

# Patrones para datos de investigación en la web a través de diferentes comunidades de investigación

Patterns for searching data on the web across different research communities

Timo Borst, Fidan Limani

*Liber quarterly*, ISSN 2213-056X, Vol. 30, n. 1, 2020

Los datos de investigación basados en un entorno web han ganado atención recientemente. Con los investigadores y las instituciones académicas publicando cada vez más sus datos en webs públicas, los flujos de trabajo tradicionales de investigación respecto a los datos de búsqueda son objeto de análisis empíricos, estudios de usuarios, y desarrollo de servicios. En el artículo se investigan estos flujos de trabajo más en detalle y se introducen tres patrones de búsqueda de datos basados en la web con la intención de servir como referencia general y como punto de partida para disciplinas específicas. Se dan varios ejemplos reales y se sugiere una combinación de estrategias genéricas y específicas. También se analizan los patrones con evidencias empíricas. Importantes proveedores de datos o agregadores de metadatos como Zenodo, Elsevier Data Search, o Google Dataset Search realizan el descubrimiento y la búsqueda de temas igual que se hace en la búsqueda documental: los términos de búsqueda se emparejan con metadatos, tales como título o descripción de un conjunto de datos. Permitir a los usuarios la búsqueda a través de disciplinas es un objetivo básico de los agregadores y proveedores de servicios de búsqueda. Una estrategia posible es referirse a clasificaciones de materias

para categorizar los conjuntos de datos a disciplinas científicas que son sugeridas y mantenidas por las comunidades. Otra estrategia podría fomentar el descubrimiento de conjuntos de datos de acuerdo con sus propiedades espaciales y temporales. Una última estrategia, la búsqueda semántica, permite descripciones estructurales de los términos que las comunidades usan para etiquetar sus conjuntos de datos. El objetivo es ir más allá de una búsqueda por palabra clave y tratar casos como sinónimos y términos más amplios, más concretos, relacionados, asociados o relevantes. A pesar de que algunos centros de datos y repositorios ya proporcionan una versión web de sus datos que incluyen interfaces para la exploración de contenido, todavía no se trata de una práctica básica de búsqueda, ya que requiere una infraestructura específica y software en términos de acceso, escalabilidad y desempeño. El proyecto GerDI proporciona una infraestructura de investigación que se centra en datos muy utilizados por las comunidades y disciplinas de investigación, y que conllevan diferentes prácticas y requisitos de gestión de datos de investigación. Este modelo es especialmente importante para los dominios científicos que trabajan con recursos tales como textos o datos que se representan habitualmente con descripciones textuales. Como parte de este patrón de búsqueda, también se exploran aplicaciones potenciales de búsqueda semántica. Se buscan unos servicios más orientados hacia la semántica y se aboga por una normalización de los metadatos. Pese a sus requisitos de búsqueda para descubrimiento, exploración y análisis, las comunidades GerDI no siempre proporcionan suficientes metadatos para que sea posible. Basándose en los metadatos disponibles para la recolección, el único servicio que se puede ofrecer a todas las comunidades es el de descubrimiento. Como resultado, debido a la falta de metadatos y servicios de apoyo, los modelos de búsqueda para exploración y análisis todavía no están representados entre los servicios de GerDI, que deberá incluirlos en futuras actualizaciones.

<https://www.liberquarterly.eu/articles/10.18352/lq.10317/>

Resumen elaborado por Antonio Rodríguez Vela