

Software para traducir textos excluyentes y sexistas a incluyentes y no sexistas a través de autómatas

Software to translate excluding and sexist texts to inclusive and non-sexist texts through automata

HERNÁNDEZ-GARCÍA, Héctor Daniel†*, NAVARRETE-ARIAS, Dulce J. y AGUILAR-OJEDA, Cristy Elizabeth

Instituto Tecnológico Superior del Occidente del Estado de Hidalgo, División de Ingeniería en Sistemas Computacionales, México.

ID 1^{er} Autor: *Héctor Daniel, Hernández-García* / ORC ID: 0000-0001-5261-8353, Researcher ID Thomson: P-4823-2018, CVU CONACYT ID: 208146

ID 1^{er} Coautor: *Dulce J., Navarrete-Arias* / ORC ID: 0000-0002-7915-068X, CVU CONACYT ID: 366071

ID 2^{do} Coautor: *Cristy Elizabeth, Aguilar-Ojeda* / ORC ID: 0000-0002-6814-6378, CVU CONACYT ID: 853247

DOI: 10.35429/JOCS.2020.22.7.1.6

Recibido 10 de Enero, 2020; Aceptado 30 de June, 2020

Resumen

El presente trabajo describe el desarrollo de un software que tiene como tarea traducir textos escritos en un lenguaje excluyente y sexista a un texto en un lenguaje incluyente y no sexista. Este software tiene tres procesos: primero extraer el texto de diferentes formatos digitales (pdf, docx o html), segundo traducir el texto a través de un autómata que implementa el "manual para el uso no sexista del lenguaje" y tercero devolver el texto traducido en un documento digital del mismo formato que el del origen. El desarrollo se realizó a través de la metodología incremental, de la cual se obtuvieron tres incrementos: el primero para extraer el texto del documento digital, el segundo para implementar el autómata y traducir el texto extraído, y el tercer incremento para devolver el texto traducido en un formato igual al documento origen. El autómata implementado fue desarrollado en la herramienta JFlex. Las pruebas realizadas son las Alpha y consistieron en traducir oraciones simples para validar la correcta implementación de las reglas descritas en el "manual para el uso no sexista del lenguaje". El objetivo es ayudar en el aprendizaje de este nuevo lenguaje que pretende incluir y reconocer a la mujer en las redacciones.

Violencia de género, Lenguaje incluyente, Autómatas

Abstract

This work describes the development of a software whose task is to translate texts written in an exclusive and sexist language to a text in an inclusive and non-sexist language. This software has three processes: first to extract the text from different digital formats (pdf, docx or html), second to translate the text through an automaton that implements the "manual for the non-sexist use of language" and third to return the translated text in a digital document of the same format as the source. The development was carried out through the incremental methodology, from which three increments were obtained: the first to extract the text from the digital document, the second to implement the automaton and translate the extracted text, and the third increment to return the translated text in a format equal to the source document. The implemented automaton was developed in the JFlex tool. The tests carried out are Alpha and consist of translating simple sentences to validate the correct implementation of the rules described in the "manual for the non-sexist use of language". The objective is to help in the learning of this new language that aims to include and recognize women in the newsrooms.

Gender Violence, Inclusive Language, Automata

Citación: HERNÁNDEZ-GARCÍA, Héctor Daniel, NAVARRETE-ARIAS, Dulce J. y AGUILAR-OJEDA, Cristy Elizabeth. Software para traducir textos excluyentes y sexistas a incluyentes y no sexistas a través de autómatas. *Revista de Sociología Contemporánea*. 2020. 7-22:1-6.

* Correspondencia del Autor (correo electrónico: hhernandez@itsoeh.edu.mx)

† Investigador contribuyendo como primer autor.

Introducción

La violencia contra la mujer es un tema abordado por diferentes organizaciones como la Organización Mundial de la Salud (OMS) que la define como "todo acto de violencia de género que resulte, o pueda tener como resultado un daño físico, sexual o psicológico para la mujer, inclusive las amenazas de tales actos, la coacción o la privación arbitraria de libertad, tanto si se producen en la vida pública como en la privada" (Organización Mundial de la Salud, 2017). Para la Organización de las Naciones Unidas (ONU) es un tema que se debe tratar y como respuesta en la asamblea de Julio de 2010 crea la entidad de ONU Mujeres que tiene como objetivo centrarse en la igualdad y empoderamiento de la mujer (ONU Mujeres, 2017). Esta entidad establece que la violencia contra las mujeres se da de diversas formas: física, sexual, psicológica y económica; donde cada una se interrelacionan y afectan a las mujeres desde que nacen hasta que son mayores. Mencionan que "las mujeres que experimentan algún tipo de violencia sufren una variedad de problemas de salud y por consecuencia disminuyen su capacidad para participar en la vida pública", afectando no solamente a ellas, también a sus familias, comunidades y naciones. Para la ONU la violencia contra las mujeres no se confina a una cultura, región o país en específico, ni a sus costumbres o religiones, se yacen en la discriminación persistente a la mujer.

México en apoyo a este tema establece el Programa Nacional para la Igualdad entre Mujeres y Hombres, durante la Secretaría de Gobernación 2008 – 2012, en el cual estipula en el objetivo 1 y estrategia 1.2 que se deberá "transformar la cultura y los procesos de gestión de la Secretaría a favor de la igualdad y equidad de género". Como propuesta a este programa se presenta el lenguaje incluyente y no sexista que tiene como objetivo crear redacciones en donde no se discrimine o se desvalore la presencia de la mujer. Como menciona María Julia en (Perez Cervera, 2011), el lenguaje es una forma de expresar nuestra concepción del mundo y refleja la sociedad en que vivimos, también menciona que el lenguaje es un instrumento de intercambio de conocimiento y cultura, los cuales van pasando de generación en generación, convirtiendo el lenguaje en una herramienta clave para el desarrollo de la sociedad.

Pero, ¿qué pasa cuando el vocabulario que se utiliza en el lenguaje, omite la presencia y los derechos de la mujer? Es aquí donde se presenta una de las maneras de discriminación hacia la mujer, ya que, a pesar de que el lenguaje español cuenta con vocabulario para hacer referencia a la mujer, no se utiliza. Esto provoca que cuando se leen textos de cualquier índole no se tenga presente la presencia femenina, haciendo un conocimiento subconsciente de que la mujer no puede figurar en labores cotidianas de la vida, donde la presencia del hombre es dominante. Una de las causas de este problema es por la costumbre de que las palabras plurales masculinas engloban tanto a mujeres como hombres, sin embargo, esto inconscientemente originan problemas en el ámbito social como lo reportado por la OMS y la ONU (Organización Mundial de la Salud, 2017) (ONU Mujeres, 2017) (ONU Mujeres, 2017). Es por estas razones que la propuesta del lenguaje incluyente y no sexista representa una oportunidad para cambiar los paradigmas que la sociedad y herencia ha dejado para incluir ahora la presencia de la mujer en cualquier texto que se redacte o se lea. Ahora bien, a pesar de ser una buena propuesta existe el problema que la sociedad actual no está educada ni preparada para la redacción incluyente y no sexista, por lo que intentar redactar en este lenguaje resulta ser confuso, cansado y hasta frustrante. A pesar de tener hoy en día herramientas ofimáticas que ayudan a la redacción, éstas no cuentan con un método para ayudar a corregir las redacciones excluyentes y sexistas, solamente se limitan a corregir errores ortográficos y gramaticales. En el mercado solamente se encuentra una solución como Themis (Themis, 2018) que ofrece una herramienta para la corrección y guía para redacción de documentos en el lenguaje incluyente y no sexista, sin embargo, esta solución es propietaria por lo que se tiene que pagar una licencia por servicio lo que resulta para algunos un gasto innecesario.

En respuesta a lo anterior, el presente trabajo presenta el desarrollo de una herramienta ofimática que recibe un documento de texto en formato docx, pdf o html, y devuelve otro archivo en el mismo formato con un texto en lenguaje incluyente y no sexista a fin de ayudar a las personas a comprender las diferencias entre un texto excluyente y sexista con un texto incluyente y no sexista.

Con esto se pretende colaborar con el cambio de paradigma en la sociedad donde la presencia de la mujer en los diferentes ámbitos de la vida cotidiana sea equitativa.

El documento se conforma de otras seis secciones. La sección “Metodología desarrollada” describe la metodología de desarrollo de software utilizado y lo desarrollado en cada una de sus etapas. La sección “Resultados” muestra y analiza los resultados obtenidos de las pruebas realizadas al software. En la siguiente sección llamada “Conclusiones” se comunica la conclusión a la que se ha llegado con base en los resultados obtenidos. En la sección “Agradecimientos” se agradece la participación de los voluntarios que trabajaron en el desarrollo del software. La sección “Trabajo a futuro” menciona las tareas que se deben realizar para mejorar el funcionamiento del software. Y en la sección “Referencias” se listan las fuentes de información consultadas para el desarrollo del software.

Metodología desarrollada

El desarrollo del software se realizó con el lenguaje Java y la metodología de desarrollo incremental (Pressman, 2010) que divide el desarrollo de software en incrementos funcionales como resultados de la ejecución de cinco etapas (comunicación, planeación, modelado, construcción y despliegue). Por lo tanto, el desarrollo se dividió en tres incrementos funcionales que se pueden apreciar en la Figura 1.

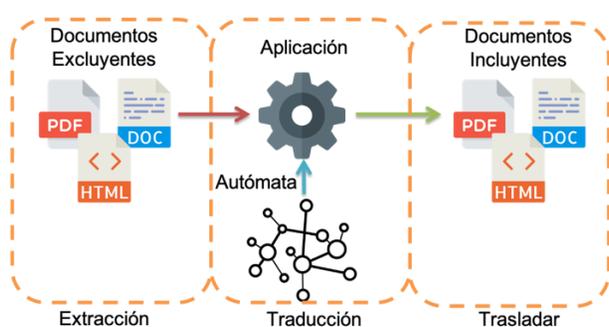


Figura 1 Funcionalidad del traductor de textos excluyente a incluyente

Fuente: Elaboración propia

La función del primer incremento es mostrar una interfaz de usuario (UI por sus siglas en inglés) que permita seleccionar un documento en formato pdf, docx o html, y extraer el texto que contiene. La interfaz de usuario desarrollada se puede observar en la Figura 2. Para este incremento se utilizaron las librerías de Java Swing (Eckel, 2010) para la interfaz de usuario, las librerías POI (Apache org, 2018) del proyecto jakarta de Apache para extraer el texto de documentos en formato docx y las librerías de PDFBox (Apache org, 2018) de Apache para extraer el texto de documentos en formato pdf.



Figura 2 Interfaz de usuario del software

Fuente: Elaboración propia

El segundo incremento tiene como tarea traducir el texto extraído del lenguaje excluyente al lenguaje incluyente con ayuda de un autómata generado en la herramienta JFlex (JFlex Team, 2018). El trabajo de este autómata es leer las oraciones a fin de determinar la estructura e identificar del “Manual para el uso no sexista del lenguaje” (Perez Cervera, 2011) la regla a implementar para traducir la oración y escribirla en un archivo de texto, el diseño general de autómata se muestra en la Figura 3. El objetivo del tercer y último incremento es trasladar el texto traducido del archivo creado a un nuevo archivo que del mismo formato que el documento origen utilizando las mismas librerías para la extracción del texto.

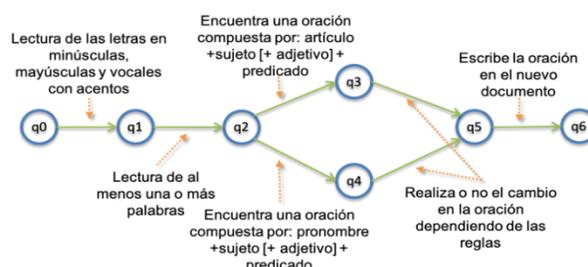


Figura 3 Estructura general del autómata

Fuente: Elaboración propia

Las pruebas realizadas son las denominadas Alpha y el escenario experimental implementado para este software consistió en los siguientes actividades: reclutar un conjunto de 11 documentos en formato docx, pdf o html, con un número variable de oraciones simples (entre 15 y 25), identificar en las oraciones las traducciones que se deben realizar, traducir cada documento con el software, identificar las traducciones realizadas en las oraciones para determinar el grado de eficiencia que el software obtuvo en un documento y reportar las eficiencias obtenidas en cada documento para determinar la eficiencia general del software. Esta eficiencia se calculó con la relación entre el número de traducciones realizadas por el software y el número de traducciones identificadas previamente.

Resultados

Para analizar los resultados obtenidos, se describe el resultado de una prueba realizada en un documento. En la Figura 4 se pueden observar las oraciones que contiene un documento reclutado utilizado como entrada al software marcando en color amarillo las partes identificadas que se deben traducir.

1. Los maestros de la escuela están dando clase.
2. Las doctoras están en cirugía.
3. Los perros están ladrando.
4. Los ingenieros trabajan.
5. Los alumnos comen en la cafetería.
6. Los ingenieros dedicados estudian mucho.
7. En la reunión estaban los ingenieros.
8. Aquellos doctores están en cirugía.
9. Las vacas comen pasto.
10. Las clases de los maestros en la universidad son buenas.
11. Se dice que los doctores en este hospital son excelentes.
12. En este trabajo los arquitectos se lucieron.
13. Aquellos alumnos realizaron un excelente proyecto final.
14. Nosotros los licenciados somos los mejores alumnos de la carrera.
15. Estos trabajos son de los alumnos de sistemas.
16. Se dice que aquellos arquitectos son los mejores.
17. Se dice que nosotros los doctores trabajamos mucho.
18. La mejor carrera es de nosotros los ingenieros en sistemas.
19. En este momento los maestros están en curso.
20. Este trabajo queda a cargo de los arquitectos.

Figura 4 Documento de prueba con texto original

Fuente: Elaboración propia

La Figura 5, muestra el resultado obtenido en la traducción realizada por el software, donde en color azul se marcan las partes que se tradujeron correctamente mientras que en color rojo se marcan las partes que no se tradujeron o se tradujeron incorrectamente.

1. Las y los maestros de la escuela están dando clase.
2. Las doctoras están en cirugía.
3. Los perros están ladrando.
4. Las y los ingenieros trabajan.
5. Las y los alumnos comen en la cafetería.
6. Las y los ingenieros dedicados estudian mucho.
7. En la reunión estaban las y los ingenieros.
8. Aquellos doctores están en cirugía.
9. Las vacas comen pasto.
10. Las clases de las y los maestros en la universidad son buenas.
11. Se dice que las y los doctores en este hospital son excelentes.
12. En este trabajo las y los arquitectos se lucieron.
13. Aquellos alumnos realizaron un excelente proyecto final.
14. Nosotros las y los licenciados somos los mejores alumnos de la carrera.
15. Estos trabajos son de las y los alumnos de sistemas.
16. Se dice que aquellas y aquellos arquitectos son los mejores.
17. Se dice que nosotros las y los doctores trabajamos mucho.
18. La mejor carrera es de aquellas y aquellos ingenieros en sistemas.
19. En este momento las y los maestros están en curso.
20. Este trabajo queda a cargo de las y los arquitectos.

Figura 5 Documento generado con la traducción del texto

Fuente: Elaboración propia

Analizando los resultados obtenidos de la prueba mostrada, podemos observar que de 20 oraciones que contiene el documento 18 se tradujeron correctamente y 2 no, lo que conlleva a tener una efectividad del 90% en el software. Este mismo proceso se realizó en los 10 documentos faltantes y el porcentaje de efectividad de cada traducción se puede ver en el Gráfico 1. Analizando los resultados se tiene un promedio de eficiencia de 85.81%, una mediana de 84%, una moda de 84% y una desviación estándar de 3.76%.

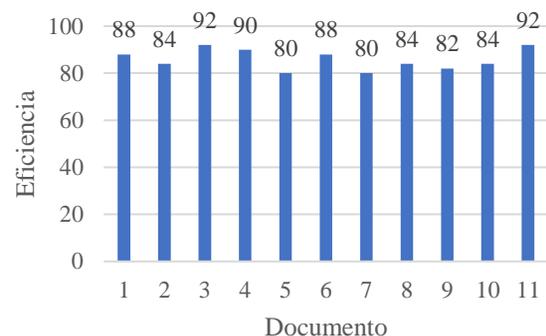


Gráfico 1 Eficiencia de traducción del software en las pruebas realizadas

Fuente: Elaboración propia

Con esta información se puede decir que la traducción de textos que realiza el software es aceptable. Los errores encontrados en las pruebas se debieron principalmente a que no se contemplaron todas las combinaciones posibles en la estructura de una oración en el idioma español ya que una oración puede construirse de diferentes maneras, esto origina que el diseño del autómata se vuelva complejo.

Por ejemplo, en la Figura 5 se puede observar que las oraciones 8 y 13, que no se tradujeron correctamente, empiezan con el pronombre demostrativo plural “Aquellos”, mientras que las oraciones 16 y 18 que tienen el mismo pronombre demostrativo se tradujeron bien por tenerlo dentro de la oración. Otro motivo que originó errores fue que no se contemplaron todos los posibles sustantivos para profesiones u oficios laborales, esto origina que si se trata de traducir una oración que contiene un sustantivo laboral que no se contempló, la traducción no se realiza.

Conclusiones

Con base en los resultados mostrados se puede concluir que el software desarrollado tiene una eficiencia en la traducción aceptable. Sin embargo, es importante mencionar que ante la complejidad del idioma español solamente se diseñó el autómata para implementar dos reglas gramaticales indicadas en el “Manual para el uso no sexista del lenguaje” (Perez Cervera, 2011) para el tratamiento de pronombres demostrativos y los artículos determinados, por lo que aún hay trabajo por hacer para enriquecer el autómata que realiza las traducciones. Ahora si hablamos del software comercial Themis (Themis, 2018), este tiene una mejor eficiencia, sin embargo, se debe considerar dos aspectos relevantes. El primero consiste en el costo de la licencia ya que Themis maneja una licencia anual lo que indica que se debe pagar una cuota anual para tener los servicios que el software ofrece. El segundo aspecto se refiere a la región donde se comercializa Themis puesto que, al ser un software europeo el idioma que utiliza es el español en castellano, lo que provoca que algunas palabras que son utilizadas coloquialmente en esa región sean plasmadas en nuestras redacciones, o viceversa, que palabras que utilizamos coloquialmente en México provoquen que la traducción no se realice correctamente. Por estas razones, el contar con un software que traduzca textos de un lenguaje excluyente y sexista a un lenguaje incluyente y no sexista es relevante ya que permite ayudar y guiar en la redacción de textos incluyentes utilizando el idioma de nuestra región.

Agradecimientos

Se desea agradecer tanto a las ingenieras Sandy Joyseth Sosa Camacho y Osiris Jaqueline Estrada Sánchez como al ingeniero Orlando Ángeles Granados por las aportaciones realizadas en el desarrollo de este software traductor durante su servicio social o residencia profesional.

Trabajo a futuro

Como trabajo a futuro se tiene el rediseño del autómata para implementar más reglas gramaticales indicadas en el “Manual para el uso no sexista del lenguaje”, abarcar las diferentes estructuras que una oración en el idioma español puede tomar y también para enriquecer la gama de sustantivos laborales que debe contemplar. Todo lo anterior para mejorar la traducción de textos excluyentes y sexistas a textos incluyentes y no sexistas.

Referencias

Apache org. (2018). *Apache PDFBox® - A Java PDF Library*. Recuperado el 6 de noviembre de 2018, de Apache PDFBox®: <https://pdfbox.apache.org>

Apache org. (2018). *Apache POI - the Java API for Microsoft Documents*. Recuperado el 6 de noviembre de 2018, de Apache POI: <https://poi.apache.org>

Eckel, B. (2010). *Piensa en Java* (4ª ed.). Madrid, España: Prentice Hall / Pearson.

JFlex Team. (2018). *JFlex The Fast Scanner Generator for Java*. (JFlex Team) Recuperado el 6 de noviembre de 2018, de <http://www.jflex.de>
ONU Mujeres. (18 de mayo de 2017). *Acerca de ONU Mujeres*. Recuperado el 6 de noviembre de 2018, de <http://www.unwomen.org/es/about-us/about-un-women>

ONU Mujeres. (5 de mayo de 2017). *Hechos y cifras: Acabar con la violencia contra mujeres y niñas*. (ONU Mujeres: ¿Qué hacemos? Poner fin a la violencia contra las mujeres) Recuperado el 6 de noviembre de 2018, de <http://www.unwomen.org/es/what-we-do/ending-violence-against-women/facts-and-figures>

ONU Mujeres. (3 de abril de 2017). *ONU Mujeres y el Gobierno de la Ciudad de México lanzan campaña contra el acoso sexual que padecen mujeres y niñas en el transporte público*. (ONU Mujeres: Noticias) Recuperado el 6 de noviembre de 2018, de <http://www.unwomen.org/es/news/stories/2017/4/press-release-un-women-and-government-of-mexico-city-launch-noesdehombres>

Organización Mundial de la Salud. (29 de noviembre de 2017). *Violencia contra la mujer*. Recuperado el 6 de noviembre de 2018, de <http://www.who.int/es/news-room/fact-sheets/detail/violence-against-women>

Perez Cervera, M. (2011). *Manual para el uso no sexista del lenguaje*. Ciudad de México: Comisión Nacional para Prevenir y Erradicar la Violencia contra las Mujeres.

Pressman, R. S. (2010). *Ingeniería del Software: un enfoque práctico*. Ciudad de México: Mc Graw Hill.

Themis. (2018). *Themis: Por un lenguaje no sexista*. (Themis) Recuperado el 6 de noviembre de 2018, de http://www.themis.es/Lenguaje_No_Sexista.html