanto, la lectura documental debe ser objetivada. Si la lectura es objetivable (en reglas), entonces es programable.

Para ello, hemos estudiado las propuestas de algunos lingüistas y analistas del discurso como Pottier, Greimas, Van Dijk. En el texto periodístico funcionan algunos de estos métodos que no son válidos en otros corpus pero aún presentan muchas restricciones, por ejemplo, si bien la gramática de casos responde a cómo produce el texto el periodista tras interrogar la realidad y los hechos o el modelo de Van Dijk nos da la clave de la estructura repetitiva del texto en prensa, existen dentro de nuestra misma aplicación un gran número de variables que impiden la extrapolación fuera de las muestras: extensión, profundidad, número de enunciados o párrafos, géneros o estilos.

Pero a pesar de lo incipiente de este tipo de investigación comenzamos a entender la lectura pragmática de otra forma. Ya hemos abandonado, por ejemplo en el nivel docente, las prácticas de interpretación textual basadas en instrumentos referentes, en la intuición o en el sentido común. Sin embargo, necesitamos conectar estos pequeños logros con los obtenidos en el nivel de representación (que desgraciadamente ha invadido y cegado, hasta ahora, la visión global del modelo) y además, conseguir equipos que permitan mecanizar las experiencias una vez maduras. La mecanización ayudaría, también, a puntualizar los métodos de validación.

Mucha información sobre nuestra aproximación en investigación documentológica se nos escapa en el marco de una comunicación. Sólo cabe desear un acercamiento entre las disciplinas, ricas y pobres, para solucionar los problemas hasta donde convergen, ya largo recorrido conjunto, y, a partir de ahí, comenzar a construir este espacio común de las disciplinas que se ocupan de practicar con el Conocimiento.

