

# Instituto tecnológico de Culiacán

# Beltrán Ruiz Gerardo Angel.

**Investigación Modelo Big-Five y corpus orientados a la personalidad.**

Contents

[Qué es un corpus 1](#_Toc129206370)

[Big Five 2](#_Toc129206371)

[MyPersonality 2](#_Toc129206372)

[Corpus B5 2](#_Toc129206373)

[Essays 3](#_Toc129206374)

[Referencias 4](#_Toc129206375)

# Qué es un corpus

En el campo de la lingüística, la palabra corpus es una palabra algo ambigua y que actualmente se utiliza en un sentido general para referirse a cualquier tipo de recopilación de textos. En realidad, para ser más exactos, en el ámbito de la recopilación de textos hay que distinguir, según el grado de especificación en los criterios de selección, al menos entre tres tipos diferentes de recopilaciones:

• Archivo/colección (informatizado) (Archive/Collection). Es un repertorio de textos en soporte informático sin buscar ningún tipo de relación entre ellos.

• Biblioteca de Textos Electrónicos (Electronic text library). Es una colección de textos en soporte informático, guardados en un formato estándar, siguiendo ciertas normas de contenido, pero sin un criterio riguroso de selección.

• Corpus Informatizado (Computer corpus). Es una recopilación de textos seleccionados según criterios lingüísticos, codificados de modo estándar y homogéneo, con la finalidad de poder ser tratados mediante procesos informáticos y destinados a reflejar el comportamiento de una o más lenguas.

# Big Five

El modelo Big Five (Goldberg, 1990) comprende cinco dimensiones fundamentales de la personalidad humana – Apertura experiencia, escrupulosidad, extraversión, amabilidad y neuroticismo, que pueden estimarse mediante el uso de una amplia gama de métodos, siendo el más común el uso de inventarios de personalidad. Entre muchos inventarios desarrollados para el modelo Big Five, la necesidad de una evaluación rápida herramienta condujo a la propuesta del inventario BFI (John et al., 1991).

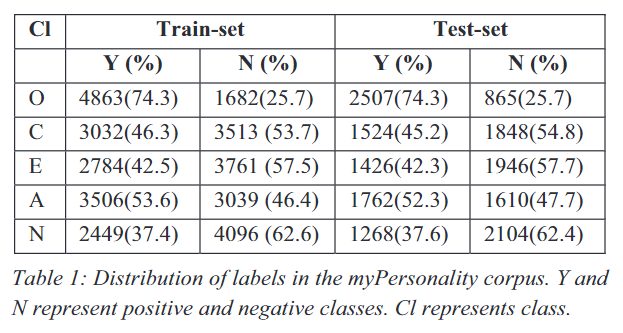
# MyPersonality

myPersonality fue una aplicación de Facebook que permitía a sus usuarios participar en investigaciones psicológicas completando un cuestionario de personalidad. También les ofreció retroalimentación sobre sus puntajes. Fue creado en 2007 por David Stillwell (entonces estudiante de posgrado en la Universidad de Nottingham y ahora profesor en la Universidad de Cambridge).

Siguiendo los principios de la ciencia abierta y la replicabilidad, se proporcionó a otros académicos datos anónimos que podrían usar para investigaciones académicas no comerciales. Esto ha permitido a académicos de muchos campos hacer descubrimientos significativos que mejoran la comprensión del comportamiento humano.

Nunca se compartieron los datos de ningún usuario de myPersonality con fines comerciales. Los creadores de myPersonality fueron acusados de no respetar la información de los usuarios por lo que decidieron dejar de dar soporte y de compartir los datos.

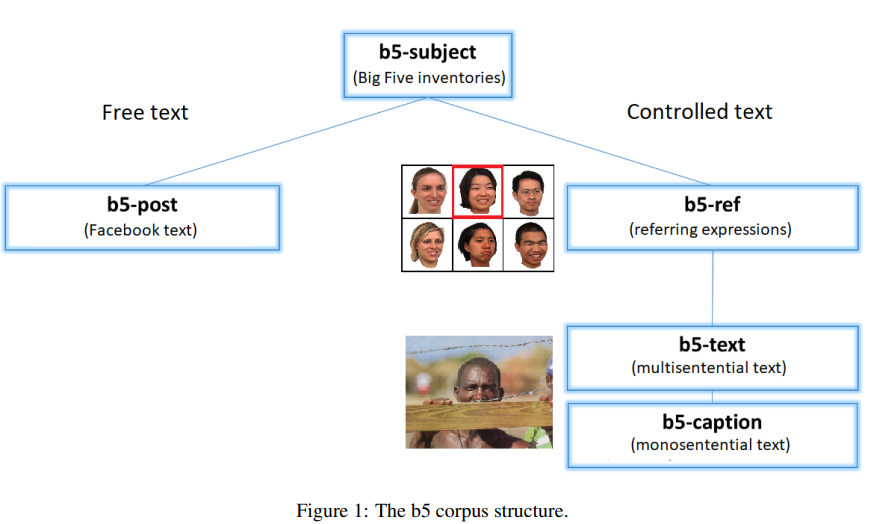
El corpus myPersonality (Alam, et al., 2021) tiene una estructura de 5 clases que cada una representa un rasgo de la personalidad basada en el Modelo Big Five y los resultados de cada clase se muestra con una Y en caso de que se presente el rasgo de la respectiva clase o N si no presenta el rasgo de la personalidad. A continuación, una imagen de ejemplo del corpus myPersonality.



# Corpus B5

El corpus b5 es un conjunto de datos que contiene textos y autoinformes inventarios de personalidad de sus autores. esto consiste en una base de conocimiento del autor llamada b5-subject y cuatro textos bases de datos (o subcorpora) denominadas b5-post, b5-ref, b5-text, y b5-título.

La siguiente figura representa la estructura del corpus B5:



el corpus b5 transmite cuatro categorías de texto divididas en dos clases generales: texto libre obtenido de las actualizaciones de estado de Facebook de cada participante, y tres tipos de texto controlado obtenido de una serie de entrevistas en persona tareas de recogida de datos. La recopilación de texto libre y controlado está motivada por el doble propósito del corpus, es decir, por nuestro objetivo a largo plazo de reutilizar los datos tanto en la comprensión del lenguaje como en los estudios de generación del lenguaje. El uso de texto libre está motivado principalmente por las necesidades específicas de ciertos tipos de aplicaciones, como el reconocimiento de personalidad a partir de texto en redes sociales, mientras que el uso requiere texto controlado para una variedad de lenguaje natural.

# Essays

Essays es un conjunto de datos creado por Pennebaker & King (2000) y contiene aproximadamente 2,400 ensayos de flujo de conciencia etiquetados con los rasgos de personalidad de los autores. El corpus está organizado de la siguiente manera:

* Identificador de los autores.
* Texto de los ensayos de flujo de conciencia de los autores.
  + Rasgos de personalidad etiquetados como (y/n) para cada uno de los cinco rasgos de personalidad: EXT (*Extraversion*), NEU (*Neuroticism*), AGR (*Agreeableness*), CON (*Conscientiousness*) y OPN (*Openness to Experience*).

# Referencias

<https://aclanthology.org/L18-1183.pdf>

<https://sites.google.com/michalkosinski.com/mypersonality>