¿Y por qué nadie me lo dijo antes?

Fabricio Gaviña

Licenciado en Bibliotecología y Documentación · Argentina

https://orcid.org/0009-0007-1776-791X

gavinafabricio@gmail.com

Resumen

Resumen

Un desarrollador de un ministerio de educación provincial construye un chatbot sobre décadas de normativa y descubre que las dificultades no son solo técnicas, sino también terminológicas. Al intentar resolverlas, reconstruye prácticas de control del vocabulario que los bibliotecarios y archivistas del propio organismo ya conocen. El artículo analiza cómo la falta de diálogo entre saberes profesionales puede limitar proyectos tecnológicos y desaprovechar conocimientos institucionales existentes.

Palabras clave: control de vocabulario; tesauros; chatbots; organización del conocimiento; administración pública; sistemas de organización del conocimiento

Abstract

A developer at a provincial ministry of education builds a chatbot using decades of regulations, only to discover that technical solutions depend on how vocabulary is organized. After struggling with ambiguities and inconsistent terminology, he unknowingly recreates vocabulary control practices already familiar to the ministry's librarians and archivists. The article explores how the lack of dialogue between professional fields can limit technological projects and leave existing institutional knowledge untapped.**Keywords:** vocabulary control; thesauri; chatbots; knowledge organization; public administration

Keywords: vocabulary control; thesauri; chatbots; knowledge organization; public administration; knowledge organization systems

La demo que salió perfecta

Juan tiene cuarenta y un años, un buen sueldo para lo que es el país y un problema que todavía no vio venir. Programa en el Ministerio de Educación de una provincia, que acumula décadas de resoluciones, decretos, expedientes y circulares escaneados en carpetas que nadie abre. Le pidieron un chatbot, porque todos los ministerios están armando uno y el de la provincia vecina ya lo

presentó en un acto. Que el docente escriba lo que necesita en una ventana de chat y el asistente le conteste con lo que corresponde. Juan sabe lo que hace, así que lo armó con lo mejor que tenía a mano, un modelo de lenguaje, un índice de documentos y un par de semanas de trabajo prolijo.

La demo salió perfecta. Escribió «inscripción a concursos de titularización» y el asistente contestó con la resolución exacta, citada y todo. Escribió «reglamento de ingreso a la docencia» y contestó con un resumen impecable del reglamento. El ministro asintió con la cara de quien entendió dos tercios de lo que vio y está seguro del tercio que le falta, y se fijó fecha de lanzamiento.

El lanzamiento tiene su auge. Durante los primeros tres meses el chatbot se usa muchísimo, hay fotos en las redes del ministerio y una nota en el diario provincial. Después el uso empieza a caer, mes a mes, y al sexto casi no lo abre nadie. El informe dice poco, salvo que la gente probó, preguntó dos o tres cosas y no volvió. Los que sí se quejan se quejan de lo mismo, que la respuesta es una ensalada de textos, un poco de una resolución, un poco de otra, un párrafo de un decreto que ya no rige, todo dicho con un aplomo envidiable.

Lee las conversaciones y no entiende. Hay una pregunta que se repite cuarenta veces por semana y siempre termina igual, el docente la reformula tres o cuatro veces y se rinde. Mira la pregunta y no le ve nada raro. Mira los documentos y están todos ahí. El sistema tiene todo lo necesario y no anda.

Su primera hipótesis es la de casi todos, que falta potencia. Reindexa todo, prueba con un modelo más grande, suma documentos que había dejado afuera. Se le va una semana y bastante plata del ministerio, y las quejas son las mismas.

La hipótesis tiene su lógica. Una revisión mexicana sobre chatbots en la administración pública encontró que buena parte de los asistentes oficiales se limita a menús de opciones y usa poca inteligencia artificial (Ramírez-Hernández et al., 2023), de modo que el remedio parece obvio, más modelo y menos menú. Ya hizo exactamente eso. Su asistente no tiene un solo menú, entiende el castellano mejor que la mayoría de sus usuarios y contesta de corrido. Y sigue sin andar. Lo que cuenta la revisión será cierto, pero a Juan le falta otra cosa.

Primer tropiezo

La pregunta que se repetía cuarenta veces dice «San Martín».

Se pone a pensar qué estará buscando la gente. En el ministerio hay una escuela San Martín, un instituto, una localidad, un departamento y, claro, el prócer, que aparece en cada resolución del calendario escolar y de los actos patrios, más un par de parajes con ese nombre que figuran en algún convenio. El chatbot elige una, la que le parece más probable, y contesta como si fuera la única. Una docente que quería los datos de contacto de la escuela San Martín de su zona recibió un resumen muy prolijo de una resolución de 1998 sobre el acto del 17 de agosto.

Hace lo que haría cualquiera con dos ojos y una tarde libre. Le agrega a cada documento un campo que diga de qué tipo de cosa se habla (persona, lugar, institución, acto), y ahora el asistente pregunta «¿qué San Martín?» antes de contestar. Es un parche, y bastante bueno. Sin darse cuenta acaba de inventar el control de homónimos.

Segundo tropiezo

La segunda familia de preguntas es más difícil de ver, y es la que más me interesa. Son las que el asistente contesta mal sin que nadie lo note. No hay un error ni una lista vacía ni un reclamo. El chatbot contesta siempre, y quien pregunta se va con algo que suena razonable, o se va con las manos vacías convencido de que lo que buscaba no existe.

Un sistema de este tipo tiene cuatro clases de usuarios. Los que encuentran lo que buscaban, los que no lo encuentran y lo saben, los que nunca se enteran de que estaba ahí, y los que se llevan una respuesta que suena bien y está mal. Del primer grupo el informe de uso cuenta todo, del segundo algo y de los otros dos nada. Son usuarios fantasma, y para el sistema, estadísticamente, no existen. En recuperación de información hay una medida para el tercer grupo, la exhaustividad, que dice cuánto de lo pertinente logra traer un sistema. Tiene una trampa conocida. Para calcularla hay que saber cuánto había para recuperar, y eso es justo lo que nadie sabe cuando el sistema falla sin hacer ruido.

Una docente escribe «reemplazo» y la resolución habla de «suplencia». Otro escribe «acumulación» y el reglamento habla de «acrecentamiento», según la década. Alguien pregunta por la «licencia por enfermedad» y los papeles dicen «inasistencia justificada por razones de salud». Una preceptora pide el «certificado de trabajo» y el trámite, en el sistema, se llama «certificación de servicios». Juan, que mide lo que puede medir, mira el informe y concluye que el chatbot dio curiosidad y después aburrió. Se equivoca, y el ministro también.

Los modelos de lenguaje ayudan con esto, claro, entienden bastante bien que reemplazo y suplencia andan cerca. Lo que no pueden saber es cómo se llama cada cosa acá, cuál de los dos términos usa la dirección que resuelve el trámite, ni qué resolución reemplazó a cuál y con qué nombre. Un modelo no tiene por qué saber cuál es la denominación vigente y autorizada en este ministerio, y que ande dando vueltas en algún documento no alcanza para que sepa cuál manda. Eso vive en la cabeza de tres empleados que se jubilan en dos años, y cuando falta, el chatbot rellena igual con lo que encuentra.

Con un modelo de lenguaje es fácil creer que la vieja discusión entre el lenguaje natural de quien pregunta y el lenguaje controlado de quien describe quedó resuelta, que ahora el sistema entiende cualquier cosa que se le escriba. Esa discusión lleva décadas abierta, y Mario Barité (2014) sostiene que no hace falta cerrarla porque las dos formas de acceso pueden convivir. Los buscadores, de

hecho, incorporaron mecanismos de control a su manera. El autocompletado y el «quizás quiso decir» aprovechan vocabularios construidos, en parte, a partir de lo que la gente consulta. Lo que ningún sistema trae de fábrica es el vocabulario particular de un ministerio.

Arma entonces una tabla. En una columna, la palabra que usa la gente. En la otra, la que usan los papeles. Empieza con veinte filas, a los dos meses tiene trescientas y a los tres descubre que las equivalencias no van en los dos sentidos. «Profesor» lleva a «docente», pero «docente» no puede llevar solo a «profesor», porque también incluye al maestro y a cualquiera que tenga título para dictar clases de lo que sea. Hay palabras más amplias y palabras más específicas, y la tabla tiene que saber cuál es cuál. Le agrega una columna para eso y el archivo termina llamándose equivalencias_ahora_si.xlsx, como un mal chiste.

Tercer tropiezo

El tercero le cuesta más, porque ahí ya hay gente de por medio.

Abre el registro de conversaciones y lee lo que la gente escribe de verdad, que no se parece en nada a lo que había imaginado. Hay faltas de ortografía, hay apodos de organismos que no figuran en ningún organigrama («la junta», «el listado», «el puntaje»), hay nombres de trámites que ya no existen y se siguen pidiendo igual, como la «foja de servicios». Algunos escriben una frase entera, otros una sola palabra y se enojan porque el chatbot no adivinó.

Con el tiempo entiende que el problema no es el código. Está parado entre dos maneras de hablar. De un lado la de quien redactó la resolución, formal, cerrada, con palabras que nadie usa en la calle. Del otro la de quien pregunta, que escribe como habla. El sistema tiene que traducir de un lado al otro, y a Juan nadie le avisó que ese era el trabajo.

La salida parece obvia, elegir una palabra oficial para cada cosa y que todo converja ahí. Arma una lista de términos controlados y de términos que remiten a ellos, y empiezan los problemas que nadie le había anticipado. La Dirección de Nivel Primario dice «cargo», la de Secundaria dice «horas cátedra» y la de Nivel Superior dice «horas terciarias». Para Liquidaciones, una «licencia» es una cosa, y para Salud Laboral otra bien distinta. Cada dirección está convencida de que su palabra es la correcta, y tiene razón, para su dirección.

Entró al ministerio para programar y una tarde se descubre mediando cuarenta minutos de discusión sobre cómo se llama una cosa. Piensa que está perdiendo el tiempo. Está haciendo la parte más difícil del proyecto, a puro olfato y con una planilla.

Lo que Juan reinventó

Lo que fue haciendo se llama control de vocabulario, y Barité (2014) lo ordena en objetivos, y casi todos ya los cumplió sin saberlo. Eligió la palabra oficial, separó los San Martín, hizo converger

reemplazo y suplencia, y averiguó que un profesor es un docente pero un docente no es solo un profesor. Le falta uno, el de relacionar. El primer docente que busque «acumulación» va a necesitar, dos clics después, el régimen de incompatibilidades, y el sistema no sabe que esas dos cosas van juntas.

Hay además una parte que no se resuelve con técnica, y es la que explica los cuarenta minutos de discusión. Diego Ferreyra (2021) propone mirar los sistemas de organización del conocimiento como infraestructuras sociales de mediación simbólica y clasificarlos según la autoridad cognitiva que organiza a cada uno. Visto así, los cuarenta minutos que tomaba por tiempo perdido eran el trabajo mismo, porque decidir cómo se llama una cosa es decidir quién manda sobre esa palabra.

El día que alguien le puso nombre

Seis meses después del lanzamiento el chatbot contesta bastante mejor, aunque ya casi no lo usa nadie, y está agotado. Tiene una planilla de mil ochocientas filas que nadie más entiende, unas reglas escritas en un cuaderno y la sospecha de que el sistema depende de él para seguir andando. Si mañana se va del ministerio, el chatbot sigue un tiempo y después empieza a degradarse en silencio.

Una tarde baja a la cocina a prepararse un café y se encuentra con dos personas apoyadas en la mesada. Las conoce de vista. Él es bibliotecario del centro de documentación del ministerio y ella archivista del archivo general, dos oficinas del subsuelo que casi nadie visita. Están en plena conversación y no se interrumpen cuando llega. Llena la pava y se queda esperando que hierva.

Sin quererlo, escucha. Hablan de si «suplencia» va como término controlado y «reemplazo» como sinónimo que remite a ella, de si «interinato» es un sinónimo o un término relacionado, de qué nota de alcance hace falta para que nadie confunda «cargo» con «horas cátedra». Hablan rápido, sin explicar nada, cada uno completa la frase del otro como se hace entre quienes comparten un código. Reconoce de qué hablan, son sus propios problemas, los de la planilla, pero no tiene ninguna de las palabras.

La pava hierve y él sigue ahí, revolviendo un café que ya está listo. Piensa en irse. Meterse le da vergüenza. Nadie lo mira con mala cara, pero cada frase supone un diccionario que no tiene. Al final dice lo único que se le ocurre, «perdón, ¿eso es algo como una tabla de equivalencias?».

Ella sonríe y le dice «eso es un tesauro». Siente alivio y un poco de bronca, las dos cosas juntas. Ella no lo corrige, lo dice como un dato. Le pregunta si su tabla tiene una columna de términos amplios y términos específicos. Juan dice que sí. Si hay una lista de los términos controlados y de los que remiten a ellos. Dice que sí. Si registró de dónde sacó cada equivalencia y quién la aprobó. Dice que no, pero que debería.

Entonces el bibliotecario le cuenta que eso tiene un nombre, un método, una norma internacional y décadas de gente que lo hizo en bibliotecas, en archivos, en museos, en ministerios, y que enfrentó

cada uno de los tres tropiezos que Juan resolvió a mano y dejó escrito cómo evitar los que todavía no le habían pasado. Le explica que hay maneras de decidir cuál es el término controlado sin romper las relaciones entre direcciones, de dejar constancia de quién aprobó cada cosa y de mantener vivo el vocabulario cuando cambia la gente. Y le dice algo que Juan no esperaba, que esa tabla no solo ayuda a los que ya buscan, también hace aparecer a los que nunca llegaban, los fantasmas del informe, que dejan de irse con cara de no haber entendido nada.

Juan se queda callado un momento. Después pregunta. «¿Y por qué nadie me lo dijo antes?»

Los dos se miran, sin culpa y sin rencor, y ella deja la taza en la mesada. «Traé esa planilla», dice.

Por qué nadie se lo dijo

La primera explicación es la más piadosa. No estaba en ninguna conversación donde alguien pudiera decírselo. Su formación fue en programación y sistemas, y las comunidades donde los desarrolladores discuten cómo armar buenos asistentes hablan de vectores, de índices y de la última versión de algo, casi nunca de qué palabra se usa para nombrar qué cosa. Esa conversación nunca incluyó a nadie que supiera.

Tampoco es cosa solo de desarrolladores. Un análisis de repositorios universitarios brasileños encontró que casi todos ofrecían para buscar por tema una simple lista alfabética de términos, y que los tesauros eran la rara excepción, con la conclusión de las autoras de que convendría integrar las palabras clave en lenguaje natural con términos de vocabularios controlados (Fujita y Tolare, 2019). Y dentro del campo el diálogo tampoco está resuelto. Barité (2025) muestra que el intercambio entre la organización del conocimiento y la terminología, que estudia cómo cada comunidad nombra sus conceptos, sigue en un estado que él mismo llama embrionario, aun tratándose de disciplinas vecinas.

La segunda explicación es peor, y le toca a las organizaciones. El ministerio tiene un centro de documentación y un archivo general en el mismo edificio, con gente que lleva años haciendo con rigor lo que Juan armó a mano, y a nadie se le cruzó consultarlos antes de encargar el chatbot. Se contratan modelos, se pagan licencias, se organiza el acto de lanzamiento, y la pregunta de quién en la casa sabe de esto no figura en ningún pliego. Al desarrollador no le importaba el nombre del oficio. Le importaba que el docente encontrara el trámite. Pero tuvo que cruzarse con ellos por casualidad, esperando que se calentara el agua, porque en el organigrama nadie había pensado en juntarlos.

Qué le diríamos mañana

Si pudiéramos hablar con Juan antes del primer tropiezo, alcanzaría con decirle que su problema es de vocabulario antes que de tecnología, que no tiene que empezar de cero y que deje constancia de todo, de quién aprobó cada término, por qué se eligió ese y no otro, cuándo se cambió. Una planilla

que solo entiende su autor se degrada en silencio el día que el autor se va, y en un ministerio ese día suele coincidir con el cambio de gestión.

Juan hizo solo, con herramientas que no eran las suyas, un trabajo que podían haberle ahorrado a pocos metros de su escritorio. El conocimiento para hacerlo existe, tiene método, normas y décadas de experiencia acumulada, y está en el mismo edificio donde se escribe el código. A nadie se le ocurrió ir a buscarlo. Mientras tanto, cada ministerio lanza el suyo con el mismo entusiasmo y el mismo informe de uso a los seis meses.

Juan ya cerró su planilla y el ministro ya cortó la cinta. En el subsuelo, el centro de documentación sigue funcionando, y nadie se acuerda de que existe.

Referencias

Barité, Mario. (2014). El control de vocabulario en la era digital: revisión conceptual. *Scire*, 20(1), 99-108. https://www.ibersid.eu/ojs/index.php/scire/article/download/4196/3766/5347

Barité, Mario. (2025). Knowledge organization and terminology: intersections, interlocutions and projections. *Informatio*, 30(1), e209. https://doi.org/10.35643/Info.30.1.14

Ferreyra, Diego. (2021). Sistemas de organización del conocimiento: propuesta de un modelo unificado de definición. *Palabra Clave (La Plata)*, 11(1), e141. https://doi.org/10.24215/18539912e141

Fujita, Mariângela Spotti Lopes y Jessica Beatriz Tolare. (2019). Vocabulários controlados na representação e recuperação da informação em repositórios brasileiros. *Informação & Informação*, 24(2), 93-125. https://doi.org/10.5433/1981-8920.2019v24n2p93

Ramírez-Hernández, Pablo, Valle-Cruz, David y Mendoza-Méndez, Rafael Valentín. (2023). Revisión de los chatbots basados en inteligencia artificial en la administración pública: hacia una arquitectura para el gobierno. *Espacios Públicos*, (60), 29-49. https://doi.org/10.36677/espaciospublicos.v23i60.21317