

## TABLA DE CONTENIDOS

	página
<b>Dedicatoria</b>	<b>I</b>
<b>Agradecimientos</b>	<b>II</b>
<b>Tabla de Contenidos</b>	<b>III</b>
<b>Índice de Figuras</b>	<b>V</b>
<b>Índice de Tablas</b>	<b>VI</b>
<b>Resumen</b>	<b>VII</b>
<b>1. Introducción</b>	<b>8</b>
1.1. Contexto . . . . .	8
1.2. Descripción del Problema . . . . .	9
1.3. Objetivos . . . . .	10
1.3.1. Objetivo General . . . . .	10
1.3.2. Objetivos Específicos . . . . .	10
1.4. Alcances del Proyecto . . . . .	11
<b>2. Antecedentes</b>	<b>12</b>
2.1. Traductores Automáticos . . . . .	12
2.1.1. Componentes lingüísticos de un traductor automático . . . . .	13
2.1.2. Falencias en las traducciones automáticas . . . . .	14
2.2. Recuperación de la Información (RI) . . . . .	16
2.2.1. Problemática de la RI . . . . .	17
2.2.2. Algoritmos de búsqueda para la recuperación de datos . . . . .	20
2.3. Métricas de similitud . . . . .	21
2.3.1. Distancia de Levenshtein . . . . .	21
2.3.2. Distancia de Indel . . . . .	24
2.3.3. Distancia de Damerau . . . . .	25

<b>3. Desarrollo</b>	<b>26</b>
3.1. Metodología de Desarrollo . . . . .	26
3.2. Terminología y Definiciones Básicas . . . . .	29
3.3. Criterios que Determinan la Similitud entre Frases . . . . .	31
3.4. Estrategias de Búsqueda por Similitud . . . . .	33
3.4.1. Búsqueda exacta . . . . .	33
3.4.2. Búsqueda exacta mediante palabras significativas . . . . .	34
3.4.3. Búsqueda aproximada basada en combinaciones de palabras con menor frecuencia de apariciones . . . . .	35
3.4.4. Búsqueda aproximada mediante combinación total de palabras	38
3.4.5. Ejemplo aplicando búsqueda aproximada de combinación total	39
3.4.6. Análisis y selección de estrategias de búsqueda . . . . .	40
3.5. Métricas de Similitud entre Frases . . . . .	43
3.5.1. Ponderación por cantidad de palabras (PL) . . . . .	43
3.5.2. Ejemplo aplicando ponderación por cantidad de palabras (PL)	45
3.5.3. Ponderación por coincidencia (PC) . . . . .	46
3.5.4. Ejemplo aplicando ponderación por coincidencia (PC) . . . . .	46
3.5.5. Ponderación por frecuencia de apariciones en la web (PF) . . . . .	47
3.5.6. Distancia de similitud con transposición (DST) . . . . .	48
3.5.7. Matriz de penalización: . . . . .	48
3.5.8. Cálculo de la mínima distancia: . . . . .	50
3.5.9. Ejemplo aplicando la distancia de similitud con transposición .	52
<b>4. Resultados Experimentales</b>	<b>54</b>
4.1. Pruebas Realizadas a las Métricas de Similitud . . . . .	55
4.2. Comparaciones entre las Métricas de Similitud . . . . .	68
4.3. Análisis Global de las Pruebas Realizadas . . . . .	73
<b>5. Conclusiones y Trabajo Futuro</b>	<b>74</b>
5.1. Conclusiones . . . . .	74
5.2. Trabajo Futuro . . . . .	77

## ÍNDICE DE FIGURAS

	página
1.1. Arquitectura de un sistema de refinamiento de traducciones automáticas.	11
2.1. Arquitectura básica de un sistema de recuperación de la información.	18
2.2. Distancia entre dos palabras utilizando la métrica de Levenshtein. . . . .	22
2.3. Ejemplo de la Distancia de Levenshtein. . . . .	22
2.4. Matriz de programación dinámica para el cálculo de la distancia de Levenshtein. . . . .	24
3.1. Ejemplo de creación de la matriz de penalización para la distancia por similitud con transposición. . . . .	49
3.2. Ejemplo de la matriz de distancia por similitud con transposición . . . . .	52
4.1. Gráfico del posicionamiento en 1er lugar de las frases a encontrar en las pruebas de nivel básico . . . . .	70
4.2. Gráfico del posicionamiento en 1er lugar de las frases a encontrar en las pruebas de nivel medio . . . . .	71
4.3. Gráfico del posicionamiento en 1er lugar de las frases a encontrar en las pruebas de nivel superior . . . . .	72
4.4. Gráfico que indica el total de triunfos y empates entre las métricas de similitud . . . . .	73

## ÍNDICE DE TABLAS

	página
2.1. Algunas ocurrencias del traductor automático LOGO. . . . .	15
2.2. Ejemplos de traducciones automáticas para la frase “ <i>I used to be in love</i> ”. . . . .	16
3.1. Frecuencias de palabras pertenecientes a la frase traducida . . . . .	37
3.2. Datos pertenecientes a las palabras de una frase de entrada para el análisis de las estrategias de búsqueda. . . . .	41
3.3. Comparación entre cuatro estrategias de búsqueda. . . . .	42
3.4. Palabras significativas de una frase de entrada y una frase candidata .	45
4.1. Resultados prueba 1: Inserción . . . . .	55
4.2. Ranking de las 10 mejores frases candidatas de la prueba 1 . . . . .	55
4.3. Resultados prueba 2: Eliminación . . . . .	57
4.4. Ranking de las 10 mejores frases candidatas de la prueba 2 . . . . .	57
4.5. Resultados prueba 3: Sustitución . . . . .	58
4.6. Ranking de las 10 mejores frases candidatas de la prueba 3 . . . . .	59
4.7. Resultados prueba 4: Transposición . . . . .	60
4.8. Ranking de las 10 mejores frases candidatas de la prueba 4 . . . . .	61
4.9. Resultados prueba 5: Inserción y Sustitución . . . . .	62
4.10. Ranking de las 10 mejores frases candidatas de la prueba 5 . . . . .	62
4.11. Resultados prueba 6: Inserción y Transposición . . . . .	64
4.12. Ranking de las 10 mejores frases candidatas de la prueba 6 . . . . .	64
4.13. Resultados prueba 10: Inserción, Transposición y Sustitución . . . . .	66
4.14. Ranking de las 10 mejores frases candidatas de la prueba 10 . . . . .	66
4.15. Resultados de ranking sobre las pruebas realizadas a las métricas de similitud . . . . .	69
4.16. Estadísticas de las pruebas de nivel básico . . . . .	70
4.17. Estadísticas de las pruebas de nivel medio . . . . .	71
4.18. Estadísticas de las pruebas de nivel superior . . . . .	72