PONTIFICIA UNIVERSIDAD CATOLICA DE VALPARAISO
FACULTAD DE INGENIERIA
ESCUELA DE INGENIERIA INFORMATICA

DETERMINACIÓN DE AUTORÍA MEDIANTE GRAFOS

Bryan Vicente Soto Astudillo
Víctor Israel Toledo Arizmendi

Profesor Guía: **Dr**. **Rodrigo Alfaro Arancibia**

Profesor Co-referente: **Dr. Héctor Allende Cid**

INFORME FINAL DEL PROYECTO
PARA OPTAR AL TÍTULO PROFESIONAL DE
INGENIERO CIVIL EN INFORMÁTICA

Marzo 2017

Índice

Lista de Figuras

Lista de Tablas

Resumen

En la siguiente investigación se busca determinar una forma eficiente de determinar la autoría de textos mediante grafos, donde la forma de determinar esto es la división de un conjunto de textos en 2 subconjuntos desiguales en cantidad, los cuales son clasificados como de “knowledge” y “testing”. Cada uno de estos cumple una función fundamental en la determinación de autoría, ya que desde los textos de conocimiento se extraen las bases que mediante algoritmos de clasificación como PageRank se analizan, comparan con los textos de testing y toma la decisión de autor para cada uno de estos.

Como aplicación fundamental de esta investigación está el área de investigación de obras literarias. También puede ayudar en cuanto a la determinación de autores de acoso en las redes sociales, determinar plagios en papers, columnas, revistas, entre muchas otras.

Palabras claves: Grafos, Autoría de textos, Plagio.

Abstract

In the following research, we try to determine an efficient way to determine the authorship of texts using graphs, where the way to determine this is the division of a set of texts into 2 unequal subsets in quantity, which are classified as "knowledge" and "testing". Each one of them fulfills a fundamental function in the determination of authorship, since from the texts of knowledge the bases are extracted that by means of algorithms of classification like PageRank are analyzed, compare with the texts of testing and takes the decision of author for each one of these.

As a fundamental application of this research is the research area of literary works. It can also help in determining the authors of harassment in social networks, determine plagiarism in papers, columns, magazines, among many others.

Key words: Graphs, Authorship of texts, Plagiarism.

1. Introducción

Desde el comienzo de la redacción de obras literarias y el posterior análisis de estas, han llevado a la necesidad de determinar la autoría para textos anónimos y confirmarla en textos en que el autor se ponga en duda. Con el paso del tiempo nace la estilometria como análisis lingüístico de textos que permite identificar patrones únicos, los cuales permiten atribuir la autoría. Además, con la evolución de la tecnología permite automatizar estos procesos permitiendo el surgimiento de los grafos, denominados redes de palabras.

El análisis y la comparación de estos grafos o redes de palabras, permiten mediante métodos y funciones estadísticas construir un grado de similitud entre un texto con autor y el texto prueba. Dentro de este proyecto de investigación, existen diferentes secciones las cuales serán descritas a continuación:

En la sección 2 se presentan objetivos a alcanzar, tanto generales como específicos.

En la sección 3 se define la problemática, además se presentan avances y aplicaciones de esta.

En la sección 4 está el marco teórico donde se presentan conceptos a utilizar, además de técnicas y métodos de determinación de autoría.

En la sección 5, se presenta la metodología de resolución.

En la sección 6 se presenta el plan de pruebas y sus resultados.

En la sección 7 se presenta los futuros trabajos basados en este proyecto.

En las secciones, 8 y 9 respectivamente, se presentan la conclusión y referencias utilizadas.

2. Objetivos

A continuación, se presenta el objetivo general y los objetivos específicos del proyecto.

2.1 Objetivo general

Determinar de forma automática la autoría de textos mediante la comparación de grafos

2.2 Objetivos específicos

- Seleccionar un algoritmo que maximice la eficiencia y efectividad en la comparación de grafos.
- Implementar el algoritmo seleccionado de comparación de grafos.
- Realizar pruebas comparativas con algoritmos clasificadores
- Entregar conclusiones con respecto al rendimiento del algoritmo seleccionado

3. Problema

Dentro de este punto se desarrolla la definición central del problema, además se describen algunos avances y finalmente las aplicaciones de la utilización de grafos en la detección de autoría.

3.1 Definición

Determinar la autoría de un determinado texto constituye un análisis comparativo de la estructura y la sintaxis de un conjunto de textos, permitiendo establecer un grado semejanza entre el texto con autor desconocido y un conjunto de textos previamente analizados. En este proceso se construye un perfil lingüístico del autor basándose en la forma que este escribe, la profundidad de su vocabulario, la omisión o la frecuencia en que utiliza ciertas palabras.

Esta problemática no es nueva y abarca diversas áreas, desde la literatura hasta la criminología, pasando por análisis de comentarios en la internet y por el análisis de autenticidad de escritos académicos.

En este proyecto se buscará la autoría de textos mediante de la construcción de grafos dirigidos que conformen redes de palabras que serán comparadas con textos previamente analizados y procesados el cual permitirá obtener un grado de similitud, entre 0 y 1, el cual indicará a qué autor perteneciera hipotéticamente.

3.2 Avances

Los avances en la determinación de autoría provienen a finales del siglo XVIII cuando en 1787 Edmond Malone propuso, mediante el análisis de la rima de los textos, que Williams Shakespeare no había escrito ninguno de las tres partes de la obra “Henry VI”. Esto puso en duda la autoría por parte de Shakespeare con respecto a sus obras. En 1812, Henry Weber identificó que la obra “Los dos nobles caballeros” de Shakespeare había sido coescrita junto al dramaturgo Inglés John Fletcher. En 1850, James Spedding mediante el estudio de la obra “Henry VI” determinó que el ya mencionado John Fletcher era el colaborador de Shakespeare.

En el año 1939 G. Undy Yule inauguró la nueva era de la determinación de autoría con su publicación “On Sentence-length as Statical Characteristic of Style in Prose” donde analizaba la longitud de las oraciones como patrón de escritura. En 1949, George Zipf encontró dos patrones, el primero indicaba que las palabras cortas eran más utilizadas que las palabras más largas. El segundo, indicaba que un pequeño número de palabras es utilizado con mucha frecuencia mientras que un gran número de palabras es ocupado con baja frecuencia. Uno de los conjuntos de investigaciones más importante fueron las realizadas por Frederick Mosteller y David Wallace en los años 1963, 1964 y 1984 usando la frecuencia de utilización de palabras como indicador para analizar los textos de “The Federalist Papers”.

En la década de los 90, D. Holmes define la Estilometría, conteniendo una gran cantidad de mediciones como el largo de frases y de palabras, la frecuencia de palabras y caracteres junto con funciones para identificar la riqueza del vocabulario.

El avance en las tecnologías de información y la aparición de internet junto con los documentos electrónicos llevo a la aparición de la determinación de autoría basada en computadoras. Esta permite un mayor manejo de volúmenes de textos, desarrollar y utilizar algoritmos que más poderosos y eficientes que maximicen los resultados entregados, además de expandir las áreas de análisis más allá de la literatura como correos electrónicos, mensajes en foros, blogs de opinión, tweets, entre otros.

3.3 Aplicaciones

Entre las principales aplicaciones se encuentran las orientadas al ámbito académico, para determinar plagios o copias indiscriminadas, en el ámbito de políticas de foros web para determinar si un usuario posee más de un perfil o aplicaciones similares para clasificar textos u otros como tweets.

- Detección automática de plagios en textos. (Cedeño, 2008)
- Reconocimiento de acuerdo a los post en foros web. (Solorio, Pillay, Raghavan, & Montes y Gómez, 2011)
- Reconocimiento de autor de acuerdo a distintos tipos de atributos. (Monroy, 2012)
- Reconocimiento de autoría mediante patrones de estilos de escritura. (Amancio, 2015).

4. Marco Teórico

En esta sección se presentan conceptos a utilizar, además de los fundamentos básicos para la construcción y comprensión del concepto de grafos, estilometria y detección de autoría principalmente.

4.1 Planteamiento Inicial

Para comenzar se necesitan definir conceptos básicos como grafos, subgrafos, supergrafos e isomorfismo de grafos.

4.1.1 Grafos

Un grafo es un TDA el cual está compuesto de nodos (Vértices) y un conjunto de conexiones llamadas aristas. Estas aristas establecen relaciones entre los nodos y pueden ser unidireccionales o bidireccionales. Esto determina la dirección de la relación entre los vértices.

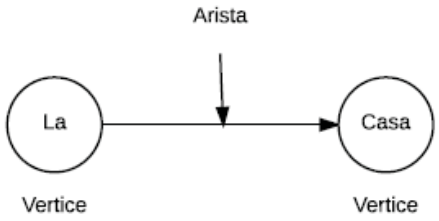

Figura 4.1 Representación básica conexión grafo.

Un grafo es definido como G= (V, E) donde V es un conjunto finito V= {1, 2, 3, 4,..., N} de vértices y E es un conjunto de aristas. Si este nodo tiene atributos este grafo pasa a ser un grafo atribuido definido como G= (V, E, l) donde l es una función de etiquetas

$l: V \rightarrow L_N$.

Si ambos vértices y aristas en un gráfico tienen atributos, el grafo es un grafo atribuido definido por G = (V, E, α, β), donde

$\alpha: V \rightarrow L_N$ y $\beta: E \rightarrow L_E$

Son vértice y arista función de etiquetas. L_N y L_E están restringidos a las etiquetas que consisten en tuplas de tamaño fijo, es decir, $L_N = R^P$, $L_E = R$, p, q E N U {0}.

4.1.2 Subgrafo y supergrafo

Sean G = (V, E, α, β) y G' = (V', E', α', β') dos grafos; G' es un subgrafo de G, G es un supergrafo de G'.
G' ⊆ G, si:

- V' ⊆ V,
- E' ⊆ E',
- α' (x) = α (x), ∀ x ∈ V',
- β' ((x,y)) = β ((x,y)), ∀ (x,y) ∈ E'

Para grafos no atribuidos, solo las dos primeras condiciones son necesarias.

4.1.3 Isomorfismo de grafos

Sean $G_1 = (V_1, E_1, \alpha_1, \beta_1)$ y $G_2 = (V_2, E_2, \alpha_2, \beta_2)$ dos grafos. Un isomorfismo de grafos entre G_1 y G_2 es una biyección entre sus vértices f: $V_1 \rightarrow V_2$ tal que:

- $\alpha_1 (x) = \alpha_2 (f(x)), \forall x \in V_1$,
- $\beta_1 ((x,y)) = \beta_2 ((f(x),f(y))), \forall (x,y) \in E_1$.

Para grafos no atribuidos $G_1' = (V_1', E_1')$ y $G_2' = (V_2', E_2')$, una biyección f: $V_1' \rightarrow V_2'$ tal que $(u,v) \in E'_1 \Leftrightarrow (f(u),f(v)) \in E'_2$, ∀ u, v $\in V_1'$ es un isomorfismo de grafos entre los dos.

Si V1=V2=⌀, entonces f es llamado isomorfismo de grafos vacío.

4.2. Estilometria

La estilometria, corresponde a la aplicación del análisis lingüístico a textos con el objetivo de reconocer patrones de redacción que son plasmadas de forma subconsciente por el autor. De esta forma, cada autor posee un estilo único de redacción, influido por el léxico y el vocabulario que ha adquirido a lo largo de su vida. Mediante el análisis de los textos se reconocen distintos marcadores lexicales de estilo, siendo divididos en dos tipos: la frecuencia de utilización de palabras y la riqueza de vocabulario.

Entre los principales usos se encuentra el que tiene relación a la determinación de autoría para documentos anónimos, la determinación de autoría de secciones de un texto construido por más de un autor y la confirmación de autoría de escritos cuya redacción este en disputa o en entredicho. Otras aplicaciones en que es usada la estilometría están enfocadas a la clasificación de textos, el análisis de estilos lingüísticos, entre otras.

En la actualidad la estilometria se basa en los recursos computacionales para reconocer las características léxicas, sintácticas, semánticas, de caracteres y las referidas a lenguaje y al contenido.

Tabla 4.1 Características Léxicas obtenidas del análisis de estilometría

	Características
Léxico	Basado en tokens(largo de palabras, largo de oraciones, etc)
	Riqueza de vocabulario
	Frecuencia de palabras
	N-gramas de palabras
	Errores
Caracteres	Tipo de caracteres (letras, dígitos, etc)
	N-gramas de caracteres (tamaño fijo)
	N-gramas de caracteres (tamaño variable)
	Métodos de compresión
Sintáctico	Parte del discurso
	Pedazos
	Estructura de frase y sentencia
	Reglas de frecuencia de re-escritura
	Errores
Semántico	Sinónimos
	Dependencias semánticas
	Funcionales

4.3 Atribución de autoría

El problema de atribución de autoría de textos es tan antiguo como la historia de la literatura. Desde la antigüedad la autoría ha sido atribuida en libros tan importantes como el antiguo testamento, cartas y evangelios canónicos, documentos imperiales de roma, entre otros.

La atribución de autoría está asociada a diferentes conceptos, como la estilometria, la cual será explicada dentro de los siguientes puntos. La atribución de autoría, con la estilometria y los grafos, se ligan directamente mediante las redes de palabras, la cual es una de las herramientas esenciales para poder determinar la atribución de autoría.

Hoy en día se investigan nuevas formas de atribuir autoría a través de herramientas que se utilizan en la lingüística forense. Esta investigación busca determinar autoría mediante la utilización de comparación de redes de palabras, mediante algoritmos de comparación de grafos. En esta temática existen 3 grandes métodos de atribución de autoría:

- **Invariante Unitaria**

Este método se consiguió tras estudiar textos de Bacon, Marlowe y Shakespeare, estudiados por Mendenhall en el siglo XIX, además, fueron estudiados los evangelios del nuevo testamento por Mascol.

La idea principal de este método está planteada bajo la relación de la longitud de palabras y la frecuencia de ocurrencias que el autor manifiesta. De las estadísticas obtenidas se buscaban propiedades invariantes del texto (Zipf), y de esta forma se encontraban características invariantes para cada autor.

- **Análisis Multivariante**

Mosteller y Wallace plantearon en 1964 el uso de frecuencia de algunas palabras funcionales como: de, para, con…. A las que, al aplicarles una metodología Bayesiana, podría producir un método fiable de atribución de autoría. Este método es aplicable a cualquier tipo de texto, ya que toma en cuenta palabras independiente del tema que se está tratando.

De esta forma se tiene un documento base, con el que posteriormente se comparan los textos que se quieren analizar y determina semejanza con el texto base.

- **Aprendizaje computacional**

Se utilizan métodos de categorización, como el Support Vector Machine Learning Method (Método de Aprendizaje de Máquina de Soporte Vectorial), este es el método más eficaz. Este método consiste en textos de entrenamiento con vectores numéricos etiquetados, los métodos de aprendizaje marcan límites entre clases para evitar malas clasificaciones.

Además, dentro de la atribución de autoría se barajan ciertas características que son útiles para la aplicación de los métodos y técnicas para la determinación de estas.

- **Sintaxis y partes de la oración**

Esta característica está definida por la frecuencia de utilización de pequeñas frases, partes de la oración o combinaciones entre estas.

- **Palabras funcionales**

Dentro de estas palabras están consideradas pronombres, preposiciones, verbos, conjunciones, entre otras. De estas se analiza su frecuencia en busca de un estilo de escritura.

- **Medidas de complejidad**

Aspectos que parecen únicos de cada autor, incluyendo el número promedio de longitud de palabras.

- **Palabras de contenido**

Analizar las preferencias léxicas del autor y su frecuencia, considerando que los autores de estas características utilizaron palabras en inglés, el ejemplo es el siguiente:

Start-Begin
Large-Big

Cuidando el manejo de textos con diferentes temas y entrenando el sistema para arrojar resultados más eficientes.

- **N-gramas**

Frecuencia de n-gramas para determinar preferencias léxicas, gramáticas y ortográficas del autor. Determina el idioma original del autor y es posible identificar textos similares.

4.2 Técnicas utilizadas

Para poder atribuir autoría se necesita comprender definiciones y conceptos como redes de palabras, además de comprender su correcta representación y propiedades.

4.2.1 Redes de palabras

4.2.1.1 Definición

Una red de palabras, corresponde a un grafo compuesto por palabras unidas por enlaces que las relacionan entre sí. Este grafo se construye de la base de un texto seleccionado, donde las palabras adyacentes son representadas por un enlace. En estas redes no se consideran los signos de puntuación, así que palabras separadas por un punto o coma no se consideran adyacentes, por lo tanto, no habrá un enlace entre ellas.

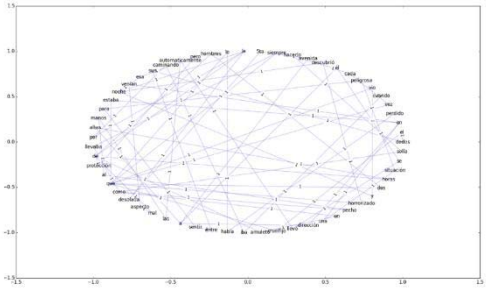

Figura 4.2 Ejemplo grafico de una red de palabras.

4.2.1.2 Representación de una red de palabras

Una red de palabras se puede representar mediante un grafo G(P,E), donde $P = \{w_i\}, (i = 1, \dots, P)$ es el conjunto de P palabras y $E = \left\{\{w_i, w_j\}\right\}$ el número de conexiones entre palabras adyacentes. Estas conexiones entre palabras pueden ser representada por una matriz P_i x P_j, donde cada P_i representa una palabra y cada P_{ij} representa un enlace entre palabras. Si hay un enlace entre la palabra i y la palabra j se representa como $P_{ij} = 1$, sino existe enlace se representa como $P_{ij} = 0$ y si hay más de un enlace se representa $P_{ij} = P_{ij} + 1$.

4.2.1.3 Propiedades de redes complejas

En las redes de palabras construidas se puede observar las siguientes propiedades que se observan en la frecuencia de ocurrencia y la longitud de palabras.

- **Ley de Zipf**

George Zipf demostró que las palabras utilizadas dentro de un texto escrito decaían en una ley de potencia mediante iban disminuyendo en el ranking de aparición, afirmando que un pequeño número de palabras es utilizado con mucha frecuencia, mientras en caso contrario un gran número de palabras es utilizado con poca frecuencia.

La ley de potencia descrita por Zipf está definida como $P_n \sim {}^{1}/_{n^a}$, donde P_n es la frecuencia de una palabra en la posición *n* de ranking y *a* la constante con un valor cercano a uno. Esta ley no dice nada sobre el sentido del texto, pero permite analizar los comportamientos de frecuencia dentro de una red de palabras.

Figura 4.3 Grafica de comportamiento de frecuencia según la Ley de Zipf

4.2.2 Algoritmos de comparación de isomorfismo

4.2.2.1 PageRank

Definido en el año 1998 por Sergei Brin y Lawrence Page, ambos fundadores del motor de búsqueda Google, corresponde a un algoritmo que describe una manera de calcular el ranking o la clasificación de páginas web que existen en la internet.

Este algoritmo representa a la internet en un grafo dirigido donde cada nodo representa un sitio en específico y los vértices cada enlace que referencia o direcciona hacia cada nodo, que a la vez puede ser representada como una matriz no negativa donde cada fila y columna representa las páginas y las casillas corresponden a las referencias o conexiones.

El ranking de cada página se calcula como:

$$Rank(u) = c \sum_{v \in L_u} \frac{Rank(v)}{N_v} + cE(u) \qquad (4.2.2.1.1)$$

Donde N_v corresponde al número de páginas que sale de la página v, L_u es el conjunto de páginas que tiene enlace a u, E(u) es un vector de rango junto con c que es una constante de normalización. Esta fórmula permite calcular el ranking de una determinada página basándose en el ranking de las páginas que las referencia. El vector de rango fue introducido para evitar que el algoritmo entrase en un ciclo infinito en el caso de que una página no tuviese conexiones hacia y desde está.

Definiendo a la constante como d igual a 0.85 y al vector como (1-d) la ecuación anterior queda como:

$$Rank(u) = d \sum_{v \in L_u} \frac{Rank(v)}{N_v} + (1 - d) \qquad (4.2.2.1.2)$$

4.2.2.1.1 PageRank e isomorfismo

La relación que se da entre el cálculo del PageRank y el isomorfismo de grafos dirigidos se basa en el valor del vector de PageRank. Dos grafos A y B serán isomorfos si:

$$PRA[i] = PRB[i], con\ i = 0, \dots, n \quad (4.2.2.1.1.1)$$

Donde $PRA\ y\ PRB$ corresponde al vector PageRank de los grafos A y B.

4.2.2.1.2 Perplejidad

Por definición, corresponde al promedio ponderado de las opciones de una variable aleatoria, se calcula como:

$$2^{H(p)} = 2^{-\sum_n p(x) \log_2 p(x)} \quad (4.2.2.1.2.1)$$

El algoritmo PageRank la convergencia determina la cantidad de iteraciones en el cálculo del ranking de páginas.

5. Metodología de solución

A continuación, se especifican las metodologías de solución al problema de determinación de autoría.

5.1 Determinación de autoría mediante PageRank

En esta metodología, se calcula el valor de PageRank de cada palabra de los textos, lo que permitirá realizar la comparación una por una de estos valores pertenecientes a los textos de "testing" y los textos de "knowledge", considerando la menor diferencia acumulada para determinar el autor.

Está conformada por las siguientes tres etapas:

- Etapa de pre-procesado: En esta etapa, se limpian los textos de los signos de puntuación y se realiza la construcción del diccionario de palabras. En la fase de pruebas, esta metodología tendrá dos configuraciones en el pre-procesado: con Stopwords y sin Stopwords, por lo que se deberá determinar que configuración entrega mejores resultados.

- PageRank: En esta etapa, se procede a calcular el vector de PageRank tanto para los textos de "knowledge" y los textos de "testing" por cada palabra perteneciente al diccionario de datos.

- Determinación de autoría: En esta etapa se procede a calcular las diferencias entre los vectores de PageRank del texto de "testing" y los textos de "knowledge" con la finalidad de seleccionar la menor diferencia acumulada entre los valores del PageRank.

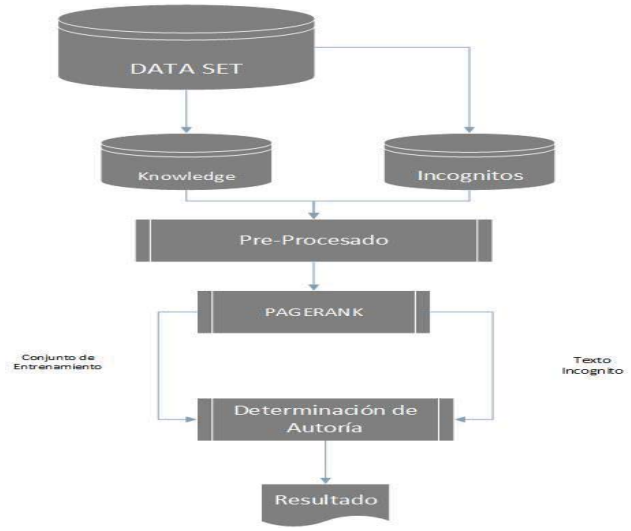

Figura 5.1 Determinación de autoría, con PageRank

5.2 Determinación de autoría mediante PageRank + Algoritmo de clasificación

La principal diferencia entre está metodología y la anterior se encuentra en las etapas de pre-procesado y de determinación de autoría. En la etapa de pre-procesado, utiliza la configuración que mantiene las Stopwords y el cálculo de la etapa de determinación de autoría es reemplazado por un de los algoritmos de clasificación mencionados más adelante.

- Etapa de pre-procesado: En esta etapa se limpian los textos de "testing" y de "knowledge" de signos de puntuación además de la construcción del diccionario de palabras. Como se menciona anteriormente se mantienen las Stopwords.

- PageRank: En esta etapa, se procede a calcular el vector de PageRank tanto para los textos de "knowledge" y los textos de "testing" por cada palabra perteneciente al diccionario de datos.
-
- Algoritmos de clasificación: En esta etapa los vectores de PageRank correspondientes a los textos de "knowledge" y de "testing" son las entradas a los algoritmos de clasificación, los cuales son: SVM, SVM Optimizado (SMO), Naive Bayes, J48 y RNN.

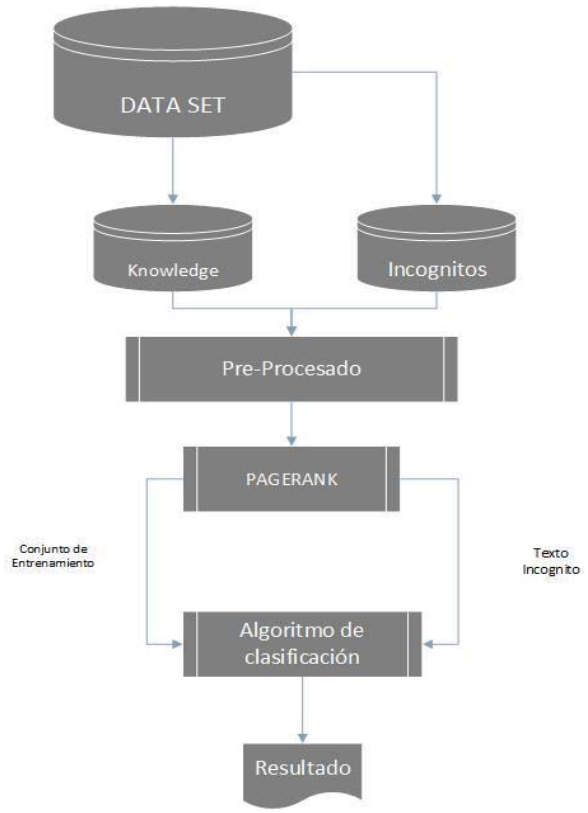

Figura 5.2 Determinación de autoría con algoritmo de clasificación

6. Experimentación

A continuación, se especifica todo lo relativo a las pruebas de experimentación.

6.1 Datos a utilizar

Los siguientes conjuntos de datos corresponde a un set de textos de opinión llamada Guardian Corpus, con un total de 844 textos de distintos tópicos y con 13 autores distintos. Para el caso de las pruebas solo será utilizado una temática en especial, específicamente el tópico de textos que hablan sobre hechos que ocurren en el mundo.

6.1.1 Caracterización de textos

Dentro de la investigación se utilizaron 5 sets de textos, luego utilizando validación cruzada de k iteraciones, se selecciona cierta cantidad de textos al azar siendo considerados como los textos de "testing" y los restantes considerados como textos de "knowledge"

Este proceso se realiza k veces, que en nuestra investigación generó 10 pruebas con los mismos 5 sets de textos, considerando siempre al azar los textos de conocimiento y prueba.

Estos textos están todos escritos en inglés y los autores presentes son los siguientes:

- Catherine Bennett.
- Polly Toynbee.
- Zoe Williams.
- Peter Preston.
- Jonathan Freedland.
- George Monbiot.
- Hugo Young.
- Martin Kettle.

En la siguiente tabla se presentan los sets de textos en combinación con los autores y su correspondiente cantidad de textos por set.

Tabla 6.1 Cantidad de textos por autor

C1	Catherine Bennett	10 Textos
	Polly Toynbee	10 Textos
	Zoe Williams	10 Textos
C2	Catherine Bennett	11 Textos
	Polly Toynbee	12 Textos
	Zoe Williams	14 Textos

C3	George Monbiot	41 Textos
	Jonathan Freedland	67 Textos
	Peter Preston	66 Textos
C4	Catherine Bennett	11 Textos
	George Monbiot	41 Textos
	Jonathan Freedland	67 Textos
	Peter Preston	66 Textos
	Polly Toynbee	12 Textos
	Zoe Williams	14 Textos
C5	George Monbiot	41 Textos
	Hugo Young	35 Textos
	Jonathan Freedland	67 Textos
	Martin Kettle	36 Textos
	Peter Preston	66 Textos

6.1.2 Descripción set C1

Este set de textos está compuesto de 3 autores que suman una totalidad de 30 textos de los cuales al realizar las pruebas se toman 9 textos de "testing" al azar, considerando tomar 3 textos de cada autor.

Este set de textos se ve descrito por la cantidad de palabras que tiene cada uno de los textos, las stopwords y los textos limpiados de estas últimas. En la siguiente tabla se describe el set mediante los promedios de palabras por cada autor y el promedio total de palabras (Detalle en el Anexo A).

Tabla 6.2 Descripción set C1

Autor	**Cantidad promedio de palabras**	**Cantidad promedio de Stopwords**	**Cantidad promedio de palabras S/ Stopwords**
Catherine Bennett	1046	472	574
Polly Toynbee	1495	651	844
Zoe Williams	785	389	395
Promedio Total	**1109**	**504**	**605**

Además, gracias a los datos podemos obtener su desviación estándar, la cual nos da a conocer que de la totalidad de textos solo 2 están fuera del rango, ósea que tienen muchas palabras o muy pocas, los otros 28 se encuentran dentro de la normalidad de la población.

6.1.3 Descripción set C2

Este set de textos está compuesto de 3 autores que suman una totalidad de 37 textos de los cuales al realizar las pruebas se toman 9 textos de "testing" al azar, considerando tomar 3 textos de Catherine Bennett, 3 textos de Polly Toynbee y 4 de Zoe Williams (Detalle en el Anexo A).

De estos textos se puede extraer la misma tabla del set anterior, la cual se presenta a continuación:

Tabla 6.3 Descripción set C2

Autor	Cantidad promedio de palabras	Cantidad promedio de Stopwords	Cantidad promedio de palabras S/ Stopwords
Catherine Bennett	997	452	545
Polly Toynbee	1446	632	814
Zoe Williams	799	394	405
Promedio Total	**1068**	**489**	**579**

Además, gracias a los datos podemos obtener su desviación estándar, la cual nos da a conocer que de la totalidad de textos solo 2 están fuera del rango, ósea que tienen muchas palabras o muy pocas, los otros 35 se encuentran dentro de la normalidad de la población.

6.1.3 Descripción set C4

Este set de textos está compuesto de 3 autores que suman una totalidad de 174 textos de los cuales al realizar las pruebas se toman 52 textos de "testing" al azar, considerando tomar 12 textos de George Monbiot, 20 textos de Jonathan Freedland y 20 de Peter Preston (Detalle en el Anexo A).

De estos textos se puede extraer la misma tabla del set anterior, la cual se presenta a continuación:

Tabla 6.4 Descripción set C3

Autor	Cantidad promedio de palabras	Cantidad promedio de Stopwords	Cantidad promedio de palabras S/ Stopwords
George Monbiot	1078	498	580
Jonathan Freedland	1184	545	638
Peter Preston	989	438	551
Promedio Total	**1085**	**493**	**592**

Además, gracias a los datos podemos obtener su desviación estándar, la cual nos da a conocer que de la totalidad de textos 36 están fuera del rango, ósea que tienen muchas palabras o muy pocas, los otros 138 se encuentran dentro de la normalidad de la población. Por lo tanto, casi el 80% de los textos se encuentra en este rango.

6.1.4 Descripción set C4

Este set de textos está compuesto de 6 autores que suman una totalidad de 211 textos de los cuales al realizar las pruebas se toman 62 textos de "testing" al azar, considerando tomar 3 textos de Catherine Bennett, 3 textos de Polly Toynbee, 4 textos de Zoe Williams, 12 textos de George Monbiot, 20 textos de Jonathan Freedland y 20 de Peter Preston (Detalle en el Anexo A).

De estos textos se puede extraer la misma tabla del set anterior, la cual se presenta a continuación:

Tabla 6.5 Descripción set C4

Autor	Cantidad promedio de palabras	Cantidad promedio de Stopwords	Cantidad promedio de palabras S/ Stopwords
Catherine Bennet	997	452	545
Polly Toynbee	1446	632	814
Zoe Williams	799	394	405
George Monbiot	1078	498	580
Jonathan Freedland	1184	545	638
Peter Preston	989	438	551
Promedio Total	**1082**	**492**	**590**

Además, gracias a los datos podemos obtener su desviación estándar, la cual nos da a conocer que de la totalidad de textos 25 están fuera del rango, ósea que tienen muchas palabras o muy pocas, los otros 186 se encuentran dentro de la normalidad de la población. Por lo tanto, casi el 80% de los textos se encuentra en este rango.

6.1.5 Descripción set C5

Este set de textos está compuesto de 5 autores que suman una totalidad de 245 textos de los cuales al realizar las pruebas se toman 72 textos de "testing" al azar, considerando tomar 12 textos de George Monbiot, 20 textos de Jonathan Freedland, 20 de Peter Preston, 10 de Hugo Young y 10 de Martin Kettle (Detalle en el Anexo A).

De estos textos se puede extraer la misma tabla del set anterior, la cual se presenta a continuación:

Tabla 6.6 Descripción set C5

Autor	Cantidad promedio de palabras	Cantidad promedio de Stopwords	Cantidad promedio de palabras S/ Stopwords
George Monbiot	1078	498	580
Jonathan Freedland	1184	545	638
Peter Preston	989	438	551
Hugo Young	1152	538	614
Martin Kettle	943	438	504
Promedio Total	**1074**	**492**	**582**

Además, gracias a los datos podemos obtener su desviación estándar, la cual nos da a conocer que de la totalidad de textos 47 están fuera del rango, ósea que tienen muchas palabras o muy pocas, los otros 198 se encuentran dentro de la normalidad de la población. Por lo tanto, casi el 80% de los textos se encuentra en este rango.

6.2 Experimentos

Los experimentos se realizaron mediante la utilización de Java, para el pre-procesado y la determinación de autoría y Python para el cálculo del PageRank.

Además, se utilizó la herramienta Weka de la Universidad de Waikato de Nueva Zelanda, para comparar los resultados de la primera metodología con otros clasificadores y cómo determinador de autoría de la segunda metodología.

Los clasificadores utilizados fueron:

- Naive Bayes,
- SVM,
- SMO (SVM optimizado),
- J48.
- RNN

Pruebas fueron realizadas en un computador con las siguientes características:

Procesador: Intel Core i7-5500 5ta Generación.
Memoria RAM: 8 Gb de RAM, 2400 GHz de frecuencia
Almacenamiento: Disco Duro 1 TB a 7200 RPM

6.3 Métricas utilizadas

Las métricas utilizadas para analizar y exponer los resultados de las pruebas son las siguientes:

- TPRate: Correspondiente a la tasa en la que un autor fue identificado correctamente.
- FNRate: Correspondiente a la tasa en la que un autor fue identificado incorrectamente.
- Recall: Corresponde a la tasa en que un autor fue identificado correctamente, siendo igual a la TPRate.
- Precision: Corresponde a que tan preciso es nuestra tasa de Recall por cada autor.
- F-Measure: Corresponde a la medida de precisión que tienen nuestras pruebas, siendo calcula como 2* $\frac{Recall * Precision}{Recall + Precision}$

6.4 Resultados

A continuación, se expondrán los resultados de las pruebas primeramente sin considerar a las Stopwords, después se darán a conocer los resultados de las pruebas considerando las Stopwords para luego realizar una comparación entre ambos casos. De esta comparación se seleccionará la configuración que tenga mejores resultados para ser comparadas con otros clasificadores.

6.4.1 Pruebas iterativas sin stopwords

Estás pruebas fueron desarrolladas bajo un cross-validation de 10 iteraciones (10-fold cross-validation).

6.4.1.1 Pruebas iterativas conjunto C1

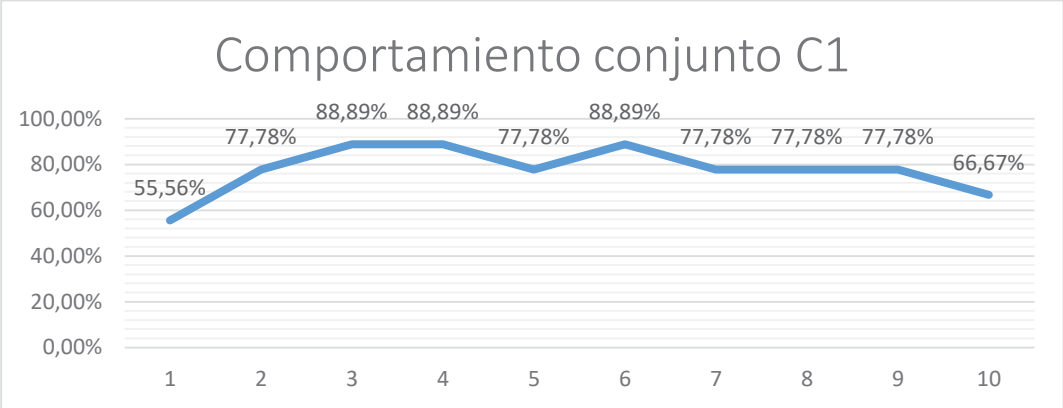

Figura 6.11 Resultado pruebas iterativas conjunto C1.

El promedio de acierto de estas pruebas fue 77.78% con una varianza de 0,010973937

Tabla 6.7 Métricas por autores de C1

Autores	TPRate	FNRate	Recall	Precision	F-Measure
Catherine Bennet	0,5	0,5	0,5	1	0,66666667
Polly Toynbee	1	0	1	0,61785714	0,76379691
Zoe Williams	0,833333333	0,166666667	0,833333333	1	0,90909091

Cómo se observa en el gráfico y en su varianza, el porcentaje de acierto se mueve entre el 77.78% y el 88.89%, salvo la P1 y la P10, cuyos valores son bajo el 67%. Por autor, Polly Toynbee fue el que mayor acierto tuvo, de un 100%, pero tuvo menor Precision debido a que todos los autores mal clasificados fueron asignados a este.

6.4.1.2 Pruebas iterativas conjunto C2

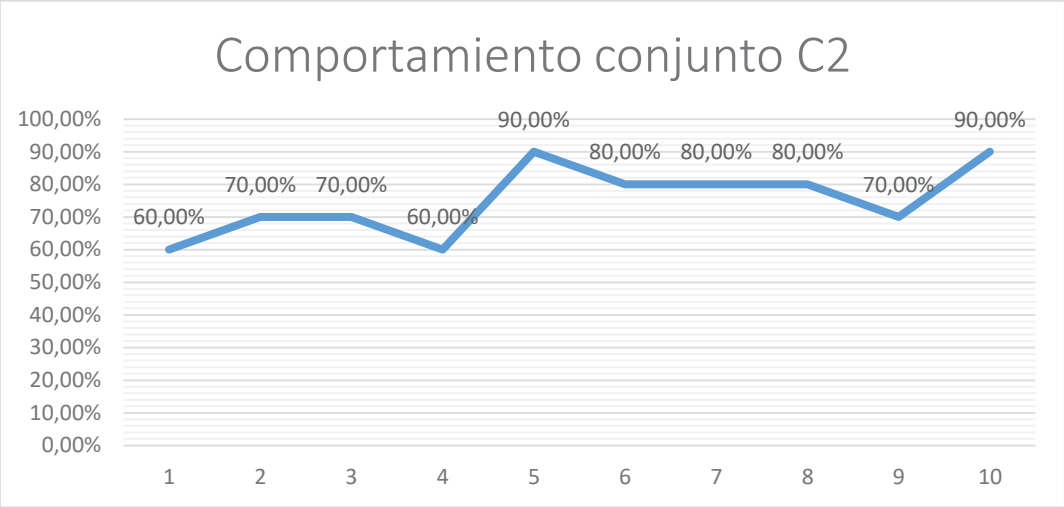

Figura 6.12 Resultado comportamiento conjunto C2

El promedio de acierto de estas pruebas fue de 75,00% con una varianza de 0,011666667

Tabla 6.8 Métricas por autores de C2

Autores	**TPRate**	**FNRate**	**Recall**	**Precision**	**F-Measure**
Catherine Bennet	0,466666667	0,533333333	0,466666667	1	0,63636364
Polly Toynbee	0,878787879	0,121212121	0,878787879	0,63571429	0,73774475
Zoe Williams	0,775	0,225	0,775	0,92	0,84129794

Como se observa en el gráfico y su varianza, el comportamiento de las pruebas fue más variable del anterior teniendo un máximo de 90% de acierto y un mínimo de un 60%. Por autor, nuevamente Polly Toynbee tuvo una baja Precision debido al ser el principal autor con asignaciones erróneas, pero también se observa que Zoe Williams disminuye su Precision, debido a que en 4 de 10 iteraciones fue asignado erróneamente un texto. Además, Catherine Bennet obtiene la tasa de acierto más baja de todas las pruebas sin Stopwords, con un 0,466666667.

6.4.1.3 Pruebas iterativas conjunto C3

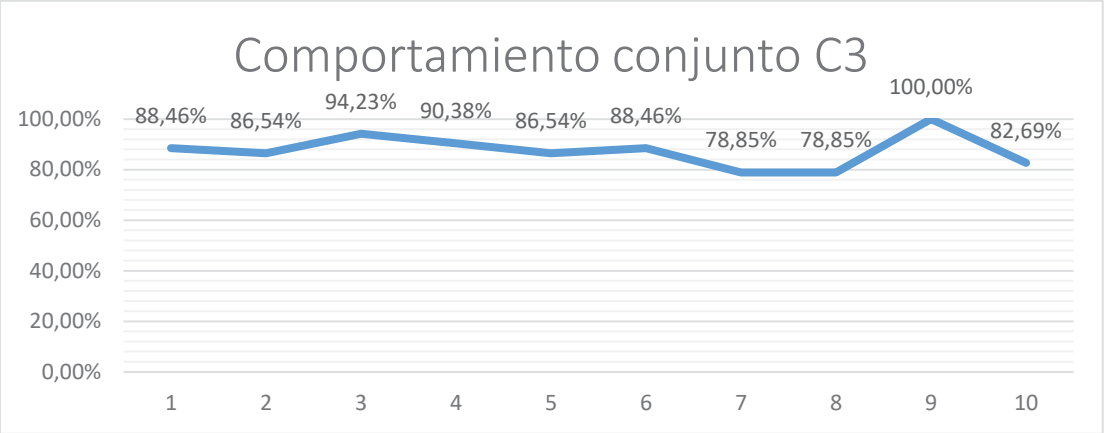

Figura 6.13 Resultado comportamiento conjunto C3.

El promedio de acierto de estas pruebas fue de 87,50% con una varianza de 0,00429405

Tabla 6.9 Métricas por autores de C3

Autores	**TPRate**	**FNRate**	**Recall**	**Precision**	**F-Measure**
George Monbiot	0,866666667	0,133333333	0,866666667	0,94153846	0,90255247
Jonathan Freedland	0,95	0,05	0,95	0,83933826	0,89124719
Peter Preston	0,805	0,195	0,805	0,98095238	0,88430876

Como se observa en el gráfico y en su varianza, los resultados de este conjunto fueron menos variables que los dos conjuntos anteriores. El acierto no bajo de un 78.85% y tuvo un pico de un 100%. Por autor, Jonathan Freedland fue quien tuvo mayor TPRate, sin embargo, fue quien presento menor Precision, siendo al que más se le acciono un texto de manera equivoca.

6.4.1.4 Pruebas iterativas conjunto C4

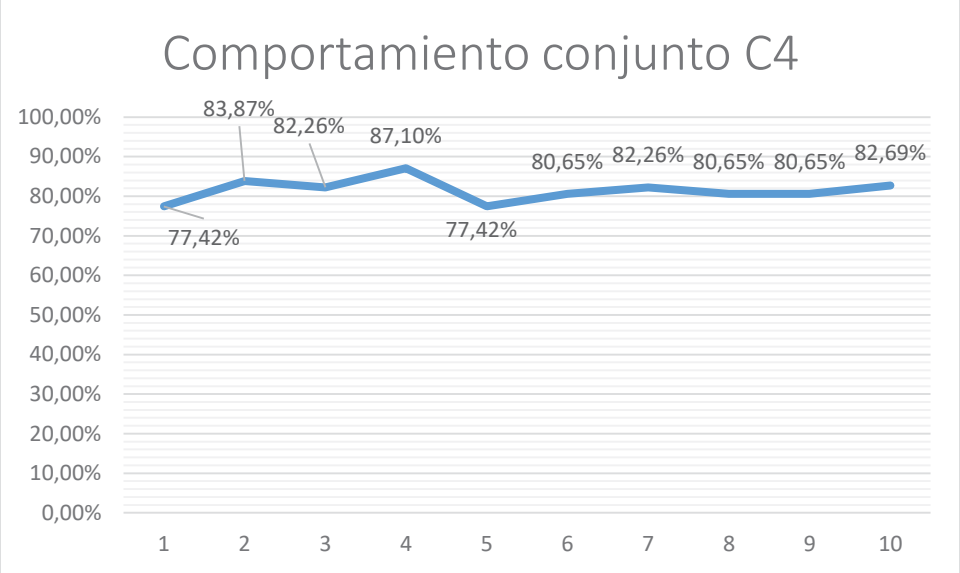

Figura 6.14 Resultado comportamiento C4

El promedio de acierto de estas pruebas fue de 81,50% con una varianza de 0,000833462

Tabla 6.10 Métricas por autores de C4

Autores	TPRate	FNRate	Recall	Precision	F-Measure
Catherine Bennet	0,333333333	0,666666667	0,333333333	1	0,5
Polly Toynbee	0,558333333	0,441666667	0,558333333	0,56666667	0,56246914
Zoe Williams	0,6	0,4	0,6	0,9962963	0,74895592
George Monbiot	0,833333333	0,166666667	0,833333333	0,92642191	0,87741549
Jonathan Freedland	0,96	0,04	0,96	0,69695456	0,80759774
Peter Preston	0,77	0,23	0,77	0,9226848	0,83945611

Como se observa en el gráfico y en su varianza, su comportamiento es el de más baja variabilidad comparado con los otros conjuntos de pruebas. Entre el mínimo y máximo acierto solo hay un 10% de diferencia. Por autores, Jonathan Freedland presenta resultados similares al conjunto C3 con 93% de acierto, pero disminuyendo su Precision. En este caso la menor Precision fue Polly Toynbee debido a que fue el autor con mayor asignación de autoría errónea.

6.4.1.5 Pruebas iterativas conjunto C5

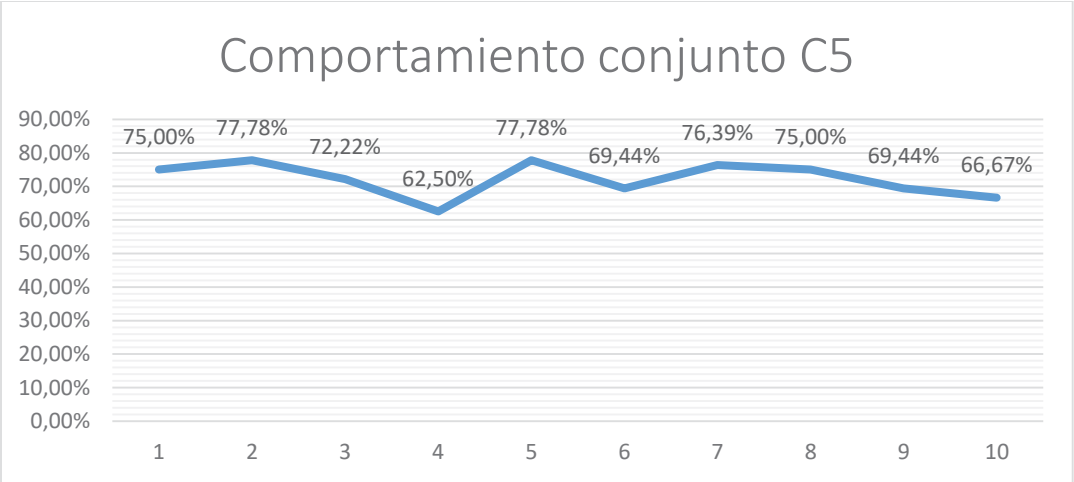

Figura 6.15 Resultado comportamiento C5

El promedio de acierto de estas pruebas fue de 72,22% con una varianza de 0,002614883

Tabla 6.11 Métricas por autores de C5

Autores	TPRate	FNRate	Recall	Precision	F-Measure
George Monbiot	0,8	0,2	0,8	0,95517094	0,8707263
Jonathan Freedland	0,865	0,135	0,865	0,73340999	0,79378838
Peter Preston	0,635	0,365	0,635	0,94869658	0,76077998
Hugo Young	0,75	0,25	0,75	0,47080029	0,57847335
Martin Kettle	0,49	0,51	0,49	0,65416667	0,5603059

Como se observa en el gráfico y en su varianza, los valores de acierto de este caso de prueba se encuentran entre un 77.78% el máximo y un 62.50% el mínimo. Por autor, se observa que Martin Kettle obtuvo una baja TPRate de 0,49 además de tener la 2da peor Precision de esté conjunto de pruebas. Hugo Young también presenta una baja Precision, obteniendo el peor valor de las pruebas sin Stopwords.

6.4.2 Pruebas iterativas con Stopwords

Estás pruebas fueron desarrolladas bajo un cross-validation de 10 iteraciones (10-fold cross-validation).

6.4.2.1 Prueba iterativas conjunto C1

Figura 6.16 Resultado comportamiento C1

El promedio de acierto de estas pruebas fue de 97,78% con una varianza de 0,002194787

Tabla 6.12 Métricas por autores de C1

Autores	TPRate	FNRate	Recall	Precision	F-Measure
Catherine Bennet	1	0	1	1	1
Polly Toynbee	0,933333333	0,06666667	0,93333333	1	0,96551724
Zoe Williams	1	0	1	0,925	0,96103896

Como se observa en el gráfico y en su varianza, los valores varían entre 100% el máximo y 88,89% el mínimo. Por autor, se observa que Catherine Bennet y Zoe Williams tuvieron el 100% de acierto a lo largo de las 10 iteraciones. A diferencia, la Precision de Zoe Williams fue 0,95 debido a que en la iteración 4 y 8 se le fue asignado incorrectamente un texto.

6.4.2.2 Prueba iterativas conjunto C2

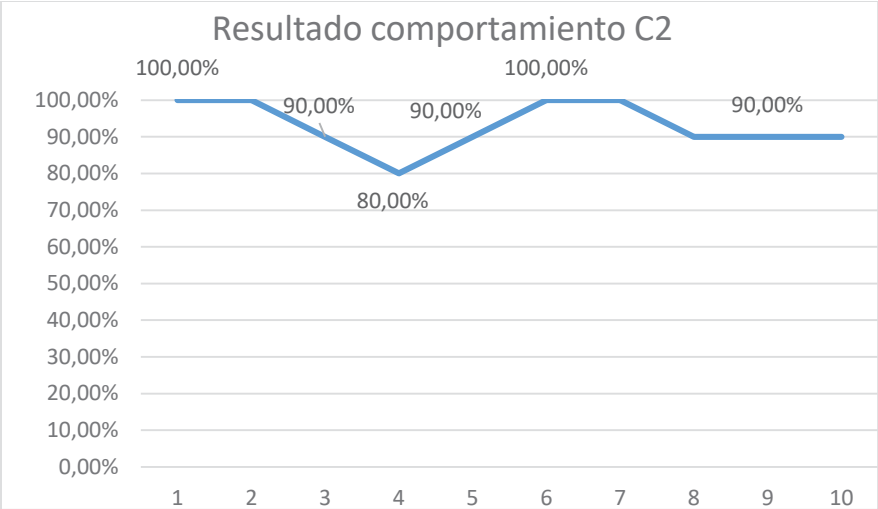

Figura 6.17 Resultado comportamiento C2

El promedio de acierto de estas pruebas fue de 93,00% con una varianza de 0,004555556

Tabla 6.13 Métricas por autores de C2

Autores	TPRate	FNRate	Recall	Precision	F-Measure
Catherine Bennet	0,933333333	0,06666667	0,93333333	1	0,96551724
Polly Toynbee	0,833333333	0,16666667	0,83333333	1	0,90909091
Zoe Williams	1	0	1	0,86666667	0,92857143

Como se observa en el gráfico y en su varianza, la varianza aumenta en comparación al conjunto C1, variando los valores de un 80% el mínimo y un 100% el máximo. Por autor, se observa que Zoe Williams mantiene el 100% de acierto, pero empeorando su Precisión debido a que fue el único autor con asignaciones equivocas. También, se observa la disminución del TPRate de Catherine Bennet y Polly Toynbee en comparación al anterior, pero se mantiene su Precision.

6.4.2.3 Prueba iterativas conjunto C3

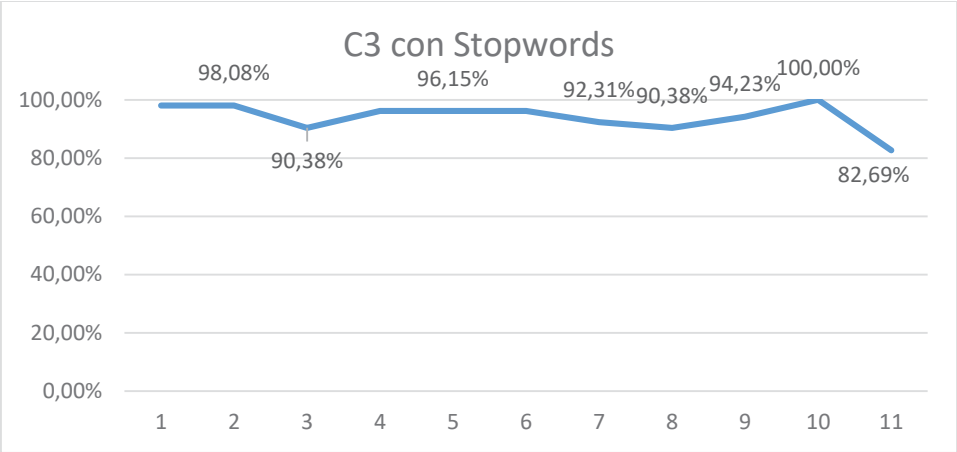

Figura 6.18 Resultado comportamiento C3

El promedio de acierto de estas pruebas fue de 93,65% con una varianza de 0,002469592

Tabla 6.14 Métricas por autores de C3

Autores	TPRate	FNRate	Recall	Precision	F-Measure
George Monbiot	0,883333333	0,11666667	0,88333333	0,98461538	0,93122855
Jonathan Freedland	1	0	1	0,88792584	0,94063636
Peter Preston	0,904545455	0,09545455	0,90454545	0,98138528	0,9414

Como se observa en el gráfico y en su varianza, la varianza disminuye a diferencia a C2 observándose valores de acierto entre 82.69% el mínimo y el 100% el máximo. Por autor, Jonathan Freedland obtuvo el 100% de acierto en las 10 iteraciones, pero teniendo la peor Precision de este conjunto que a diferencia de los anteriores conjuntos, estas determinaciones de autor equivocas se distribuyeron entre los 3 autores.

6.4.2.4 Prueba iterativas conjunto C4

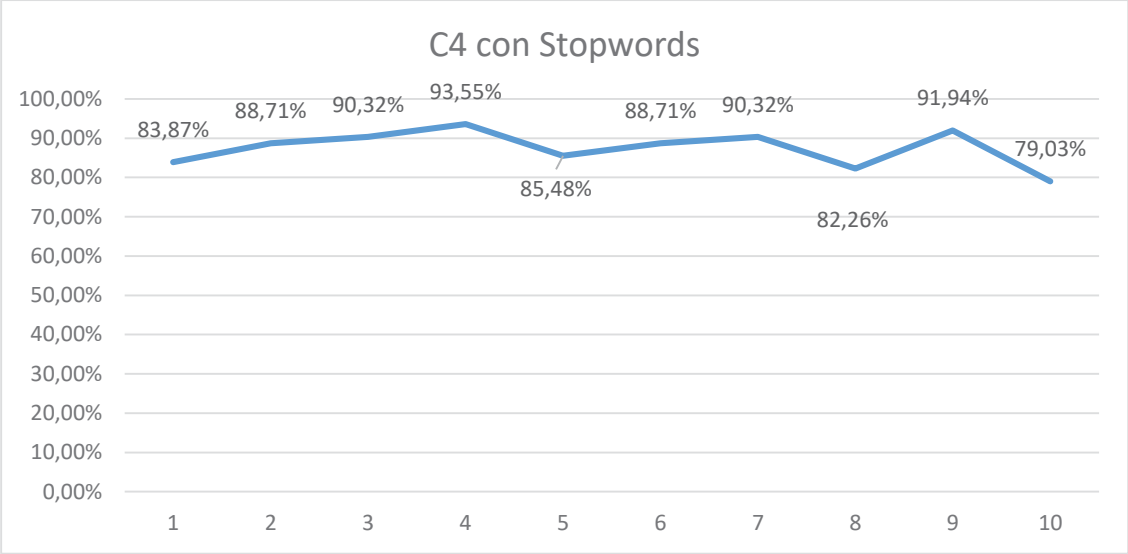

Figura 6.19 Resultado comportamiento C4

El promedio de acierto de estas pruebas fue de 87,42% con una varianza de 0,002127414

Tabla 6.15 Métricas por autores de C4

Autores	TPRate	FNRate	Recall	Precision	F-Measure
Catherine Bennet	0,666666667	0,33333333	0,66666667	1	0,8
Polly Toynbee	0,2	0,8	0,2	1	0,33333333
Zoe Williams	0,85	0,15	0,85	0,96	0,90165746
George Monbiot	0,833333333	0,16666667	0,83333333	0,98461538	0,90267983
Jonathan Freedland	1	0	1	0,78319602	0,87841831
Peter Preston	0,91	0,09	0,91	0,92910785	0,91945466

Como se observa en el gráfico y en su varianza, la variancia es similar al conjunto C3 teniendo un 91.94% de máximo y un mínimo de 82.26%. Por autor, se observa el peor TPRate promedio de todas las pruebas con un 0,2 para Polly Toynbee. Nuevamente Jonathan Freedland obtiene el 100% de acierto para las 10 iteraciones también teniendo la peor Precision, que salvo este autor este indicador no baja del 0.92.

6.4.2.5 Prueba iterativas conjunto C5

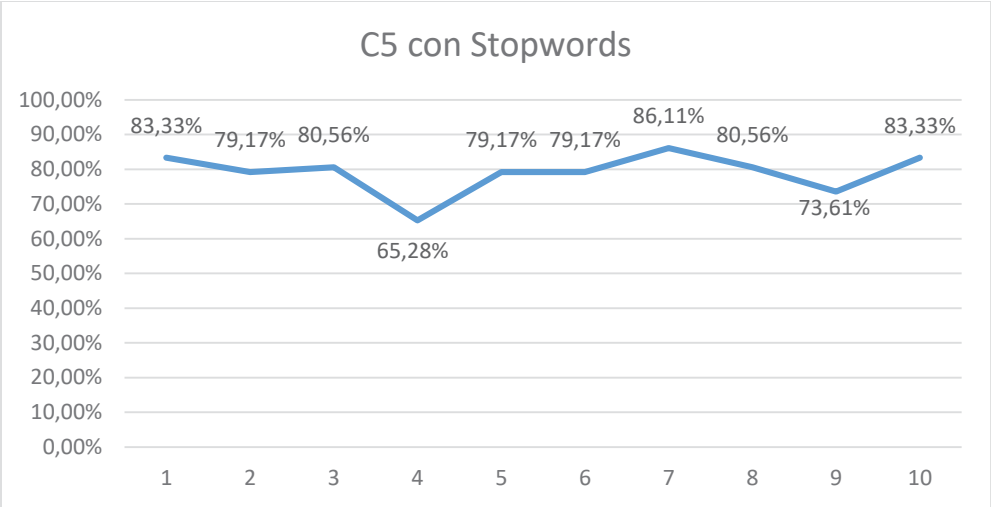

Figura 6.20 Resultado comportamiento C5

El promedio de acierto de estas pruebas fue de 79,03% con una varianza de 0,003448645

Tabla 6.16 Métricas por autores de C5

Autores	TPRate	FNRate	Recall	Precision	F-Measure
George Monbiot	0,725	0,275	0,725	0,98571429	0,83549061
Jonathan Freedland	0,855	0,145	0,855	0,8317632	0,84322154
Peter Preston	0,915	0,085	0,915	0,95324675	0,93373188
Hugo Young	0,81	0,19	0,81	0,69937943	0,75063609
Martin Kettle	0,47	0,53	0,47	0,77024544	0,58378019

Como se observa en el gráfico y en su varianza, la varianza aumento con respecto a C4 teniendo valores entre 65.28% el mínimo y un 86.11% el máximo. Por autor, se observa que Martin Kettle nuevamente no supera el 50% de la misma forma que el C5 sin Stopwords, por lo que la poca cantidad de palabras de sus textos puede ser un factor a considerar, también siendo el segundo autor con más asignaciones de autoría equivocas, solo siendo superado por Hugo Young.

6.4.3 Comparativa de resultados de pruebas con Stopwords y sin Stopwords

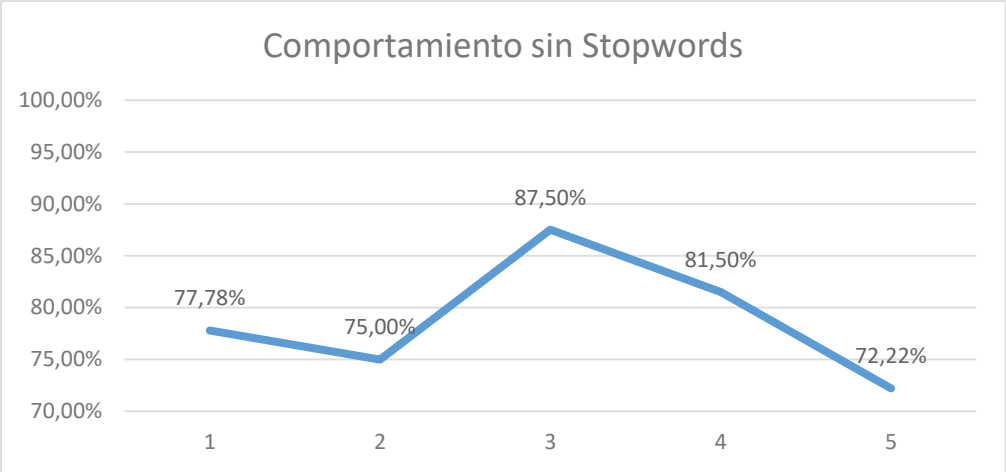

Figura 6.21 Resultados sin Stopwords

En este gráfico se observa el comportamiento sin una tendencia clara de las pruebas sin Stopwords a lo largo de los 5 conjuntos de pruebas.

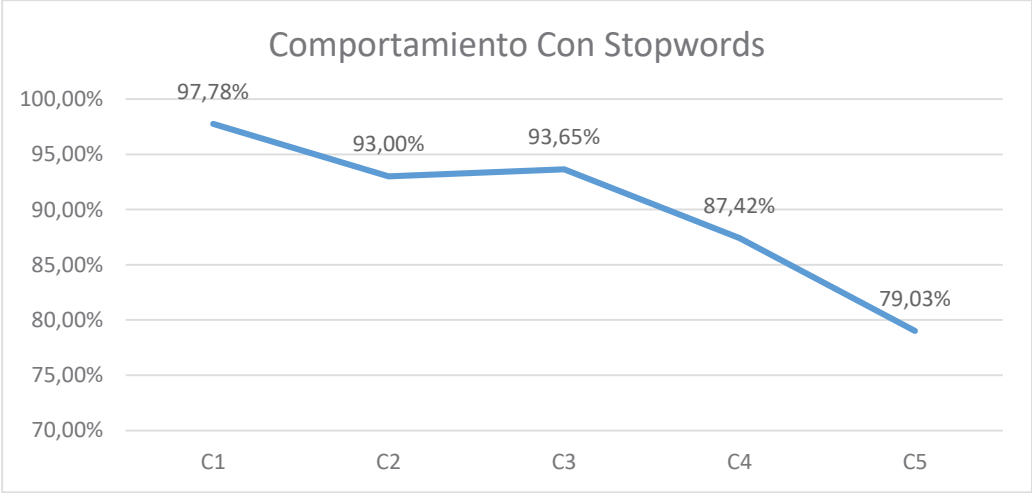

Figura 6.22 Resultados con Stopwords

Mientras que, en este gráfico, se observa una tendencia a la disminución del porcentaje de acierto a lo largo que aumenta la cantidad de textos de “knowledge” y textos de “testing”. En ambos gráficos el comportamiento de C3 a C5 es de una disminución en proporciones similares, mientras que de C1 a C2 el aumento de palabras con la inclusión de las Stopwords permite mejores resultados en términos de acierto y observar una tendencia.

Por lo tanto, se concluye que utilizar Stopwords es la configuración optima al momento de realizar el Pre-Procesado del PageRank.

6.4.4 Comparativa resultados PageRank con Stopwords y otros clasificadores

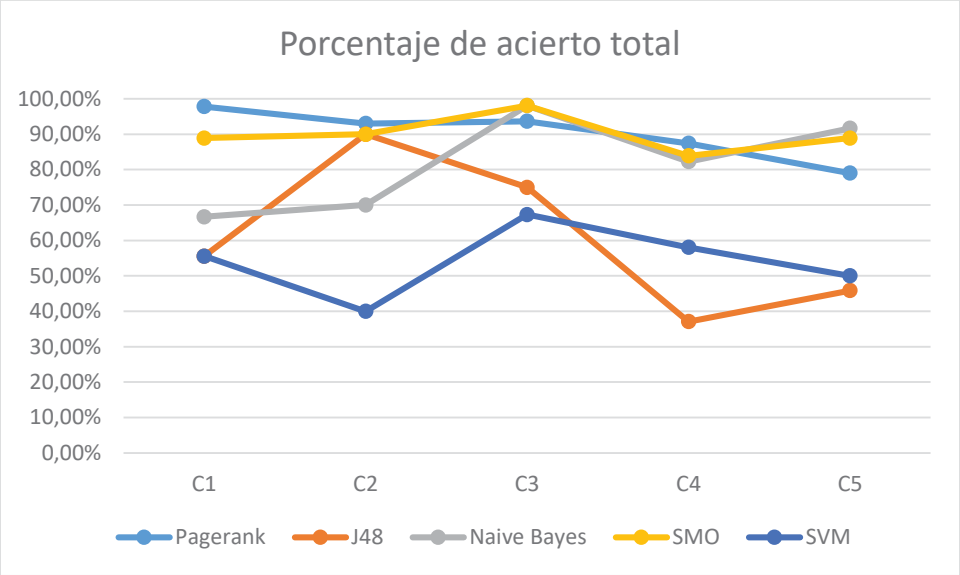

Figura 6.23 Porcentaje de acierto promedio total

En cuanto la comparación de la metodología de PageRank y otros clasificadores se observa que la metodología utilizada en este proyecto posee una tendencia decreciente a lo largo que aumenta la cantidad de textos. En cuanto porcentaje de aciertos PageRank obtiene resultados similares a SMO ambos no menores al 78%. Para Naive Bayes y SMO sus porcentajes de aciertos aumentan a medida que aumentan la cantidad de textos de "knowledge" y de "testing".

En conclusión, la metodología PageRank con la configuración con Stopwords en el Pre-Procesado obtiene resultados que permiten que está sea utilizada cuando la cantidad de textos a procesar de "knowledge" y de "testing" es reducida, obteniendo mejores resultados que otros clasificadores cuyo porcentaje de acierto aumenta a medida que la cantidad de textos a procesar aumenta.

6.4.5 Resultados metodología PageRank + Otros clasificadores.

Está metodología fue pensada como una alternativa a la metodología PageRank con Stopwords donde en vez de calcular la diferencia entre vectores de PageRank, estos entraban a un clasificador proporcionado con Weka.

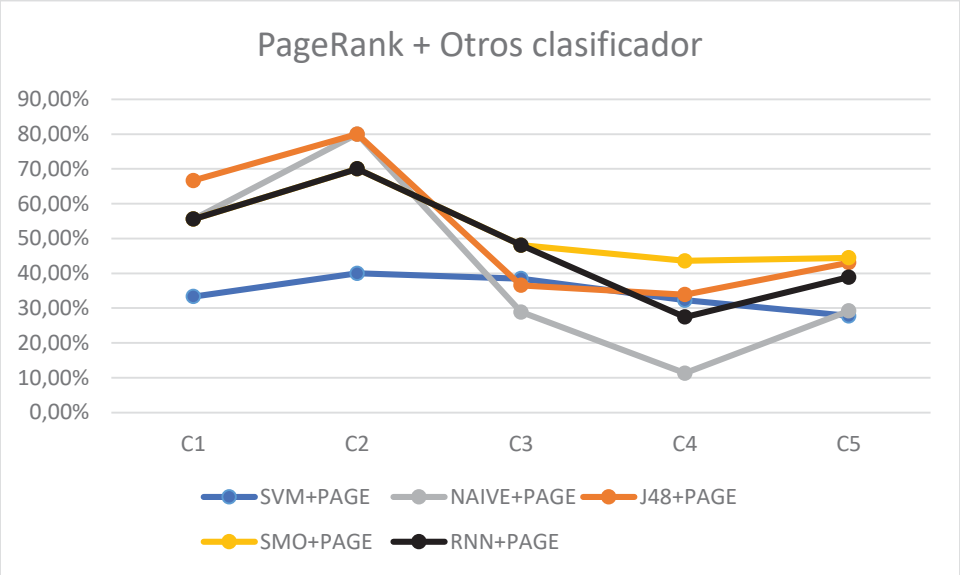

Figura 6.24 Porcentaje de acierto PageRank + Otro clasificador

Se observa a través del gráfico que entre C1 a C2 se observa un aumento en el porcentaje de acierto, siendo esté el valor más algo para todas las combinaciones. Entre todas las combinaciones la única que presento baja variabilidad fue PageRank + SVM y una tendencia de descendente entre C2 a C5, en cambio las otras metodologías se observa un aumento en el porcentaje de acierto entre C4 a C5.

En comparación a la metodología de PageRank sin clasificador y los otros clasificadores en solitario, el porcentaje de acierto disminuyo, por lo que se puede concluir que está metodología es contraproducente para los resultados obtenidos solo con PageRank y de los otros clasificadores por si solos.

7. Trabajos Futuros

Entre los posibles trabajos futuros observados para nuestro proyecto, se consideran los siguientes:

- Implementar pruebas para conjuntos de datos con temas de sus textos variados, debido a que en esta investigación se consideró textos en los que se escribían de temas similares.

- Revisar las configuraciones del PageRank, con la finalidad de que los resultados obtenidos en este proyecto mejoren.

- Probar con otros algoritmos de clasificación que combinados con PageRank entreguen mejores resultados que los obtenidos en este proyecto.

8. Conclusiones

Basándose en lo anteriormente presentado, determinamos que se necesitan una gran variedad de conceptos y conocimientos para poder construir un determinador de autoría, con un rango alto de asertividad. Existe más de una manera de llegar a una solución acertada dependiendo del algoritmo comparativo de grafos que se pueda implementar, variando principalmente en los tiempos de ejecución.

En la actualidad existen distintas formas de determinar autoría, dentro de las que están las redes neuronales, máquinas de soporte vectorial, entre otras. También variando en el uso de la estilometria en la determinación de estilos literarios, categorización de textos, publicaciones, tweets, posts, emails, entre otros.

La utilización de PageRank no se encuentra acotada al internet, sino que también puede ser utilizada para el cálculo de isomorfismo entre grafos dirigidos compuesto por palabras, posiciones geográficas entre otros. Dentro de lo que se puede determinar con PageRank es la dispersión de resultados obtenidos, dependiendo de los textos y de la cantidad de textos de conocimiento asociados. Los resultados obtenidos sugieren que la metodología PageRank puede ser utilizada para la determinación de autoría, teniendo un potencial de utilización sobre algoritmos especializados en la clasificación como SVM y Naive Bayes.

También, se determinó que utilizar PageRank combinado con otros algoritmos de clasificación no obtiene tan buenos resultados que utilizar solo el PageRank o estos algoritmos por separado. Además, se realizó un análisis del comportamiento de las pruebas al mantener las Stopwords dentro de los textos utilizados, observándose un incremento en el promedio de acierto, siendo la mejor opción al momento de utilizar el PageRank.

9. Referencias

Amancio, D. R. (2015). Autorship recognition via fluctuation analysis of network topology and word intermittency. University of Sao Paulo, Sao Paulo.

Anderson, O., Corney, M., & Mohay, G. (2001). Mining e-mail content for author identification forensics. ACM SIGMOD, 55-64.

Baldi. P, F. P. (2003). Modeling the Internet and the Web.

Cárdenas, J. M. (2011). Topological Cpmplexity in Natural and Formal Languages.
Cárdenas, J., Olivares, G., & Alfaro, R. (2014). Clasificación automática de textos usando redes de palabras. Signos, 346-364.

Cedeño, L. a. (2008). Detección automática de plagio en texto. Tesis, Universidad Politecnica de Valencia, Valencia.

Cristianini, N. &.-T. (2002). Introduction to support vector machines: And other kernel-based learning methods. Cambridge: University of Cambridge.

E., S. (2009). A survey on modern authorship attribution methods. Journal of the American Society for Information Science and Technology.

Erciyes, K (2014) Complex Networks: An Algorithmic Perspective 174-181.

Figueroa, C. Z. (2000). categorización automática de documentos en español.

Hair, J. A. (1999). Análisis multivariante. Madrid: Prentice-hall.

Hunter, S (2013). A Novel Method of Network Text Analysis.

Jurafsky, D. &. (2000). Speech and language processing: An introduct ion to natural language. New Jersey: Prentice-Hall.

Levitan., S. A. (2005). Measuring the usefulness. In Proceedings of the Conference of the Association.

M.C. De Marneffe, B. M. (2006). Generating typed dependency parses from phrase structure parses.

Mendenhall, T. C. (1887). The characteristic curves of composition. Science, IX, 237–49.

Milgran. (1967). The Small World Problem.

Monroy, A. P. (2012). Atribución de Autoría utilizando Distintos tipos de Características A través de una Nueva Representación. INAOE.

Namata, G. M. (2009). A Pipeline Approach to Graph Identification.

Sapkota, U., Solorio, T., Montes-y-Gómez, M., Bethard, S., & Rosso, P. (2014). Cross-Topic Authorship Attribution: Will Out-Of-Topic Data Help? Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, 1228–1237.

Sebastiani, F. (2002). Machine Learning in Automated Text Categorization.

Solé, R. F. (2001). The Small-World of Human Language.

Solorio, T., Pillay, S., Raghavan, S., & Montes y Gómez, M. (2011). Modality Specific Meta Features of Authorship Attribution in Web Forums Post. Proceeding of the 5th International Joint conference of natural language processing (págs. 156-164). Chiang Mai, Thailand: AFNLP.

Stamatatos , E., & Houvaradas, J. (2006). N - Gram feature selection for author identification. In Proceedings of the 12th International Conference on Artificial Intelligence (págs. 77-86). LNCS.

Stamatatos, E. (2008). A Survey of Modern Authorship Attribution Methods. greece: University of the Aegean.

Stolcke, A. (2002). SRILM - an extensible language modeling toolkit.

V. Keselj, F. P. (2003). N-gram based author profiles for authorship attribution. En In Proceedings of the Pacific Association for Computational Linguistics (págs. 255–264).

Vadnick, C. &. (1995). Support vector networks. Machine Learning.

Vapnick, V. (2000). The nature of statistical learning theory. New York: Springer.

William Grieve, J. (2002) Quantitative Authorship Attribution: A History and an Evaluation of Techniques.

Xinbo Gao Æ Bing Xiao Æ Dacheng Tao Æ Xuelong Li (2007). A surbey of Graph Edit Distance.

Zipf, G. L. (1965). Human behavior and the principle of least effort. Addison-Wesley.
S. Brin, L. Page (1999). The PageRank Citation Ranking: Bringing Order to the Web.

C.Augeri (2008). On Graph Isomorphism and the PageRank Algoritm.

P. Fernandez (2004). El Secreto de Google y la Algebra Lineal.

Anexos

Anexo A: Detalle caracterización textos

Detalle del set C1

Nombre del texto	Cantidad de Palabras	Cantidad de Stopwords	Cantidad de Palabras S/ stopwords
catherinebennett_World_001	1107	473	634
catherinebennett_World_002	1299	592	707
catherinebennett_World_003	550	263	287
catherinebennett_World_004	1217	540	677
catherinebennett_World_005	1098	480	618
catherinebennett_World_006	1244	606	638
catherinebennett_World_007	1164	520	644
catherinebennett_World_008	1295	577	718
catherinebennett_World_009	1244	552	692
catherinebennett_World_010	242	115	127
Promedio Catherine Bennett	1046	472	574
pollytoynbee_World_001	1146	492	654
pollytoynbee_World_002	1162	501	661
pollytoynbee_World_003	4484	2032	2452
pollytoynbee_World_004	1167	505	662
pollytoynbee_World_005	1143	489	654
pollytoynbee_World_006	1164	499	665
pollytoynbee_World_007	1157	475	682

pollytoynbee_World_008	1136	473	663
pollytoynbee_World_009	1179	541	638
pollytoynbee_World_010	1212	504	708
Promedio Polly Toynbee	1495	651	844
zoewilliams_World_001	515	259	256
zoewilliams_World_002	654	323	331
zoewilliams_World_003	832	418	414
zoewilliams_World_004	882	425	457
zoewilliams_World_005	827	426	401
zoewilliams_World_006	757	365	392
zoewilliams_World_007	831	402	429
zoewilliams_World_008	838	426	412
zoewilliams_World_009	901	458	443
zoewilliams_World_010	811	392	419
Promedio Zoe Williams	785	389	395
Promedio Total	1109	504	605

Detalle del set C2

Nombre del texto	Cantidad de Palabras	Cantidad de Stopwords	Cantidad de Palabras S/ stopwords
catherinebennett_World_001	1107	473	634
catherinebennett_World_002	1299	592	707
catherinebennett_World_003	550	263	287
catherinebennett_World_004	1217	540	677
catherinebennett_World_005	1098	480	618
catherinebennett_World_006	1244	606	638
catherinebennett_World_007	1164	520	644
catherinebennett_World_008	1295	577	718
catherinebennett_World_009	1244	552	692
catherinebennett_World_010	242	115	127

catherinebennett_World_011	507	259	248
Promedio Catherine Bennet	997	452	545
pollytoynbee_World_001	1146	492	654
pollytoynbee_World_002	1162	501	661
pollytoynbee_World_003	4484	2032	2452
pollytoynbee_World_004	1167	505	662
pollytoynbee_World_005	1143	489	654
pollytoynbee_World_006	1164	499	665
pollytoynbee_World_007	1157	475	682
pollytoynbee_World_008	1136	473	663
pollytoynbee_World_009	1179	541	638
pollytoynbee_World_010	1212	504	708
pollytoynbee_World_011	1193	528	665
pollytoynbee_World_012	1207	548	659
Promedio Polly Toynbee	1446	632	814
zoewilliams_World_001	515	259	256
zoewilliams_World_002	654	323	331
zoewilliams_World_003	832	418	414
zoewilliams_World_004	882	425	457
zoewilliams_World_005	827	426	401
zoewilliams_World_006	757	365	392
zoewilliams_World_007	831	402	429
zoewilliams_World_008	838	426	412
zoewilliams_World_009	901	458	443
zoewilliams_World_010	811	392	419
zoewilliams_World_011	795	399	396
zoewilliams_World_012	885	430	455
zoewilliams_World_013	788	363	425
zoewilliams_World_014	870	431	439
Promedio Zoe Williams	799	394	405
Promedio Total	1068	489	579

Detalle del set C3

Nombre del texto	Cantidad de Palabras	Cantidad de Stopwords	Cantidad de Palabras S/ stopwords
georgemonbiot_World_001	1205	543	662
georgemonbiot_World_002	1141	531	610
georgemonbiot_World_003	1234	582	652
georgemonbiot_World_004	1192	564	628
georgemonbiot_World_005	1188	568	620
georgemonbiot_World_006	1158	568	590
georgemonbiot_World_007	1194	584	610
georgemonbiot_World_008	1161	557	604
georgemonbiot_World_009	1180	556	624
georgemonbiot_World_010	1173	559	614
georgemonbiot_World_011	1154	574	580
georgemonbiot_World_012	1165	573	592
georgemonbiot_World_013	1182	542	640
georgemonbiot_World_014	1176	535	641
georgemonbiot_World_015	1162	555	607
georgemonbiot_World_016	1179	536	643
georgemonbiot_World_017	1173	535	638
georgemonbiot_World_018	1176	534	642
georgemonbiot_World_019	1173	545	628
georgemonbiot_World_020	1204	552	652

georgemonbiot_World_021	1184	565	619
georgemonbiot_World_022	1126	487	639
georgemonbiot_World_023	1211	564	647
georgemonbiot_World_024	1225	569	656
georgemonbiot_World_025	1139	496	643
georgemonbiot_World_026	1184	530	654
georgemonbiot_World_027	1179	540	639
georgemonbiot_World_028	1194	595	599
georgemonbiot_World_029	1186	545	641
georgemonbiot_World_030	1181	519	662
georgemonbiot_World_031	758	347	411
georgemonbiot_World_032	890	394	496
georgemonbiot_World_033	738	328	410
georgemonbiot_World_034	771	355	416
georgemonbiot_World_035	758	349	409
georgemonbiot_World_036	757	353	404
georgemonbiot_World_037	803	353	450
georgemonbiot_World_038	1121	472	649
georgemonbiot_World_039	733	302	431
georgemonbiot_World_040	714	319	395
georgemonbiot_World_041	781	336	445
Promedio George Monbiot	1078	498	580
jonathanfreedland_World_001	630	297	333

jonathanfreedland_World_002	1319	612	707
jonathanfreedland_World_003	642	295	347
jonathanfreedland_World_004	607	277	330
jonathanfreedland_World_005	913	401	512
jonathanfreedland_World_006	1070	501	569
jonathanfreedland_World_007	1005	472	533
jonathanfreedland_World_008	1236	568	668
jonathanfreedland_World_009	1239	564	675
jonathanfreedland_World_010	1111	513	598
jonathanfreedland_World_011	604	293	311
jonathanfreedland_World_012	1242	566	676
jonathanfreedland_World_013	1248	588	660
jonathanfreedland_World_014	772	373	399
jonathanfreedland_World_015	1233	559	674
jonathanfreedland_World_016	1259	567	692
jonathanfreedland_World_017	1255	573	682
jonathanfreedland_World_018	1250	530	720
jonathanfreedland_World_019	1261	591	670
jonathanfreedland_World_020	1291	604	687
jonathanfreedland_World_021	1233	561	672
jonathanfreedland_World_022	892	421	471
jonathanfreedland_World_023	1227	568	659
jonathanfreedland_World_024	1229	564	665

jonathanfreedland_World_025	1261	602	659
jonathanfreedland_World_026	1240	582	658
jonathanfreedland_World_027	1221	555	666
jonathanfreedland_World_028	1178	516	662
jonathanfreedland_World_029	1196	561	635
jonathanfreedland_World_030	1188	581	607
jonathanfreedland_World_031	1200	525	675
jonathanfreedland_World_032	1023	461	562
jonathanfreedland_World_033	1186	541	645
jonathanfreedland_World_034	1266	574	692
jonathanfreedland_World_035	1810	850	960
jonathanfreedland_World_036	1189	582	607
jonathanfreedland_World_037	1197	544	653
jonathanfreedland_World_038	1648	763	885
jonathanfreedland_World_039	1175	528	647
jonathanfreedland_World_040	1186	548	638
jonathanfreedland_World_041	1203	566	637
jonathanfreedland_World_042	1204	571	633
jonathanfreedland_World_043	1193	541	652
jonathanfreedland_World_044	1198	582	616
jonathanfreedland_World_045	1194	533	661
jonathanfreedland_World_047	1166	585	581
jonathanfreedland_World_048	1167	516	651

jonathanfreedland_World_049	1206	557	649
jonathanfreedland_World_050	1228	553	675
jonathanfreedland_World_051	1209	553	656
jonathanfreedland_World_052	1220	529	691
jonathanfreedland_World_053	1205	531	674
jonathanfreedland_World_054	1207	535	672
jonathanfreedland_World_055	1249	602	647
jonathanfreedland_World_056	839	378	461
jonathanfreedland_World_057	2016	930	1086
jonathanfreedland_World_058	1157	512	645
jonathanfreedland_World_059	790	352	438
jonathanfreedland_World_060	1200	525	675
jonathanfreedland_World_061	2107	995	1112
jonathanfreedland_World_062	1201	535	666
jonathanfreedland_World_095	1188	542	646
jonathanfreedland_World_096	1203	570	633
jonathanfreedland_World_097	1175	565	610
jonathanfreedland_World_098	1222	562	660
jonathanfreedland_World_099	1185	552	633
jonathanfreedland_World_100	1235	579	656
Promedio Jonathan Freedland	1184	545	638
peterpreston_World_001	338	145	193
peterpreston_World_002	758	342	416

peterpreston_World_003	798	359	439
peterpreston_World_004	772	314	458
peterpreston_World_005	751	330	421
peterpreston_World_006	772	301	471
peterpreston_World_007	935	399	536
peterpreston_World_008	940	402	538
peterpreston_World_009	927	392	535
peterpreston_World_010	936	382	554
peterpreston_World_011	983	438	545
peterpreston_World_012	962	433	529
peterpreston_World_013	974	432	542
peterpreston_World_014	1089	488	601
peterpreston_World_015	974	430	544
peterpreston_World_016	1277	531	746
peterpreston_World_017	953	401	552
peterpreston_World_018	798	353	445
peterpreston_World_019	953	421	532
peterpreston_World_020	963	418	545
peterpreston_World_021	971	464	507
peterpreston_World_022	746	345	401
peterpreston_World_023	953	397	556
peterpreston_World_024	945	420	525
peterpreston_World_025	900	403	497

peterpreston_World_026	924	382	542
peterpreston_World_027	793	341	452
peterpreston_World_028	908	421	487
peterpreston_World_029	941	417	524
peterpreston_World_030	984	453	531
peterpreston_World_031	986	457	529
peterpreston_World_032	978	442	536
peterpreston_World_033	1085	488	597
peterpreston_World_034	852	391	461
peterpreston_World_035	1134	505	629
peterpreston_World_036	1148	523	625
peterpreston_World_037	937	414	523
peterpreston_World_038	1150	500	650
peterpreston_World_039	1138	494	644
peterpreston_World_040	1138	534	604
peterpreston_World_041	890	435	455
peterpreston_World_042	1155	539	616
peterpreston_World_043	1151	515	636
peterpreston_World_044	1121	510	611
peterpreston_World_045	1137	502	635
peterpreston_World_046	1131	479	652
peterpreston_World_047	1135	554	581
peterpreston_World_048	1088	475	613

peterpreston_World_049	1138	555	583
peterpreston_World_050	1140	508	632
peterpreston_World_051	1826	790	1036
peterpreston_World_052	684	287	397
peterpreston_World_053	1118	527	591
peterpreston_World_054	1593	619	974
peterpreston_World_055	1164	544	620
peterpreston_World_056	1312	578	734
peterpreston_World_057	670	274	396
peterpreston_World_058	610	256	354
peterpreston_World_059	513	217	296
peterpreston_World_060	581	254	327
peterpreston_World_061	1130	476	654
peterpreston_World_062	1119	502	617
peterpreston_World_063	1116	477	639
peterpreston_World_064	1118	537	581
peterpreston_World_065	1116	505	611
peterpreston_World_066	1072	480	592
Promedio Peter Preston	989	438	551
Promedio Total	1085	493	592

Detalle del set C4

Nombre del texto	Cantidad de Palabras	Cantidad de Stopwords	Cantidad de Palabras S/ stopwords
catherinebennett_World_001	1107	473	634
catherinebennett_World_002	1299	592	707
catherinebennett_World_003	550	263	287
catherinebennett_World_004	1217	540	677
catherinebennett_World_005	1098	480	618
catherinebennett_World_006	1244	606	638
catherinebennett_World_007	1164	520	644
catherinebennett_World_008	1295	577	718
catherinebennett_World_009	1244	552	692
catherinebennett_World_010	242	115	127
catherinebennett_World_011	507	259	248
Promedio Catherine Bennett	997	452	545
georgemonbiot_World_001	1205	543	662
georgemonbiot_World_002	1141	531	610
georgemonbiot_World_003	1234	582	652
georgemonbiot_World_004	1192	564	628
georgemonbiot_World_005	1188	568	620
georgemonbiot_World_006	1158	568	590
georgemonbiot_World_007	1194	584	610
georgemonbiot_World_008	1161	557	604

georgemonbiot_World_009	1180	556	624
georgemonbiot_World_010	1173	559	614
georgemonbiot_World_011	1154	574	580
georgemonbiot_World_012	1165	573	592
georgemonbiot_World_013	1182	542	640
georgemonbiot_World_014	1176	535	641
georgemonbiot_World_015	1162	555	607
georgemonbiot_World_016	1179	536	643
georgemonbiot_World_017	1173	535	638
georgemonbiot_World_018	1176	534	642
georgemonbiot_World_019	1173	545	628
georgemonbiot_World_020	1204	552	652
georgemonbiot_World_021	1184	565	619
georgemonbiot_World_022	1126	487	639
georgemonbiot_World_023	1211	564	647
georgemonbiot_World_024	1225	569	656
georgemonbiot_World_025	1139	496	643
georgemonbiot_World_026	1184	530	654
georgemonbiot_World_027	1179	540	639
georgemonbiot_World_028	1194	595	599
georgemonbiot_World_029	1186	545	641
georgemonbiot_World_030	1181	519	662
georgemonbiot_World_031	758	347	411

georgemonbiot_World_032	890	394	496
georgemonbiot_World_033	738	328	410
georgemonbiot_World_034	771	355	416
georgemonbiot_World_035	758	349	409
georgemonbiot_World_036	757	353	404
georgemonbiot_World_037	803	353	450
georgemonbiot_World_038	1121	472	649
georgemonbiot_World_039	733	302	431
georgemonbiot_World_040	714	319	395
georgemonbiot_World_041	781	336	445
Promedio George Monbiot	1078	498	580
jonathanfreedland_World_001	630	297	333
jonathanfreedland_World_002	1319	612	707
jonathanfreedland_World_003	642	295	347
jonathanfreedland_World_004	607	277	330
jonathanfreedland_World_005	913	401	512
jonathanfreedland_World_006	1070	501	569
jonathanfreedland_World_007	1005	472	533
jonathanfreedland_World_008	1236	568	668
jonathanfreedland_World_009	1239	564	675
jonathanfreedland_World_010	1111	513	598
jonathanfreedland_World_011	604	293	311
jonathanfreedland_World_012	1242	566	676

jonathanfreedland_World_013	1248	588	660
jonathanfreedland_World_014	772	373	399
jonathanfreedland_World_015	1233	559	674
jonathanfreedland_World_016	1259	567	692
jonathanfreedland_World_017	1255	573	682
jonathanfreedland_World_018	1250	530	720
jonathanfreedland_World_019	1261	591	670
jonathanfreedland_World_020	1291	604	687
jonathanfreedland_World_021	1233	561	672
jonathanfreedland_World_022	892	421	471
jonathanfreedland_World_023	1227	568	659
jonathanfreedland_World_024	1229	564	665
jonathanfreedland_World_025	1261	602	659
jonathanfreedland_World_026	1240	582	658
jonathanfreedland_World_027	1221	555	666
jonathanfreedland_World_028	1178	516	662
jonathanfreedland_World_029	1196	561	635
jonathanfreedland_World_030	1188	581	607
jonathanfreedland_World_031	1200	525	675
jonathanfreedland_World_032	1023	461	562
jonathanfreedland_World_033	1186	541	645
jonathanfreedland_World_034	1266	574	692
jonathanfreedland_World_035	1810	850	960

jonathanfreedland_World_036	1189	582	607
jonathanfreedland_World_037	1197	544	653
jonathanfreedland_World_038	1648	763	885
jonathanfreedland_World_039	1175	528	647
jonathanfreedland_World_040	1186	548	638
jonathanfreedland_World_041	1203	566	637
jonathanfreedland_World_042	1204	571	633
jonathanfreedland_World_043	1193	541	652
jonathanfreedland_World_044	1198	582	616
jonathanfreedland_World_045	1194	533	661
jonathanfreedland_World_047	1166	585	581
jonathanfreedland_World_048	1167	516	651
jonathanfreedland_World_049	1206	557	649
jonathanfreedland_World_050	1228	553	675
jonathanfreedland_World_051	1209	553	656
jonathanfreedland_World_052	1220	529	691
jonathanfreedland_World_053	1205	531	674
jonathanfreedland_World_054	1207	535	672
jonathanfreedland_World_055	1249	602	647
jonathanfreedland_World_056	839	378	461
jonathanfreedland_World_057	2016	930	1086
jonathanfreedland_World_058	1157	512	645
jonathanfreedland_World_059	790	352	438

jonathanfreedland_World_060	1200	525	675
jonathanfreedland_World_061	2107	995	1112
jonathanfreedland_World_062	1201	535	666
jonathanfreedland_World_095	1188	542	646
jonathanfreedland_World_096	1203	570	633
jonathanfreedland_World_097	1175	565	610
jonathanfreedland_World_098	1222	562	660
jonathanfreedland_World_099	1185	552	633
jonathanfreedland_World_100	1235	579	656
Promedio Jonathan Freedland	1184	545	638
peterpreston_World_001	338	145	193
peterpreston_World_002	758	342	416
peterpreston_World_003	798	359	439
peterpreston_World_004	772	314	458
peterpreston_World_005	751	330	421
peterpreston_World_006	772	301	471
peterpreston_World_007	935	399	536
peterpreston_World_008	940	402	538
peterpreston_World_009	927	392	535
peterpreston_World_010	936	382	554
peterpreston_World_011	983	438	545
peterpreston_World_012	962	433	529
peterpreston_World_013	974	432	542

peterpreston_World_014	1089	488	601
peterpreston_World_015	974	430	544
peterpreston_World_016	1277	531	746
peterpreston_World_017	953	401	552
peterpreston_World_018	798	353	445
peterpreston_World_019	953	421	532
peterpreston_World_020	963	418	545
peterpreston_World_021	971	464	507
peterpreston_World_022	746	345	401
peterpreston_World_023	953	397	556
peterpreston_World_024	945	420	525
peterpreston_World_025	900	403	497
peterpreston_World_026	924	382	542
peterpreston_World_027	793	341	452
peterpreston_World_028	908	421	487
peterpreston_World_029	941	417	524
peterpreston_World_030	984	453	531
peterpreston_World_031	986	457	529
peterpreston_World_032	978	442	536
peterpreston_World_033	1085	488	597
peterpreston_World_034	852	391	461
peterpreston_World_035	1134	505	629
peterpreston_World_036	1148	523	625

peterpreston_World_037	937	414	523
peterpreston_World_038	1150	500	650
peterpreston_World_039	1138	494	644
peterpreston_World_040	1138	534	604
peterpreston_World_041	890	435	455
peterpreston_World_042	1155	539	616
peterpreston_World_043	1151	515	636
peterpreston_World_044	1121	510	611
peterpreston_World_045	1137	502	635
peterpreston_World_046	1131	479	652
peterpreston_World_047	1135	554	581
peterpreston_World_048	1088	475	613
peterpreston_World_049	1138	555	583
peterpreston_World_050	1140	508	632
peterpreston_World_051	1826	790	1036
peterpreston_World_052	684	287	397
peterpreston_World_053	1118	527	591
peterpreston_World_054	1593	619	974
peterpreston_World_055	1164	544	620
peterpreston_World_056	1312	578	734
peterpreston_World_057	670	274	396
peterpreston_World_058	610	256	354
peterpreston_World_059	513	217	296

peterpreston_World_060	581	254	327
peterpreston_World_061	1130	476	654
peterpreston_World_062	1119	502	617
peterpreston_World_063	1116	477	639
peterpreston_World_064	1118	537	581
peterpreston_World_065	1116	505	611
peterpreston_World_066	1072	480	592
Promedio Peter Preston	989	438	551
pollytoynbee_World_001	1146	492	654
pollytoynbee_World_002	1162	501	661
pollytoynbee_World_003	4484	2032	2452
pollytoynbee_World_004	1167	505	662
pollytoynbee_World_005	1143	489	654
pollytoynbee_World_006	1164	499	665
pollytoynbee_World_007	1157	475	682
pollytoynbee_World_008	1136	473	663
pollytoynbee_World_009	1179	541	638
pollytoynbee_World_010	1212	504	708
pollytoynbee_World_011	1193	528	665
pollytoynbee_World_012	1207	548	659
Promedio Polly Toynbee	1446	632	814
zoewilliams_World_001	515	259	256
zoewilliams_World_002	654	323	331
zoewilliams_World_003	832	418	414
zoewilliams_World_004	882	425	457
zoewilliams_World_005	827	426	401
zoewilliams_World_006	757	365	392
zoewilliams_World_007	831	402	429

zoewilliams_World_008	838	426	412
zoewilliams_World_009	901	458	443
zoewilliams_World_010	811	392	419
zoewilliams_World_011	795	399	396
zoewilliams_World_012	885	430	455
zoewilliams_World_013	788	363	425
zoewilliams_World_014	870	431	439
Promedio Zoe Williams	799	394	405
Promedio Total	1082	492	590

Detalle del set C5

Nombre del texto	Cantidad de Palabras	Cantidad de Stopwords	Cantidad de Palabras S/ stopwords
georgemonbiot_World_001	1205	543	662
georgemonbiot_World_002	1141	531	610
georgemonbiot_World_003	1234	582	652
georgemonbiot_World_004	1192	564	628
georgemonbiot_World_005	1188	568	620
georgemonbiot_World_006	1158	568	590
georgemonbiot_World_007	1194	584	610
georgemonbiot_World_008	1161	557	604
georgemonbiot_World_009	1180	556	624
georgemonbiot_World_010	1173	559	614
georgemonbiot_World_011	1154	574	580
georgemonbiot_World_012	1165	573	592
georgemonbiot_World_013	1182	542	640
georgemonbiot_World_014	1176	535	641
georgemonbiot_World_015	1162	555	607

georgemonbiot_World_016	1179	536	643
georgemonbiot_World_017	1173	535	638
georgemonbiot_World_018	1176	534	642
georgemonbiot_World_019	1173	545	628
georgemonbiot_World_020	1204	552	652
georgemonbiot_World_021	1184	565	619
georgemonbiot_World_022	1126	487	639
georgemonbiot_World_023	1211	564	647
georgemonbiot_World_024	1225	569	656
georgemonbiot_World_025	1139	496	643
georgemonbiot_World_026	1184	530	654
georgemonbiot_World_027	1179	540	639
georgemonbiot_World_028	1194	595	599
georgemonbiot_World_029	1186	545	641
georgemonbiot_World_030	1181	519	662
georgemonbiot_World_031	758	347	411
georgemonbiot_World_032	890	394	496
georgemonbiot_World_033	738	328	410
georgemonbiot_World_034	771	355	416
georgemonbiot_World_035	758	349	409
georgemonbiot_World_036	757	353	404
georgemonbiot_World_037	803	353	450
georgemonbiot_World_038	1121	472	649

georgemonbiot_World_039	733	302	431
georgemonbiot_World_040	714	319	395
georgemonbiot_World_041	781	336	445
Promedio George Monbiot	1078	498	580
hugoyoung_World_001	1191	590	601
hugoyoung_World_002	1188	592	596
hugoyoung_World_003	1175	562	613
hugoyoung_World_004	1200	554	646
hugoyoung_World_005	1150	517	633
hugoyoung_World_006	1220	596	624
hugoyoung_World_007	1162	568	594
hugoyoung_World_008	1194	586	608
hugoyoung_World_009	1173	549	624
hugoyoung_World_010	964	478	486
hugoyoung_World_011	1166	527	639
hugoyoung_World_012	1136	509	627
hugoyoung_World_013	1201	575	626
hugoyoung_World_014	1185	567	618
hugoyoung_World_015	1159	570	589
hugoyoung_World_016	1200	560	640
hugoyoung_World_017	1154	495	659
hugoyoung_World_018	1281	597	684
hugoyoung_World_019	1213	594	619
hugoyoung_World_020	1147	520	627
hugoyoung_World_021	1179	516	663
hugoyoung_World_022	1160	538	622
hugoyoung_World_023	1142	508	634
hugoyoung_World_024	1125	529	596
hugoyoung_World_025	1137	519	618
hugoyoung_World_026	1133	498	635
hugoyoung_World_027	1138	497	641
hugoyoung_World_028	1168	524	644
hugoyoung_World_029	1101	482	619
hugoyoung_World_030	1137	546	591
hugoyoung_World_031	1152	540	612
hugoyoung_World_032	1192	564	628
hugoyoung_World_033	816	395	421
hugoyoung_World_034	1148	537	611
hugoyoung_World_035	1118	524	594
Promedio Hugo Young	1152	538	614
jonathanfreedland_World_001	630	297	333

jonathanfreedland_World_002	1319	612	707
jonathanfreedland_World_003	642	295	347
jonathanfreedland_World_004	607	277	330
jonathanfreedland_World_005	913	401	512
jonathanfreedland_World_006	1070	501	569
jonathanfreedland_World_007	1005	472	533
jonathanfreedland_World_008	1236	568	668
jonathanfreedland_World_009	1239	564	675
jonathanfreedland_World_010	1111	513	598
jonathanfreedland_World_011	604	293	311
jonathanfreedland_World_012	1242	566	676
jonathanfreedland_World_013	1248	588	660
jonathanfreedland_World_014	772	373	399
jonathanfreedland_World_015	1233	559	674
jonathanfreedland_World_016	1259	567	692
jonathanfreedland_World_017	1255	573	682
jonathanfreedland_World_018	1250	530	720
jonathanfreedland_World_019	1261	591	670
jonathanfreedland_World_020	1291	604	687
jonathanfreedland_World_021	1233	561	672
jonathanfreedland_World_022	892	421	471
jonathanfreedland_World_023	1227	568	659
jonathanfreedland_World_024	1229	564	665

jonathanfreedland_World_025	1261	602	659
jonathanfreedland_World_026	1240	582	658
jonathanfreedland_World_027	1221	555	666
jonathanfreedland_World_028	1178	516	662
jonathanfreedland_World_029	1196	561	635
jonathanfreedland_World_030	1188	581	607
jonathanfreedland_World_031	1200	525	675
jonathanfreedland_World_032	1023	461	562
jonathanfreedland_World_033	1186	541	645
jonathanfreedland_World_034	1266	574	692
jonathanfreedland_World_035	1810	850	960
jonathanfreedland_World_036	1189	582	607
jonathanfreedland_World_037	1197	544	653
jonathanfreedland_World_038	1648	763	885
jonathanfreedland_World_039	1175	528	647
jonathanfreedland_World_040	1186	548	638
jonathanfreedland_World_041	1203	566	637
jonathanfreedland_World_042	1204	571	633
jonathanfreedland_World_043	1193	541	652
jonathanfreedland_World_044	1198	582	616
jonathanfreedland_World_045	1194	533	661
jonathanfreedland_World_047	1166	585	581
jonathanfreedland_World_048	1167	516	651

jonathanfreedland_World_049	1206	557	649
jonathanfreedland_World_050	1228	553	675
jonathanfreedland_World_051	1209	553	656
jonathanfreedland_World_052	1220	529	691
jonathanfreedland_World_053	1205	531	674
jonathanfreedland_World_054	1207	535	672
jonathanfreedland_World_055	1249	602	647
jonathanfreedland_World_056	839	378	461
jonathanfreedland_World_057	2016	930	1086
jonathanfreedland_World_058	1157	512	645
jonathanfreedland_World_059	790	352	438
jonathanfreedland_World_060	1200	525	675
jonathanfreedland_World_061	2107	995	1112
jonathanfreedland_World_062	1201	535	666
jonathanfreedland_World_095	1188	542	646
jonathanfreedland_World_096	1203	570	633
jonathanfreedland_World_097	1175	565	610
jonathanfreedland_World_098	1222	562	660
jonathanfreedland_World_099	1185	552	633
jonathanfreedland_World_100	1235	579	656
Promedio Jonathan Freedland	1184	545	638
martinkettle_World_001	1098	519	579
martinkettle_World_002	1102	526	576
martinkettle_World_003	1152	542	610
martinkettle_World_004	1169	583	586
martinkettle_World_005	1298	659	639

martinkettle_World_006	1143	579	564
martinkettle_World_007	1024	405	619
martinkettle_World_008	353	158	195
martinkettle_World_009	1245	573	672
martinkettle_World_010	1179	579	600
martinkettle_World_011	1131	517	614
martinkettle_World_012	1295	638	657
martinkettle_World_013	1134	528	606
martinkettle_World_014	1193	590	603
martinkettle_World_015	1236	567	669
martinkettle_World_016	1138	531	607
martinkettle_World_017	553	279	274
martinkettle_World_018	852	421	431
martinkettle_World_019	1179	516	663
martinkettle_World_020	1246	577	669
martinkettle_World_021	1138	533	605
martinkettle_World_022	800	376	424
martinkettle_World_023	795	367	428
martinkettle_World_024	15	3	12
martinkettle_World_025	787	354	433
martinkettle_World_026	874	414	460
martinkettle_World_027	716	323	393
martinkettle_World_028	870	392	478
martinkettle_World_029	838	367	471
martinkettle_World_030	771	359	412
martinkettle_World_031	757	331	426
martinkettle_World_032	713	324	389
martinkettle_World_033	740	286	454
martinkettle_World_034	845	343	502
martinkettle_World_035	680	300	380
martinkettle_World_036	875	425	450
Promedio Martin Kettle	943	438	504
peterpreston_World_001	338	145	193
peterpreston_World_002	758	342	416
peterpreston_World_003	798	359	439
peterpreston_World_004	772	314	458
peterpreston_World_005	751	330	421
peterpreston_World_006	772	301	471
peterpreston_World_007	935	399	536

peterpreston_World_008	940	402	538
peterpreston_World_009	927	392	535
peterpreston_World_010	936	382	554
peterpreston_World_011	983	438	545
peterpreston_World_012	962	433	529
peterpreston_World_013	974	432	542
peterpreston_World_014	1089	488	601
peterpreston_World_015	974	430	544
peterpreston_World_016	1277	531	746
peterpreston_World_017	953	401	552
peterpreston_World_018	798	353	445
peterpreston_World_019	953	421	532
peterpreston_World_020	963	418	545
peterpreston_World_021	971	464	507
peterpreston_World_022	746	345	401
peterpreston_World_023	953	397	556
peterpreston_World_024	945	420	525
peterpreston_World_025	900	403	497
peterpreston_World_026	924	382	542
peterpreston_World_027	793	341	452
peterpreston_World_028	908	421	487
peterpreston_World_029	941	417	524
peterpreston_World_030	984	453	531

peterpreston_World_031	986	457	529
peterpreston_World_032	978	442	536
peterpreston_World_033	1085	488	597
peterpreston_World_034	852	391	461
peterpreston_World_035	1134	505	629
peterpreston_World_036	1148	523	625
peterpreston_World_037	937	414	523
peterpreston_World_038	1150	500	650
peterpreston_World_039	1138	494	644
peterpreston_World_040	1138	534	604
peterpreston_World_041	890	435	455
peterpreston_World_042	1155	539	616
peterpreston_World_043	1151	515	636
peterpreston_World_044	1121	510	611
peterpreston_World_045	1137	502	635
peterpreston_World_046	1131	479	652
peterpreston_World_047	1135	554	581
peterpreston_World_048	1088	475	613
peterpreston_World_049	1138	555	583
peterpreston_World_050	1140	508	632
peterpreston_World_051	1826	790	1036
peterpreston_World_052	684	287	397
peterpreston_World_053	1118	527	591

peterpreston_World_054	1593	619	974
peterpreston_World_055	1164	544	620
peterpreston_World_056	1312	578	734
peterpreston_World_057	670	274	396
peterpreston_World_058	610	256	354
peterpreston_World_059	513	217	296
peterpreston_World_060	581	254	327
peterpreston_World_061	1130	476	654
peterpreston_World_062	1119	502	617
peterpreston_World_063	1116	477	639
peterpreston_World_064	1118	537	581
peterpreston_World_065	1116	505	611
peterpreston_World_066	1072	480	592
Promedio Peter Preston	989	438	551
Promedio Total	1074	492	582