

UNIVERSIDAD ESTATAL A DISTANCIA
VICERRECTORÍA ACADÉMICA ESCUELA DE
CIENCIAS EXACTAS Y NATURALES CARRERA
INGENIERÍA INFORMÁTICA
CÁTEDRA INGENIERÍA DE SOFTWARE

ASIGNATURA

03084 BASES DE DATOS II

TAREA 2

VALOR: 15% (1.5)

ESTUDIANTE:

Francisco Campos Sandi

CÉDULA:

114750560

I CUATRIMESTRE 2025

Tabla de contenido

GLOSARIO	4
INTRODUCCIÓN	5
DESARROLLO.....	6
SOFTWARES PROPIETARIOS.....	6
SOFTWARE DE FUENTE ABIERTA.....	9
COSTOS ESTIMADOS DE LOS SISTEMAS DE MINERÍA DE TEXTO ABIERTOS Y PROPIETARIOS	12
CONCLUSIONES.....	13
REFERENCIAS.....	14
ANEXO.....	15
MAPA CONCEPTUAL SOBRE MINERÍA DE TEXTO.....	14

Índice de ilustraciones

Ilustración 1: Clear Forest	6
Ilustración 2 Costos estimados de los sistemas de minería	12
Ilustración 3 Mapa Conceptual de la Minería de Texto	15

Índice de tablas

Tabla 1: Software abierto vs Software Propietario.....	11
Tabla 2 Comparación entre software abierto y software propietario	16

Índice de anexos

ANEXO 1: MAPA CONCEPTUAL SOBRE MINERÍA DE TEXTO	15
ANEXO 2: COMPARACIÓN ENTRE SOFTWARE ABIERTO Y SOFTWARE PROPIETARIO	16

Glosario

- **Aika:** Es una herramienta de minería de texto que es de código abierto, lo que significa que se puede modificar y usar sin pagar.
- **IBM Intelligent Miner:** Es una herramienta de minería de texto de IBM que funciona con licencia de pago.
- **IBM SPSS Predictive Analytics Suite:** Otra herramienta de IBM para minería de texto, pero enfocada en análisis predictivo.
- **LingPipe:** Otra herramienta de código abierto para trabajar con minería de texto y procesamiento de lenguaje natural.
- **Minería de texto:** Es el proceso de analizar grandes cantidades de texto para extraer información útil y relevante. Se usa mucho en la investigación y el análisis de datos.
- **NLP (Procesamiento de Lenguaje Natural):** Es una rama de la inteligencia artificial que permite a las computadoras entender y generar lenguaje humano. Se usa en chatbots, traducción automática, etc.
- **Redes neuronales:** Son modelos matemáticos inspirados en el cerebro humano y se usan en inteligencia artificial para reconocer patrones, hacer predicciones y aprender de los datos.
- **Software de fuente abierta:** Son programas cuyo código es público, lo que permite que cualquiera lo modifique o lo distribuya libremente. Ejemplo: Linux o LibreOffice.
- **Software propietario:** Son programas que tienen restricciones de uso y requieren una licencia para poder usarlos. Ejemplo: Windows o Microsoft Office.

Introducción

El tema principal de la tarea se centra en el análisis y comparación de herramientas de minería de texto, tanto de fuente abierta como propietarias. Este estudio incluye un enfoque detallado sobre las características, beneficios y aplicaciones de cada herramienta en diferentes contextos empresariales y académicos. La selección de estos sistemas obedece a su relevancia en el campo del análisis de datos no estructurados, destacando su impacto en la toma de decisiones estratégicas y en la generación de nuevos conocimientos.

La tarea se desarrolla con el propósito de ofrecer un panorama integral sobre las soluciones de minería de texto disponibles, permitiendo a las empresas y organizaciones costarricenses elegir la alternativa más adecuada según sus necesidades. Además, busca fomentar el uso de tecnologías innovadoras en el análisis de datos textuales, contribuyendo a la optimización de procesos y el descubrimiento de patrones significativos.

El desarrollo de la tarea está organizado en etapas específicas que incluyen la identificación de herramientas representativas, la evaluación de sus características y beneficios, así como el análisis de casos de empresas que ya emplean estas soluciones. A través de una comparación estructurada, se pretende brindar una visión clara y fundamentada que oriente a los tomadores de decisiones en su selección.

El método empleado en el desarrollo de la tarea se fundamenta en la investigación documental y el análisis comparativo. Se recopilan datos provenientes de fuentes confiables, como publicaciones académicas y documentos empresariales, para garantizar la validez de los hallazgos. Sin embargo, el proceso puede enfrentar limitaciones como la disponibilidad de información actualizada sobre las herramientas seleccionadas, la variabilidad de costos según los proveedores y la falta de acceso a implementaciones prácticas en un entorno costarricense. Estas restricciones se consideran al presentar las recomendaciones finales.

Desarrollo

Softwares propietarios

1. Clear Forest

Para qué se utiliza: Clear Forest es una herramienta diseñada para el análisis y la visualización de colecciones de documentos. Su principal objetivo es extraer información significativa de grandes volúmenes de texto no estructurado, facilitando la identificación de patrones y relaciones ocultas en los datos. Esto la convierte en una solución ideal para empresas que manejan grandes bases de datos textuales.

Ilustración 1: Clear Forest



Fuente: Tomada de ComparaSoftware, (2022)

Beneficios: Entre sus ventajas destacan la capacidad de procesar grandes cantidades de datos textuales de manera eficiente, la generación de visualizaciones claras y comprensibles, y la posibilidad de integrar sus resultados en estrategias empresariales. Estas características permiten a las empresas tomar decisiones informadas basadas en datos.

Empresa que la utiliza: Clear Forest ha sido utilizada por empresas como Procter & Gamble para analizar grandes volúmenes de datos textuales y optimizar sus estrategias de mercado.

Recomendaciones para empresas costarricenses:

1. Identificar áreas clave donde el análisis de texto pueda aportar valor, como en la atención al cliente o el análisis de tendencias del mercado.
2. Capacitar al personal en el uso de herramientas de minería de texto para maximizar su potencial y garantizar una implementación efectiva.

2. IBM Intelligent Miner Data Mining Suite

Para qué se utiliza: Esta suite, integrada en InfoSphere Warehouse, incluye herramientas avanzadas de minería de datos y textos. Es ideal para descubrir patrones complejos y relaciones en grandes bases de datos empresariales, facilitando la toma de decisiones estratégicas.

El documento "Minería de datos con Intelligent Miner" de IBM (2022) describe cómo esta herramienta permite a las empresas analizar grandes volúmenes de datos para identificar patrones y obtener información valiosa. Intelligent Miner utiliza técnicas avanzadas de minería de datos y modelos predictivos para mejorar la toma de decisiones empresariales. Además, integra funciones como la creación, prueba y aplicación de modelos de minería, y ofrece capacidades para consultas complejas que van más allá de los métodos tradicionales. También destaca su capacidad para predecir comportamientos y optimizar estrategias comerciales mediante el análisis de datos históricos y actuales.

Beneficios:

Ofrece algoritmos avanzados para el análisis de datos, permite la integración con bases de datos existentes y mejora la calidad de las decisiones empresariales al proporcionar insights profundos y precisos.

Empresa que la utiliza:

Bancos internacionales han empleado esta herramienta para analizar perfiles de clientes y optimizar campañas de marketing dirigidas.

Recomendaciones para empresas costarricenses:

1. Evaluar la compatibilidad de la herramienta con las bases de datos existentes para garantizar una integración fluida.
2. Utilizar los insights generados para personalizar productos y servicios, mejorando la experiencia del cliente.

3. IBM SPSS Predictive Analytics Suite

Para qué se utiliza: Esta suite combina análisis predictivo y descriptivo, minería de datos y gestión de decisiones. Es ampliamente utilizada en sectores como la medicina, la gestión de capital humano y la prevención de delitos.

Beneficios: Permite analizar datos estructurados y no estructurados, optimizar decisiones operativas y predecir comportamientos futuros. Además, su interfaz visual facilita su uso incluso para usuarios sin conocimientos avanzados en estadística.

Empresa que la utiliza: Universidades como Stanford emplean esta herramienta para analizar datos académicos y mejorar la gestión educativa.

Recomendaciones para empresas costarricenses:

1. Implementar la herramienta en áreas clave como la gestión de recursos humanos para identificar patrones de desempeño y retención.
2. Aprovechar su capacidad predictiva para anticipar tendencias del mercado y ajustar estrategias empresariales

Software de fuente abierta

1. Aika

Aika es una biblioteca de código abierto que se especializa en el minado de patrones frecuentes en texto. Esta herramienta utiliza conceptos avanzados de redes neuronales y desarrollos gráficos para procesar datos textuales. Esto permite extraer información valiosa de conjuntos de datos complejos. Su enfoque avanzado facilita el análisis y la visualización de datos, brindando resultados efectivos "biblioteca open source para minado de patrones frecuentes en texto, utilizando conceptos de redes neuronales y desarrollos gráficos"(Joyanes Aguilar, 2019, p.273). Entre sus beneficios destacan la capacidad de manejar grandes volúmenes de datos, la precisión en los resultados gracias al uso de redes neuronales y la posibilidad de personalizarla debido a su naturaleza de código abierto.

Algunas instituciones de investigación han adoptado Aika para desarrollar modelos predictivos y analizar patrones en textos masivos.

Recomendaciones:

1. Capacitar a los equipos en conceptos básicos de redes neuronales para optimizar el uso de Aika.
2. Aplicar la herramienta en análisis de mercado o proyectos académicos para maximizar su utilidad.

2. LingPipe

Esta herramienta consiste en una suite de bibliotecas escritas en Java, pensadas específicamente para tareas de análisis lingüístico del lenguaje humano. LingPipe se aplica comúnmente en tareas como detección de entidades, análisis de sentimientos y clasificación de texto. Esto la hace muy adecuada para sectores enfocados en el análisis del procesamiento de

lenguaje natural (NLP). Su flexibilidad y capacidad para manejar múltiples idiomas la destacan en este campo "es una suite de bibliotecas escritas en Java para el análisis lingüístico del lenguaje humano"(Joyanes Aguilar, 2019, p.273). Entre sus beneficios se encuentran su escalabilidad, lo que permite manejar grandes proyectos de texto, y su enfoque especializado, que puede adaptarse a las necesidades de diversos sectores.

Empresas tecnológicas han adoptado LingPipe para sistemas de análisis de sentimientos en redes sociales.

Recomendaciones:

1. Usar LingPipe en análisis de sentimientos de clientes en redes sociales para medir la percepción de marcas.
2. Integrar la herramienta en sistemas automáticos de clasificación de correos electrónicos y documentos internos.

3. S-EM (Spy-EM)

S-EM es un sistema de clasificación que aprende con ejemplos positivos, incluso en ausencia de etiquetas negativas. Este enfoque lo hace ideal para datos limitados o incompletos. Al utilizar algoritmos como Naïve Bayes y Expectación-Maximización (EM), S-EM permite identificar patrones de manera eficiente. Su capacidad para trabajar en escenarios con datos escasos proporciona soluciones valiosas en tareas de clasificación de texto "sistema de clasificación de texto que aprende con ejemplos positivos y sin etiquetas"(Joyanes Aguilar, 2019, p.273). Entre sus ventajas destaca la posibilidad de realizar análisis de texto con ejemplos mínimos, asegurando resultados confiables en proyectos con datos iniciales restringidos.

Ha sido implementado en instituciones académicas para analizar datos no estructurados y clasificar textos de manera más accesible.

Recomendaciones:

1. Utilizar S-EM en proyectos de análisis de datos internos para una mejor organización empresarial.
2. Implementarlo en análisis de datos de clientes para identificar comportamientos y patrones clave.

Tabla 1: Software abierto vs Software Propietario

Aspecto	Software Abierto	Software Propietario
Ejemplos de herramientas	Aika, LingPipe, S-EM	Clear Forest, IBM Intelligent Miner, IBM SPSS
Costo	Generalmente gratuito, pero puede requerir inversión en personalización y soporte técnico.	Suele ser costoso, con licencias periódicas y costos de mantenimiento.
Flexibilidad	Alta, permite modificaciones y personalización según las necesidades del usuario.	Limitada, depende de las capacidades definidas por el proveedor.
Facilidad de uso	Requiere conocimientos técnicos más avanzados para su implementación y personalización.	Generalmente más accesible, con interfaces gráficas diseñadas para usuarios sin experiencia técnica.
Comunidad y soporte	Amplia comunidad de desarrolladores que brinda soporte y actualizaciones.	Soporte técnico especializado proporcionado por el proveedor.
Propiedad y control	El usuario tiene control total sobre el código y los datos procesados.	La gestión del software y los datos procesados está regulada por las condiciones del proveedor.

Seguridad	Depende de la configuración realizada por el usuario, lo que puede ser un desafío para garantizar altos estándares.	Ofrecen altos estándares de seguridad integrados según la normativa del proveedor.
Adaptabilidad	Excelente para proyectos específicos que requieren soluciones a medida.	Más adecuada para usos generalizados o estándares dentro de una industria.
Ejemplos de usuarios	Instituciones académicas, proyectos de investigación, startups tecnológicas.	Grandes corporaciones, bancos, instituciones educativas internacionales.
Beneficios principales	Bajo costo inicial, flexibilidad en personalización, apropiado para innovación.	Fácil de implementar, con soporte especializado y amplia integración empresarial.

Fuente: Elaboración Propia (2025).

Costos estimados de los sistemas de minería de texto abiertos y propietarios

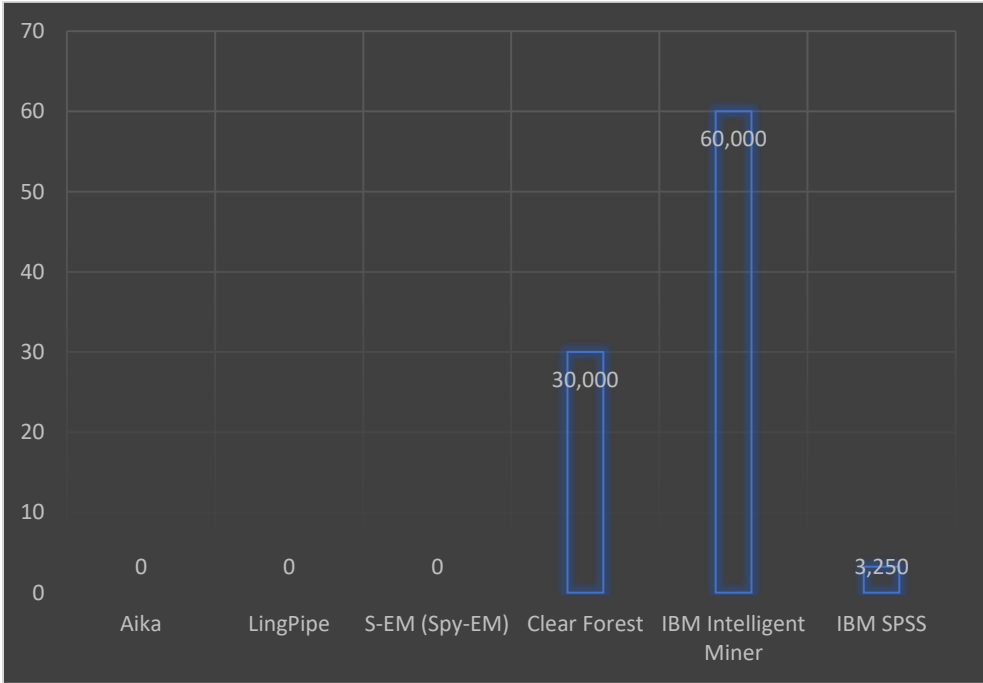


Ilustración 2 Costos estimados de los sistemas de minería

Fuente: Elaboración Propia (2025).

Conclusiones

En conclusión, el análisis realizado sobre herramientas de minería de texto, tanto de fuente abierta como propietarias, evidencia que ambas opciones presentan ventajas significativas dependiendo del contexto en que se utilicen. Las herramientas de código abierto ofrecen flexibilidad y menores costos, lo que las hace ideales para proyectos académicos o de investigación. Por otro lado, las herramientas propietarias, aunque implican una mayor inversión, destacan por su facilidad de uso y soporte técnico especializado, siendo más adecuadas para entornos empresariales estructurados.

Asimismo, se concluye que la selección de una herramienta debe estar orientada a las necesidades específicas de la empresa u organización. Factores como el presupuesto disponible, el nivel de personalización requerido y la experiencia del equipo en tecnologías de análisis deben guiar la decisión. Esta alineación asegura un uso eficiente de los recursos y maximiza el impacto de la implementación de las soluciones de minería de texto.

Finalmente, se reconoce que una de las principales limitaciones en el desarrollo de este tipo de proyectos es la disponibilidad de información actualizada y el acceso a implementaciones prácticas en el contexto local. Para superar este desafío, se sugiere fomentar la capacitación en tecnologías emergentes y establecer alianzas con expertos en minería de datos. Esto garantizará una adecuada transferencia de conocimiento y permitirá aprovechar al máximo las herramientas disponibles.

Referencias

Arriola Navarrete, O., Tecuatl Quechol, G., & González Herrera, G. (2011). Software propietario vs software libre: Una evaluación de sistemas integrales para la automatización de bibliotecas. Investigación bibliotecológica, 25(54), 37-70.

Capetillo, J. (2021, marzo 1). Software propietario y software libre: ¿Qué debemos saber? CEESA, S.A. <https://ceesa.com/software-propietario-y-software-libre/>

ComparaSoftware. (2022). Clear Software: Precios, funciones y opiniones [Imagen]. <https://d3v6byorcue2se.cloudfront.net/wp-content/uploads/2018/07/logoClear-Software.png>

IBM. (s.f.). Precios - IBM SPSS Statistics. IBM - United States. <https://www.ibm.com/es-es/products/spss-statistics/pricing>

Joyanes Aguilar, L. (2019). INTELIGENCIA DENEGOCIOSY ANALÍTICA DEDATOS Una visión global de Business Intelligence & Analytics. Alfaomega Grupo Editor Argentino.

Minería de datos con Intelligent Miner. (2022). IBM - United States. <https://www.ibm.com/docs/es/db2/11.1?topic=api-data-mining-intelligent-miner>

Software libre y Software privativo: Servicio de Promoción y Apoyo a la Investigación, Innovación y Transferencia: UPV. (s. f.). Recuperado 08 de marzo de 2025, de <https://www.upv.es/entidades/I2T/info/890613normalc.html>

Anexo

Anexo 1: Mapa conceptual sobre Minería de texto

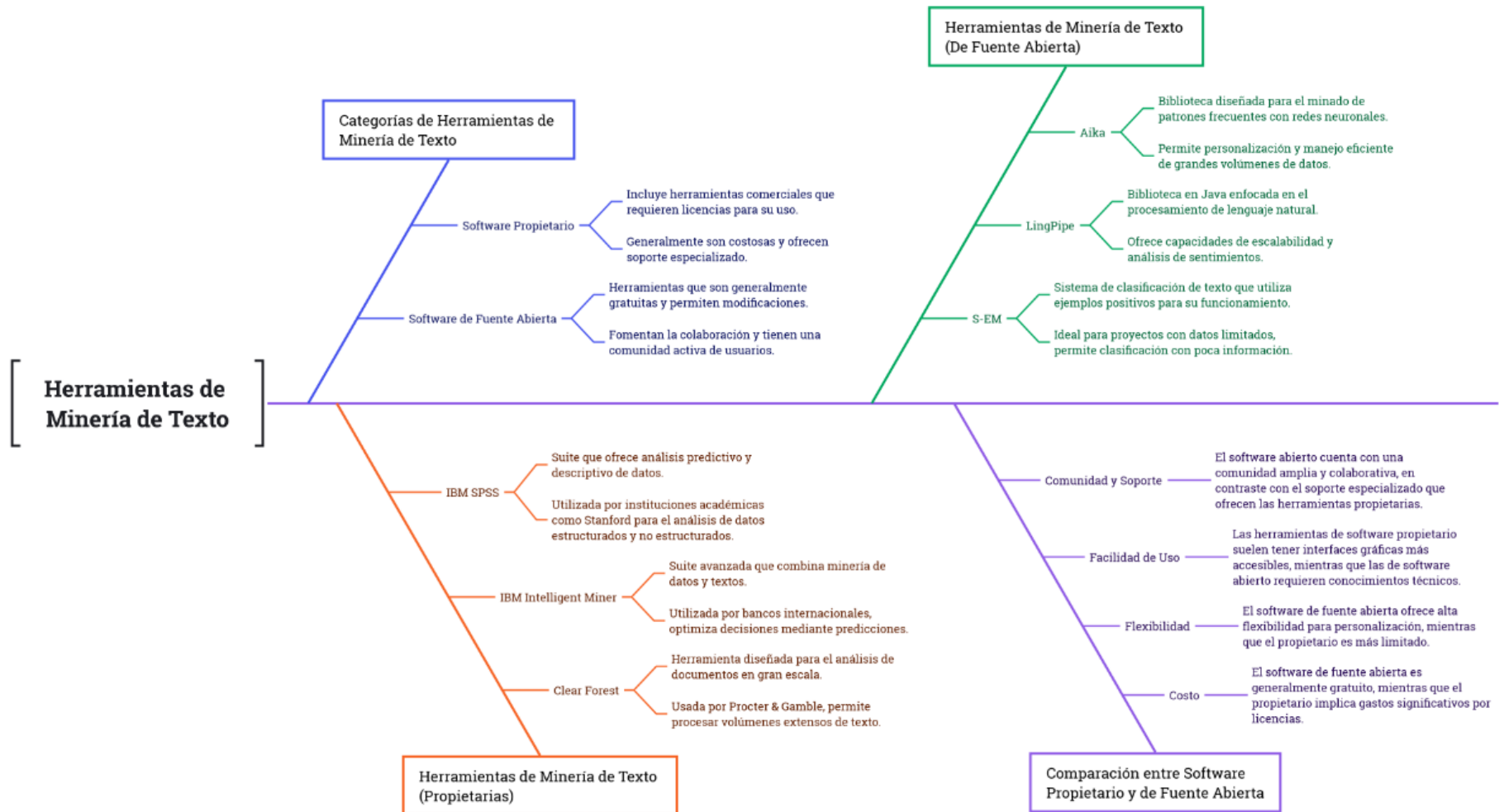


Ilustración 3 Mapa Conceptual de la Minería de Texto

Elaboración propia (2025).

Anexo 2: Comparación entre software abierto y software propietario

Tabla 2 Comparación entre software abierto y software propietario

Aspecto	Software abierto	Software propietario
Acceso al código fuente	Acceso completo al código fuente, permitiendo modificarlo según las necesidades de los colaboradores en comunidad.	Código cerrado o privado, no se permite el acceso ni la modificación del software.
Licencia	Licencias libres o de código abierto sin pagar por su uso (por ejemplo, Libre Office).	Licencias comerciales, generalmente con costos asociados.
Mantenimiento y actualizaciones	Dependiente de la comunidad o de los contribuyentes. Las actualizaciones pueden ser irregulares.	Mantenimiento y actualizaciones regulares garantizadas por el proveedor.
Personalización	Altamente personalizable, se puede modificar el software para adaptarlo a tus necesidades específicas.	Personalización limitada, muchas veces solo es posible dentro de los parámetros del software.

Compatibilidad	Suelen ser más flexible en diferentes plataformas y sistemas operativos.	Generalmente optimizado para plataformas específicas, con menos flexibilidad.
Ejemplos conocidos	Linux, Apache, WordPress, GIMP, LibreOffice.	Microsoft Windows, Adobe Photoshop, Microsoft Office, AutoCAD.
Soporte técnico	Soporte generalmente a través de foros y comunidades en línea.	Soporte profesional y técnico directo del proveedor o empresa.
Escalabilidad	Puede requerir esfuerzo adicional para escalar, dependiendo de la implementación.	Generalmente diseñado para escalar fácilmente según las necesidades empresariales.
Requisitos de hardware	En ocasiones necesita menos recursos, adaptándose a configuraciones más diversas.	Puede ser más exigente en términos de hardware y recursos del sistema.
Seguridad	Depende de la comunidad y los desarrolladores, puede ser vulnerable si no se actualiza adecuadamente.	Seguridad más robusta debido a auditorías profesionales y pruebas exhaustivas.

Elaboración propia (2025).