

Nombre _____ Fecha _____

Actividad NUMB3RS: Estilometría

En "Charla asesina" el FBI ha conseguido una colección de diálogos de charlas en línea de un sospechoso de asesinato. Charlie propone usar el Análisis lingüístico estadístico para identificar al autor. El estilo y el lenguaje ofrecen algo así como una huella digital de la escritura. Charlie dice que "pensemos en un joyero que ensarta un collar. Escoge ciertas cuentas, decide qué patrón seguirá para ensartarlas según su estilo personal. Otros joyeros tienen estilos diferentes y exhiben diferentes patrones. Igualmente, los diferentes despachadores de IM exhibirán su propio estilo, o sea ciertos patrones del lenguaje que les son propios".

La estilometría, que es la práctica de aplicar el análisis estadístico al estilo lingüístico, se ha empleado para despejar dudas sobre la autoría, notablemente respecto de los *Federalist Papers*, cartas de la Guerra Civil norteamericana y las obras de Shakespeare. Se pueden analizar obras con muchas estadísticas diferentes. Esta actividad se concentrará en dos estadísticas: longitud de las oraciones y longitud de los párrafos. Como el análisis de diálogos de las charlas en línea presenta más dificultades, las muestras para esta actividad son artículos cortos escritos en prosa.

La situación

Se descubre un escrito anónimo. Tu objetivo es aplicar el análisis estadístico para determinar a cuál de dos muestras conocidas se parece más el escrito anónimo. Para ahorrar tiempo y trabajo tedioso, tu profesor puede pedir que se repartan las preguntas 1a a 3c. Cada parte te pide que cuentes el número de palabras en un documento. A continuación damos algunas pautas:

- Si puedes abrir el archivo en un procesador de palabras, quizá el computador evalúe algunos de estos valores durante o después de una verificación gramatical. Quizá necesites activar este atributo. Las siguientes reglas concuerdan con un programa así y se emplearon para generar los resultados presentados en esta actividad. La clase puede acordar sus propias reglas para aplicar en forma uniforme.
- Una palabra con guión es una sola palabra.
- Una abreviatura es una palabra.
- Un URL (dirección de un sitio Web) es una palabra.
- Un número es una palabra, aun al comienzo de una lista (en la muestra, "2 Everyone procrastinates" es una oración de 3 palabras).
- En la Muestra A, "Q: Así, ..." es una oración de dos palabras.

1. Compila estadísticas de la Muestra A, escrita por Julia Latino.

- a. Anota el número de palabras por oración en la Muestra A. Ingresas el dato en la lista L₁ en la calculadora.
- b. Anota el número de palabras por párrafo en la Muestra A. Ingresas el dato en la lista L₄.

2. Compila estadísticas de la Muestra B, escrita por Aaron Moser.
 - a. Anota el número de palabras por oración en la Muestra B. Ingresas el dato en la lista L_2 .
 - b. Anota el número de palabras por párrafo en la Muestra B. Ingresas el dato en la lista L_5 .

3. La Muestra C fue escrita por Julia o por Aaron.
 - a. Anota el número de palabras por oración en la Muestra C. Ingresas el dato en la lista L_3 .
 - b. Anota el número de palabras por párrafo en la Muestra C. Ingresas el dato en la lista L_6 .

4.
 - a. Prepara una gráfica de caja y bigotes para el número de palabras por oración. Los datos se encuentran en L_1 , L_2 y L_3 . Usa tres gráficas separadas para que la calculadora muestre los tres diagramas en una pantalla, con la Muestra A arriba, seguida de la Muestra B y luego la Muestra C.
 - b. Los números de palabras por oración, ¿sugieren quién podría ser el autor de la Muestra C? En tal caso, ¿quién es el autor? Explica.

5.
 - a. Prepara una gráfica de caja y bigotes para el número de palabras por párrafo. Los datos se encuentran en L_4 , L_5 y L_6 . La calculadora mostrará las tres gráficas en una pantalla, con la Muestra A arriba, seguida de la Muestra B y luego la Muestra C.
 - b. Los números de palabras por párrafo, ¿sugieren quién podría ser el autor de la Muestra C? En tal caso, ¿quién es el autor? Explica.

6. La muestra de escritura que Charlie analiza en el episodio es diálogo de charlas en línea
¿Qué estadísticas podrías compilar al tratar de establecer la "huella digital" del autor de una muestra así?

El objeto de esta actividad es dar a los estudiantes un vistazo breve y sencillo de un tema matemático muy extenso. TI y NCTM lo invitan a usted y a sus estudiantes a aprender más sobre este tema con las extensiones que se ofrecen abajo y con su propia investigación independiente.

Extensiones

Esta actividad se puede aplicar a muchas situaciones diversas. Explora el tema más a fondo con otras muestras de escritura. Ten presente que el verdadero análisis estilométrico exige muestras muy grandes, por lo cual los resultados no siempre serán claros. A continuación se sugieren algunas áreas de exploración.

- Ensayá esta actividad con muestras de escritura de tu clase o del periódico de la escuela. Experimenta con otras estadísticas. ¿Son ciertas estadísticas mejores indicadores que otras? Puedes ahorrar tiempo usando un procesador de palabras en el computador para evaluar algunos de estos valores durante o después de una verificación gramatical.
- Compara obras de diversos autores o de diferentes géneros para determinar si algunas estadísticas sirven como "huellas digitales" de la obra. Las estadísticas que sirven para este fin serían las categorías léxicas de las palabras, puntuación y longitud de capítulos.
- ¿Puedes distinguir entre autores de charlas en línea? Algunas estadísticas serían abreviaturas de computador como LOL o BRB, longitud de respuestas y tiempo entre respuestas.
- Investiga más el uso de la estilometría para analizar los *Federalist Papers*, cartas de la guerra civil norteamericana y las obras de Shakespeare.

Recursos adicionales

Para un artículo instructivo con detalles sobre el tema de la autoría en un libro de la serie Oz, visita el sitio Web Science News: <http://www.sciencenews.org/articles/20031220/bob8.asp>

En el siguiente sitio Web se encuentra un artículo de la BBC sobre estilometría aplicada a la música: <http://news.bbc.co.uk/1/hi/5083986.stm?ls>