

به نام خدا

گزارش میزگرد تخصصی: بررسی آثار و پیامدهای سرقت علمی و راه‌های جلوگیری از آن: پایان‌نامه‌ها و رساله‌ها

(مقابله نرم‌افزاری با سرقت علمی)

یکشنبه اول تیرماه ۱۳۹۳ ساعت ۱۴-۱۶

پژوهشگاه علوم و فناوری اطلاعات ایران

سومین میزگرد تخصصی با عنوان «بررسی آثار و پیامدهای سرقت علمی در پایان‌نامه‌ها و رساله‌ها: مقابله نرم‌افزاری با سرقت علمی» در پژوهشگاه علوم و فناوری اطلاعات ایران در تاریخ اول تیرماه ۱۳۹۳ برگزار شد که طی آن سخنرانان دیدگاه‌های خود را ارائه کردند. این میزگرد با مشارکت جناب آقایان دکتر مهدی بهنیا، مهندس امیر پارسی و مهندس عمار جلالی منش برگزار شد.

در ابتدا دکتر محمود بابایی، سرپرست معاونت پژوهشی ایرانداک، با مروری بر محورهای گفتگو در دو میزگرد گذشته، تأکید کردند به دلیل گستردگی بحث سرقت علمی، تمرکز این میزگرد فقط برپایان‌نامه‌ها و رساله‌های دکتری است.

سپس آقای مهندس عمار جلالی منش به پیوند ارجاعات با پدیده سرقت علمی پرداختند که بر پایه آن‌ها می‌توان استراتژی‌های مختلفی برای کاهش سرقت علمی برگزینیم. از جمله: پیگرد قانونی، ایجاد محیط مشارکت به جای تنبیه، آموزش روش‌های ارجاع و نقل قول، استفاده از نرم‌افزارهای پشتیبان ارجاع که فرایند ارجاع را تسهیل می‌کند و نرم‌افزارهای کنترل سرقت علمی. همچنین افزودند:

دو نوع سرقت علمی عمدی (خواسته) و غیر عمدی (ناخواسته) وجود دارد. سرقت علمی ناخواسته به دلیل عدم ارجاع بسیاری از پژوهشگران به دلیل اطلاع نداشتن از الگوها، روش‌ها و ابزارهای ارجاع رخ می‌دهند. فرایند ارجاع در صورتی که به وسیله سیستم‌های نرم‌افزاری پشتیبانی نگردد بسیار وقت گیر خواهد بود و افرادی که به تازگی وارد عرصه پژوهش می‌گردند به دلیل ناآشنایی با مصادیق سرقت علمی و ناتوانی در تبعیت از فرایند ارجاع و نقل قول سارق علمی می‌شوند.

از جمله سختی‌های ارجاع می‌توان به این موارد اشاره کرد که در زبان فارسی از شیوه‌های استاندارد ارجاع در مستندات علمی مانند پایان‌نامه‌ها و رساله‌ها تبعیت نمی‌شود، عدم امکان استخراج ارجاعات به شکل ماشینی، ذخیره نشدن اطلاعات کتابشناختی در قالب‌های استاندارد و قابل اشاعه به وسیله ناشران و وبسایت‌های نمایه‌ساز مقالات و مستندات علمی و عدم

پشتیبانی نرم افزارها از مشکلات شیوه‌های استنادی زبان فارسی است. هرچه ارجاعات در زبان فارسی کم شود، خود سبب به خطر افتادن تولید علم به زبان فارسی خواهد شد.

استانداردهای اطلاعات کتابشناختی و شیوه‌نامه ارجاع شامل استانداردهای نرم‌افزاری بر اساس استاندارد (XML)، تبعیت دانشگاه‌ها، ناشران و وبسایت‌های نمایه‌ساز از این استانداردها و دسترسی از درگاه‌های یگانه به اطلاعات کتابشناختی استاندارد است. نرم‌افزارهای مدیریت ارجاع نرم‌افزاری است که قابلیت پشتیبانی از استانداردهای اطلاعات کتابشناختی، پشتیبانی از شیوه‌های مختلف ارجاع در مجلات و ...، پشتیبانی از نرم‌افزارهای واژه‌پرداز، امکانات مدیریت ارجاع و مستندات را داشته باشد و دارای امکانات جستجو در سایت‌های ناشران و نمایه‌سازان نیز باشد.

سپس مهندس امیر پارسی با تمرکز بر نرم‌افزارهای کشف تشابه بحث خود را ارائه کردند.

وی با اشاره به دو روشی که برای تشخیص تشابه‌های متنی وجود دارد به شرح آن‌ها پرداخت: در روش اول، متن موردنظر را با متون مرجع (که نوعاً در پایگاه‌های اطلاعاتی قرار دارند) مقایسه می‌کنند و در روش دوم، با تحلیل متن و بدون مقایسه این کار انجام می‌شود (تحلیل شیوه‌ی نوشتار نویسنده).

روش‌های تشابه‌یابی در متن با روش مقایسه مبتنی بر کپی دقیق، بازنویسی و ترجمه اجرا می‌شوند. منابع مرجع که در تشابه‌یابی موردتوجه است شامل: پایگاه‌های اطلاعاتی داخلی، اینترنت (مثل ایرانداک و ...)، پایگاه‌های اطلاعاتی ساماندهی‌شده خارج از کشور (الزویر، اسکوپوس و ...) است.

نرم‌افزارهای تشابه‌یاب خارجی مزایایی به شرح زیر دارند: کیفیت سرویس‌دهی، دقت، سرعت، سرویس‌های متنوع، بی‌نیاز کردن ما از هزینه‌های طراحی و دسترسی به منابع خارجی

مهندس پارسی نارسایی‌های اساسی نرم‌افزار را خارجی را نیز بدین شرح برشمردند: عدم دسترسی به منابع داخلی، متون، اصطلاح‌نامه‌ها و سایر ابزارهای تحلیل در زبان‌شناسی فارسی، نادیده گرفتن ویژگی‌های زبان فارسی، عدم اطمینان به آن‌ها در سرعت مطالب و ایده‌های نو (مقالات، طرح‌های پیشنهادی، ...) و عدم اطمینان به آن‌ها در شرکت در تحریرها علیه ایران، فراهم شدن زمینه برای جاسوسی علمی و نبود امکان توسعه با گسترش نیازهای کشور.

در رابطه با نرم‌افزارهای داخلی می‌توان گفت، آنچه معایب سیستم‌های خارجی محسوب می‌شد، محاسن نرم‌افزارهای داخلی است.

باین حال، نرم‌افزارهای داخلی معایبی هم دارند. از جمله: به‌طورجدی راه‌اندازی نشده‌اند، از بلوغ کافی برخوردار نیستند، هم در ویژگی‌های کارکردی، هم در ویژگی‌های غیرکارکردی، برای دسترسی به برخی منابع خارجی، با قدری مشکل مواجه‌اند و تعدادشان چندان قابل توجه نیست. بنابراین گزینه‌های چندانی پیش رو نیست. در دسترسی به منابع، با انواع مختلف

فیلترینگ،، نقص در فرهنگ سازی، کمبود منابع مالی در بخش خصوصی برای توسعه و در آخر با مشکلاتی که دیگر نرم افزارها در سطح کشور با آنها مواجه اند، از جمله کپی رایت، روبرو هستند..

آقای مهندس پارسی اشاره کردند که محدودیت هایی وجود دارد که هر دو گروه نرم افزاری داخلی و خارجی با آنها مواجه اند: منابع داخلی سازمان یافته نیستند، به دلیل کاربرد کدپیج های متنوع فارسی، جستجو در برخی منابع مثل فایل های پی.دی.اف. که مورد استفاده هستند، با دشواری روبروست، حجم قابل توجهی از منابع با فرمت های دیگر هستند. بخشی از منابع به صورت اسکن شده هستند که تبدیل آنها به متن و نهایتاً تحلیل آنها، به سادگی میسر نیست.

آنچه در ایرانداک انجام شده است:

تشابه یابی در متن به روش کپی دقیق در اینترنت و پایگاه داخلی ایرانداک صورت گرفته و امکان استفاده از پایگاه های اطلاعاتی دیگر در سراسر کشور در صورت همکاری سازمان های دارنده اطلاعات طراحی شده است و این سیستم هم اکنون بر روی شبکه وب، به صورت آزمایشی قرار داده شده و بانام تجاری ثبت شده است.

از محدودیت ها فعلی این نرم افزار امکان جستجو فقط در چکیده ها است و البته به زودی امکان جست و جو در متن کامل پایان نامه هایی که به شکل فایل ورد ارسال شده اند وجود دارد و مقالات و منابع خارجی داوری شده به آن افزوده خواهد شد و محدودیت دیگر اینکه سرویس استفاده از اینترنت، گر چه مراحل آزمودن را پشت سر گذاشته اما پس از خرید این سرویس از منابع خارجی، قابل استفاده توسط همگان خواهد بود.

آنچه باید در این زمینه انجام شود نیز عبارت است از: افزودن متن پایان نامه، عقد قرارداد با پایگاه های خارجی برای استفاده از آنها به عنوان مرجع، استفاده از اصطلاح نامه های طراحی شده در ایرانداک در الگوریتم های باز نویسی، توسعه روز افزون منابع به شکل رایگان یا غیر آن.

در ادامه **دکتر مهدی بهنیا فر** به ارائه مطالب خود پرداختند که گزیده ای از آن، بدین قرار است:

مقوله سرقت علمی، تنها یکی از اقسام سوء رفتارهای پژوهشی است و در میان عوامل مختلف سرقت علمی، می توان به فقر استنادی تأکید بیشتری کرد. نسبت به حوزه های دیگر، در حوزه علوم انسانی و اسلامی نقش زبان بسیار برجسته تر است و بحث امروز بیشتر معطوف به تولیدات علمی فارسی و عربی را بیشتر مد نظر دارد.

ابزارهای مورد تأکید این بحث، ابزارهای ایجابی (مقام پژوهش) در مقابل ابزارهای کشف سرقت است که فرد حین پژوهش از این ابزارها استفاده خواهد کرد شامل ابزارهای استناددهی، ابزارهای سازمان دهی دانش و سامانه های مشابهت یابی را

می‌توان هم به‌عنوان ابزار مقام پژوهش و هم ابزار کشف سرقت نام برد و دسته دوم، ابزارهای سلبی (مقام کشف سرقت) شامل سامانه‌های مشابهت یابی، سامانه‌های کشف تقلب و سامانه‌های مختلط (یعنی ابتدا ترجمه و سپس مشابهت یابی) هستند.

باید توجه داشت که بین مشابهت و تقلب تفاوت وجود دارد و نمی‌توان هر مشابهتی را به‌عنوان تقلب محسوب کرد.

ابزارهای بومی استناددهی وجود واقعی و مؤثر ندارند. **ابزار استنادی بومی** ابزاری است که کاربری آسان و رابط کاربری فهم‌پذیر داشته باشد، کلیدهای استناددهی به اقلام اطلاعاتی بومی (نه فقط شامل پایان‌نامه و مقاله بلکه تمام اقلام اطلاعاتی از جمله روزنامه و حتی فیلم، عکس و...) را داشته باشد، اساساً زبان‌های راست به چپ (و خصوصاً فارسی) را پشتیبانی کند، شیوه‌نامه استنادی بومی را (با مشخصات خاص خود) پشتیبانی کند. نرم‌افزار استناددهی «پژوهیار» (۱۳۸۹-۱۳۹۲) گامی است که در این راه برداشته شده است، ابزار استنادی فارسی است که در مواجهه با سرقت علمی، رویکردی ایجابی دارد نه سلبی. از جمله امکانات جانبی آن می‌توان این موارد را ذکر کرد:

مجموعه‌سازی از منابع مختلف اطلاعاتی؛

سازمان‌دهی دانش؛

هم‌زمان‌سازی اطلاعات کاربر.

از جمله مشکلات «پژوهیار» و هر نرم‌افزار بومی دیگر در فضای کاربری و انسانی می‌توان گفت از کمبودهایی دو حوزه رنج می‌برد:

که مورد اول حوزه دانش کاربران (تأکید بحث بر حوزه علوم انسانی) و دیگری مهارت کاربران است، در حوزه علوم انسانی دانش استناددهی ما پایین است و اینکه ضرورت‌های اخلاقی و علمی منتهی به استناددهی به‌خوبی درک نشده است و در حوزه مهارت کاربران، کاربران لزومی بر مهارت کار با نرم‌افزارهای این حوزه را نمی‌بینند.

نکته قابل توجه دیگر اینکه دانشوران ایرانی علوم انسانی در مقام تولید علم با کیفیات سروکار دارند تا کمیات سنجش‌پذیر، با «برهان» (به معنای ریاضی و محاسباتی) کمتر سروکار دارند و بیش از دیگر رشته‌ها با منابع پژوهشی چندزبانه در ارتباط هستند.

دانشوران علوم انسانی در بهره‌گیری از ابزارهای مجازی وب، کمتر آگاهی دارند و از نرم‌افزارهای پردازش متن (ابزارهای استناددهی، خطایاب‌های املائی و نگارشی) کمتر استفاده می‌کنند، گاهی به‌طور ناخواسته در مواجهه با تولید علمی دیگران

به استناددهی اهمیت کمتری می‌دهند و گاهی هم برای سرقت علمی مصادیق چندانی قائل نیستند. از علل عدم استفاده از استناددهی در دوره پست‌مدرن هم می‌توان به مشاع دانستن دانش اشاره کرد.

«پژوهیار» در زنجیره خدمات پژوهشی فارسی از کمبودهای زیرساختی رنج می‌برد که شامل موارد زیر است: فقدان شیوه‌نامه‌های استنادی بومی تخصصی، فقدان بانک‌های فراداده‌ای قابل اتکا، کمبود پایگاه‌های علمی استاندارد فارسی قابل درون‌برد.

دکتر بهنیا فردر رابطه با شیوه‌نامه‌ها استناددهی اشاره کردند که شیوه‌نامه‌های بومی استنادی هم وجود مؤثر و توزیع شده‌ای ندارند با این حال در آینده، از شیوه‌نامه‌های موجود مانند «شیوه‌نامه ایران» که از انتشارات پژوهشگاه علوم و فناوری اطلاعات ایران است، استفاده خواهد شد. شیوه‌نامه استنادی بومی یعنی شیوه‌نامه‌ای که افزون بر استانداردهای رایج دیگر زبان‌ها و ملل، مقتضیات زبان‌های راست به چپ به خصوص فارسی را پشتیبانی کند و کلیدواژه‌های استنادی فارسی در آن جانمایی شده باشد، نحوه استناد به اقلام اطلاعاتی بومی ما را روشن کند، نحوه استناد به قرآن کریم، متون حدیثی و دعایی (ازجمله نهج‌البلاغه، صحیفه سجاده و...)، آثار فارسی بومی (دیوان‌های شعر فارسی)، آثار هنری بومی (صنایع دستی، نسخ خطی و...) و .. را برای ما روشن سازد و شیوه استناد فارسی در یک رشته استنادی شامل دیگر زبان‌ها را تعیین کند.

کاستی‌ها در حوزه عملکرد این نرم‌افزار و دیگر نرم‌افزارهای مشابهت یاب فارسی وجود دارد: کاستی‌های حوزه انسانی، کاستی‌های حوزه قضایی، کاستی‌های حوزه بانک اطلاعاتی. کاستی‌های حوزه انسانی در لایه تشخیص سرقت با ارتقاء و بهبود الگوریتم‌های ماشینی و مهارت انسانی سوء رفتارهای پژوهشی هم پیچیده‌تر می‌شوند و سبب افزایش مهارت کاربران در دور زدن ماشین شده است و نهایتاً سبب توسعه دستبردهای علمی از لایه لفظ به لایه معنا شده است.

کاستی‌های حوزه قضایی این است که قوانین موضوعی ما ماشینی نیستند تفکیکی میان اقسام سوء رفتارهای پژوهشی دیده نشده از این رو، کشف ماشینی و تفکیک شده آن‌ها هم، عقوبت و مجازاتی خاصی را در پی ندارد.

گاهی مشابهت را در زبانی و گاهی بین زبانی باید مطرح کرد و برای اینکه مشابهت یابی بین زبانی صورت بگیرد ما باید بتوانیم دادگان در زبانی، دادگان موازی و پیکره‌های مناسبی برای این کار داشته باشیم.

و در پایان نتیجه‌گیری شد که دسترسی به عنوان یک ارزش باید ترویج گردد و این دسترسی‌پذیری الزاماً به معنای به خطر افتادن مالکیت فرد یا سازمان دارنده متن نیست و نکته دیگر اینکه ما باید رویکردهای پرتالی را در تعاملات بین سازمانی توسعه دهیم.

جلسه پس از پرسش و پاسخ، در ساعت ۱۶ پایان یافت.