

پژوهشگران قصد دارند روندهای فکری در arXiv را دنبال کنند / اریک هند منبع: Nature (24 February 2012)

گینزپارک می‌گوید که او این ابزار را برای ثبت روندهایی به کار خواهد برد که ممکن است انتظارش را در فیزیک انرژی‌های بالا داشته باشد: نبض فزاینده‌ی مقالات در ارتباط با بوزون هیگز، یا شروع افت شمار مقالات در ارتباط با ابرتقارن، نظریه‌ی که شاید دوران اوجش به سر آمده است (نگاه کنید به برخورد نظریه‌ی زیبا با داده‌های کوبنده‌ی ذرات^۱). او می‌گوید این ابزار می‌تواند از طریق مشخص کردن زمان‌های ابداع واژه‌های جدید و هم‌گرایی و واگرایی عرصه‌های علمی موجب پیش‌رفت عرصه‌ی نوپای «تبارشناسی اطلاعات^۲» شود.

او اضافه می‌کند: «در آینده بعضی مورخان هنگامی که به عقب برگردند و به این لحظه نگاه کنند به‌شکلی باورنکردنی. مجذوب خواهند شد. آن‌چه ثبت می‌شود در واقع جنبش‌های فکری در جامعه‌اند.» گینزپارک می‌گوید سیاست‌گذاران علم حتی می‌توانند با **کرم کتاب** عرصه‌های جدیدی را که نیاز به تأمین بودجه دارند یا آن عرصه‌های در حال افول را که بودجه‌ی کم‌تری لازم دارند شناسایی کنند.

بعد از تحلیل arXiv اشمیت و همکارانش برای بررسی‌های بعدی‌شان در مضیقه‌ی داده نخواهند بود. نمونه‌ی مورد توجه عامه مانند روزنامه‌هاست بلکه منابع علمی مانند PubMedCentral، مخزن بر خط حاوی ۲/۳ میلیون مقاله مربوط به زیست‌پزشکی را هم در بر می‌گیرد.

Researchers aim to chart intellectual trends
Eric Hand

مترجم: محمدرضا بهاری



ژان باتیست میشل (جلوی شکل) می‌خواهد arXiv را برای بررسی چه‌گونگی تغییرات زبان علمی به کار ببرد. (عکس از کریس استیب دانشگاه هاروارد) برگرفته از وبگاه نیچر.

«Culturomics: Word play». اما پایگاه داده‌های گوگل گرفتار محدودیتی عمده است: از آنجا که بسیاری از این آثار مشمول انحصار حق نشر (کپی‌رایت) هستند نمی‌توان کاربران را به منابع اصلی آنها ارجاع داد.

در نتیجه پژوهشگران توجه‌شان را به این معطوف کرده‌اند که رابط نرم‌افزار و داده‌ها را با مجموعه داده‌هایی سازگار کنند که مانند arXiv دسترسی به آنها آزاد است. آنها شکل جدید نرم‌افزارشان را که **کرم کتاب**^۷ نامیده‌اند برای بررسی یک‌میلیون کتاب **کتابخانه‌ی آزاد**^۸ به کار برده‌اند که فاقد انحصار حق نشرند. به نرم‌افزار جدید قابلیت تفکیک کتاب برحسب موضوع و محل انتشار نیز اضافه شده است. پژوهشگران تاکنون با این نرم‌افزار داده‌های یک ماه arXiv را آزموده‌اند اما در نظر دارند مجموعه‌ی کامل داده‌های arXiv را در هفته‌های آینده به آزمایش خود اضافه کنند. هیجان‌زدگی میشل نه فقط به‌خاطر آن است که که گروه جدیدی از کاربران این ابزار را خواهند آزمود بلکه این هم هست که دانش نهفته در arXiv متفاوت است. او می‌گوید: «الگوهای که از این مجموعه به‌دست می‌آید شاید کاملاً با آن‌چه از فرهنگ همه‌پسند (پاپ) به‌دست می‌آید متفاوت باشد.»

یکی از خالقان **کرم کتاب**، بنجامین اشمیت^۹، دانشجوی دکترای تاریخ در دانشگاه پرینستون نیوجرسی، می‌خواهد در مقالات arXiv در باره‌ی قوانین ریاضی حاکم بر بازگانی، صفتهایی را که در ارتباط با معادله‌ی بلک-اسکولز^{۱۰} آمده است بکاود تا معلوم کند قبل و بعد از سقوط بازار در ۲۰۰۸ تغییری کرده‌اند یا نه – این معادله برای تعیین قیمت اوراق بهادار مشتق‌شده^{۱۱} به کار می‌رود.

تیم «فرهنگ‌سنجی^۱» از بخش کتاب گوگل به پیش‌نوشت‌های علمی رو می‌آورد.

پل گینزپارک^۲ در نظر دارد هفته‌ی آینده که به جلسه‌ی انجمن فیزیک آمریکا در بوستون ماساچوست می‌رود حافظه‌ی فلش ۶۴ گیگابایتی شامل حدود ۷۴۰۰۰ مقاله‌ی ذخیره‌شده در arXiv را با خودش ببرد. arXiv مخزن بزرگ پیش‌نویس مقالات است که او در سال ۱۹۹۱ ایجادش کرده و اداره‌اش را به دانشگاه کورنل در ایالتاکیونیورک سپرده است.

گینزپارک این داده‌ها را در اختیار پژوهشگران «رصدخانه‌ی فرهنگی^۳» وابسته به دانشگاه هاروارد در کمبریج ماساچوست خواهد گذاشت. آنها می‌خواهند متن کامل مقالات را به عبارت‌های تشکیل‌دهنده‌شان تجزیه کنند تا ببینند بس آمد هر واژه یا عبارت خاص نسبت به باقی کلمات و عبارت چه‌قدر است – که معیاری از درجه‌ی شیوع اصطلاحات در این مقاله‌ها خواهد بود.

هدف‌شان این است که arXiv را به ابزار جدیدی برای تشخیص مقالات اصل و بدیع در فیزیک، ریاضیات، و علوم کامپیوتر مجهز کنند و به مورخان امکان بدهند که روندها را در این مجموعه که در بیست سال اخیر گردآوری شده است تشخیص بدهند.

ژان باتیست میشل^۴ یکی از مدیران رصدخانه‌ی فرهنگی و پژوهشگر دوره‌ی پساکرتاری روان‌شناسی در هاروارد این پرسش را مطرح می‌کند: «چه طور می‌توانید نقطه‌ی وقوع تحولی علمی را پیدا کنید؟» و پاسخش این است که «می‌توانید به خواننده کمک کنید پی‌برد مربوط‌ترین مقاله‌ها در چه زمانی واقع شده‌اند که البته همیشه کار سختی بوده است.»

بازی با کلمات

تیم رصدخانه‌ی فرهنگی قبلاً هم به‌خاطر اعمال روش مشابه برای ۵ میلیون کتاب گوگل تحسین برانگیخته است. پژوهشگران با استفاده از ابزاری که اسمش را تماشاچی n-گرم^۵ نامیده‌اند حساب کرده‌اند که با چه آهنگی افعال بی‌قاعده به‌شکل افعال باقاعده صرف می‌شوند و راه‌یابی «گفتار تجارت‌خانه‌ای» به زبان عادی محاوره را دنبال کرده‌اند (نگاه کنید به

1. culturomics
2. Paul Ginsparg
3. Culture Observatory
4. Jean-Baptiste Michel
5. n-gram viewer
6. <http://www.nature.com/news/2011/110617/full/474436a.html>
7. Bookworm
8. Open Library
9. Benjamin Schmidt
10. Black-Scholes equation
11. derivatives
12. <http://www.nature.com/news/2011/110228/full/471013a.html>
13. Information genealogy