

# **SOBRE LA IMPORTANCIA DEL PENSAMIENTO COMPUTACIONAL Y LA LITERATURA**

## **Introducción**

Tanto la literatura como la programación utilizan el lenguaje escrito. En el caso de la literatura, los escritores que poseen un vocabulario extenso cuentan con una percepción diferente de la realidad de la que tienen aquellos con un vocabulario poco desarrollado. Nuestro vocabulario afecta nuestra percepción y comprensión de la realidad, y esto se refleja en lo que escribimos. De igual forma, la profundidad de conocimientos de programación, con sus códigos y convenciones, también afecta la percepción y la imaginación de los individuos. Para poder imaginar nuevas formas de literatura digital es importante comprender qué es lo que la computadora hace, qué es lo que la computadora puede llegar a hacer y tener la habilidad para determinar lo que la computadora hará. La mayoría de las personas que utilizan una computadora para escribir saben qué es lo que la computadora hace. El uso de la computadora para estas personas está basado en procesos previamente diseñados, que no se pueden alterar. Los procesadores de palabras tienen un número específico de opciones que el escritor puede usar durante el proceso creativo. Por ejemplo, la interacción con el programa le permite al escritor realizar las tareas básicas del procesador de palabras, como convertir texto en negritas, cortar y pegar texto, o llevar a cabo operaciones avanzadas, como la consulta de sinónimos o la corrección de estructuras gramaticales. Sin embargo, la mayoría de las personas que utilizan una computadora para escribir carecen de conocimientos de programación y de la habilidad para determinar los procesos computacionales que otras personas pueden activar. Sin el marco de referencia que los conocimientos de programación brindan, la creación de nueva literatura en computadoras está básicamente determinada por experiencias establecidas por los medios impresos tradicionales. En otras palabras, la autoría y el medio de presentación de lo escrito cambia, pero las convenciones literarias establecidas por los medios impresos se mantienen.

A diferencia de los procesos de escritura tradicionales, que están basados en la manipulación de medios físicos, los procesos de escritura en computadora están mediados por electricidad. Los procesos que se utilizan en la producción y distribución de escritos físicos —ya sea en forma de libros, revistas, periódicos, etc.— también son diferentes de aquellos que se usan en la producción y distribución de escritos digitales. Cuando el escritor utiliza medios físicos, por lo general, trabaja de forma individual y sólo al finalizar proporciona su obra a un editor. El editor produce un número determinado de copias físicas y él mismo las pone a disposición de los lectores o las entrega a un distribuidor que se encarga de esa tarea. Si el número de lectores interesados que acaba obteniendo una copia excede el número de ejemplares impresos, el editor puede volver a surtir su inventario a través de un nuevo proceso de producción, el cual conlleva nuevos costos. En contabilidad estos costos son conocidos como costos variables porque el costo total de producción varía dependiendo del número de unidades producidas.

Durante las primeras dos décadas de uso público de la red mundial de computadoras, el sector editorial ha presenciado básicamente procesos de convergencia de los medios tradicionales hacia los medios digitales. Esta convergencia está impulsada por los beneficios económicos inherentes a los medios digitales, en los cuales, una vez que la primera copia de un producto editorial se ha generado, el costo de reproducción de un gran número de ejemplares es casi nulo. De la misma forma, el costo de distribución de ejemplares es insignificante. La computadora y las redes, como medios de creación y distribución, alteran considerablemente el costo de los procesos tradicionales

de producción de textos. Este hecho, de carácter económico, continuará provocando que las industrias editoriales adopten los medios digitales en sus procesos de producción y distribución de publicaciones.

La transferencia de publicaciones de los medios físicos a los medios digitales conlleva otros tipos de cambios que aún no han sido asimilados por la mayoría de las editoriales y de los escritores. La alteración de los costos de producción es sólo una de las transformaciones que ocurren al convertir los escritos de átomos a bytes. La computadora, como medio programable basado en flujos de electricidad, afecta la creación y la experiencia de la literatura. Por ejemplo, los mismos procesos digitales que permiten la reproducción y la distribución de obras a través de impulsos eléctricos, también permiten la creación de textos colectivos producidos por escritores ubicados en diferentes puntos geográficos. Este tipo de colaboración se puede llevar a cabo a través de la utilización de un sólo documento en el que múltiples escritores escriben de forma abierta y simultánea. Esta tecnología existe en el mercado hoy en día y es gratuita. No obstante, la autoría colectiva de escritos literarios es poco común y no existen convenciones de trabajo que se puedan heredar de los medios impresos tradicionales y que faciliten la realización, en grupo, de obras literarias. Los medios físicos también difieren de los medios digitales en la conceptualización de las obras como obras finitas. La impresión física de un escrito requiere que su autor determine el momento en el que la creación de la obra ha concluido. Por lo general, las publicaciones impresas también condicionan la conceptualización de las obras como unidades de contenidos discretas (en el sentido matemático de la palabra). En el caso del entorno digital, la terminación de una obra no es un hecho absolutamente necesario.

Como se mencionó al principio de este escrito, otro aspecto en que las publicaciones digitales difieren sustancialmente de las publicaciones impresas se encuentra en la habilidad del escritor para programar la creación y la experiencia que el lector pueda tener de sus textos literarios. El propósito de esta introducción a esta compilación de la poesía de José Juan Tablada, con hiperíndices y el sistema *poesía densa*, es explicar los procesos de programación que se llevaron a cabo para generar una nueva experiencia de lectura de textos tradicionales. Aunque en este caso los procesos de programación no están directamente relacionados con la creación literaria de los textos, son procesos creativos que poseen un carácter literario por la experiencia de lectura que proponen.

### **Pensamiento computacional y el objetivo de la programación**

Según la Sociedad Internacional para la Tecnología en la Educación, el *pensamiento computacional* es un proceso orientado a la solución de problemas que incluye las siguientes actividades, aunque no se limita a ellas:

- Formular problemas de manera que sea posible usar computadoras y otras herramientas para solucionarlos.
- Organizar datos de manera lógica y analizarlos.
- Representar datos mediante abstracciones, como modelos y procesos de simulación.
- Automatizar soluciones mediante el pensamiento algorítmico (conjunto ordenado y finito de operaciones).
- Identificar, analizar e implementar posibles soluciones con el objeto de encontrar la combinación de acciones y recursos más eficiente y efectiva.
- Generalizar y transferir los procesos de solución a una gran diversidad de problemas.<sup>1</sup>

---

<sup>1</sup> Traducción realizada por EDUTEKA (<http://www.eduteka.org/modulos/9/272/2082/1>) Liga consultada el 8 de enero de 2014. Documento original de la International Society for Technology in Education (<http://www.iste.org/learn/computational-thinking/ct-operational-definition>).

El pensamiento computacional es un tipo de conciencia en la que se tiene conocimiento de lo que la computadora puede hacer. La consecuencia inmediata de esta conciencia es que el trabajo a realizar se estructure de forma tal que la computadora pueda ser utilizada, eficazmente, con determinados fines. El sistema de lectura *poesía densa* fue concebido bajo este modo de pensar y el objetivo que se planteó fue ofrecer una forma alternativa de secuenciación de la lectura —diferente de la lectura lineal o de aquella que es inducida al proporcionar un índice tradicional— que entregará al lector un beneficio funcional y, a la vez, una experiencia estética. Para lograr este objetivo, se planeó crear hiperligas en un cuerpo de textos, vínculos electrónicos que permitieran una secuenciación de la lectura basada en las palabras mismas. Por ejemplo, si el lector oprime la palabra “viento” en un texto, el programa mostrará otro texto donde la palabra “viento” es utilizada. Esta secuenciación de los textos, guiada por las palabras, provee una experiencia de lectura que permite descubrir el uso del lenguaje de un autor específico, la amplitud de su vocabulario, la frecuencia de determinadas palabras, los contextos en que se lleva a cabo su utilización, etc.

Ante la necesidad de contar con un volumen grande de textos, que permitiera mostrar las virtudes del pensamiento computacional, las antologías me parecieron ser los objetos ideales sobre los que se podría implementar el nuevo sistema de lectura. Como lo explica detalladamente Rodolfo Mata en el texto introductorio. Él se encontraba realizando la transcripción parcial de la poesía de José Juan Tablada a texto electrónico, con el apoyo del equipo que había conformado para el proyecto *Letra e imagen: literatura mexicana en CD-ROM e internet* (Conacyt 25096-H), y aceptó extender la tarea para incluir la totalidad de la producción del poeta vanguardista mexicano.

Acordamos entonces trabajar sobre ese corpus para la sección *José Juan Tablada en Poesía Densa* del CD-ROM *José Juan Tablada: letra e imagen* (2003).<sup>2</sup> En él, a la versión de los poemas que reunió Héctor Valdés en el volumen *Obras I. Poesía* (1971), Mata y su equipo agregaron 49 poemas recopilados por varios investigadores y no reunidos en libro, para sumar un total de 531 textos. En la presente versión se incorporaron los 19 poemas del volumen *Li-po y otros poemas*, antes omitidos, en versiones reescritas de manera lineal que van acompañadas con imágenes pop-up de los poemas originales. De esta manera, se tiene la referencia de la disposición gráfica y se logra una apreciación completa del vocabulario tabladiano. Se trata de alrededor de 700 páginas de poemas en una edición tradicional en papel. Otras peculiaridades importantes de esta nueva edición electrónica son la incorporación de un hiperíndice cronológico convencional, siguiendo la organización ideada por Valdés, y un hiperíndice por palabra; la eliminación de la “navegación aleatoria” de la versión de 2003, pues consideramos que el uso de estos índices era más útil; y una nueva interface gráfica. Los detalles acerca de la revisión de los textos y su transcripción los aborda Mata.

A continuación describiré el método utilizado para la creación del programa *José Juan Tablada en poesía densa*, en 2003, el cual no difiere conceptualmente del realizado para esta nueva versión. El propósito primordial es presentarle al lector el tipo de pensamiento computacional utilizado en el proyecto. Cabe aclarar que las estrategias tomadas durante el diseño, la creación de metadatos y la programación de la interface gráfica en 2003 reflejan tanto las posibilidades computacionales como mis conocimientos computacionales al principio del siglo. Las decisiones que tomé hoy en día sobre cómo programar la experiencia de lectura no son las mismas. El ambiente computacional al momento de escribir estas palabras es exponencialmente más rápido, más capaz, y más extenso. No obstante, como ya dije, la lógica es muy parecida y los ejemplos de programación de 2003 son más fáciles de presentar. Mi escritura está orientada a los escritores y académicos sin conocimientos

---

<sup>2</sup> La aplicación se puede encontrar en la siguiente ubicación en línea, que debe ser consultada con un navegador que permita el uso del plugin Shockwave (Chrome no lo permite, Firefox y Explorer sí): <http://www.hypergraphia.com/tablada/>. Corresponde a la primera versión, la cual fue publicada en el CD-ROM *José Juan Tablada: letra e imagen (poesía, prosa, obra gráfica y varia documental)* (2003).

computacionales, esperando que el documento en sí sea una invitación cordial a un reto intelectual: aprender un lenguaje de programación.

### Método

Computacionalmente hablando, siempre existen múltiples formas para alcanzar un objetivo planteado. A continuación describiré el tipo de procesos algorítmicos que se llevaron a cabo para la realización de *poesía densa* utilizando la poesía completa de José Juan Tablada.

El primer paso está basado en la obtención y el estudio del corpus poético completo en forma textual. En este caso se comienza con todos los poemas presentados en forma secuencial, en un sólo documento electrónico con cientos de “páginas”. El primer poema del corpus, titulado “Poemas dispersos”, empieza de la siguiente forma:

#### POEMAS DISPERSOS

[1888-1914]

A ...

Ese cantar de tu garganta de oro  
que cruza el aire en sosegado vuelo,  
es lo que cantan en divino coro  
los ángeles del cielo.

[...]

El poema continúa hasta su fin y, al terminar, se presenta el segundo poema titulado “Nupcial”:

#### NUPCIAL

Arrobada en su erótico embeleso  
abre la flor de su corola el broche,  
de su lejano amante siente el beso.  
Envuelta entre las auras de la noche.

[...]

Se llega al fin de este poema y se prosigue al tercer poema de forma secuencial, etc.

Sabiendo que la poesía completa se encuentra en un sólo documento de texto, se puede escribir un algoritmo sencillo para enumerar todas las palabras en el corpus poético de José Juan Tablada. La computadora puede generar como resultado una lista de todas las palabras y su ubicación correspondiente en el corpus siguiendo la convención *palabra<sub>n</sub>*, donde *n* es el número absoluto de esa palabra dentro del conjunto. Por ejemplo:

palabra<sub>1</sub> = “POEMAS”  
palabra<sub>2</sub> = “DISPERSOS”  
[...]  
palabra<sub>13</sub> = “que”  
palabra<sub>14</sub> = “cruza”  
palabra<sub>15</sub> = “el”  
palabra<sub>16</sub> = “aire”  
[...]  
palabra<sub>301</sub> = “NUPCIAL”

palabra<sub>302</sub> = “Arrobada”

palabra<sub>303</sub> = “en”

palabra<sub>304</sub> = “su”

palabra<sub>305</sub> = “erótico”

palabra<sub>306</sub> = “embeleso”

Al hacer esto, cada palabra tiene un número asignado. Si se quisiera programar la búsqueda de una palabra específica en el corpus, lo que se puede pedir a la computadora es que compare, una a una, la palabra que se está buscando con todas las palabras existentes en el documento. En el algoritmo para realizar este proceso, donde  $n$  inicialmente es igual a 1, la computadora empezaría por leer la primera palabra y la compararía con la palabra que se está buscando; si el resultado de la comparación es positivo, el proceso se detiene y muestra la ubicación de la palabra; si el resultado de la comparación es negativo, se le añade 1 a  $n$ , el nuevo valor de  $n$  es 2, y por lo tanto se lee la segunda palabra del corpus y la comparación se vuelve a realizar. Por ejemplo, si se decide buscar por la palabra “decora”, el algoritmo descrito generaría los siguientes procesos:

Número $n$	Palabra correspondiente	Palabra buscada	¿Igual?	Acción si No: $n = n + 1$ , volver a comparar. Acción si Sí: Detener y mostrar la palabra
1	POEMAS	decora	No	$n = 1 + 1 = 2$
2	DISPERSOS	decora	No	$n = 2 + 1 = 3$
↓	↓	↓	↓	↓
13	que	decora	No	$n = 13 + 1 = 14$
14	cruza	decora	No	$n = 14 + 1 = 15$
15	el	decora	No	$n = 15 + 1 = 16$
16	aire	decora	No	$n = 16 + 1 = 17$
↓	↓	↓	↓	↓
301	NUPCIAL	decora	No	$n = 301 + 1 = 302$
302	Arrobada	decora	No	$n = 302 + 1 = 303$
303	en	decora	No	$n = 303 + 1 = 304$
304	su	decora	No	$n = 304 + 1 = 305$
305	erótico	decora	No	$n = 305 + 1 = 306$
306	embeleso	decora	No	$n = 306 + 1 = 307$
↓	↓	↓	↓	↓
2883	decora	decora	Sí	Detener y presentar número $n$ (2883)

El mismo proceso de este algoritmo se puede representar a través de un diagrama de flujo:

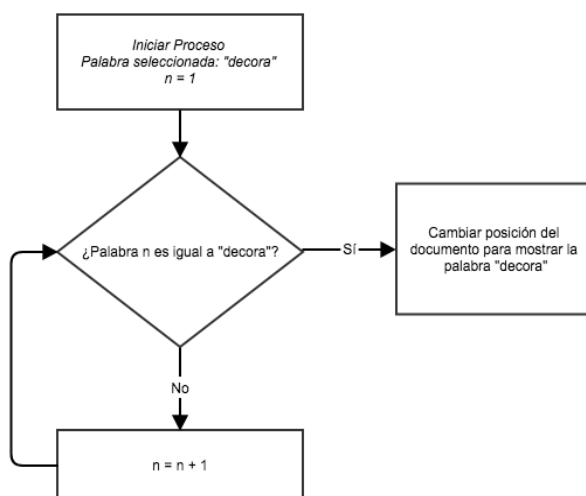


Figura 1. Algoritmo de búsqueda de una palabra.

### Desarrollo de los metadatos para la obra de José Juan Tablada

La numeración de las palabras en el ejemplo anterior es el preámbulo necesario para describir el desarrollo de metadatos que se llevó a cabo para *poesía densa*. Los metadatos son básicamente descripciones de datos o información, es decir, datos sobre datos o información sobre información. La información descriptiva luego es utilizada para alcanzar diferentes propósitos. En este caso, la creación de metadatos tiene el propósito principal de eliminar procesos computacionales largos que impedirían la utilización fluida del programa. Volviendo al ejemplo anterior, si un lector quisiera navegar la publicación a través del uso de la palabra “viento”, se le podría pedir a la computadora que iniciara un proceso de búsqueda palabra por palabra. Al encontrar la primera aparición de “viento”, el programa se detendría y sólo se reanudaría cuando el lector decidiera continuar la lectura e ir a la siguiente aparición. En volúmenes grandes de información, como es el caso de la poesía completa de Tablada, este tipo de navegación podría resultar en experiencias de lectura con largas pausas. Por el contrario, los metadatos agilizarían este tipo de lectura.

Ya que *poesía densa* se desarrolló a comienzos del siglo XXI, cuando las computadoras comerciales rápidas alcanzaban apenas los 800 Mhz, fue indispensable diseñar una estructura de metadatos que pudiera contribuir a que la navegación del corpus poético fuera más rápida y eficiente para agilizar la lectura de la obra. Debido a esta limitación de velocidad, se desarrolló un algoritmo para documentar la presencia de cada una de las palabras en cada uno de los poemas. El programa, en vez de recuperar la posición absoluta de la palabra en el corpus total de la obra, como se mostró anteriormente, registró el lugar relativo de cada palabra en cada poema. En el caso de la poesía completa de Tablada, se cuenta con 550 poemas diferentes y la programación se diseñó y se escribió para correr desde el poema<sub>1</sub> hasta el poema<sub>550</sub>, donde cada uno de los poemas tiene  $n$  número de palabras. Siguiendo la notación palabra<sub>m,n</sub> se obtiene el siguiente resultado

palabra<sub>1,1</sub> = “POEMAS”  
 palabra<sub>1,2</sub> = “DISPERSOS”  
 [...]
   
palabra<sub>1,13</sub> = “que”  
 palabra<sub>1,14</sub> = “cruza”  
 palabra<sub>1,15</sub> = “el”  
 palabra<sub>1,16</sub> = “aire”

[...]

palabra<sub>2,1</sub> = "NUPCIAL"

palabra<sub>2,2</sub> = "Arrobada"

palabra<sub>2,3</sub> = "en"

palabra<sub>2,4</sub> = "su"

palabra<sub>2,5</sub> = "erótico"

palabra<sub>2,6</sub> = "embeleso"

Si se quisiera hacer una nueva búsqueda, recorriendo todas las palabras del corpus poético con esta nueva forma de separar el texto, el proceso tendría un diagrama de flujo como el mostrado en la Figura 2.

Para la computadora, el proceso es similar y cumple el mismo objetivo, sólo que se desarrolla con antelación. Si se busca la palabra "decora", la máquina no tendrá que producir en ese momento la búsqueda en todo el corpus textual sino que va a contar con el "mapa" de los poemas en los que la palabra se ubica.

Una vez que el algoritmo comenzó a correr, en 2003, el proceso de creación de metadatos llevó aproximadamente 90 horas en una computadora con un procesador Pentium III que corría a 800 Mhz. En 2018, el programa demoró menos de 10 horas, con un procesador Intel Core i7, a 2.6 Mhz, y realizó alrededor de 1,567 millones de comparaciones. El resultado fue una lista con todas las palabras diferentes utilizadas en el corpus y los números de los respectivos poemas donde éstas aparecen. El ejemplo siguiente muestra un fragmento pequeño de toda la lista de palabras procesadas:

lujuria 66 109 151 193 197 200 370

lujurias 10 139

lujuriosa 144 194

lujuriosas 8 135

lujurioso 66

lujuriosos 65

lumbos 40

lumbre 41(2) 46 62 66 69 78 117 119 156(2) 171(2) 193 206 336 350 364 377(2)

lumbres 116 159 160 169 181 330

lunia 190 193 328 450

luminar 139

luminosa 78(2) 87 94 105 119 122 126 156 157 160(2) 195 204 399

luminosas 46 66(3) 111 155

luminoso 25 41 66 95 96 112(2) 119 123 134 179 193 354 364 510 514

luminosos 40 41 368

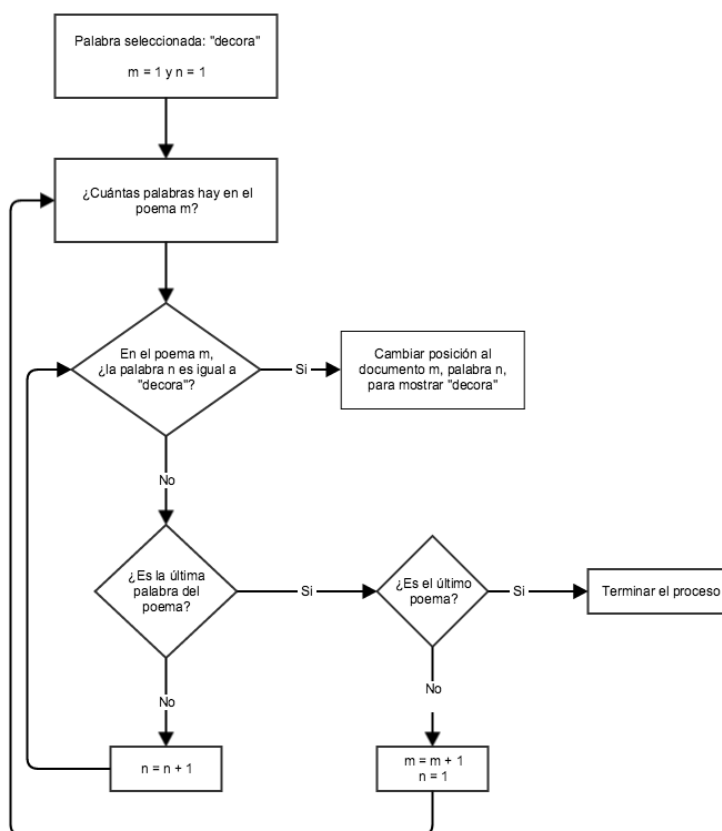


Figura 2. Algoritmo para creación de metadatos.

Esto quiere decir que la palabra “lujuria” aparece en los poemas 66, 109, 151, 193, 197, 200 y 370; “lujuriosa”, en los poemas 144 y 194; “lumbres”, en los poemas 116, 159, 160, 169, 181 y 330; etc. En varias instancias se pueden ver números seguidos por otro número dentro de paréntesis, éste número se refiere al número de repeticiones de una palabra dentro del mismo poema.

La lista con la ubicación de todas las palabras se dividió en 26 listas, una por cada letra del abecedario. Esta información, organizada gráficamente (véase la Figura 3), nos permite apreciar que, en el corpus de los poemas de Tablada, se encontraron 645 palabras que empiezan con la letra “a”, 434 palabras que empiezan con la letra “e”, 41 palabras que empiezan con la letra “q”, etc. En la ejecución computacional del algoritmo y en la conformación final de las listas, se excluyeron las preposiciones, las conjunciones y otras clases de palabras que se juzgaron poco significativas, desde el punto de vista de la variabilidad y riqueza del vocabulario poético de Tablada. También es importante señalar que, como se puede apreciar en el ejemplo citado, se prefirió mantener la separación de palabras flexionadas, como el género y número de adjetivos y sustantivos. Rodolfo Mata explica más detalladamente estas decisiones, en la introducción a la edición electrónica con hiperíndices de la *Poesía reunida de José Juan Tablada*.

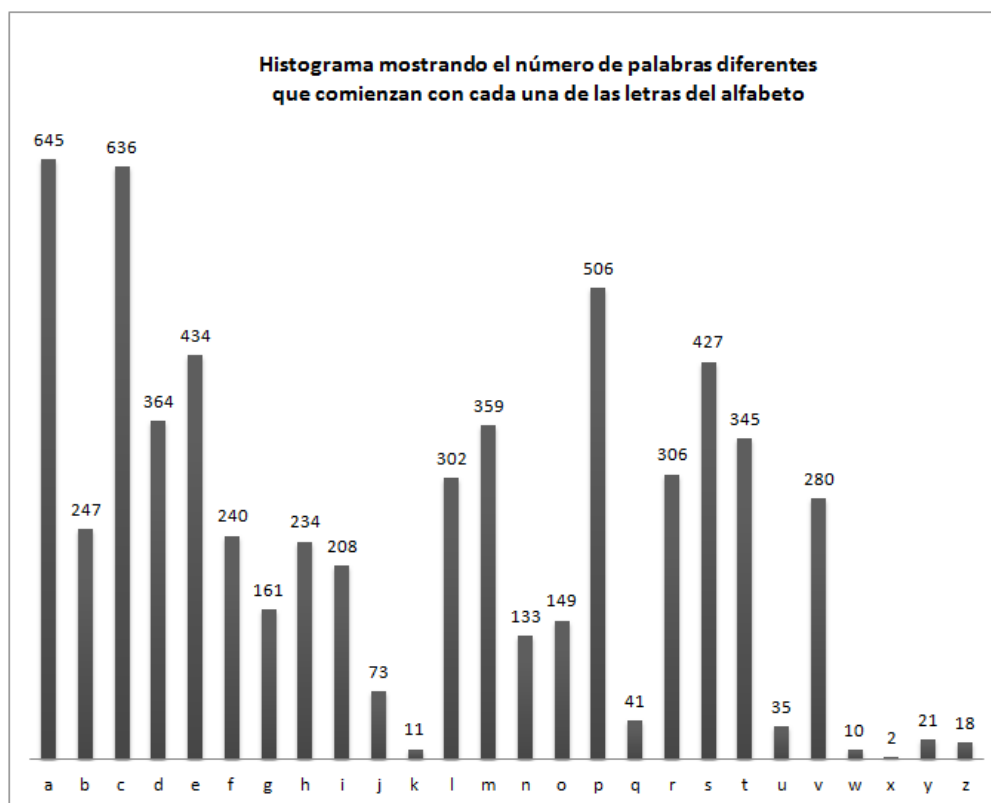


Figura 3. Histograma de frecuencia de palabras por letra inicial.

Los metadatos, la información sobre los poemas en el corpus poético, ahora están compuestos por 26 listas. En cada lista se encuentran todas las palabras utilizadas por el autor, empezando con la letra correspondiente, y cada una de las palabras contiene el número del poema, o los poemas, donde ésta aparece. Esta reestructuración del contenido en listados alfabéticos, además de proporcionarnos estas estadísticas, facilita en gran medida la navegación hipertextual de la obra de forma rápida e eficiente, pues ya no se tienen que hacer todas las comparaciones cada vez que se busca una palabra determinada, sino que se tiene un listado de todas las palabras diferentes en el corpus poético, con su localización en cada poema.

### Presentación utilizando los metadatos

Con los metadatos fue posible crear una navegación hipertextual que utiliza todas las palabras en el volumen de poesía (menos las exclusiones de preposiciones, conjunciones y otras clases de palabras anteriormente mencionadas). Esta navegación se logró a través de la creación de un proceso que se lleva a cabo cada vez que el usuario oprime en la pantalla la palabra de un poema. En la Figura 4 se muestra la pantalla del programa con el principio del poema “Flores de Invernadero”, en una ventana, mientras que en la ventana contigua aparece el “Hiperíndice de poemas” —desde el cual ha sido llamado el poema—, índice que no existía en la versión de 2003, pues sólo se manejaba una navegación desde el cuerpo de los poemas. En la ventana con el poema “Flores de Invernadero” se puede notar cómo las palabras que lo integran están coloreadas de gris o de verde. Las palabras en verde indican que están hiperligadas a la misma palabra en algún otro lugar de la poesía completa de José Juan Tablada. Si el usuario oprime, por ejemplo, la palabra “pálido”, el programa inicia un algoritmo que lo llevará al siguiente poema donde la palabra “pálido” vuelve a aparecer.



Figura 4. Pantalla de navegación con hiperíndice de poemas

El algoritmo cuenta con los siguientes pasos (si es posible, recomendando abrir la aplicación e interactuar con ella conforme se lee la secuencia de procesos que se llevan a cabo):

- 1) Si hay un clic, registrar el color de la palabra que fue oprimida. Si la palabra es gris oscuro, no hacer nada. Si la palabra es de color azul claro, lo cual significa que está hiperligada, registrar la palabra oprimida, "pálido", y registrar el número del poema oprimido, en este caso 61. Estos dos datos serán utilizados para seleccionar el siguiente poema.
- 2) Reconocer la primera letra de la palabra seleccionada. En este caso, la "p" de la palabra "pálido".
- 3) Ir al listado con todas las palabras que empiezan con la primera letra de la palabra oprimida. Determinar el número de palabras en la lista. En el caso de la poesía completa en *poesía densa* de José Juan Tablada, existen 506 palabras diferentes empezando con la letra "p". Esto se determina de forma muy rápida, gracias al proceso de creación de metadatos citado anteriormente, el cual llevó un buen tiempo de indización.
- 4) Compara la palabra 1, de las 506, con la palabra "pálido". Si las palabras no son iguales, compara la siguiente palabra, la palabra 2, y así hasta llegar a la 506. Si son iguales, avanza al siguiente paso del algoritmo.
- 5) Obtener de los metadatos todas las ubicaciones de la palabra "pálido" dentro del corpus. En este caso, el resultado es el siguiente: pálido aparece en los poemas 2, 6, 7(2), 9, 10, 17, 24, 29, 43(2), 50, 59, 66(2), 94, 101, 111(2), 114, 118, 119, 139, 143, 145, 150, 171 y 349.
- 6) El siguiente paso es buscar el próximo poema donde la palabra "pálido" aparezca. Se comienza por calcular el número de ocasiones en las que la palabra "pálido" aparece en el corpus, en este caso, 24 veces que corresponden a poemas diferentes (del 61 al 389). Una vez que se calculó el número de ocurrencias, se utiliza el número del poema "Flores de invernadero" (61) para determinar el siguiente poema que se mostrará al usuario.
- 7) El algoritmo busca el número de poema donde la palabra "pálido" se oprimió (poema 61) y selecciona el siguiente poema donde existe la misma palabra. En este caso, se selecciona el

poema 65, el cual es una oda nocturna titulada “El telón rojo”, donde la palabra “pálido” es utilizada de el siguiente verso: “y el horizonte pálido se inflama.”

- 8) El algoritmo dirige la navegación al poema seleccionado, cuenta el número total palabras en el nuevo poema y comienza un nuevo proceso para marcar de rojo la palabra que fue seleccionada (esto ayuda a que el usuario/lector pueda localizarla de forma rápida).

El proceso en general de buscar las palabras de un poema en los listados, para determinar si se deben colorear o no, se logra en una fracción de segundo. El proceso anterior de creación de metadatos, proceso que llevó mucho más tiempo para completarse, es el que permite establecer, de forma rápida e imperceptible para el lector, qué palabras deben contener ligas, cuando éste lee la obra de José Juan Tablada.

La nueva versión del programa permite una segunda forma de acceder al corpus de poemas, a través de las palabras ordenadas en un índice alfabético interactivo. En este caso, se le ofrece al lector que estudia la obra de José Juan Tablada el listado completo de palabras y sus respectivos metadatos generados. En el momento que el lector elige una palabra y pulsa el número del poema que indica su ubicación en el corpus, el paso número 7 del algoritmo descrito anteriormente hace aparecer en la ventana correspondiente el poema seleccionado y colorea de azul la palabra en cuestión, mientras las demás permanecen siguiendo las reglas de colores ya descritas.



Figura 5. Pantalla de navegación con hiperíndice de palabras

### El proyecto hoy

El concepto de esta forma de lectura de un corpus poético surgió en 1998 y se montó en línea sólo en 2003. Más de una década más tarde, el campo de la computación ha evolucionado aceleradamente. No sólo las computadoras se han vuelto más rápidas y más poderosas, también los lenguajes de programación, las bases de datos y las interfaces gráficas han evolucionado exponencialmente. En los días que corren contamos con inigualables recursos en línea para realizar tareas computacionales complejas. Estos recursos, comúnmente identificados como “servicios en la nube”, se encuentran establecidos en múltiples servidores en Internet. La multiplicidad de

servidores involucrados permite que las tareas computacionales sean repartidas entre ellos, sin importar que se ubiquen en diferentes puntos geográficos. Hoy en día, cualquier persona puede rentar este tipo de herramientas computacionales de gran poder, y operarlas desde una computadora (una terminal) conectada a Internet.

En diciembre de 2010, Google anunció que daba acceso libre a una base de datos de su propiedad.<sup>3</sup> Esta base de datos, llamada Google Books Ngram Viewer, contiene alrededor de 500 mil millones de palabras de más de 5.2 millones de libros en inglés, francés, español, alemán, chino y ruso, publicados entre 1473 y 2008.<sup>4</sup> Es interesante ver cómo los n-gramas han sido conceptualizados, ya que no sólo están formados por palabras específicas, sino que también están integrados por secuencias de palabras. Por ejemplo, un 1-gram (o unigrama) puede ser “love” y un 3-gram (o trigram) puede ser “I love you”. Esta conceptualización de los metadatos que se generarán sobre volúmenes enormes de texto permite un estudio más profundo del lenguaje. Las últimas versiones del Ngram de Google cuentan con metadatos relacionados con la utilización gramatical de las palabras dentro de su contexto y con posibilidades de búsqueda filtradas por operaciones de suma o resta de n-gramas, consideración de mayúsculas o minúsculas e inserción de signos “comodines” (wildcards).<sup>5</sup>

Al hacer una prueba con el trigram “José Juan Tablada” obtenemos el resultado mostrado en la Figura 6:

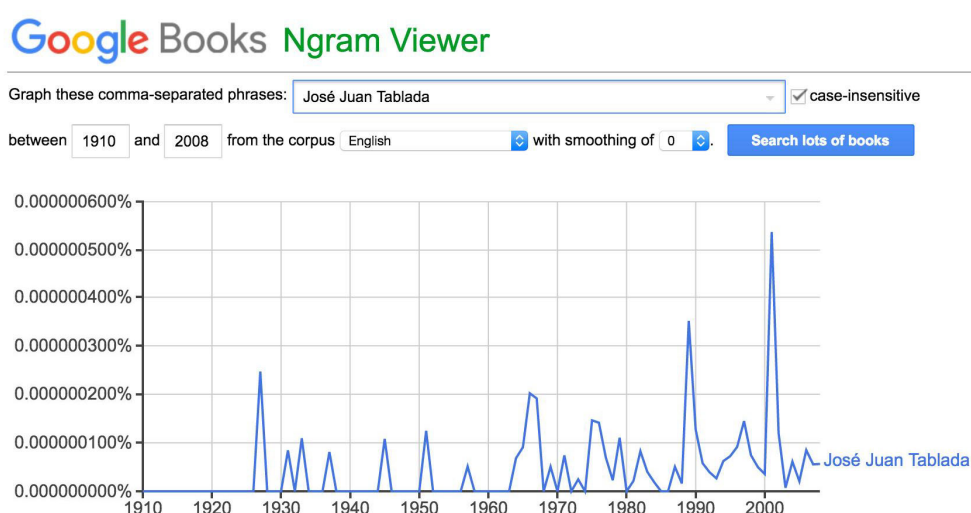


Figura 6. Ocurrencia del nombre “José Juan Tablada” en los libros en inglés.

La gráfica muestra la frecuencia de aparición de las tres palabras juntas en los libros en inglés digitalizados por Google. En el eje de las abscisas (x) aparece el año y en el de las ordenadas (y) se registra el porcentaje que resulta de comparar el número de veces que aparece el 3-gram “José Juan Tablada” con el total de los Ngrams de los libros en inglés digitalizados por Google.

Para estimular y facilitar el análisis secundario de esta base de datos, Google también abrió al público la descarga de la información, a través de internet. Todos los Ngrams se pueden descargar bajo una licencia Creative Commons Atribución.<sup>6</sup> La descarga completa es muy grande, por lo que

<sup>3</sup> [http://www.nytimes.com/2010/12/17/books/17words.html?\\_r=0](http://www.nytimes.com/2010/12/17/books/17words.html?_r=0)

<sup>4</sup> <https://books.google.com/ngrams/>

<sup>5</sup> <https://books.google.com/ngrams/info>

<sup>6</sup> Las licencias de Creative Commons son una serie de contratos que han sido estandarizados para permitir que los autores cedan y resguarden diferentes aspectos de los derechos que tienen sobre sus obras. Es una

Google proporciona un recurso en línea donde la información se ha separado en módulos más pequeños.<sup>7</sup> En los metadatos del 3-gram que empiezan con las letras “jo”<sup>8</sup> se puede leer:

José Juan Tablada_NOUN	1927	3	1
José Juan Tablada_NOUN	1931	1	1
José Juan Tablada_NOUN	1933	1	1
José Juan Tablada_NOUN	1937	1	1
José Juan Tablada_NOUN	1945	1	1
José Juan Tablada_NOUN	1951	2	1
José Juan Tablada_NOUN	1957	1	1
José Juan Tablada_NOUN	1964	2	1
José Juan Tablada_NOUN	1965	3	3
José Juan Tablada_NOUN	1966	7	7
José Juan Tablada_NOUN	1967	7	1
José Juan Tablada_NOUN	1969	2	1
José Juan Tablada_NOUN	1971	3	3
José Juan Tablada_NOUN	1973	1	1
José Juan Tablada_NOUN	1975	6	3
José Juan Tablada_NOUN	1976	6	2
José Juan Tablada_NOUN	1977	3	3
José Juan Tablada_NOUN	1978	1	1
José Juan Tablada_NOUN	1979	5	3
José Juan Tablada_NOUN	1981	1	1
José Juan Tablada_NOUN	1982	4	4
José Juan Tablada_NOUN	1983	2	2
José Juan Tablada_NOUN	1984	1	1
José Juan Tablada_NOUN	1987	3	3
José Juan Tablada_NOUN	1988	1	1
José Juan Tablada_NOUN	1989	23	2
José Juan Tablada_NOUN	1990	9	7
José Juan Tablada_NOUN	1991	4	3
José Juan Tablada_NOUN	1992	3	1
José Juan Tablada_NOUN	1993	2	2
José Juan Tablada_NOUN	1994	5	5
José Juan Tablada_NOUN	1995	6	3
José Juan Tablada_NOUN	1996	8	6
José Juan Tablada_NOUN	1997	13	4
José Juan Tablada_NOUN	1998	7	6
José Juan Tablada_NOUN	1999	5	3
José Juan Tablada_NOUN	2000	4	4
José Juan Tablada_NOUN	2001	61	8
José Juan Tablada_NOUN	2002	15	4
José Juan Tablada_NOUN	2003	1	1

---

respuesta al problema que trajeron las nuevas tecnologías digitales al facilitar y acelerar la producción, reproducción y publicación de contenido por individuos que disponían de pocos recursos para gestionar derechos autorales y que encontraban, principalmente en internet, obra de determinados autores que deseaban utilizar.

<sup>7</sup> <http://storage.googleapis.com/books/ngrams/books/datasetv2.html>

<sup>8</sup> La ubicación en línea de los metadatos del 3-ngram que empiezan con las letras “jo” (18.59 GB de tamaño) es: <http://storage.googleapis.com/books/ngrams/books/googlebooks-eng-all-3gram-20120701-jo.gz> .

José Juan Tablada_NOUN	2004	9	8
José Juan Tablada_NOUN	2005	3	3
José Juan Tablada_NOUN	2006	13	9
José Juan Tablada_NOUN	2007	9	7
José Juan Tablada_NOUN	2008	11	5

Es interesante ver que el formato para los metadatos generados por el Ngram Viewer es similar al formato desarrollado para la poesía completa de José Juan Tablada. En *poesía densa*, se realizó también un listado donde la primera entrada es cada palabra y después vienen datos numéricos sobre la ubicación de la palabra en el corpus (los números de los poemas). En el listado anterior, se muestran todas las ocurrencias donde aparece el ngram (el nombre del poeta) seguido por los metadatos. El primer dato describe la función gramatical del ngram (noun), información que se añadió en la segunda versión del Ngram producida en 2012 y que no estaba incluida en la primera versión de 2009. Las siguientes 3 columnas con números describen dónde aparece “José Juan Tablada”: El primer dato es el año de las publicaciones revisadas, el segundo se refiere al número de páginas donde aparece el 3-gram, y el tercer dato corresponde al número de libros donde éste aparece. Se pueden hallar detalles de esta información mediante una búsqueda directa de la cadena “José Juan Tablada” en Google Books. Solamente hay que tener cuidado con los parámetros que se eligen pues, aunque Google Ngrams y Google Books funcionan con la misma base de información, no la manejan y muestran de la misma manera. Es decir, no hay una relación directa perfectamente transparente. Ojalá pronto llegue el día en que al oprimir un punto en una gráfica como la anterior, lleguemos directo a un listado de las publicaciones que albergan el ngrama buscado.

El concepto del proyecto de *poesía densa* hoy en día sigue vigente. Al igual que el Ngram se utiliza en los estudios literarios dentro de las humanidades digitales, es posible imaginar un proceso similar de digitalización de obras literarias en español, la creación subsecuente de metadatos, y la utilización de la nueva información para múltiples tipos de lecturas y enfoques de estudio de las obras.<sup>9</sup>

## Conclusión

El pensamiento computacional, como conciencia de lo que la computadora puede hacer, plantea un campo de interacción creativa entre la literatura y los medios computacionales. Este cruce de disciplinas permite desarrollar el potencial de herramientas y metodologías cuya aplicación se encontraba antes limitada. Conocer la lógica computacional proporciona a los autores y a los editores la capacidad de alcanzar el potencial del nuevo medio de comunicación. Sin embargo, para que esta interacción sea apropiada, es necesario que haya más personas interesadas en ambas disciplinas, lo cual no es común. Ya que los procesos computacionales están formulados, en gran medida, como procesos matemáticos, se requiere de individuos que puedan navegar con suficiente habilidad tanto en un mar de letras como en un mar de números.

Sin conocimientos sobre lógica computacional, programación y matemáticas básicas, el escritor de hoy en día sólo puede usar la computadora para reproducir las actividades que ya realizaba con los medios impresos. La máquina de escribir es reemplazada por el teclado y el libro impreso es proyectado, con la mayoría de las convenciones que arrastra, en una pantalla electrónica. A pesar de que la computadora es programable —lo cual la convierte en la herramienta de comunicación más poderosa de la historia—, es utilizada por el medio editorial primordialmente para suplantar a los libros, dejando intactas muchas de las tradiciones establecidas por la imprenta. Si se carece de pensamiento computacional, la habilidad para imaginar nueva literatura queda considerablemente

<sup>9</sup> La siguiente es una búsqueda en la plataforma Google Scholar donde se puede percibir la cantidad de veces que “Google Ngram” aparece en textos académicos:

[https://scholar.google.com/scholar?q=%22Google+Ngram%22&btnG=&hl=en&as\\_sdt=0%2C5](https://scholar.google.com/scholar?q=%22Google+Ngram%22&btnG=&hl=en&as_sdt=0%2C5)

mermada: o se imagina lo que se escribe como un producto literario tradicional, sólo que distribuido en un medio electrónico, haciendo caso omiso de las demás peculiaridades propias del medio; o se imagina lo que se escribe como un documento en el que el medio electrónico avasalla la expresión literaria y la distorsiona, opacando gran parte de sus rasgos estéticos o pretendiendo suplantarlos. La ausencia de pensamiento computacional limita la creación de nueva literatura.

Una vez que se cuenta con conocimientos de computación, los procesos imaginativos cambian y el número de opciones creativas crece. La computadora, como medio programable, brinda a los autores posibilidades insospechadas; no sólo por lo que pueden hacer por sí mismos, sino también por las oportunidades de interconexión entre su obra personal y el universo de obras digitales presente en Internet. Aunque parezca un poco trivial, improviso aquí la siguiente propuesta, intuita ejercitando el pensamiento computacional, para que sirva de ejemplo: Con los conocimientos de programación necesarios, es posible “conectar” cada palabra del corpus de la poesía reunida de José Juan Tablada con una imagen del servicio Flickr.<sup>10</sup> Cuando se escoja una palabra en un poema, un programa podría hacer una búsqueda en Flickr de todas la fotos descritas con esa palabra, y seleccionar aquella imagen que hubiera recibido el mayor número de votos.<sup>11</sup> Mediante un filtro rudimentario se dejarían pasar sólo imágenes de calidad alta y finalmente, la imagen seleccionada se añadiría difuminada como fondo de lectura del texto. En otras palabras, con las herramientas disponibles hoy en día, sería posible mostrar una imagen seleccionada popularmente, relacionada con cada una de las palabras del corpus de poesía. El contar con el pensamiento computacional permite imaginar procesos posibles que pueden ayudar en el estudio o la creación de obras literarias sacando provecho del potencial del nuevo medio.

Una última aclaración es necesaria. La lógica computacional está relacionada con la programación y el funcionamiento general del nuevo medio y no debe ser confundida con el alfabetismo computacional, el cual se refiere a las habilidades básicas de un individuo para hacer uso específico de los procesos previamente programados por otras personas. En otras palabras, el pensamiento computacional es un tipo de conciencia y no es equivalente a conocer a fondo tal o cual aplicación, plataforma o técnica de programación. La inclusión de cursos de computación para estudiantes y académicos por igual, ayudaría a desatar un arroyo de posibilidades creativas que impulsaría tanto al estudio académico como a la imaginación literaria.

**DIEGO BONILLA**

Communication Studies  
California State University, Sacramento

---

<sup>10</sup> El servicio en línea Flickr (<https://www.flickr.com>), que cuenta con cientos de millones de fotografías, no sólo proporciona licencias Creative Commons, sino también una forma de acceso a su base de datos a través de programación (<https://www.flickr.com/services/api/>), para hacer posible la extracción automática de imágenes siguiendo criterios específicos de selección.

<sup>11</sup> En Flickr, este “ranking” se va creando con las marcas de una foto como favorita por los millones de usuarios.