

GUILLERMO DE JORGE-BOTANA, INVESTIGADOR DE LA UNED

## “La tecnología que hemos desarrollado analiza textos como lo haría una persona”

Para procesar el volumen de llamadas telefónicas de las grandes compañías, las centralitas cuentan con un sistema automatizado de atención al cliente que, con frecuencia exaspera al usuario. Guillermo de Jorge-Botana, investigador de Psicología Evolutiva y de la Educación de la UNED, junto a otro experto de la Universidad Autónoma de Madrid, desarrolla tecnologías semánticas basadas en la comprensión humana. Estas herramientas ayudan a solucionar esos problemas telefónicos, además de evaluar exámenes de forma automática o incluso a medir el nivel de plagio entre textos.

Laura Chaparro |  
divulgaUNED

20/11/2012 10:03 CEST



El investigador en Psicología, Guillermo de Jorge. Imagen: UNED.

**Los modelos que desarrolláis están basados en la comprensión humana.  
¿En qué consisten?**

Son modelos que parten de la psicología cognitiva. De lo que se trata es de analizar los textos tal y como lo haría una persona, integrando lo que está explícito en un documento con el conocimiento previo. En psicología cognitiva se dice que lo leído en el texto se incorpora con los contenidos de la memoria a largo plazo mediante los mecanismos de la memoria operativa.

**Y este principio lo aplicáis a Gallito, uno de vuestros programas. ¿Cómo funciona?**

Gallito es como una navaja suiza, hace muchas cosas. Primero convierte las palabras de los textos en números. A partir de aquí, se pueden realizar diferentes análisis. Lo más simple es comprobar cómo las palabras se parecen unas a otras, aunque también se pueden llevar a cabo tareas más complejas, como resumir un texto en sus ideas principales, medir la cantidad de información que tienen las palabras en los contextos en los que aparecen, hacer listados de las que más se parecen... Además, en el ámbito académico, Gallito preevalúa las respuestas de los exámenes siguiendo varios criterios: si la respuesta se corresponde con lo que se pregunta, si el texto guarda coherencia entre párrafos y frases que lo componen o si esa respuesta está plagada de otras redactadas por otros alumnos.

**¿Por qué se llama Gallito?**

Un amigo me comentó que una vez le preguntaron a Valle Inclán sobre algo de música y que él respondió: "yo solo sé de música lo que es el 'pasodoble de gallito' y lo que no es el 'pasodoble de gallito'". Y de ahí, el nombre. De todas formas, viene a cuento porque lo que hacen los gallos y las gallinas es picar el grano, y nuestro programa picotea el texto hasta que lo procesa.

**Además de Gallito, tenéis un categorizador semántico, del que este programa forma parte. ¿En qué consiste?**

Sí, es un sistema que usa los componentes procesados por Gallito para categorizar llamadas telefónicas o correos electrónicos. El objetivo final es que, cuando tengas que hacer una llamada a un sistema de voz automatizado, puedas usar tu propio lenguaje sin necesidad de repetir palabras de un menú o pulsar teclas. El sistema en cuestión tiene el reto de

reconocer lo que se ha dicho y digitalizarlo. Después, tiene que 'enrutarlo', que significa categorizarlo en un tema. Lo que hemos hecho nosotros es insertar la tecnología semántica de Gallito en el proceso de enrutamiento completo de una llamada.

### ¿Y por qué abundan tanto los sistemas de '[pulse 1](#)'?

Porque es lo más sencillo. El problema de decir frases a la máquina es el ruido ambiental y que las oraciones se pueden expresar de una forma muy variable. Por eso siempre va a haber una tasa de error, que se puede ir controlando a posteriori.

---

"El problema de decir frases a la máquina es el ruido ambiental y que las oraciones se pueden expresar de forma variable"

### ¿Qué ventajas tiene vuestro programa?

Las tecnologías que hemos desarrollado son más flexibles y también más económicas. Estamos poniendo en el mercado un producto barato y que es muy flexible para todo tipo de sistemas.

### ¿Por qué son caras las otras herramientas?

Porque suelen ser propietarias, es decir, tienes que pagar los derechos a grandes compañías. Además, no suelen ser tecnologías que usen todo con XML, un formato estándar.

### ¿Quiénes están mostrando interés por vuestras aplicaciones: investigadores o empresas?

Respecto al enrutador, se han mostrado interesados proveedores de tecnología que quieren de abaratar costes en los servicios de telecomunicaciones que ofrecen a sus clientes. En cuanto a Gallito, vale para todos pero está teniendo muy buena acogida sobre todo entre los investigadores porque sirve para sus experimentos. El programa controla las

palabras en algunas dimensiones. Por ejemplo, si quieres abstractas o concretas, polisémicas o monosémicas, la similitud que tienen unas con otras... Tiene utilidad en este campo pero también se puede utilizar para minería de datos textual.

### **¿Para buscadores?**

Sí, es más o menos la misma tecnología. La policía podría utilizar esta herramienta para procesar atestados. En grandes textos de ese tipo se pueden encontrar algunos patrones léxicos comunes y hacer una especie de estudio exploratorio para complementar con otros estudios estadísticos que buscan patrones en la conducta de los delincuentes.

---

"La policía podría utilizar esta herramienta para procesar atestados"

### **¿Esos datos se les escaparían de otra forma?**

Sí, porque son, por así decirlo, rasgos latentes, que no están a la vista pero que de repente aparecen. Por ejemplo, estás buscando en un determinado tipo de delito un rasgo que se produce en un momento dado: un incendio y que hay quema de rastrojos. Observas lo que tiene que ver con quema de rastrojos y te aparecen muchos términos, frases... Así vas acotando el corpus de investigación.

### **¿Tiene similitudes con las herramientas SEO?**

En cierta manera, sí. Al final la filosofía es la misma, porque SEO y SEM se basan en encontrar palabras claves que tengan mucha audiencia. Estas estrategias se podrían llegar a desarrollar con nuestros programas, porque hay índices numéricos en Gallito que dicen qué palabras aportan más información que otras.

### **A la hora de comercializarlo, te estás ayudando de Internet. ¿Es útil?**

Sin duda. Nosotros tenemos dos blogs. Uno es del grupo de interés ([E](#)

[semántico](#)), y el otro se llama [The contact center](#).

Derechos: **Creative Commons**

TAGS

TEXTOS | LLAMADAS | PLAGIO | BUSCADORES | CATEGORIZADOR |  
GALLITO | PALABRAS | UNED | DIVULGAUNED |

### Creative Commons 4.0

Puedes copiar, difundir y transformar los contenidos de SINC. [Lee las condiciones de nuestra licencia](#)