

Búsqueda con texto libre en AXIS Camera Station Pro

Enero 2025

Resumen

AXIS Camera Station Pro incluye varias herramientas de búsqueda forense de vídeo preinstaladas. Entre ellas se incluyen la depuración de líneas temporales, la búsqueda de datos y la búsqueda inteligente con objetos preclasificados y búsqueda de texto libre.

La búsqueda de texto libre le permite buscar cualquier objeto en movimiento describiéndolo con sus propias palabras. La libertad para crear filtros de búsqueda detallados con una gran variedad de atributos descriptivos permite encontrar más rápidamente las secuencias pertinentes.

La función de búsqueda de texto libre se basa en la correspondencia texto-imagen proporcionada por un modelo de base de código abierto preentrenado que ha sido optimizado por Axis para entornos de vigilancia. La búsqueda puede aplicarse a una sola cámara o a varias al mismo tiempo.

Se compara una representación numérica de su consulta de texto libre con representaciones numéricas de imágenes de objetos en movimiento detectados. Los resultados se muestran como miniaturas, incluidos el nombre de la cámara, la hora y la fecha, ordenados por relevancia para su consulta de búsqueda.

Con la búsqueda de texto libre, utilizamos la IA para aumentar la precisión y eficacia de nuestras soluciones de búsqueda forense y, en última instancia, mejorar el proceso de toma de decisiones a cargo de personas. Con el fin de cumplir las normas legales y éticas, la función de búsqueda incluye una función de moderación independiente desarrollada por Axis que restringe el uso de palabras ofensivas en las consultas de búsqueda. Todas las búsquedas quedan registradas y son visibles para los administradores, lo que permite hacer un seguimiento y adoptar acciones correctivas en caso de uso indebido.

Índice

1	Introducción	4
2	Segundo plano: búsqueda forense en AXIS Camera Station Pro	4
3	¿Cómo funciona la búsqueda de texto libre?	4
4	Construcción de consultas de búsqueda	5
	4.1 Directrices para prompts	6
	4.2 Moderación de consultas	6
	4.3 Registro de consultas de búsqueda	6
5	Correspondencia entre texto e imagen	6
	5.1 Modelo de base optimizado para la vigilancia	7
	5.2 Vectores de características	7
6	Información general del proceso	7
7	Uso responsable de la IA	9

1 Introducción

La herramienta de búsqueda de texto libre de AXIS Camera Station Pro le permite buscar grabaciones de vídeo utilizando sus propias palabras en lugar de filtros predefinidos.

En este documento técnico se explica cómo funciona el método de búsqueda y se ofrecen algunas directrices sobre cómo utilizarlo. También se describe la función de moderación y el registro de consultas que existen para garantizar el cumplimiento de las normas legales y éticas.

2 Segundo plano: búsqueda forense en AXIS Camera Station Pro

AXIS Camera Station Pro incorpora varias herramientas de búsqueda forense de vídeo preinstaladas, que incluyen depuración de líneas de tiempo, búsqueda de datos, así como búsqueda inteligente con objetos preclasificados y búsqueda de texto libre.

La función de búsqueda inteligente utiliza metadatos de escena generados por el dispositivo Axis. Los metadatos incluyen el tipo de objeto (persona, tipo de vehículo u objeto desconocido) para los objetos en movimiento, junto con atributos como el color de la ropa y del vehículo, las matrículas, la velocidad, la ubicación y la fecha y hora.

En el caso de los dispositivos con capacidades analíticas limitadas, la función de búsqueda se basa en la detección de movimiento en el dispositivo combinada con la clasificación de objetos realizada en el servidor AXIS Camera Station Pro. La búsqueda forense en AXIS Camera Station Pro es, por tanto, una solución híbrida en la que las capacidades de los dispositivos periféricos se utilizan en la medida de lo posible, pero se complementan con datos del servidor cuando es necesario.

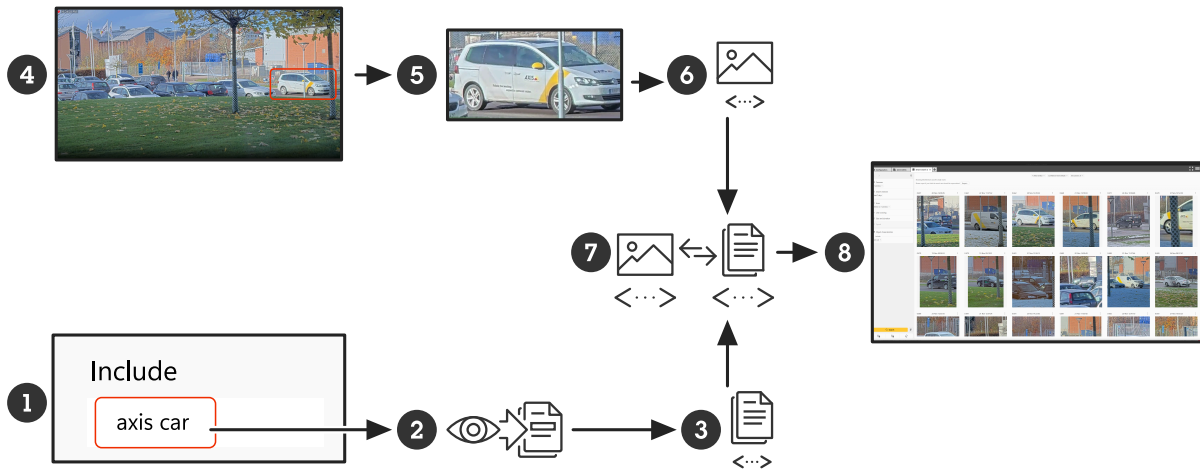
Tradicionalmente, las búsquedas con metadatos de escenas tenían que realizarse con filtros de búsqueda predefinidos. Con ellos, se eligen descriptores de objetos fijos de una lista, incluido el tipo de objeto (como «vehículo»), el tipo de vehículo (si procede, como «coche»), el color (como «azul»), etc. En cambio, el nuevo método de búsqueda de texto libre le permite crear su propio filtro de búsqueda.

Aunque la búsqueda preclasificada ofrece resultados de gran precisión, no puede detectar nuevos tipos de objetos que no estén predefinidos. Para hacer frente a esta limitación, la búsqueda de texto libre le ofrece la libertad y flexibilidad de buscar utilizando sus propias palabras. Puede describir con mayor detalle cualquier objeto en movimiento con lenguaje natural y asociaciones para acotar su búsqueda y obtener más resultados.

3 ¿Cómo funciona la búsqueda de texto libre?

Se compara una representación numérica de su consulta de texto libre con representaciones numéricas de imágenes de objetos en movimiento detectados. El resultado de esta correspondencia texto-imagen se

presenta y ordena por la mejor correspondencia. Los resultados se muestran como miniaturas, incluidos el nombre de la cámara, la hora y la fecha, ordenados por relevancia para su consulta de búsqueda.



Información general simplificada del proceso de búsqueda de texto libre. Tenga en cuenta que los pasos del 4 al 6 tienen lugar continuamente, incluso cuando no esté buscando, para crear vectores de características de todos los objetos en movimiento detectados.

- 1 *Escriba su consulta de búsqueda de texto libre.*
- 2 *Un módulo de moderación impide el uso de palabras ofensivas y poco éticas.*
- 3 *El modelo de base crea una representación numérica (un vector de características) de la consulta de búsqueda.*
- 4 *Una cámara detecta el movimiento en una escena.*
- 5 *La cámara selecciona una imagen recortada para representar el objeto en movimiento.*
- 6 *El modelo de base crea un vector de características del objeto tras analizar su forma, patrones, color, etc.*
- 7 *Se comparan los dos vectores de características.*
- 8 *El resultado de la comparación se ordena según la mejor coincidencia y se presenta como miniaturas.*

La búsqueda de texto libre puede aplicarse a una sola cámara o a varias al mismo tiempo.

Para limitar el alcance de su búsqueda de texto libre, puede combinarla con otras funciones de búsqueda inteligente, como la búsqueda por similitud o la búsqueda por tiempo, utilizando un tipo de búsqueda tras otro.

4 Construcción de consultas de búsqueda

Puede buscar cualquier objeto en movimiento y cualquier tipo de vehículo. Siga las directrices para obtener resultados óptimos.

Tenga en cuenta que solo debe buscar objetos en movimiento. La búsqueda de objetos inmóviles no funcionará en la mayoría de los casos.

Las frases de búsqueda se moderan y registran para evitar comportamientos de búsqueda poco éticos.

4.1 Directrices para prompts

- Describe las situaciones como usted describiría una imagen. El modelo se alimenta de imágenes fijas, por lo que la búsqueda de acciones (como caerse, correr o robar) puede resultar difícil, ya que requeriría más contexto.
- Al igual que otros modelos multimodales, el modelo de búsqueda de texto libre funciona bien con descriptores como objetos y colores, pero es menos adecuado para recuentos («tres personas»), jerga o señales emocionales («hombre con aspecto enfadado»). La descripción del objeto no debe ser subjetiva, demasiado imprecisa ni incluir detalles demasiado específicos.
- Combine varios atributos de objeto utilizando y: «persona con gorro rojo y mochila».
- Describa texto, logotipos de texto o nombres de marca: «furgoneta con el texto Axis».
- No se centre en describir entornos. El procesamiento se realiza en imágenes recortadas de objetos, lo que significa que el modelo podría no ver lo que rodea a los objetos. Por lo tanto, es posible que los descriptores de escenas o entornos amplios (como «ciudad», «urbano», «parque», «jardín», «lago» o «playa») no den buenos resultados.
- Experimente con otras formulaciones si no está satisfecho con el resultado.
- La consulta de texto libre solo admite el inglés.

4.2 Moderación de consultas

Hemos implementado una moderación de consultas basada en prácticas habituales con el fin de garantizar un filtrado eficaz. El modelo de moderación es un modelo de procesamiento del lenguaje natural que comprueba la consulta para restringir el texto ofensivo. Comprueba cadenas de texto completas en busca de contenido dañino, inapropiado u ofensivo. Además, hemos mejorado estas capacidades con medidas propias, que incluyen una lista personalizada de categorías y palabras de búsqueda prohibidas. Cuando una consulta contiene palabras o frases de esta lista, rechazamos la búsqueda para mantener un entorno de búsqueda seguro y garantizar resultados éticos. Si no está de acuerdo con que se bloquee una determinada palabra o desea sugerir el bloqueo de una palabra, puede enviar comentarios de usuario anónimos a Axis.

4.3 Registro de consultas de búsqueda

AXIS Camera Station Pro mantiene un registro de auditoría de las operaciones del usuario. Los registros de auditoría no solo llevan un seguimiento de las operaciones específicas y la identidad del usuario, sino que también conservan cualquier dato utilizado en la operación. Esto significa que todas las búsquedas de los usuarios, incluidas las solicitudes de búsqueda, quedan registradas. Los administradores pueden utilizar los registros para identificar comportamientos de búsqueda inadecuados entre los usuarios, señalar los mensajes de búsqueda poco éticos y adoptar medidas correctivas.

Tenga en cuenta que los datos de vídeo no se comparten con Axis. Sus datos permanecen en su servidor.

5 Correspondencia entre texto e imagen

La posibilidad de buscar metadatos de vídeo mediante consultas de texto libre amplía considerablemente las posibilidades de búsqueda, pasando de una lista predefinida de atributos a criterios de búsqueda casi ilimitados. En AXIS Camera Station Pro, esta función se basa en un modelo de base de código abierto,

entrenado en miles de millones de pares imagen-texto y perfeccionado por Axis para entornos de vigilancia con el fin de mejorar el rendimiento.

5.1 Modelo de base optimizado para la vigilancia

El modelo de base es un modelo de texto-imagen entrenado con grandes conjuntos de datos de pares texto-imagen. Es un modelo «zero-shot» que empareja texto con imágenes relevantes. Un modelo «zero-shot» es un tipo de modelo de inteligencia artificial (IA) capaz de reconocer y clasificar objetos o conceptos sin datos de entrenamiento previos. En otras palabras, el modelo puede realizar tareas sin haber visto antes ningún ejemplo de la tarea. Esta capacidad es crucial para garantizar un rendimiento óptimo a la hora de emparejar lenguaje natural con imágenes.

El modelo se entrenó con una gran cantidad de combinaciones de texto e imagen y funciona con una red neuronal de más de 2500 millones de parámetros. En Axis, hemos utilizado nuestro propio y exclusivo material de formación para perfeccionar aún más este modelo, mejorando su capacidad para interpretar imágenes con vistas de cámaras de vigilancia y objetos convencionales. Esto significa que hemos optimizado el modelo para entornos de vigilancia.

5.2 Vectores de características

Cuando se realiza una búsqueda de texto libre, el modelo de base crea un vector de características de la consulta de búsqueda.

El modelo de base también produce continuamente descripciones de cada objeto rastreado en las escenas y crea vectores de características para representarlos. Cada objeto está representado por un único vector de características, que se almacena en nuestra base de datos. Esto agiliza la búsqueda, puesto que los vectores de características ya están precalculados y disponibles en la base de datos.

Ambos tipos de vectores de características se introducen en el motor de comparación de vectores para determinar la distancia de similitud entre su consulta de búsqueda y todos los vectores de características disponibles que representan objetos detectados.

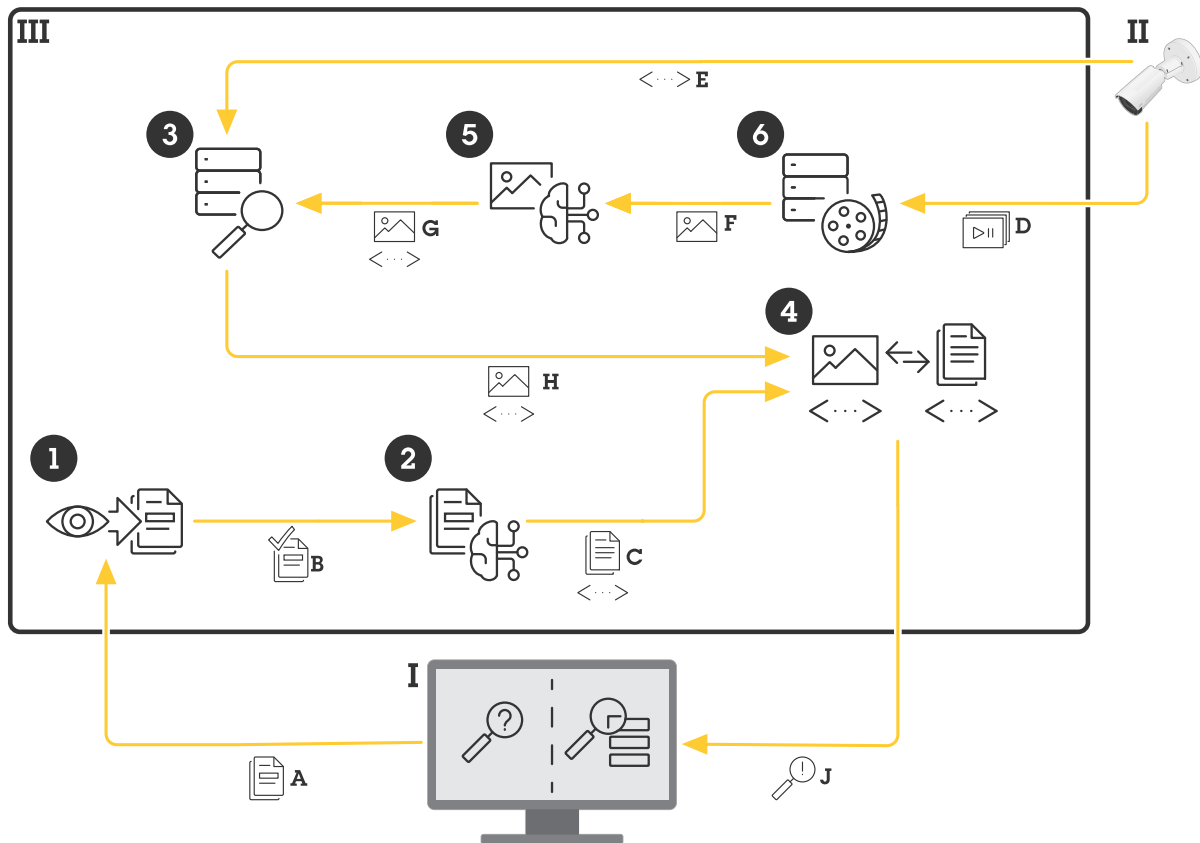
Un vector de características es una representación numérica de texto o imágenes. Por tanto, los vectores de características de personas u objetos son solo representaciones abstractas del aspecto de la persona o el objeto. Los vectores de características no contienen ninguna información interpretable por el ser humano sobre rasgos (como el color del pelo o de la ropa) que pueda asignarse de forma explícita a una persona concreta o utilizarse para su identificación. Los vectores de características solo pueden utilizarse para comparaciones con otros vectores de características.

6 Información general del proceso

El diagrama de información general del proceso muestra los principales pasos del proceso, incluidos los lugares donde estos se producen y el tipo de datos que genera cada paso.

Obsérvese que el bucle superior del diagrama, que incluye la cámara (1), el almacenamiento de grabaciones (6), el modelo de base (5) y la base de datos de búsqueda (3) es un proceso que tiene lugar de forma

constante para crear vectores de características de todos los objetos en movimiento detectados, y no solo cuando se realiza una búsqueda.



Principales localizaciones (I-III) para el proceso de búsqueda de texto libre

- I Cliente de AXIS Camera Station Pro
- II Cámara(s)
- III Servidor de AXIS Camera Station Pro

Principales etapas del proceso (1-6)

- 1 Moderación de las consultas de búsqueda
- 2 Modelo de base (texto)
- 3 Buscar en la base de datos
- 4 Comparación de vectores
- 5 Modelo de base (imagen)
- 6 Almacenamiento de grabaciones

Tipo de datos o resultado (A-J)

- A Cadena de texto
- B Cadena de texto
- C Vector de características (texto)
- D Vídeo
- E Metadatos
- V Imágenes
- G Vectores de características (imagen)
- H Vectores de características (imagen)
- J Resultados de la búsqueda

(I) **Cliente de AXIS Camera Station Pro:** Aquí puede escribir su consulta de búsqueda y obtener resultados ordenados.

(II) **Cámara(s):** La búsqueda de texto libre funciona en cámaras Axis con AXIS OS 5.51 o posterior, pero cuanto mejor sea la cámara, mejores resultados obtendrá. Los dispositivos más antiguos proporcionan metadatos menos granulares basados únicamente en la detección de movimiento. Los dispositivos más nuevos producen AXIS Scene Metadata, que incluye la clasificación de objetos. La detección y el seguimiento de objetos en movimiento de la cámara sirven para encontrar una imagen representativa de cada objeto detectado, reduciendo así el número de imágenes que se van a analizar en el servidor.

(III) **Servidor de AXIS Camera Station Pro:** Aquí se procesan y almacenan todos los metadatos y datos de vídeo de las cámaras. Antes de realizar una búsqueda de texto libre, el servidor debe (para cada objeto en movimiento detectado) descodificar el vídeo y extraer una imagen del objeto detectado. A continuación, el modelo de base procesa esta imagen para crear el vector de características. Estas operaciones son bastante costosas en términos de capacidad de procesamiento, por lo que, para mejorar el rendimiento, los vectores de características se guardan en una base de datos que permite realizar búsquedas rápidas en el futuro. Si su servidor tiene capacidad de sobra, le recomendamos encarecidamente que habilite el procesamiento en segundo plano de los vídeos de sus cámaras más importantes para que las búsquedas sean mucho más rápidas.

(1) **Moderación de las consultas de búsqueda:** El modelo de moderación comprueba la consulta para restringir el contenido ofensivo.

(2) **Modelo de base (texto):** El modelo de base crea una representación numérica (vector de características) de la cadena de texto de la consulta de búsqueda moderada.

(3) **Buscar en la base de datos:** La base de datos de búsqueda contiene metadatos completos de AXIS Scene Metadata o metadatos creados por el servidor, incluidos datos de clasificación de objetos con atributos, tiempo, posición y vectores de características.

(4) **Comparación de vectores:** La representación del vector de características de la cadena de texto de la consulta de búsqueda se compara con las representaciones del vector de características de las imágenes de objetos detectadas en vídeo.

(5) **Modelo de base (imagen):** El modelo de base crea representaciones numéricas (vectores de características) de cada pista de objeto en el vídeo grabado. Se trata de un proceso continuo que tiene lugar también cuando no se está buscando.

(6) **Almacenamiento de grabaciones:** Aquí es donde se almacena el vídeo de la cámara y donde el modelo de base obtiene sus imágenes.

7 Uso responsable de la IA

Con la búsqueda de texto libre, utilizamos la IA para aumentar la precisión y eficacia de nuestras soluciones de búsqueda forense y, en última instancia, mejorar el proceso de toma de decisiones a cargo de personas.

La responsabilidad y la rendición de cuentas son pilares esenciales del enfoque de Axis con respecto a la IA. Y su aplicación pasa por asegurarnos de que los sistemas de IA que creamos se ajusten a principios éticos, cumplan la legislación y permitan gestionar los riesgos. Axis proporciona herramientas que permiten a nuestros clientes estar seguros acerca de la integridad de sus operaciones. La función de búsqueda de texto libre de AXIS Camera Station Pro incluye un modelo de clasificación de texto perfeccionado para la moderación de «prompts» o instrucciones de texto. Hemos desarrollado este modelo para moderar las consultas de búsqueda con el fin de ayudarle a evitar un uso poco ético.

La búsqueda de texto libre se conecta a los servicios en la nube de Axis una vez a la semana para comprobar si es necesario actualizar los modelos de IA con vistas a cumplir los nuevos reglamentos o requisitos. Si falla la conexión, las operaciones de búsqueda de texto libre no estarán disponibles hasta que se restablezca.

Para seguir cumpliendo las normas legales y éticas en la aplicación de la IA, nuestros productos proporcionan controles de acceso basados en credenciales de autenticación de usuarios y permisos de acceso. De esta forma, nuestros clientes pueden imponer a los usuarios el cumplimiento de las políticas de funcionamiento.

Acerca de Axis Communications

Axis abre la puerta a un mundo más seguro e inteligente creando soluciones para mejorar la seguridad y los resultados comerciales. Desde su condición de líder en tecnologías de red, Axis ofrece soluciones de videovigilancia, control de acceso, intercomunicación y audio, complementadas por aplicaciones de analítica inteligente y asociadas a una formación de altísima calidad.

Axis cuenta con unos 4.000 empleados propios en más de 50 países y colabora con socios tecnológicos e integradores de sistemas de todo el mundo para hacer llegar sus soluciones a los clientes. Axis fue fundada en 1984 y su sede central se encuentra en Lund (Suecia).