

<http://physicsweb.org/article/news/4/5/6>

2000/05/12

داده، همه جا داده

بر اساس مقاله‌ی یک گروه از پژوهش‌گران ایالات متحده در نیچر [1] این هفته، مقدار زیادی داده‌ی علمی دارد تلف می‌شود، چون دولت‌های و سازمان‌های تأمین‌کننده‌ی هزینه‌های علمی به اندازه‌ی کافی برای پای‌گاه‌داده‌های بزرگ سرمایه‌گذاری نمی‌کنند. ریچرد فایرستون [2] از آزمایش‌گاه ملی لاورنس برکلی [3]، و هم‌کارانش می‌گویند پای‌گاه‌های داده را سنتاً آدم‌ها یا گروه‌های کوچک متعهد اداره می‌کند. اما مقدار زیادی داده دارد از بین می‌رود، چون برای نگهداری پای‌گاه‌های خاص‌ی تعهدی وجود ندارد. یک دانش‌پیشه‌ی دیگر ایالات متحده (سیرگی برین [4] از دانش‌گاه ستنفرد [5]) ادعا کرده ” ظرف پنج سال جست‌وجوگر وب (به شکل‌ی که می‌شناسیم) از بین خواهد رفت“ [6].

مثال‌ی از داده‌های گم‌شده، نتایج آزمایش‌های یون‌سنگین در شتاب‌دهنده‌ی بیولک [7] در برکلی است. کار این شتاب‌دهنده در 1993 تمام شد اما بسیاری از داده‌هایش (که در پژوهش‌های نوترینوهای خورشیدی، هسته‌زایی، و پرتوهای کیهانی مهم اند) هرگز به هیچ شکل‌ی منتشر نشد. فایرستون و هم‌کارانش می‌گویند: ” دانش‌پیشه‌ها باید چندین دهه صبر کنند تا این داده‌ها دوباره سنجیده شوند.“ آن‌ها از طرح درازمدت انبارش و انتشار داده‌های دست‌گاه 600 میلیون دلاری برخورددهنده‌ی یون‌سنگین‌نسبیتی (ریک) [8] در آزمایش‌گاه ملی بروک‌هیون [9]، و نیز طرح مربوط به شتاب‌دهنده‌ی سیبف [10] در تأسیسات شتاب‌دهنده‌ی ملی تامس جفرسون [11] هم انتقاد کرده اند، اما طرح ساختن رصدخانه‌های مجازی برای تحلیل داده‌های حاصل از تله‌سکوپ‌ها و ماه‌واره‌ها [12] را ستوده اند.

دیکن باتلر [13]، در یک مقاله‌ی مربوط به همین موضوع می‌نویسد انبارش داده

کافی نیست. باید بتوان جست‌وجو هم کرد یا از آن نمایه تهیه کرد. فقط 50% از میلیارد‌ها صفحه‌ی وب فهرست شده‌اند، و انتظار می‌رود طی دو سال آینده بیش از 100 میلیارد صفحه به این مجموعه افزوده شود. به گفته‌ی باتلر، تنه‌اراه بارآوری نسبی جدیدی از جست‌وجوگرها برای دانش‌پیشه‌ها است. او می‌نویسد، با ابزارهایی مثل اکس‌ام‌ال [14] (جانشین اچ‌تی‌ام‌ال [15]) باید بشود جست‌وجو را به مقاله‌های علمی محدود کرد، و البته ظهور خبرده‌های تخصصی علمی هم کمک می‌کند. الگوریتم‌های جدید جست‌وجو (که تعداد رجوع به صفحه‌ها را هم در نظر می‌گیرند) هم قاعدتاً کمک می‌کنند کاربرها مستقیماً به آن چه به کارشان مربوط است برسند. اولین دوسر نمونه‌ی این جست‌وجوگرهای پیش‌رفته (گوگل [16] و ریسرچ‌ایندکس [17]) همین حالا هم روی وب‌اند.

- [1] Nature
- [2] Richard Firestone
- [3] Lawrence Berkeley National Laboratory
- [4] Sergey Brin
- [5] Stanford University
- [6] Nature **405** 117; Nature **405** 112
- [7] Bevelac
- [8] Relativistic Heavy-Ion Collider (RHIC)
- [9] Brookhaven National Laboratory
- [10] CEBAF
- [11] Thomas Jefferson National Accelerator Facility
- [12] physicsweb.org/article/news/4/4/8
- [13] Declan Butler
- [14] XML (eXtensible Markup Language)
- [15] HTML (HyperText Markup Language)
- [16] Google
- [17] ResearchIndex