

## MARCO CONCEPTUAL Y METODOLÓGICO

### Contexto de la investigación

La serie “*Informes sobre Herramientas Gerenciales*” está estructurado por 138 documentos técnicos que buscan ofrecer un análisis bibliométrico y estadístico de datos longitudinales sobre el comportamiento y evolución de una selección de 23 grupos de herramientas gerenciales desde la perspectiva de 5 bases de datos diferentes (Google Trends, Google Books Ngram, Crossref.org, encuestas sobre usabilidad y satisfacción de Bain & Company) en el contexto de una investigación de IV Nivel<sup>1</sup> sobre la “*Dicotomía ontológica en las «modas gerenciales»: Un enfoque proto-meta-sistémico desde las antinomias ingénitas del ecosistema transorganizacional*”, llevada a cabo por Diomar Añez, como parte de sus estudios doctorales en Ciencias Gerenciales en la Universidad Latinoamericana y del Caribe (ULAC).

En este contexto, el presente estudio se inscribe en el debate académico sobre la naturaleza y dinámica de las denominadas «modas gerenciales» que se conceptualizan, *prima facie*, como innovaciones de carácter tecnológico-administrativo –que se manifiestan en forma de herramientas, técnicas, tendencias, filosofías, principios o enfoques gerenciales o de gestión<sup>2</sup>– y que exhiben potenciales patrones de adopción y declive aparentemente cíclicos en el ámbito organizacional. No obstante, la mera existencia de estos patrones cíclicos, así como su interpretación como “modas”, son objeto de controversia. La investigación doctoral que enmarca esta serie de informes propone trascender la mera descripción fenomenológica de estos ciclos, para indagar en sus fundamentos causales; por lo cual, se exploran dimensiones onto-antropológicas y microeconómicas que podrían subyacer a la emergencia, difusión y eventual obsolescencia (o persistencia) de estas innovaciones<sup>3</sup>. Es decir, se parte de la premisa de que las organizaciones contemporáneas se caracterizan por tensiones inherentes y constitutivas, antinomias

<sup>1</sup> En el contexto latinoamericano, se considera un nivel equivalente a la formación de posgrado avanzada, similar al nivel de Doctor que corresponde al nivel 4 del Marco Español de Cualificaciones para la Educación Superior (MECES), y que se alinea con el nivel 8 del Marco Europeo de Cualificaciones (EQF). En el sistema norteamericano, se asocia con el grado de Ph.D. (Doctor of Philosophy), que implica una formación rigurosa en investigación. Es decir, los estudios doctorales se asocian con competencias avanzadas en investigación y una especialización profunda en un área de conocimiento.

<sup>2</sup> Cfr. Añez Barrios, D. G. (2023). El laberinto de las modas gerenciales: ¿ventaja trivial o cambio forzado en empresas disruptivas? *CIID Journal*, 4(1), 1-21. <https://scispace.com/pdf/el-laberinto-de-las-modas-gerenciales-ventaja-trivial-o-2ihewu3i.pdf>

<sup>3</sup> Cfr. Añez Barrios, D. G. (2023). ¿Racionalidad o subjetividad en las modas gerenciales?: una dicotomía microeconómica compleja. *CIID Journal*, 4(1), 125-149. <https://dialnet.unirioja.es/servlet/articulo?codigo=9662429>

entre, v. gr., la necesidad de estabilidad y la exigencia de innovación, o entre la continuidad de las prácticas establecidas y la disrupción generada por nuevas tecnologías y modelos de gestión.

Dado lo anterior, se postula que la perdurabilidad –o, por el contrario, la efímera popularidad– de una herramienta gerencial podría no depender exclusivamente de su eficacia intrínseca (medida en términos de resultados objetivos), sino adicionalmente de su potencial capacidad para mediar en estas tensiones organizacionales. Siendo así, ¿una herramienta que mitigue las antinomias inherentes a la organización podría tener una mayor probabilidad de adopción sostenida, mientras que una herramienta que las exacerbe podría ser percibida como una “moda pasajera”? Ahora bien, antes de poder abordar esta temática, es imprescindible establecer si, efectivamente, existe un patrón identificable que rijan el comportamiento en la adopción y uso de herramientas gerenciales que lleve a su similitud con una “moda”; es decir, se requiere evidencia que sustente (o refute) la premisa *a priori* de que estas herramientas presentan “ciclos de auge y declive”. Por tanto, para abordar esta cuestión preliminar, se hace necesario llevar a cabo este análisis para detectar si existen patrones sistemáticos que justifiquen la caracterización de estas herramientas como “modas”; y profundizar sobre la existencia de otros mecanismos causales subyacentes.

Para abordar esta temática con plena pertinencia, resulta metodológicamente imperativo establecer que el propósito primordial de estos informes es detectar y caracterizar patrones sistemáticos en las fuentes de datos disponibles, para determinar si existe una base empírica que valide, matice o refute la caracterización de estas herramientas como «modas» en términos de su difusión y adopción, o si, por el contrario, su trayectoria se ajusta a otros modelos de comportamiento; por tanto, constituyen una fase exploratoria y descriptiva de naturaleza cuantitativa previa a la teorización, a fin de establecer la existencia, magnitud y forma del fenómeno a estudiar. Por tanto, los informes no buscan explicar causalmente estos patrones, sino documentarlos de manera precisa y sistemática y, por consiguiente, constituyen un aporte original e independiente al campo de la investigación de las ciencias gerenciales y de la gestión, proporcionando una base de datos y análisis cuantitativos sin precedentes en cuanto a su alcance y detalle.

La investigación doctoral, en contraste, adopta una aproximación metodológica eminentemente cualitativa, con el propósito de explorar en profundidad las perspectivas, motivaciones e intereses involucrados en la adopción y el uso de estas herramientas. Se busca así trascender la mera descripción cuantitativa de los patrones de auge y declive, para indagar en los mecanismos causales y procesos sociales subyacentes; partiendo de la premisa de que las «modas gerenciales» no son fenómenos aleatorios o irracionales, sino que responden a una compleja interrelación de factores contextuales,

organizacionales y cognitivos que, al converger, determinan la perdurabilidad (o el abandono) de una herramienta, más allá de su sola eficacia organizacional intrínseca o percibida. En última instancia, se busca comprender cómo las circunstancias contextuales, las estructuras de poder, las redes sociales y los procesos de legitimación dan forma a la percepción del valor y la utilidad de las herramientas gerenciales, modulando su trayectoria y determinando si se consolidan como prácticas establecidas o se desvanecen como modas pasajeras, y explorando cómo las antinomias organizacionales influyen en este proceso. Independientemente de los patrones específicos observados en los datos cuantitativos, la tesis explorará las tensiones organizacionales, los factores culturales y las dinámicas de poder que podrían influir en la adopción y el abandono de herramientas gerenciales.

**Nota relevante:** Si bien los informes técnicos y la tesis doctoral abordan la misma temática general, es necesario aclarar que lo hacen desde perspectivas metodológicas muy distintas pero complementarias. Los informes proporcionan una base empírica cuantitativa, mientras que la tesis ofrece una interpretación cualitativa y una profundización teórica. *Los informes técnicos, por lo tanto, sirven como punto de partida empírico, proporcionando un contexto cuantitativo y un anclaje descriptivo para la posterior investigación cualitativa, pero no predeterminan ni condicionan las conclusiones de la tesis doctoral.* Ambos componentes son esenciales para una comprensión holística del fenómeno de las modas gerenciales, y su combinación dialéctica representa una contribución original y significativa al campo de la investigación en gestión. *La tesis se apoya en los informes, pero los trasciende y los contextualiza, sin que sus hallazgos sean vinculantes para el desarrollo de la misma.*

### **Objetivo de la serie de informes**

El objetivo central de esta serie de informes técnicos es proporcionar una base empírica para el análisis del fenómeno de las innovaciones tecnológicas administrativas (herramientas gerenciales), de las que se dicen exhiben un comportamiento similar al fenómeno de las modas. A través de un enfoque cuantitativo y el análisis de datos provenientes de múltiples fuentes, se examina el comportamiento de 23 grupos de herramientas de gestión (cada uno potencialmente compuesto por una o más herramientas específicas). Los informes buscan identificar tendencias, patrones cíclicos, y la posible influencia de factores contextuales en la adopción y percepción de este grupo de herramientas para proporcionar un análisis particular, permitiendo una comprensión profunda de su evolución y uso desde bases de datos distintas.

### **Sobre los autores y contribuciones**

Este informe es producto de una colaboración interdisciplinaria que integra la experticia en las ciencias sociales y la ingeniería de software:

**Diomar Añez:** Investigador principal. Su formación multidisciplinaria (Estudios base en Filosofía, Comunicación Social, con posgrados en Valoración de Empresas, Planificación Financiera y Economía), y su formación doctoral en Ciencias Gerenciales; junto con más de 25 años de experiencia en consultoría organizacional en diversos sectores: aporta el rigor conceptual y académico. Es responsable del marco teórico, la selección de las herramientas gerenciales, y la significación de los datos, con un enfoque en los lineamientos para la trama interpretativa de los resultados, centrándose en la comprensión de las dinámicas subyacentes a la adopción y el abandono de las herramientas gerenciales en moda.

**Dimar Añez:** Programador en Python. Con formación en Ingeniería en Computación y Electrónica, y una vasta experiencia en análisis de datos, desarrollo de *software*, y con experticia en *machine learning*, ciencia de datos y *big data*. Ha liderado múltiples proyectos para el diseño e implementación de soluciones de sistemas, incluyendo análisis estadísticos en Python. Gestionó la extracción automatizada de datos, realizó su preprocesamiento y limpieza, aplicó las técnicas de modelado estadístico, y desarrolló las visualizaciones de resultados, garantizando la precisión, confiabilidad y escalabilidad del análisis.

### Estructura de los Informes

La serie completa consta de 138 informes. Cada uno se centra en el análisis de un grupo de herramientas utilizando una única fuente de datos para cada informe. Los 23 grupos de herramientas que se han establecido, se describen a continuación:

#	GRUPO DE HERRAMIENTAS	DESCRIPCIÓN CONCISA	HERRAMIENTAS INTEGRADAS
1	REINGENIERÍA DE PROCESOS	Rediseño radical de procesos para mejoras drásticas en rendimiento, optimizando y transformando procesos existentes.	Reengineering, Business Process Reengineering (BPR)
2	GESTIÓN DE LA CADENA DE SUMINISTRO	Coordinación y optimización de flujos de bienes, información y recursos desde el proveedor hasta el cliente final.	Supply Chain Integration, Supply Chain Management (SCM)
3	PLANIFICACIÓN DE ESCENARIOS	Creación de modelos de futuros alternativos para apoyar la toma de decisiones estratégicas y desarrollar planes de contingencia.	Scenario Planning, Scenario and Contingency Planning, Scenario Analysis and Contingency Planning
4	PLANIFICACIÓN ESTRATÉGICA	Proceso sistemático para definir la dirección y objetivos a largo plazo, estableciendo una visión clara y estrategias para alcanzar metas.	Strategic Planning, Dynamic Strategic Planning and Budgeting
5	EXPERIENCIA DEL CLIENTE	Gestión de interacciones con clientes para mejorar satisfacción y lealtad, creando experiencias positivas.	Customer Satisfaction Surveys, Customer Relationship Management (CRM), Customer Experience Management
6	CALIDAD TOTAL	Enfoque de gestión centrado en la mejora continua y satisfacción del cliente, integrando la calidad en todos los aspectos organizacionales.	Total Quality Management (TQM)
7	PROPÓSITO Y VISIÓN	Definición de la razón de ser y aspiración futura de la organización, proporcionando una dirección clara.	Purpose, Mission, and Vision Statements

#	GRUPO DE HERRAMIENTAS	DESCRIPCIÓN CONCISA	HERRAMIENTAS INTEGRADAS
8	BENCHMARKING	Proceso de comparación de prácticas propias con las mejores organizaciones para identificar áreas de mejora.	Benchmarking
9	COMPETENCIAS CENTRALES	Capacidades únicas que otorgan ventaja competitiva.	Core Competencies
10	CUADRO DE MANDO INTEGRAL	Sistema de gestión estratégica que mide el desempeño desde múltiples perspectivas (financiera, clientes, procesos internos, aprendizaje y crecimiento).	Balanced Scorecard
11	ALIANZAS Y CAPITAL DE RIESGO	Mecanismos de colaboración y financiación para impulsar el crecimiento e innovación.	Strategic Alliances, Corporate Venture Capital
12	OUTSOURCING	Contratación de terceros para funciones no centrales.	Outsourcing
13	SEGMENTACIÓN DE CLIENTES	División del mercado en grupos homogéneos para adaptar estrategias de marketing.	Customer Segmentation
14	FUSIONES Y ADQUISICIONES	Combinación de empresas para lograr sinergias y crecimiento.	Mergers and Acquisitions (M&A)
15	GESTIÓN DE COSTOS	Control y optimización de costos en la cadena de valor.	Activity Based Costing (ABC), Activity Based Management (ABM)
16	PRESUPUESTO BASE CERO	Metodología de presupuestación que justifica cada gasto desde cero.	Zero-Based Budgeting (ZBB)
17	ESTRATEGIAS DE CRECIMIENTO	Planes y acciones para expandir el negocio y aumentar la cuota de mercado.	Growth Strategies, Growth Strategy Tools
18	GESTIÓN DEL CONOCIMIENTO	Proceso de creación, almacenamiento, difusión y aplicación del conocimiento organizacional.	Knowledge Management
19	GESTIÓN DEL CAMBIO	Proceso para facilitar la adaptación a cambios organizacionales.	Change Management Programs
20	OPTIMIZACIÓN DE PRECIOS	Uso de modelos y análisis para fijar precios que maximicen ingresos o beneficios.	Price Optimization Models
21	LEALTAD DEL CLIENTE	Estrategias para fomentar la retención y fidelización de clientes.	Loyalty Management, Loyalty Management Tools
22	INNOVACIÓN COLABORATIVA	Enfoque que involucra a múltiples actores (internos y externos) en el proceso de innovación.	Open-Market Innovation, Collaborative Innovation, Open Innovation, Design Thinking
23	TALENTO Y COMPROMISO	Gestión para atraer, desarrollar y retener a los mejores empleados.	Corporate Code of Ethics, Employee Engagement Surveys, Employee Engagement Systems

## Fuentes de datos y sus características

Se utilizan cinco fuentes de datos principales, cada una con sus propias características, fortalezas y limitaciones:

- **Google Trends (Indicador de atención mediática):** Como plataforma de análisis de tendencias de búsqueda, proporciona datos en tiempo real (o con mínima latencia) sobre la frecuencia relativa con la que los usuarios consultan términos específicos. Este índice de frecuencia de búsqueda actúa como un proxy de la atención mediática y la curiosidad pública en torno a una herramienta de gestión determinada. Un incremento abrupto en el volumen de búsqueda puede señalar la emergencia de una moda gerencial, mientras que una tendencia sostenida a lo largo del tiempo sugiere una mayor consolidación. No obstante,

es crucial reconocer que Google Trends no discrimina entre las diversas intenciones de búsqueda (informativa, académica, transaccional, etc.), lo que introduce un posible sesgo en la interpretación de los datos. Los datos de Google Trends se utilizan como un indicador de la atención pública y el interés mediático en las herramientas gerenciales a lo largo del tiempo.

- **Google Books Ngram (Corpus lingüístico diacrónico):** Ofrece acceso a un compuesto por la digitalización de millones de libros, lo que permite cuantificar la frecuencia de aparición de un término específico a lo largo de extensos períodos. Un incremento gradual y sostenido en la frecuencia de un término sugiere su progresiva incorporación al discurso académico y profesional. Fluctuaciones (picos y valles) pueden reflejar períodos de debate, controversia o resurgimiento de interés. Para la interpretación de los datos de *Ngram Viewer* debe considerarse las limitaciones inherentes al corpus (v. g., sesgos de idioma, género literario, disciplina, etc.) así como la ausencia de contexto de uso del término. Los datos de *Ngram Viewer* se utilizan para analizar la presencia y evolución de los términos relacionados con las herramientas gerenciales en la literatura publicada.
- **Crossref.org (Repositorio de metadatos académicos):** Constituye un repositorio exhaustivo de metadatos de publicaciones (artículos, libros, actas de congresos, etc.); cuyos datos permiten evaluar la adopción, difusión y citación de un concepto dentro de la literatura científica revisada por pares. Un incremento sostenido en el número de publicaciones y citas asociadas a una herramienta de gestión sugiere una creciente legitimidad académica y una consolidación teórica. La diversidad de autores, afiliaciones institucionales y revistas indexadas puede indicar la amplitud de la adopción del concepto. Sin embargo, es importante reconocer que Crossref no captura el contenido completo de las publicaciones, ni mide directamente su impacto o calidad intrínseca. Los datos de Crossref se utilizan para evaluar la producción académica y la legitimidad científica de las herramientas gerenciales.
- **Bain & Company - Usabilidad (Penetración de mercado):** Se trata de un indicador basado en encuestas a ejecutivos y gerentes, que proporciona una medida cuantitativa de la penetración de mercado de una herramienta de gestión específica. Este indicador refleja el porcentaje de organizaciones que reportan haber adoptado la herramienta en su práctica empresarial. Una alta usabilidad sugiere una amplia adopción, mientras que una baja usabilidad indica una penetración limitada. No obstante, es crucial reconocer que este indicador no captura la profundidad, intensidad o efectividad de la implementación de la herramienta dentro de cada organización. El porcentaje de usabilidad se utiliza como una medida de la adopción declarada de las herramientas gerenciales en el ámbito empresarial.
- **Bain & Company - Satisfacción (Valor percibido):** Este índice también basado en encuestas a ejecutivos y gerentes, mide el valor percibido de una herramienta de gestión desde la perspectiva de los usuarios. Generalmente expresado en una escala numérica, refleja el grado de satisfacción que expresan los usuarios sobre el uso de la herramienta, considerando su utilidad, facilidad de uso y cumplimiento de expectativas. Una alta puntuación sugiere una experiencia de usuario positiva y una percepción de valor elevada. Sin

embargo, es fundamental reconocer la naturaleza subjetiva de este indicador y su potencial sensibilidad a factores contextuales y expectativas individuales. La combinación de la usabilidad y la satisfacción dan un panorama de adopción. El índice de satisfacción se utiliza como una medida de la percepción subjetiva del valor y la experiencia del usuario con las herramientas gerenciales.

## Entorno tecnológico y software utilizado

La presente investigación se apoya en un conjunto de herramientas de software de código abierto, seleccionadas por su robustez, flexibilidad y capacidad para realizar análisis estadísticos avanzados y visualización de datos. El entorno tecnológico principal se basa en el lenguaje de programación Python (versión 3.11), junto con una serie de bibliotecas especializadas. A continuación, se detallan los componentes clave:

- *Python* (`== 3.11`)<sup>4</sup>: Lenguaje de programación principal, elegido por su versatilidad, amplia adopción en la comunidad científica y disponibilidad de bibliotecas especializadas en análisis de datos. Se utilizó un entorno virtual de Python (venv) para gestionar las dependencias del proyecto y asegurar la consistencia entre diferentes entornos de ejecución.
- *Bibliotecas de Análisis de Datos*:
- *Bibliotecas principales de Análisis Estadístico*
  - *NumPy* (`numpy==1.26.4`): Paquete de computación científica, proporciona objetos de arreglos N-dimensionales, álgebra lineal, transformadas de Fourier y capacidades de números aleatorios.
  - *Pandas* (`pandas==2.2.3`): Biblioteca para manipulación y análisis de datos, ofrece objetos *DataFrame* para manejo eficiente de datos, lectura/escritura de diversos formatos y funciones de limpieza, transformación y agregación.
  - *SciPy* (`scipy==1.15.2`): Biblioteca avanzada de computación científica, incluye módulos para optimización, álgebra lineal, integración, interpolación, procesamiento de señales y más.
  - *Statsmodels* (`statsmodels==0.14.4`): Paquete de modelado estadístico, proporciona clases y funciones para estimar modelos estadísticos, pruebas estadísticas y análisis de series temporales.
  - *Scikit-learn* (`scikit-learn==1.6.1`): Biblioteca de *machine learning*, ofrece herramientas para preprocesamiento de datos, reducción de dimensionalidad, algoritmos de clasificación, regresión, *clustering* y evaluación de modelos.
- *Análisis de series temporales*
  - *Pmdarima* (`pmdarima==2.0.4`): Implementación de modelos ARIMA, incluye selección automática de parámetros (`auto_arima`) para pronósticos y análisis de series temporales.

<sup>4</sup> El símbolo “==” refiere a la versión exacta de una biblioteca o paquete de software, generalmente en el ámbito de la programación en Python cuando se trabaja con herramientas de gestión de dependencias como `pip` o `requirements.txt` para asegurar que no se instalará una versión más reciente que podría introducir cambios o errores inesperados. Otros símbolos en este contexto: (i) “>=” (mayor o igual que): permite versiones iguales o superiores a la indicada. (ii) “<=” (menor o igual que): permite versiones iguales o inferiores. (iv) “!=” (diferente de): Excluye una versión específica.

— *Bibliotecas de visualización*

- *Matplotlib* (*matplotlib==3.10.0*): Biblioteca integral para gráficos 2D, crea figuras de calidad para publicaciones y es la base para muchas otras bibliotecas de visualización.
- *Seaborn* (*seaborn==0.13.2*): Basada en matplotlib, ofrece una interfaz de alto nivel para crear gráficos estadísticos atractivos e informativos.
- *Altair* (*altair==5.5.0*): Basada en Vega y Vega-Lite, diseñada para análisis exploratorio de datos con una sintaxis declarativa.

— *Generación de reportes*

- *FPDF* (*fpdf==1.7.2*): Generación de documentos PDF, útil para crear reportes estadísticos.
- *ReportLab* (*reportlab==4.3.1*): Mejor que FPDF, soporta diseños y gráficos complejos (PDF).
- *WeasyPrint* (*weasyprint==64.1*): Convierte HTML/CSS a PDF, útil para crear reportes a partir de plantillas HTML.

— *Integración de IA y Machine Learning*

- *Google Generative AI* (*google-generativeai==0.8.4*): Cliente API de IA generativa de Google, para procesamiento de lenguaje natural de resultados estadísticos y generación de *insights*.

— *Soporte para procesamiento de datos*

- *Beautiful Soup* (*beautifulsoup4==4.13.3*): Parseo de HTML y XML, útil para web *scraping* de datos para análisis.
- *Requests* (*requests==2.32.3*): Biblioteca HTTP para realizar llamadas a APIs y obtener datos.

— *Desarrollo y pruebas*

- *Pytest* (*pytest==8.3.4*, *pytest-cov==6.0.0*): Framework de pruebas que asegura el correcto funcionamiento de las funciones estadísticas.
- *Flake8* (*flake8==7.1.2*): Herramienta de *linting* de código para mantener la calidad del código.

— *Bibliotecas de Utilidad*

- *Tqdm* (*tqdm==4.67.1*): Biblioteca de barras de progreso (cálculos estadísticos de larga duración).
- *Python-dotenv* (*python-dotenv==1.0.1*): Gestión de variables de entorno, útil para configuración.

— *Clasificación por función estadística*

- *Estadística descriptiva*: NumPy, pandas, SciPy, statsmodels
- *Estadística inferencial*: SciPy, statsmodels
- *Análisis de series temporales*: statsmodels, pmdarima, pandas
- *Machine learning*: scikit-learn
- *Visualización*: Matplotlib, Seaborn, Plotly, Altair
- *Generación de reportes*: FPDF, ReportLab, WeasyPrint

— *Replicabilidad*: El *pipeline* completo de análisis de esta investigación, desde la ingesta de datos crudos hasta la generación de visualizaciones finales, ha sido implementado en Python y disponible en GitHub:



<https://github.com/Wise-Connex/Management-Tools-Analysis/>. Este repositorio encapsula todos los *scripts* empleados, junto con un «requirements.txt» para la replicación del entorno virtual (*venv/conda*), con instrucciones en el «README.md» para el *setup* y la ejecución del *workflow*, y la configuración de *linters* para asegurar la calidad y consistencia del código. Se ha priorizado la modularidad y la parametrización de los *scripts* para facilitar su mantenimiento y extensión. Esta apertura total del «codebase» garantiza la transparencia del proceso computacional y la replicabilidad *bit-a-bit* de los resultados, para que la comunidad de desarrolladores y científicos de datos puedan realizar *forks*, proponer *pull requests* con mejoras o adaptaciones, y desarrollar investigaciones o aplicaciones derivadas.

- *Repositorio*: La colección integral de conjuntos de datos primarios (*raw data*) y procesados que sustentan esta investigación se encuentra curada y disponible en el repositorio Harvard Dataverse<sup>5</sup>, de la Universidad epónima, accesible en <https://dataverse.harvard.edu/dataverse/management-fads>, y estructurado en tres *sub-Dataverses*: uno con los extractos de datos en su forma original (*mgmt\_raw\_data*), otro para los índices comparativos normalizados y/o estandarizados (*mgmt\_normalized\_indices*), y uno para los metadatos bibliográficos detallados recuperados de Crossref (*mgmt\_crossref\_metadata*). En cada *sub-Dataverse*, los datos de las 23 herramientas se organizan en *Datasets* individuales. Los datos cuantitativos se proporcionan en formato CSV y los metadatos bibliográficos en formato JSON estructurado, y encapsulados en archivos comprimidos. Cada *Dataset* está acompañado de metadatos exhaustivos, conformes con el esquema Dublin Core<sup>6</sup>, que describen la procedencia, la estructura de los datos, las metodologías de procesamiento aplicadas e información contextual para su interpretación y reutilización. El control de versiones y la asignación de *Identificadores de Objeto Digital (DOI)*, asegura la trazabilidad y reproducibilidad de los hallazgos de la investigación, diseñada para potenciar la confiabilidad de las conclusiones presentadas y facilitar la reutilización crítica, la replicación y la integración de estos datos en futuras investigaciones promoviendo así el desarrollo del conocimiento en las ciencias gerenciales.
- *Justificación de la elección tecnológica*: La elección del conjunto de códigos y bibliotecas se basa en:
  - *Código abierto y comunidad activa*: Python y las bibliotecas son de código abierto, con comunidades de usuarios y desarrolladores activas, lo que garantiza soporte, actualizaciones y transparencia.
  - *Flexibilidad y extensibilidad*: Python permite adaptar y extender las funcionalidades existentes, así como integrar nuevas herramientas según sea necesario.
  - *Rigor científico*: Las bibliotecas utilizadas implementan métodos estadísticos confiables y ampliamente aceptados en la comunidad científica.
  - *Reproducibilidad*: La disponibilidad del código fuente y la descripción detallada de la metodología garantizan la reproducibilidad de los análisis.

<sup>5</sup> Su gestión se lleva a cabo mediante una colaboración entre la *Biblioteca de Harvard*, el *Departamento de Tecnología de la Información de la Universidad de Harvard (HUIT)* y el *Instituto de Ciencias Sociales Cuantitativas (IQSS) de Harvard*. El repositorio forma parte del Proyecto Dataverse.

<sup>6</sup> Se trata de un estándar de metadatos definido por la *Dublin Core Metadata Initiative (DCMI)* (<http://purl.org/dc/terms/>), que combina elementos simples (15 propiedades originales, *ISO 15836-1*) y calificados (propiedades y clases avanzadas, *ISO 15836-2*) para optimizar la descripción semántica de recursos, garantizando interoperabilidad con estándares globales y cumplimiento con los *principios FAIR* (Encontrable, Accesible, Interoperable, Reutilizable) para facilitar la persistencia de citas, el descubrimiento en múltiples plataformas y la inclusión en índices de citas de datos, apoyando la gestión de datos de investigación en entornos de ciencia abierta.

## ALCANCES METODOLÓGICOS DEL ANÁLISIS

### Procedimientos de análisis

El presente informe se sustenta en un sistema de análisis estadístico modular replicable, implementado en el lenguaje de programación Python, aprovechando su flexibilidad, extensibilidad y la disponibilidad de bibliotecas especializadas en análisis de datos y modelado estadístico. Se trata de un sistema, diseñado *ex profeso* para este estudio, que automatiza los procesos de extracción, preprocesamiento, transformación, análisis (modelos ARIMA, descomposición de Fourier) y visualización de datos provenientes de cinco fuentes heterogéneas identificadas previamente para caracterizar la existencia o prevalencia de modelos de patrones temporales, tendencias, ciclos y posibles relaciones en el comportamiento de las herramientas gerenciales, con el fin último de discriminar entre comportamientos efímeros (“modas”) y estructurales (“doctrinas”) mediante criterios cuantitativos.

#### 1. Extracción, preprocesamiento y armonización de datos:

Se implementaron rutinas *ad hoc* para la extracción automatizada de datos de cada fuente, utilizando técnicas de *web scraping* (para Google Trends y Google Books Ngram), interfaces de programación de aplicaciones (APIs) (para Crossref.org) y la importación y procesamiento de datos proporcionados en formatos estructurados (basado en las investigaciones publicadas) (en el caso de *Bain & Company*) donde, adicionalmente, los datos de “Satisfacción” fueron estandarizados mediante *Z-scores* para facilitar su análisis.

Los datos en bruto fueron sometidos a un proceso de preprocesamiento, que incluyó:

- *Transformación*: Normalización y estandarización de variables (cuando fue necesario para la aplicación de técnicas estadísticas específicas), conversión de formatos de fecha y hora, y creación de variables derivadas (v.gr., tasas de crecimiento, diferencias, promedios móviles).
- *Validación*: Verificación de la consistencia y coherencia de los datos, así como de la integridad de los metadatos asociados.
- *Armonización temporal*: Debido a la heterogeneidad en la granularidad temporal de las fuentes de datos, se implementó un proceso de armonización para obtener una base de datos temporalmente consistente.
  - La interpolación se realizó con el objetivo de armonizar la granularidad temporal de las diferentes fuentes de datos, permitiendo la identificación de posibles relaciones y desfases temporales entre las variables. Se reconoce que la interpolación introduce un grado de estimación en los datos, y

que la extrapolación implica un grado de predicción, y que los valores resultantes no son observaciones directas. Se recomienda por ello interpretar los resultados derivados de datos interpolados/extrapolados con cautela, especialmente en los análisis de alta frecuencia (como el análisis estacional).

- Un requisito fundamental para el análisis longitudinal y modelado econométrico subsiguiente fue la armonización de las distintas series temporales a una granularidad mensual uniforme. El objetivo de esta armonización fue crear una base de datos con una granularidad temporal común (mensual) que permitiera la potencial comparación directa y análisis conjunto de las series temporales provenientes de las diferentes fuentes (en la Tesis Doctoral). Dado que los datos originales provenían de fuentes diversas con frecuencias de reporte heterogéneas, se implementó un protocolo de preprocesamiento específico para cada fuente. Este proceso incluyó:
  - **Google Trends:** Se utilizaron los datos recuperados directamente de la plataforma *Google Trends* para el intervalo temporal comprendido entre enero de 2004 y febrero de 2025, basados en los términos de búsquedas predefinidos.
    - Dada la extensión plurianual de este período, *Google Trends* inherentemente agrega y proporciona los datos con una granularidad mensual. No se realiza ninguna agregación temporal o cálculo de promedios a posteriori; y la serie de tiempo mensual es la resolución nativa ofrecida por la plataforma para rangos de esta magnitud. La métrica obtenida es el Índice de Interés de Búsqueda Relativo (*Relative Search Interest - RSI*). Este índice no cuantifica el volumen absoluto de búsquedas, sino que mide la popularidad de un término de búsqueda específico en una región y período determinados, en relación consigo mismo a lo largo de ese mismo período y región.
    - La normalización de este índice la realiza *Google Trends* estableciendo el punto de máxima popularidad (el pico de interés de búsqueda) para el término dentro del período consultado (enero 2004 - febrero 2025) como el valor base de 100. Todos los demás valores mensuales del índice se calculan y expresan de forma proporcional a este punto máximo.
    - Es fundamental interpretar estos datos como un indicador de la prominencia o notoriedad relativa de un tema en el buscador a lo largo del tiempo, y no como una medida de volumen absoluto o cuota de mercado de búsquedas. Los datos se derivan de un muestreo anónimo y agregado del total de búsquedas realizadas en Google.

- **Google Books Ngram:** Se utilizaron datos extraídos del *corpus* de *Google Books Ngram Viewer*, correspondientes a la frecuencia de aparición de términos (n-gramas) predefinidos dentro de los textos digitalizados. Los datos cubren el período anual desde 1950 hasta 2019 en el idioma inglés, basados en los términos de búsqueda.
  - La resolución temporal nativa proporcionada por *Google Books Ngram Viewer* para estos datos es estrictamente anual. En consecuencia, no se realizó ninguna interpolación ni estimación intra-anual; el análisis opera directamente sobre la serie de tiempo anual original. Es fundamental destacar que las cifras proporcionadas por *Google Books Ngram* representan frecuencias relativas. Para cada año, la frecuencia de un *n-grama* se calcula como su número de apariciones dividido por el número total de *n-gramas* presentes en el *corpus* de *Google Books* correspondiente a ese año específico. Este cálculo inherente normaliza los datos respecto al tamaño variable del *corpus* a lo largo del tiempo.
  - Dado que estas frecuencias relativas anuales pueden resultar en valores numéricos muy pequeños, dificultando su manejo e interpretación directa, se aplicó un procedimiento de normalización adicional a la serie de tiempo anual (1950-2019) obtenida. De manera análoga a la metodología de *Google Trends*, esta normalización consistió en establecer el año con la frecuencia relativa más alta dentro del período analizado como el valor base de 100. Todas las demás frecuencias relativas anuales fueron reescaladas proporcionalmente respecto a este valor máximo.
  - Este paso de normalización adicional transforma la escala original de frecuencias relativas (que pueden ser del orden de  $10^{-5}$  o inferior) a una escala más intuitiva con base a 100, facilitando el análisis visual y comparativo de la prominencia relativa del término a lo largo del tiempo, sin alterar la dinámica temporal subyacente.
- **Crossref:** Para evaluar la dinámica temporal de la producción científica en áreas temáticas específicas, se utilizó la infraestructura de metadatos de *Crossref*. El proceso metodológico comprendió las siguientes etapas clave:
  - *Recuperación inicial de datos:* Se ejecutaron consultas predefinidas contra la base de datos de *Crossref*, orientadas a identificar registros de publicaciones cuyos títulos contuvieran los términos de búsqueda de interés. Paralelamente, se cuantificó el volumen total de publicaciones registradas en *Crossref* (independientemente del tema) para cada mes dentro del mismo intervalo

temporal (enero 1950 - diciembre 2024). Esta fase inicial recuperó un conjunto amplio de metadatos potencialmente relevantes.

- *Refinamiento local y creación del sub-corpus*: Los metadatos recuperados fueron procesados en un entorno local. Se aplicó una segunda capa de filtrado mediante búsquedas booleanas más estrictas, nuevamente sobre los campos de título, para asegurar una mayor precisión temática y conformar un sub-corpus de publicaciones altamente relevantes para el análisis.
- *Curación y deduplicación*: El sub-corpus resultante fue sometido a un proceso de curación de datos estándar en bibliometría. Fundamentalmente, se eliminaron registros duplicados basándose en la identificación única proporcionada por los *Digital Object Identifiers* (DOIs). Esto garantiza que cada publicación distinta se contabilice una sola vez. Se omitieron los registros sin DOIs.
- *Agregación temporal y cuantificación mensual*: A partir del sub-corpus final, curado y deduplicado, se procedió a la agregación temporal para obtener una serie de tiempo mensual. Para cada mes calendario dentro del período de análisis (enero 1950 - diciembre 2024), se realizó un conteo directo del número absoluto de publicaciones cuya fecha de publicación registrada (utilizando la mejor resolución disponible en los metadatos) correspondía a dicho mes. Esto generó una serie de tiempo de volumen absoluto de producción científica sobre el tema.
  - Utilizando el conteo absoluto relevante y el conteo total de publicaciones en Crossref para el mismo mes (obtenido en el paso 1), se calculó la participación porcentual de las publicaciones relevantes respecto al total general ( $\text{Conteo Relevante} / \text{Conteo Total}$ ). Esto generó una serie de tiempo de volumen relativo, indicando la proporción de la producción científica total que representa el tema de interés cada mes.
- *Normalización del volumen de publicación*: La serie resultante de conteos mensuales relativos fue posteriormente normalizada. Siguiendo una metodología análoga a la empleada para otros indicadores de tendencia (como *Google Trends*), se identificó el mes con el mayor número de publicaciones dentro de todo el período analizado. Este punto máximo se estableció como valor base de 100. Todos los demás conteos se reescalaron de forma proporcional a este pico. El resultado es una serie de tiempo mensual normalizada que presenta la intensidad relativa de la producción científica registrada, facilitando la identificación de tendencias y picos de actividad en una escala comparable. No se aplicó ninguna técnica de interpolación.

- ***Bain & Company - Usabilidad:*** Para el análisis de la Usabilidad de herramientas gerenciales, se utilizaron datos provenientes de las encuestas periódicas "*Management Tools & Trends*" de Bain & Company. El procesamiento de estos datos, para adaptarlos a un análisis mensual y normalizado, implicó las siguientes consideraciones y pasos metodológicos:
  - *Naturaleza de los datos fuente:*
    - *Métrica:* El indicador primario es el porcentaje de Usabilidad reportado para cada herramienta gerencial evaluada.
    - *Fuente y disponibilidad:* Los datos se extrajeron directamente de los informes publicados por Bain, siguiendo el orden cronológico de aparición de las encuestas. Es crucial notar que Bain típicamente reporta sobre un subconjunto de herramientas (el "*top*"), no sobre la totalidad de herramientas existentes o potencialmente evaluadas.
    - *Periodicidad:* La publicación de estos datos es irregular, generalmente con una frecuencia bianual o trianual, resultando en una serie de tiempo original con puntos de datos dispersos.
    - *Contexto de la encuesta:* Se reconoce que cada oleada de la encuesta puede haber sido administrada a un número variable de encuestados y potencialmente a cohortes con características distintas. Aunque la metodología exacta de encuesta no es pública, se valora la longevidad de la encuesta y su enfoque en directivos y gerentes. Sin embargo, se debe considerar la posibilidad de sesgos inherentes a la perspectiva de una consultora como Bain.
    - *Cobertura temporal variable:* La disponibilidad de datos para cada herramienta específica varía significativamente; algunas tienen registros de larga data, mientras que otras aparecen solo en encuestas más recientes o de corta duración.
  - *Pre-procesamiento y agrupación semántica:* Dada la evolución de las herramientas gerenciales y los posibles cambios en su nomenclatura o alcance a lo largo del tiempo, se realizó un agrupamiento semántico.
    - Se identificaron herramientas que representan extensiones, evoluciones o variantes cercanas de otras, y sus respectivos datos de Usabilidad fueron combinados o asignados a una categoría conceptual unificada para crear series de tiempo más coherentes y extensas.

- *Normalización de los datos originales:* Posterior a la estructuración y agrupación semántica, se aplicó un procedimiento de normalización a los puntos de datos de Usabilidad (%) originales y dispersos para cada herramienta (o grupo de herramientas).
  - Para cada herramienta/grupo, se identificó el valor máximo de Usabilidad (%) reportado en cualquiera de las encuestas disponibles para esa herramienta específica a lo largo de todo su historial registrado. Este valor máximo se estableció como la base 100.
  - Todos los demás puntos de datos de Usabilidad (%) originales para esa misma herramienta/grupo fueron reescalados proporcionalmente respecto a su propio máximo histórico. El resultado es una serie de tiempo dispersa, ahora en una escala normalizada de 0 a 100 para cada herramienta, donde 100 representa su pico histórico de usabilidad reportada.
- *Interpolación temporal para estimación mensual:* Con el fin de obtener una serie de tiempo mensual continua a partir de los datos normalizados y dispersos, se aplicó una interpolación temporal.
  - Se seleccionó la técnica de interpolación mediante *splines cúbicos*. Este método ajusta funciones polinómicas cúbicas por tramos entre los puntos de datos normalizados conocidos, generando una curva suave que pasa exactamente por dichos puntos. Se eligió esta técnica por su capacidad para capturar potenciales dinámicos no lineales en la tendencia de usabilidad entre las encuestas publicadas, lo que fundamenta la explicación de que los cambios en la usabilidad, reflejan ciclos de adopción y abandono, por lo cual tienden a ser progresivos, evolutivos y se manifiestan de manera suavizada dentro de las organizaciones a lo largo del tiempo.
  - Los *splines cúbicos* genera una curva suave (continua en su primera y segunda derivada, salvo en los extremos) que pasa exactamente por dichos puntos y es capaz de capturar aceleraciones o desaceleraciones en la adopción/abandono que podrían perderse con métodos más simples como la interpolación lineal.
  - Dada la naturaleza dispersa de los datos originales (puntos bianuales/trianuales) y la necesidad de una perspectiva temporal continua para analizar las tendencias subyacentes de adopción y abandono de estas

herramientas – procesos inherentemente cualitativos que evolucionan en el tiempo debido a múltiples factores– se requirió generar una serie de tiempo mensual completa a partir de los puntos de datos normalizados.

- **Protocolo de adherencia a límites (Clipping Post-Interpolación):** Se reconoció que la interpolación con *splines cúbicos* puede, en ocasiones, generar valores que exceden ligeramente el rango de los datos originales (fenómeno de *overshooting*).
  - Para asegurar la validez conceptual de los datos mensuales estimados en la escala normalizada, se implementó un mecanismo de recorte (*clipping*) después de la interpolación. Todos los valores mensuales interpolados resultantes fueron restringidos al rango “mínimo” y “máximo” de la serie. Esto garantiza que para los datos de usabilidad estimada no se generen otros máximos y mínimos fuera de los “máximos” y “mínimos” de la serie.
  - El resultado final de este proceso es una serie de tiempo mensual, estimada, normalizada (base 100) y acotada para la Usabilidad de cada herramienta (o grupo semántico de herramientas) gerencial analizada, derivada de los informes periódicos de Bain & Company y sujeta a las limitaciones y supuestos metodológicos descritos.
- **Bain & Company - Satisfacción:** Se procesaron los datos de “Satisfacción” con herramientas gerenciales, también provenientes de las encuestas periódicas “*Management Tools & Trends*” de Bain & Company. La “Satisfacción”, típicamente medida en una escala tipo Likert de 1 (Muy Insatisfecho) a 5 (Muy Satisfecho), requirió un tratamiento específico para su estandarización y análisis temporal.
  - **Naturaleza de los datos fuente y pre-procesamiento inicial:**
    - **Métrica:** El indicador primario es la puntuación de Satisfacción (escala original ~1-5).
    - **Características de la fuente:** Se reitera que las características fundamentales de la fuente de datos (periodicidad irregular, reporte selectivo “top”, variabilidad muestral, potencial sesgo de consultora, cobertura temporal variable por herramienta) son idénticas a las descritas para los datos de Usabilidad.
    - **Agrupación semántica:** De igual manera, se aplicó el mismo proceso de agrupación semántica para combinar datos de herramientas conceptualmente relacionadas o evolutivas.



- *Estandarización de “Satisfacción” mediante Z-Scores:*
  - *Razón y método:* Dada la naturaleza a menudo restringida del rango en las puntuaciones originales de Satisfacción (escala 1-5) y para cuantificar la desviación respecto a un punto de referencia significativo, se optó por estandarizar los datos originales dispersos mediante la transformación *Z-score*.
  - *Parámetros de estandarización:* La transformación se aplicó utilizando parámetros poblacionales justificados teóricamente:
    - *Media poblacional ( $\mu = 3.0$ ):* Se adoptó  $\mu=3.0$  basándose en la interpretación estándar de las *escalas Likert* de 5 puntos, donde “3” representa el punto de neutralidad o indiferencia teórica. El *Z-score* resultante,  $(X - 3.0) / \sigma$ , mide así directamente la desviación respecto a la indiferencia. Esta elección proporciona un *benchmark* estable y conceptualmente más significativo que una media muestral fluctuante, especialmente considerando la selectividad de los datos publicados por Bain.
    - *Desviación estándar poblacional ( $\sigma = 0.891609$ ):* Para mantener la coherencia metodológica, se utilizó una  $\sigma$  estimada en 0.891609. Este valor no es la desviación estándar convencional alrededor de la media muestral, sino la raíz cuadrada de la varianza muestral insesgada calculada respecto a la media poblacional fijada  $\mu=3.0$ , utilizando un conjunto de referencia de 201 puntos de datos (de 23 herramientas compendiadas en los 138 informes):  $\sigma \approx \sqrt{\sum (x_i - 3.0)^2 / (n - 1)}$  con  $n=201$ . Esta  $\sigma$  representa la dispersión típica estimada alrededor del punto de indiferencia (3.0), basada en la variabilidad observada en el *pool* de datos disponible, asegurando consistencia entre numerador y denominador del *Z-score*.
- *Transformación a escala de índice intuitiva (Post-Estandarización):* Tras la estandarización a *Z-scores*, estos fueron transformados a una escala de índice más intuitiva para facilitar la visualización y comunicación.
  - *Definición de la Escala:* Se estableció que el punto de indiferencia ( $Z=0$ , correspondiente a  $X=3.0$ ) equivaliera a un valor de índice de 50.
  - *Determinación del multiplicador:* El factor de escala (multiplicador del *Z-score*) se fijó en 22. Esta decisión se basó en el objetivo de que el valor

máximo teórico de satisfacción ( $X=5$ ), cuyo *Z-score* es  $(5-3)/0.891609 \approx +2.243$ , se mapeara aproximadamente a un índice de 100 ( $50 + 2.243 * 22 \approx 99.35$ ).

- *Fórmula y rango resultante*: La fórmula de transformación final es: Índice =  $50 + (Z\text{-score} \times 22)$ . En esta escala, la indiferencia ( $X=3$ ) es 50, la máxima satisfacción teórica ( $X=5$ ) es aproximadamente 100 ( $\sim 99.4$ ), y la mínima satisfacción teórica ( $X=1$ ,  $Z \approx -2.243$ ) se traduce en  $50 + (-2.243 * 22) \approx 0.65$ . Esto crea un rango operativo efectivo cercano a  $[0, 100]$ . Se prefirió esta escala  $[50 \pm \sim 50]$  sobre otras como las Puntuaciones T ( $50 + 10*Z$ ) por su mayor amplitud intuitiva al mapear el rango teórico completo (1-5) de la satisfacción original.
- *Interpolación temporal para estimación mensual*:
  - *Método*: La serie de puntos de datos discretos, ahora expresados en la escala de Índice de Satisfacción, requiere ser transformada en una serie temporal continua para el análisis mensual.
  - *Justificación de la interpolación*: Esta necesidad surge porque la Satisfacción, tal como es medida, refleja opiniones y percepciones de valor fundamentalmente cualitativas por parte de directivos y gerentes. Se parte del supuesto de que estas percepciones no permanecen estáticas entre las encuestas, sino que evolucionan continuamente a lo largo del tiempo. Esta evolución está influenciada por una multiplicidad de factores, muchos de ellos subjetivos, como experiencias acumuladas, resultados percibidos de la herramienta, cambios en el entorno competitivo, tendencias de gestión, etc. Por lo tanto, la interpolación se aplica para estimar la trayectoria más probable de esta dinámica perceptual subyacente entre los puntos de medición discretos disponibles.
  - *Selección y justificación de splines cúbicos*: Para realizar esta estimación mensual, se empleó el mismo procedimiento de interpolación temporal mediante *splines cúbicos*. La elección específica de este método se refuerza al considerar la naturaleza de los cambios de opinión y percepción. Se percibe que estos cambios tienden a ser progresivos y evolutivos, manifestándose generalmente de manera suavizada en las valoraciones agregadas. Los *splines cúbicos* son particularmente adecuados para representar esta dinámica, ya que generan una curva

suave que conecta los puntos conocidos y es capaz de modelar inflexiones no lineales. Esto permite capturar cómo las valoraciones subjetivas pueden acelerar, desacelerar o estabilizarse gradualmente en respuesta a los factores percibidos, ofreciendo una representación potencialmente más fiel que métodos lineales que asumirían una tasa de cambio constante entre encuestas.

- *Protocolo de adherencia a límites (Clipping Post-Interpolación):*
  - *Aplicación:* Finalmente, se aplicó un mecanismo de recorte (*clipping*) a los valores mensuales interpolados del Índice de Satisfacción. Los valores fueron restringidos al rango teórico operativo de la escala de índice, para corregir posibles sobreimpulsos (*overshooting*) de los *splines* y garantizar la validez conceptual de los resultados.
  - El producto final de este proceso es una serie de tiempo mensual, estimada, transformada a un índice de satisfacción (centro 50), y acotada, para cada herramienta (o grupo semántico) gerencial. Esta serie representa la evolución estimada de la satisfacción relativa a la indiferencia, derivada de los datos de Bain & Company mediante la secuencia metodológica descrita.

## 2. Análisis Exploratorio de Datos (AED):

Antes de aplicar técnicas de modelado formal, se realiza un Análisis Exploratorio de datos (AED) para cada herramienta gerencial y cada fuente de datos seleccionada. Este análisis sirve como base para los modelos posteriores y proporciona *insights* iniciales sobre los patrones temporales. La aplicación se centra en el análisis de tendencias temporales y comparaciones entre diferentes períodos, utilizando principalmente visualizaciones de series temporales y gráficos de barras para comunicar los resultados.

El AED implementado incluye:

— *Estadística descriptiva:*

- Cálculo de promedios móviles para diferentes períodos (1, 5, 10, 15, 20 años y datos completos).
- Identificación de valores máximos y mínimos en las series temporales.
- Análisis de tendencias para evaluar la dirección y magnitud de los cambios a lo largo del tiempo.
- Cálculo de tasas de crecimiento para diferentes períodos.

— *Visualización:*

- Generación de gráficos de series temporales que muestran la evolución de cada herramienta gerencial a lo largo del tiempo.
- Creación de gráficos de barras comparativos de promedios para diferentes períodos temporales.

- Visualización de tendencias con líneas de regresión superpuestas para identificar patrones de crecimiento o decrecimiento.
- *Análisis de tendencias. Implementación de análisis de tendencias para evaluar:*
  - Tendencias a corto plazo (1 año).
  - Tendencias a medio plazo (5-10 años).
  - Tendencias a largo plazo (15-20 años o más).
  - Comparación entre diferentes períodos para identificar cambios en la dirección de las tendencias.
  - Clasificación de tendencias como “creciente”, “decreciente” o “estable” basada en umbrales predefinidos.
  - Generación de afirmaciones interpretativas sobre las tendencias observadas.
- *Interpolación y manejo de datos faltantes:*
  - Aplicación de técnicas de interpolación (cúbica, B-spline).
  - Suavizado de datos utilizando promedios móviles para reducir el ruido y destacar tendencias subyacentes.
- *Normalización de datos:*
  - Implementación de normalización de conjuntos de datos para permitir potenciales comparaciones entre diferentes fuentes.
  - Combinación de datos normalizados de múltiples fuentes para análisis integrado

### **3. Modelado de series temporales:**

El núcleo del análisis implementado se centra en el modelado de series temporales, utilizando técnicas específicas para identificar patrones, tendencias y ciclos en la adopción de herramientas gerenciales: Análisis ARIMA (*Autoregressive Integrated Moving Average*). Se implementan modelos ARIMA que permite analizar y pronosticar tendencias futuras en la adopción de herramientas gerenciales. La selección de parámetros ARIMA (p,d,q) se realiza principalmente mediante funciones que automatizan la selección de los mejores parámetros. Aunque los parámetros predeterminados utilizados son (p=0, d=1, q=2), se permite la selección automática de parámetros óptimos basándose en el *Criterio de Información de Akaike* (AIC). Se advierte que el código no implementa explícitamente pruebas de diagnóstico para verificar la adecuación de los modelos o la ausencia de autocorrelación residual.

- *Análisis de descomposición estacional:*
  - Se implementa la descomposición estacional para separar las series temporales en componentes de tendencia, estacionalidad y residuo, permitiendo identificar patrones cíclicos en los datos.
  - La descomposición se realiza con un modelo aditivo o multiplicativo, dependiendo de las características de los datos.
  - Los resultados se visualizan en gráficos que muestran cada componente por separado, facilitando la interpretación de los patrones estacionales.

— *Análisis espectral (Análisis de Fourier):*

- Se implementa el análisis de Fourier descomponiendo las series temporales en sus componentes de frecuencia. Este análisis permite identificar ciclos dominantes en los datos, incluso aquellos que no son estrictamente periódicos.
- La implementación incluye la visualización de periodogramas que muestran la importancia relativa de cada frecuencia.
- Los resultados se presentan tanto en términos de frecuencia como de período (años), facilitando la interpretación de los ciclos identificados.

— *Técnicas de suavizado y procesamiento de datos:*

- Se aplican modelos de suavizado mediante promedios móviles que reduce el ruido y destaca tendencias subyacentes.
- Se utilizan técnicas de interpolación (lineal, cúbica, B-spline) para manejar datos faltantes y crear series temporales continuas.
- Estas técnicas se utilizan como preparación para el modelado y para mejorar la visualización de tendencias.

— *Análisis de tendencias:*

- Se implementa un análisis detallado de tendencias que evalúa la dirección y magnitud de los cambios a lo largo de diferentes períodos temporales.
- Este análisis complementa los modelos formales, proporcionando interpretaciones cualitativas de las tendencias observadas.
- La aplicación genera afirmaciones interpretativas sobre las tendencias, clasificándolas como “creciente”, “decreciente” o “estable” basándose en umbrales predefinidos.

— *Integración con IA Generativa:*

- Se integran modelos de IA generativa (a través de *google.generativeai*) para enriquecer el análisis de series temporales.
- Se utilizan modelos de lenguaje para generar interpretaciones contextuales de los patrones identificados en los datos.
- Estas interpretaciones se complementan los resultados de los modelos estadísticos, proporcionando *insights* adicionales sobre las tendencias observadas.

El enfoque de modelado implementado se centra en la identificación de patrones temporales y la generación de pronósticos, con un énfasis particular en la visualización e interpretación de resultados. Se combinan técnicas estadísticas tradicionales (ARIMA, análisis de Fourier, descomposición estacional) con enfoques modernos de análisis de datos e IA generativa para proporcionar un análisis integral de las tendencias en la adopción de herramientas gerenciales.

#### **4. Integración y visualización de resultados:**

Se implementa un sistema de integración y visualización de resultados que combina diferentes análisis para cada fuente de datos y herramienta gerencial. Este sistema se centra en la generación de informes visuales y textuales que facilitan la interpretación de los hallazgos, mediante la integración de resultados, y generando informes que incorporan visualizaciones, análisis estadísticos y texto interpretativo. Para ello, se convierte el contenido HTML/Markdown a PDF, en un formato estructurado.

— *Bibliotecas de visualización:*

- Se utiliza múltiples bibliotecas de visualización de manera complementaria para crear visualizaciones óptimas según el tipo de análisis:
  - *Matplotlib*: Para gráficos estáticos, incluyendo series temporales y gráficos de barras.
  - *Seaborn*: Para visualizaciones estadísticas mejoradas.

— *Tipos de visualizaciones implementadas:*

- *Series temporales*: Se generan gráficos de líneas que muestran la evolución temporal de las variables clave para cada herramienta gerencial. Se visualizan con diferentes niveles de suavizado para destacar tendencias subyacentes y configurados con formatos consistentes.
- *Gráficos comparativos*: Se generan gráficos de barras que comparan promedios para diferentes períodos temporales (1, 5, 10, 15, 20 años y datos completos). Estos gráficos utilizan un esquema de colores consistente para facilitar la comparación y en un formato estandarizado.
- *Descomposiciones estacionales*: Se generan visualizaciones de descomposición estacional. Estos gráficos muestran las componentes de tendencia, estacionalidad y residuo de las series temporales.
- *Análisis espectral*: Se generan espectrogramas que muestran la densidad espectral de las series temporales. Estos gráficos identifican las frecuencias dominantes en los datos, permitiendo detectar ciclos no evidentes en las visualizaciones directas.

— *Exportación y compartición de resultados*: Se permite guardar las visualizaciones como archivos de imagen independientes que pueden ser compartidos y archivados, facilitando la distribución de los resultados, mediante nombres únicos basados en las herramientas analizadas.

— *Transparencia y reproducibilidad*: El código está estructurado de manera que facilita la reproducibilidad. Las funciones están bien documentadas y los parámetros utilizados en los análisis son explícitos, permitiendo la replicación de los resultados. Se mantiene un registro de los análisis realizados, que se incluye en los informes generados.

El sistema está diseñado para facilitar la interpretación de patrones complejos en la adopción de herramientas gerenciales, utilizando una combinación de visualizaciones, análisis estadísticos y texto interpretativo generado tanto mediante IA como algorítmicamente.

## 5. Justificación de la elección metodológica

La elección de Python como lenguaje de programación y el enfoque en el modelado de series temporales se justifican por las siguientes razones:

- *Rigor*: Las técnicas de modelado de series temporales (ARIMA, descomposición estacional, análisis espectral) son métodos estadísticos sólidos y ampliamente aceptados para el análisis de datos longitudinales.
- *Flexibilidad*: Python y sus bibliotecas ofrecen una gran flexibilidad para adaptar los análisis a las características específicas de cada fuente de datos y cada herramienta gerencial.
- *Reproducibilidad*: El uso de un lenguaje de programación y la disponibilidad del código fuente garantizan la reproducibilidad de los análisis (Disponible en: <https://github.com/Wise-Connex/Management-Tools-Analysis/>)
- *Automatización*: Permite un flujo de trabajo automatizado.
- *Relevancia para el objeto de estudio*: Las técnicas seleccionadas son particularmente adecuadas para identificar patrones temporales, ciclos y tendencias, que son fundamentales para el estudio de las “modas gerenciales”.

Se eligió un enfoque cuantitativo para este estudio debido a la disponibilidad de datos numéricos longitudinales de múltiples fuentes, lo que permite la aplicación de técnicas estadísticas para identificar patrones y tendencias y un análisis sistemático y replicable de grandes volúmenes de datos. *Un enfoque más cualitativo, está reservado para el trabajo de investigación doctoral supra mencionado.*

Si bien el presente estudio se centra en la identificación de patrones y tendencias, es importante reconocer que no se pueden establecer relaciones causales definitivas a partir de los datos y las técnicas utilizadas, y es posible que existan variables omitidas o factores de confusión que influyan en los resultados. Para explorar posibles relaciones causales, se requerirían estudios adicionales con diseños experimentales o cuasi-experimentales, o el uso de técnicas econométricas avanzadas (v.gr., modelos de ecuaciones estructurales, análisis de causalidad de Granger) que permitan controlar por variables de confusión y establecer la dirección de la causalidad.

**NOTA METODOLÓGICA IMPORTANTE:**

- Los 138 informes técnicos que componen este estudio han sido diseñados para ser autocontenidos y proporcionar, cada uno, una descripción completa de la metodología utilizada; es decir, cada informe técnico está diseñado para que se pueda entender de forma independiente. Sin embargo, el lector familiarizado con la metodología general puede centrarse en las secciones que varían entre informes, optimizando así su tiempo y esfuerzo. Esto implica, necesariamente, la repetición de ciertas secciones en todos los informes. Para evitar una lectura redundante, se recomienda al lector lo siguiente:
  - Si ya ha revisado en informes previos las secciones **"MARCO CONCEPTUAL Y METODOLÓGICO"** y **"ALCANCES METODOLÓGICOS DEL ANÁLISIS"** en cualquiera de los informes, puede omitir su lectura en los informes subsiguientes, ya que esta información es idéntica en todos ellos. Estas secciones proporcionan el contexto teórico y metodológico general del estudio.
- La variación fundamental entre los informes se encuentra en los siguientes apartados:
  - La sección **"BASE DE DATOS ANALIZADA EN EL INFORME TÉCNICO"**, el contenido es específico para cada una de las cinco bases de datos utilizadas (Google Trends, Google Books Ngram Viewer, CrossRef, Bain & Company - Usabilidad, Bain & Company - Satisfacción). Dentro de cada base de datos, los 23 informes correspondientes de cada uno sí comparten la misma descripción de la base de datos. Es decir, hay cinco versiones distintas de esta sección, una para cada base de datos.
  - La sección **"GRUPO DE HERRAMIENTAS ANALIZADAS: INFORME TÉCNICO"** contiene elementos comunes a todos los informes de la misma herramienta gerencial, y presenta información de esta para ser analizada (nombre, descriptores lógicos, etc.).
  - La sección **"PARAMETRIZACIÓN PARA EL ANÁLISIS Y EXTRACCIÓN DE DATOS"** contiene elementos comunes a todos los informes de una misma base de datos (por ejemplo, la metodología general de Google Trends), pero también elementos específicos de cada herramienta (por ejemplo, los términos de búsqueda, el período de cobertura, etc.).



## BASE DE DATOS ANALIZADA EN EL INFORME TÉCNICO 20-GB

<i>Fuente de datos:</i>	<b>GOOGLE BOOKS NGRAM ("ARCHIVO HISTÓRICO")</b>
<i>Desarrollador o promotor:</i>	<b>Google LLC</b>
<i>Contexto histórico:</i>	Lanzado en 2010, Ngram Viewer se basa en el proyecto Google Books, iniciado en 2004, que ha digitalizado millones de libros de bibliotecas de todo el mundo.
<i>Naturaleza epistemológica:</i>	Frecuencias relativas de n-gramas (secuencias de n palabras) en un corpus diacrónico de libros digitalizados por Google. La frecuencia relativa se calcula como el número de ocurrencias del n-grama dividido por el número total de palabras en el corpus para un año dado, ajustado por un factor de escala. La unidad básica de análisis es el n-grama, considerado como un proxy lingüístico de un concepto o idea.
<i>Ventana temporal de análisis:</i>	Desde 1800 a 2022, es el período disponible más amplio, según la última actualización. La cobertura y la calidad de los datos pueden variar. Para los análisis realizados se ha delimitado a un marco de temporal desde 1950 a 2025.
<i>Usuarios típicos:</i>	Académicos (humanidades digitales, lingüística, historia, sociología), investigadores, escritores, lexicógrafos, público en general interesado en la evolución del lenguaje y las ideas.

<b><i>Relevancia e impacto:</i></b>	Proporciona una perspectiva diacrónica única de la evolución conceptual y terminológica en la literatura publicada. Su impacto radica en su capacidad para rastrear la emergencia, difusión y declive de ideas a lo largo de extensos períodos. Ampliamente utilizado en humanidades digitales, lingüística computacional, historia cultural y estudios de la ciencia. Su confiabilidad como reflejo del discurso escrito es alta dentro de los límites de su corpus, pero no es una medida directa de adopción o impacto en la práctica.
<b><i>Metodología específica:</i></b>	Utilización de descriptores lógicos (combinaciones booleanas de palabras clave) para identificar n-gramas relevantes para cada herramienta gerencial. Análisis longitudinal de series temporales de frecuencias relativas, identificando tendencias de largo plazo, puntos de inflexión, picos y valles mediante técnicas de análisis de series temporales y modelado de curvas de crecimiento.
<b><i>Interpretación inferencial:</i></b>	Los datos de Ngram Viewer deben interpretarse como un reflejo de la presencia, evolución y prominencia de un concepto en la literatura publicada, no como una medida directa de su adopción, implementación o impacto en la práctica organizacional.
<b><i>Limitaciones metodológicas:</i></b>	Sesgos inherentes al corpus: sobrerrepresentación de libros en inglés, publicaciones académicas y obras de editoriales establecidas, con subrepresentación de literatura gris, publicaciones en idiomas minoritarios y temas marginales. Ausencia de análisis contextual: Ngram Viewer solo registra la frecuencia, no el sentido o la valencia (positiva, negativa, neutra) del uso del término. Retraso en la incorporación de obras al corpus digitalizado. Posible evolución semántica de los términos a lo largo del tiempo, dificultando comparaciones directas en períodos extensos. Presencia de errores derivados del proceso de Reconocimiento Óptico de Caracteres (OCR) en la digitalización de textos antiguos.

<b><i>Potencial para detectar “Modas”:</i></b>	Moderado potencial para detectar "modas" en el largo plazo, pero con limitaciones importantes. La naturaleza retrospectiva y agregada de los datos permite identificar tendencias de uso de términos a lo largo de décadas o siglos, pero la latencia inherente a la publicación y digitalización de libros, así como los sesgos del corpus, dificultan la detección de fenómenos de corta duración. Un auge y declive rápido en la frecuencia de un término podría indicar una "moda", pero se requiere un análisis contextual cuidadoso para descartar otras explicaciones (cambios terminológicos, eventos específicos que impulsaron la publicación de libros sobre el tema, etc.). Mayor potencial para identificar tendencias de largo plazo y la persistencia (o no) de un concepto en el discurso escrito.
--	--

## GRUPO DE HERRAMIENTAS ANALIZADAS: INFORME TÉCNICO 20-GB

<i>Herramienta Gerencial:</i>	<b>OPTIMIZACIÓN DE PRECIOS (PRICE OPTIMIZATION)</b>
<i>Alcance conceptual:</i>	<p>La Optimización de Precios es un proceso analítico y un conjunto de técnicas que buscan determinar el precio óptimo para un producto o servicio, con el objetivo de maximizar un resultado específico (generalmente, los ingresos, los beneficios o la cuota de mercado). No se trata simplemente de fijar un precio basado en el costo más un margen, sino de utilizar datos y modelos para comprender cómo la demanda y los ingresos responden a diferentes niveles de precios. La optimización de precios considera una variedad de factores, incluyendo:</p> <ul style="list-style-type: none"> <li>• <b>Costos:</b> Los costos de producción, distribución y comercialización del producto o servicio.</li> <li>• <b>Demanda:</b> La cantidad de producto o servicio que los clientes están dispuestos a comprar a diferentes precios (elasticidad de la demanda).</li> <li>• <b>Competencia:</b> Los precios de los productos o servicios de la competencia.</li> <li>• <b>Segmentos de clientes:</b> Las diferentes sensibilidades al precio de los distintos grupos de clientes.</li> <li>• <b>Objetivos de la empresa:</b> Los objetivos estratégicos de la empresa (por ejemplo, maximizar la cuota de mercado, maximizar los beneficios a corto plazo, etc.).</li> </ul>

	<ul style="list-style-type: none"> <li>• Canales de distribución: Los diferentes precios que se pueden aplicar en diferentes canales (por ejemplo, online vs. offline).</li> <li>• Ciclo de vida del producto: Las diferentes estrategias de precios que se pueden aplicar en las diferentes etapas del ciclo de vida del producto.</li> <li>• Factores externos: Condiciones económicas, estacionalidad, promociones, etc.</li> </ul> <p>La optimización de precios puede ser un proceso complejo y dinámico, especialmente en entornos con alta competencia, rápida innovación y acceso a grandes volúmenes de datos (big data).</p>
<b>Objetivos y propósitos:</b>	- Comunicación: Facilitar la comunicación y el debate sobre el futuro dentro de la organización.
<b>Circunstancias de Origen:</b>	<p>La optimización de precios, en sus formas más básicas, existe desde que existe el comercio. Sin embargo, el desarrollo de técnicas y modelos sofisticados de optimización de precios es más reciente, impulsado por:</p> <ul style="list-style-type: none"> <li>• Investigación Operativa: El desarrollo de modelos matemáticos y algoritmos para la optimización.</li> <li>• Economía: El estudio de la demanda, la elasticidad y la fijación de precios.</li> <li>• Marketing: La investigación sobre el comportamiento del consumidor y la segmentación del mercado.</li> <li>• Tecnología de la Información: La disponibilidad de grandes volúmenes de datos sobre precios, ventas y clientes, y el desarrollo de software especializado para la optimización de precios.</li> </ul>
<b>Contexto y evolución histórica:</b>	<ul style="list-style-type: none"> <li>• Siglo XX: Desarrollo de los primeros modelos de optimización de precios en la investigación de operaciones y la economía.</li> <li>• Décadas de 1970 y 1980: Aplicación de técnicas de optimización de precios en industrias como la aviación (yield management).</li> <li>• Década de 1990 y posteriores: Auge de la optimización de precios en diversos sectores, impulsado por el crecimiento del comercio electrónico, la disponibilidad de datos y el desarrollo de software especializado.</li> </ul>

<p><b><i>Figuras claves (Impulsores y promotores):</i></b></p>	<ul style="list-style-type: none"> <li>• Investigadores en investigación de operaciones, economía y marketing: Han desarrollado modelos y técnicas para la optimización de precios.</li> <li>• Empresas de software: Han desarrollado software especializado para la optimización de precios (por ejemplo, PROS, Vendavo, Zilliant, Pricefx).</li> <li>• Empresas pioneras: Empresas en sectores como la aviación, el comercio minorista, la hotelería y el comercio electrónico han sido pioneras en la adopción de técnicas de optimización de precios.</li> </ul>
<p><b><i>Principales herramientas gerenciales integradas:</i></b></p>	<p>La Optimización de Precios, como proceso, utiliza una variedad de modelos, técnicas y herramientas:</p> <p>a. Price Optimization Models (Modelos de Optimización de Precios):</p> <p>Definición: Modelos matemáticos y algoritmos que se utilizan para determinar los precios óptimos.</p> <p>Objetivos: Los mencionados anteriormente para el grupo en general.</p> <p>Origen y promotores: Investigación de operaciones, economía, marketing, empresas de software.</p>
<p><b><i>Nota complementaria:</i></b></p>	<p>La optimización de precios es un proceso continuo y dinámico, que requiere un seguimiento constante de los resultados, un análisis de los datos y una adaptación a los cambios del entorno.</p>

## PARAMETRIZACIÓN PARA EL ANÁLISIS Y EXTRACCIÓN DE DATOS

<b>Herramienta Gerencial:</b>	<b>OPTIMIZACIÓN DE PRECIOS</b>
<b>Términos de Búsqueda (y Estrategia de Búsqueda):</b>	Price Optimization + Pricing Optimization + Dynamic Pricing Models + Optimal Pricing + Dynamic Pricing
<b>Criterios de selección y configuración de la búsqueda:</b>	<p>Corpus: English (general)</p> <p>Case Insensitive: Desactivado</p> <p>Suavizado: 0 (Sin suavizado)</p>
<b>Métrica e Índice (Definición y Cálculo)</b>	<p>La métrica utilizada por Google Books Ngram Viewer es la frecuencia relativa, calculada de la siguiente manera:</p> <p>Frecuencia Relativa = (Número de apariciones del término / Total de palabras en el corpus para el año) <math>\times</math> 100</p> <p>Esta métrica refleja la proporción de apariciones de los términos de búsqueda (o conjunto de términos) en relación con el número total de palabras en el corpus de libros en inglés para cada año. Un valor más alto indica una mayor prominencia relativa del término en el corpus de libros en inglés en ese año. Es importante destacar que esta métrica mide la frecuencia de uso en la literatura publicada, no la popularidad general del término.</p>

<p><b><i>Período de cobertura de los Datos:</i></b></p>	<p>Marco Temporal: 1950-2022 (Seleccionado para cubrir un amplio período de desarrollo de la gestión empresarial, incluyendo el auge de la informática y la globalización).</p>
<p><b><i>Metodología de Recopilación y Procesamiento de Datos:</i></b></p>	<ul style="list-style-type: none"> <li>- La interpretación de los datos de Google Books Ngram Viewer se centra en las tendencias de frecuencia relativa a lo largo del tiempo.</li> <li>- Estos datos provienen del corpus de libros digitalizados por Google Books.</li> <li>- Las fluctuaciones en la frecuencia relativa indican cambios en la aparición, uso y relevancia de los términos en la literatura publicada, reflejando potencialmente la evolución del discurso académico y profesional en torno a las herramientas gerenciales.</li> <li>- La amplia disponibilidad de datos permite un análisis diacrónico (a través del tiempo) contextualizado en la evolución de la literatura y el lenguaje.</li> </ul>
<p><b><i>Limitaciones:</i></b></p>	<p>Los datos de Google Books Ngram Viewer presentan varias limitaciones importantes:</p> <ul style="list-style-type: none"> <li>- La representatividad está restringida al subconjunto específico de libros digitalizados por Google Books, que no es una muestra aleatoria de toda la literatura publicada.</li> <li>- Existen sesgos inherentes hacia obras en inglés y publicaciones de grandes editoriales, lo que subrepresenta otros idiomas y obras de menor circulación o de editoriales más pequeñas.</li> <li>- El proceso de digitalización de Google Books no es aleatorio; puede haber sesgos en la selección de libros a digitalizar.</li> <li>- La digitalización de textos a través de Reconocimiento Óptico de Caracteres (OCR) puede introducir errores en los datos.</li> <li>- La frecuencia de uso en libros no es un indicador directo de la importancia, el impacto o la efectividad de una herramienta gerencial.</li> </ul>



	<ul style="list-style-type: none"> <li>- Ngram Viewer no proporciona información sobre el contexto en el que se utilizan los términos (por ejemplo, si se mencionan de forma positiva, negativa o neutral).</li> <li>- La evolución terminológica y los cambios en las convenciones de citación a lo largo del tiempo pueden afectar la consistencia longitudinal del análisis.</li> <li>- Sesgos Idiomáticos y Geográficos: Los resultados pueden sobrerrepresentar a ciertas poblaciones de autores.</li> </ul>
<b><i>Perfil inferido de Usuarios (o Audiencia Objetivo):</i></b>	<p>Refleja patrones de uso del lenguaje, tendencias académicas y de publicación, e intereses reflejados en la literatura y el conocimiento registrado en libros.</p> <p>Los usuarios típicos de Google Books Ngram Viewer son investigadores, historiadores, lingüistas y otros profesionales interesados en el análisis textual y la evolución del pensamiento a través del tiempo.</p>

***Origen o plataforma de los datos (enlace):***

— [https://books.google.com/ngrams/graph?content=Price+Optimization+%2B+Pricing+Optimization+%2B+Dynamic+Pricing+Models+%2B+Optimal+Pricing+%2B+Dynamic+Pricing&year\\_start=1950&year\\_end=2022&corpus=en&smoothing=0](https://books.google.com/ngrams/graph?content=Price+Optimization+%2B+Pricing+Optimization+%2B+Dynamic+Pricing+Models+%2B+Optimal+Pricing+%2B+Dynamic+Pricing&year_start=1950&year_end=2022&corpus=en&smoothing=0)