DCO

dco.uab.cat

Data Compression Observatory (DCO)

Qué es

El DCO es una red de investigadores, grupos de investigación y otras entidades relacionadas con la compresión de datos.

Formar parte del DCO no cuesta nada, no exige exclusividad, no obliga a ninguna dedicación mínima y no compromete a la institución de cada miembro.

Quién puede adherirse

La adhesión está abierta a cualquiera. Quien se adhiere comparte el propósito de la red, trata con discreción el material que dentro de ella se comparte con carácter no público, y no actúa ni habla en nombre del DCO sin el consenso de sus miembros.

Adhesión en castellano · English adherence

Las hojas de adhesión firmadas pueden enviarse a miguel.hernandez@uab.cat.

Pilares

El DCO se propone abordar cuatro problemas, y otros tantos espacios que cubrir.

1. Seguimiento curado del estado del arte

El estado del arte en compresión de datos (lo publicado y lo que se usa en la práctica) aparece repartido entre revistas de perfiles distintos y repositorios, privados o públicos, de instituciones y agencias. Los criterios sobre reproducibilidad o sobre el uso de conjuntos de datos varían de un foro a otro, de modo que localizar lo verdaderamente relevante en un área concreta, y saber qué parte de ello puede comprobarse, exige un esfuerzo considerable.

El DCO aspira a un canal curado y accesible, organizado por categorías (por ejemplo, compresión de imagen de un tipo determinado), que reúna las contribuciones relevantes, verificadas y con sus experimentos disponibles, y al que cualquiera pueda suscribirse. Y a una revisión por pares cualificada que respalde esa selección, profesionalizada y dignificada como el trabajo que es.

2. Evitar la repetición de resultados negativos

Recorrer un camino que otros ya han recorrido sin éxito es una pérdida de esfuerzo evitable. Para que la investigación sea eficiente hace falta saber qué no ha funcionado.

El DCO aspira a un repositorio curado de resultados negativos, con un valor a la par con el de los positivos: los caminos descartados son un elemento más del espacio de búsqueda que recorre la investigación.

3. Estándares

Este pilar complementa al primero. En cada ecosistema conviven estándares *de iure* y *de facto* que compiten o cooperan entre sí, y buena parte de ellos se publica fuera del circuito académico, de modo que queda al margen de esa observación. Son además, en muchos casos, difíciles de entender y de implementar, y quienes deciden sobre compresión no siempre son expertos en compresión: diseñan una etapa de compresión dentro de un sistema mayor, en cualquier ámbito.

El DCO aspira a incorporarlos al seguimiento, por categorías de aplicación, con comparativas homogéneas de rendimiento, con las implementaciones disponibles de cada uno y con la documentación complementaria que ayude a entenderlos. Y a promover su uso, por ejemplo los estándares del CCSDS, y a divulgación y herramientas que permitan decidir acertadamente sin ser experto. Todo ello desde la visión más imparcial posible, también cuando los miembros de la red participan en el desarrollo de estándares, en su vertiente académica o aplicada.

4. Marco común de evaluación

Este pilar es transversal a los tres anteriores. Sin datos preparados en común y sin protocolos acordados, los resultados no son comparables ni resultan convincentes fuera de quien los publica.

El DCO aspira a un repositorio dedicado a la verificación, la evaluación y la comparación de técnicas de compresión en todas las categorías cubiertas, con conjuntos de datos localizados, curados y preparados, publicados y mantenidos disponibles a largo plazo (esa preparación es lo que garantiza que los resultados sean relevantes), y con mecanismos consensuados que permitan sostener que los *benchmarks* aplicados son verdaderamente útiles, tanto para la academia como para la industria. Y a guías para la creación y la distribución de los resultados de los proyectos de investigación (experimentos reproducibles, datos intermedios, código, configuraciones) y, más adelante, un repositorio abierto de contenedores.

Origen

La iniciativa parte de Miguel Hernández-Cabronero, del DEIC (Departament d'Enginyeria de la Informació i de les Comunicacions) de la Universitat Autònoma de Barcelona.