

سرقت علمی: از پیشگیری تا تشخیص

رضیه اسماعیل پور^۱، زهرا عزیزخانی^۲، دکتر سید مصطفی فخر احمد^۳

تاریخ دریافت: ۱۳۹۵/۱۰/۱ تاریخ پذیرش: ۱۳۹۶/۶/۱۰

از صفحه ۱۲۳ تا ۱۴۴

چکیده

زمینه و هدف: سرقت علمی یکی از سوءرفتارهای پژوهشی است که از دیرباز مورد نكوهش بوده است. جامعه علمی که آگاهانه یا ناآگاهانه دچار آن بوده، می‌بایست هرچه سریعتر علّاجی برای آن بجوید. آگاهی رسانی و تدوین برنامه‌های ساختارمند آموزشی در زمینه جلوگیری از سرقت علمی و توسعه نرم‌افزارهای تشخیص و ردیابی سرقت علمی از راهکارهای موثر در مبارزه با این سوء رفتار علمی می‌باشد. از فواید استفاده از این نرم‌افزارهای تشخیص می‌توان به سرعت بخشی به فرایند داوری مقالات، تسهیل ارزیابی تکالیف درسی دانشجویان، رفع معضل فروش مقالات و پایان نامه‌ها اشاره کرد.

روش: نوشتار حاضر با روشی مروری به مطالعه پدیده سرقت علمی و راه‌های مبارزه با آن از دو جنبه اخلاق انتشار علمی و روش‌های تشخیص پرداخته است.

نتیجه‌گیری: ضرورت پرداختن به آموزش جلوگیری از سرقت علمی امروزه بر کسی پوشیده نیست. تولید محتواهای مناسب آموزشی، تدریس درس سرقت علمی و راه‌های پیشگیری از آن در دوران دانشجویی، گنجانیدن برنامه عملی مجازات، نقش کتابداران و کتابخانه‌ها در امر آموزش می‌تواند از راهکارهای آموزشی باشد. توسعه روش‌های هوشمند تشخیص سرقت علمی و پیاده‌سازی نرم‌افزارهای ردیابی از دیگر تمهیدات مبارزه با سرقت علمی به شمار می‌رود.

واژگان کلیدی: اخلاق انتشار علمی، سرقت علمی، پیشگیری از سرقت علمی، تشخیص سرقت علمی

۱. دانشجوی دکتری علم اطلاعات و دانش‌شناسی دانشگاه شیراز، مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری،

negaresma@gmail.com

۲. کارشناس ارشد علم اطلاعات و دانش‌شناسی مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری،

zahra_a1383@yahoo.com

۳. عضو هیئت علمی دانشکده مهندسی کامپیوتر دانشگاه شیراز، fakhrahmad@shirazu.ac.ir

مقدمه

رضایت خالق و سپس خدمت به خلق، انگیزه‌های اصلی نگارش از دیدگاه ابن خلدون به شمار می‌روند که به نقل از حرّی، انحراف از این دو انگیزه مهم موجب بروز بی‌اخلاقی‌های علمی می‌گردد (حرّی، ۱۳۹۰). محقق می‌بایست با هدف اجتماعی کردن اندیشه خود دست به قلم برد و با در نظر گرفتن وقت خوانندگان موجب اتلاف وقت و سرگردانی آنان برای یافتن آثار معتبر علمی نگردد. پرهیز از سوء رفتارهای پژوهشی و در نظر گرفتن اصول اخلاق انتشار علمی از وظایف اصلی پژوهشگران و نویسندگان است.

از آسیب‌های وارد شده به فعالیتهای علمی و پژوهشی، سرقت علمی است. بنا به گفته محققان، افزایش تعداد پژوهشگران، رشد بودجه‌های پژوهشی و ازدیاد تعداد نشریات منجر به ایجاد فشاری بر جامعه علمی جهت انتشار مقاله می‌شود (آی‌نتنیکیت، ۲۰۱۲، بطحایی، ۱۳۹۰، مومنی و عزیزی، ۱۳۹۱) و پژوهشگران تحت تاثیر شعار "بنویس و زنده بمان" با شتاب به سمت انتشار مقالات می‌روند (بهادری، ۱۳۹۱). برخی نیز نظام‌های ارزشگذاری علمی را علت این گونه سوء رفتارهای پژوهشی می‌دانند. سوق محققان به انتشار بیشتر بدون دغدغه در انداختن طرحی نو^۱ نشان از اثرگذاری رویکرد کثرت‌گرایی این گونه نظام‌ها دارد (حرّی، ۱۳۹۰ ص ۵). این نظام‌ها با شعار هر چه پر تولیدتر علمی تر، جامعه علمی را تشویق به تولید بیشتر مقاله می‌کنند (همان، ص ۲۲). داوری اردکانی معتقد است ارزشگذاری اشتباه موجب این سوء رفتار پژوهشی می‌گردد، وی معتقد است تقلب علمی زمانی صورت می‌گیرد که بگوییم فردی که دارای ۱۰۰ مقاله است از فردی که ۹۰ مقاله نوشته عالم‌تر است (همان ص ۷۷).

کمیت‌ه اخلاق نشر^۲ سرققت علمی را چنین تعریف می‌کند: استفاده از کار دیگران بدون ذکر منبع در هر مرحله‌ای از برنامه ریزی، تحقیق، نگارش یا چاپ مقاله (موسوی دوست ۱۳۸۹). این سوء رفتار پژوهشی به اندازه نكوهیده است که رویگ^۳ سرققت علمی در محیط دانشگاهی را نوعی جنایت می‌داند (رویگ، ۲۰۰۶). علاوه بر وجهه "غیراخلاقی بودن"، سوء رفتارهای پژوهشی تبعاتی مانند: از دست دادن بودجه پژوهشی، محدودیت در فعالیت‌های پژوهشی و از دست دادن شغل را با خود به همراه دارد (ای پی ای، ۲۰۱۳).

سرققت علمی انواع مختلفی دارد که در متون متعدد تقسیم‌بندی‌ها و تعاریف گوناگونی را از آن ارائه داده‌اند. به طور خلاصه انواع آن را می‌توان به شرح زیر بیان کرد:

- کپی کردن کلمه به کلمه متن^۴
- تفسیر کردن (بیان یک ایده با کلمات متفاوت، تغییر گرامر، استفاده از کلمات مترادف برای بیان یک محتوا) بدون ارجاع دادن

1. Publish or Perish
2. Committee On Publication Ethics (COPE)
3. Roig
4. Copy-Paste or Verbatim

- تغییر کلمات همراه با حفظ ساختار جمله از یک متن بدون استناد دادن به آن
- سرقت علمی ترجمه شده (ترجمه بین زبانی و استفاده از ترجمه متن بدون ارجاع به متن اصلی)
- سرقت علمی هنری (ارائه یک اثر در قالب رسانه‌های متفاوت مثل متن، تصویر، صدا)
- سرقت ایده: استفاده از ایده‌ها و مطالبی که دانش عمومی هستند، سرقت علمی محسوب نمی‌شود به طور مثال قانون جاذبه. سرقت ایده بکر شخص و عدم ارجاع به آن است که سرقت علمی محسوب می‌شود (مومنی و عزیزی، ۱۳۹۱).
- سرقت کدهای برنامه‌نویسی^۱ (استفاده و بکارگیری کدها و الگوریتم‌های برنامه‌نویسی بدون اجازه یا ارجاع)
- ارجاع نادرست (ارجاع دادن به منابع اشتباه یا منابعی که اصلاً وجود ندارند و یا عدم ارائه اطلاعات دقیق و بروز در مورد منابع استناد شده)
- نقل جمله‌ای از شخصی بدون گذاردن آن در علامت نقل قول (گیومه) یا استفاده ناصحیح از علائم نقل قول
- ارائه اطلاعات ناصحیح از یک منبع (ذکر منبع و ارائه اطلاعات اشتباه از آن)
- کپی کردن کلمات و ایده‌های فراوان از یک منبع به گونه‌ای که قسمت اعظمی از اثر شما را تشکیل دهد، خواه به آن استناد داده باشید یا خیر.
- تحریف و دستکاری در داده‌ها^۲

روش

در این نوشتار تلاش شده است تا با اتخاذ روشی مروری به پدیده سرقت علمی و راه‌های مبارزه با آن از جنبه اخلاق انتشار علمی و روش‌های تشخیصی پرداخته شود. مبارزه با این پدیده شوم از جنبه‌های مختلف مورد توجه پژوهشگران با تخصص‌های گوناگون قرار گرفته است. با مطالعه آثار این حوزه می‌توان پی برد که سرقت علمی از دو جنبه زیر مورد توجه پژوهشگران قرار گرفته است:

● جلوگیری از سرقت علمی^۳

● تشخیص سرقت علمی^۴

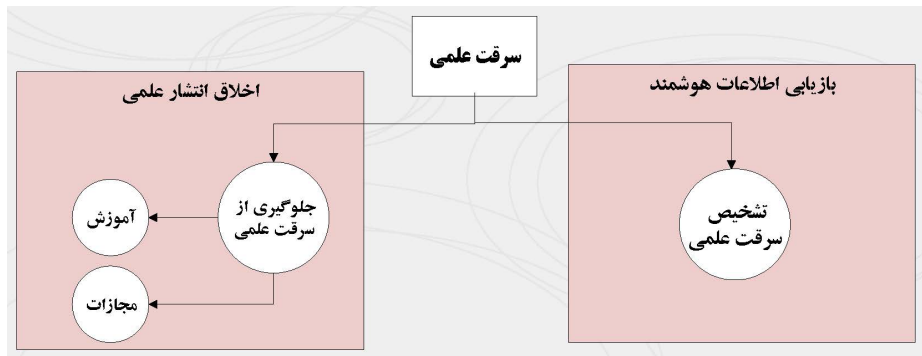
مبحث جلوگیری از سرقت علمی در مباحث اخلاق پژوهش یا اخلاق انتشار علمی مورد بررسی قرار می‌گیرد و مبحث توسعه روش‌های تشخیص سرقت علمی در حوزه باز یابی اطلاعات هوشمند مورد مطالعه قرار می‌گیرد.

1. Source Code plagiarism

۲. در بررسی‌های انجام شده توسط نشریه بریتیش مدیکال در سال ۲۰۱۲ که در بین ۲۷۰۰ محقق انگلیسی صورت گرفت، نشان داده شد که از هر ۷ دانشمند انگلیسی ۱ نفر به تغییر و دستکاری در داده‌های تحقیق خود جهت انتشار آنها پرداخته است (آی تنتیکیت، ۲۰۱۲).

3. Plagiarism Prevention

4. Plagiarism Detection



شکل ۱- تقسیم بندی مباحث مرتبط با سرقت علمی

لازم به ذکر است که هدف نگارندگان آشنا کردن مخاطب با روش‌های مبارزه با این سوء رفتار مضر در جامعه علمی می‌باشد و ضرورت توجه به تحقیق و پژوهش در جهت ارائه راهکارهای جلوگیری از سرقت علمی را نشان می‌دهد.

جلوگیری از سرقت علمی

مبحث "جلوگیری از سرقت علمی" به بررسی انواع سرقت علمی، راه‌های مبارزه با آن و ارائه آموزش‌های لازم جهت جلوگیری از ارتکاب به سرقت علمی می‌پردازد (سیندهو، توماس، ایدیکولا، ۲۰۱۱). مبحث جلوگیری از سرقت علمی به حوزه اخلاق نشر و فعالیت‌های انجمن کوپ^۱ نیز مرتبط است. امروزه نشریات در تلاشند تا با عضو شدن در این انجمن و داشتن شرایط لازم، با گذاردن علامت آن در وب سایت خود، لزوم رعایت اخلاق انتشار علمی را به نویسندگان و محققان یادآوری کنند.

آموزش ساختارمند در زمینه جلوگیری از سرقت علمی در اکثر دانشگاه‌های معتبر دنیا پیاده‌سازی شده است. شکل ۲ نمونه‌ای از آموزش جلوگیری از سرقت علمی و آموزش کار با نرم‌افزار تشخیص سرقت علمی ترتیبی در صفحه مربوط به کتابخانه دانشگاه برادفورد را نشان می‌دهد. در این سایت مطالب مرتبط با سرقت علمی و چگونگی مبارزه با آن به صورت سازماندهی شده در دسترس دانشجویان قرار گرفته است. در صفحه کتابخانه، علاوه بر کلیه‌های آموزشی، اطلاعات جامعی در مورد نحوه استناد دهی صحیح، شکل درست ارجاعات و استنادات، سبک‌های مختلف استنادی و سبک پذیرفته شده در دانشگاه ارائه می‌گردد. همچنین در این

۱. COPE: Committee On Publication Ethics انجمنی است متشکل از ویراستاران و ناشران مجلات علمی که در زمینه رعایت اخلاق نشر فعالیت دارد. ارائه توصیه‌های لازم جهت برخورد با سوء رفتارهای پژوهشی از خدمات این انجمن می‌باشد. (publicationethics.org)

دانشگاه درسی با نام پانز^۱ (جلوگیری از دستبرد علمی برای دانشجویان جدیدالورود) در سرفصل دروس اجباری دانشگاه گنجانده شده است که کتابداران متخصص در دانشگاه موظف به تدریس این درس می‌باشند (جورج، کاستیگان، اوهارا، ۲۰۱۳). آنها با شعار "علاج واقعه را قبل از وقوع باید کرد" به ارائه آموزشهای لازم به دانشجویان می‌پردازند. چنانچه دانشجویی این درس را با نمره قابل قبول نگذراند مجبور به اخذ مجدد این واحد درسی می‌باشد. همچنین گزارشاتی که نرم‌افزارهای معتبر تشخیص سرقت علمی تولید می‌کنند (مانند گزارش اصالت^۲ ترنیتین) به عهده دستیاران اساتید و یا کتابداران کتابخانه‌های دانشگاهی می‌باشد. سایت دانشگاه هاروارد^۳، دانشگاه آکسفورد^۴ و کتابخانه دانشگاه ماریلند از نمونه‌های دیگری هستند که به ارائه محتواهای آموزشی در زمینه جلوگیری از سرقت علمی به نحو مفصلی پرداخته‌اند.



شکل ۲- آموزش جلوگیری از سرقت علمی از طریق درگاه کتابخانه دانشگاه برادفورد

هر چند که آموزش بر مجازات اولویت دارد، اعمال مجازات از دیگر راهبردهای جلوگیری از سرقت علمی است. بدین معنا که ابتدا سرقت علمی و نحوه مبارزه با آن آموزش داده می‌شود و سپس در صورت اینکه دانشجو بعد از گذراندن این دوره، مرتکب سرقت علمی شد، مجازات به صورت پلکانی از گذراندن مجدد درس تا اخراج

1. PANS: Plagiarism Avoidance for New Students
2. Originality Report
3. <http://isites.harvard.edu/>
4. <https://weblearn.ox.ac.uk/portal/hierarchy/info/plag>

از دانشگاه در حق وی اعمال می‌گردد^۱. به طور کلی آموزش اخلاق و مهارت‌های آن نوعی جلوگیری از سرقت علمی به حساب می‌آید زیرا در محیطی که پایبندی به اخلاق در آن جدی گرفته می‌شود افراد تمایل کمتری به نقض حقوق دیگران و منفعت طلبی پیدا می‌کنند. علاوه بر این، پرداختن به امر آموزش هزینه کمتری را در بر دارد و برای افراد تحت تعلیم نگرش مادام‌العمر به همراه دارد. یومانز^۲ معتقد است جلوگیری از سرقت علمی در مدارس و دانشگاه‌ها فقط به این علت نیست که باعث خدشه در یادگیری و آموزش دانش‌آموزان و دانشجویان می‌گردد بلکه سرقت علمی در کلاس و محیط آموزشی زمینه‌ساز ارتکاب سرقت علمی در محل کار، دولت، دنیای مجازی و انتشار مقالات علمی می‌گردد.

سرقت علمی می‌تواند به دو دسته آگاهانه و ناآگاهانه نیز تقسیم شود. برخی بر این باورند که سرقت علمی معمولاً ناآگاهانه رخ می‌دهد و فرد از اینکه در حین پژوهش مرتکب سرقت علمی شده است آگاه نیست (جورج و دیگران، ۲۰۱۳). اما ندانستن اصول اخلاق پژوهش، مجوزی برای ارتکاب به سرقت علمی به دست‌خاطی نمی‌دهد (منصوریان، ۱۳۹۳). بنابراین محقق می‌بایست با فراگیری اصول تحقیق و اخلاق علمی راه هر گونه سوء رفتار پژوهشی را ببندد. در زمینه آموزش جلوگیری از سرقت علمی طرفداران بحث آموزش معتقدند که علاج واقعه را باید قبل از وقوع کرد و با ارائه آموزش‌های لازم از وقوع آن پیشگیری کرد. بحث آموزش نیز می‌تواند از دو جنبه آموزش رسمی و غیر رسمی مورد بحث قرار گیرد:

آموزش رسمی: پیشگیری از سرقت علمی می‌بایست در بدو امر با فرهنگ‌سازی بنیادی و آشنا کردن دانش‌آموزان با مصادیق اخلاقی و اخلاق علمی در سطح مدرسه آغاز گردد. در آموزش رسمی آکادمیک نیز در قالب دروس روش تحقیق با پر بار کردن و به روز کردن نحوه آموزش و بکارگیری نرم‌افزارهای نوین مدیریت منابع علمی مانند آندنوت، زوترو و پژوهیار، آموزش نحوه استناد دهی و نگارش مقالات ادامه یابد.

به عنوان نمونه سامانه تولید خودکار منابع برون‌متنی^۳ که در مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری طراحی و پیاده‌سازی شده است می‌تواند به دانشجویان و نویسندگان مقالات در یافتن شکل صحیح استناد با الگوهای متعارف مجلات کمک کند. با تسهیل ساخت ارجاع مقاله می‌توان پیش‌بینی کرد که مشکل استناد دهی به مجلات را برطرف سازد (فلاحتی فومنی، محمد رضا، ۱۳۹۴). شکل ۳.

1. http://www.bradford.ac.uk/library/help/plagiarism/types-of-plagiarism#buying_assignments

2. Youmans

3. Reference Generator : <http://refgen.ricest.ac.ir/>



شکل ۳- سامانه تولید خودکار منابع برون متنی مرکز منطقه ای اطلاع رسانی علوم و فناوری

آموزش غیر رسمی: برگزاری منظم و دوره‌ای کارگاه‌های آموزشی، سخنرانی‌های علمی، تولید محتوا در زمینه پیشگیری از سرقت علمی می‌تواند در آگاه‌سازی دانشجویان متمر ثمر باشد. در اکثر دانشگاه‌های پیشرفته با اینکه دستورالعمل‌های متعددی نیز در این زمینه چاپ و توزیع می‌گردد، اما طی آموزش‌های رسمی نیز این مسائل به دانشجویان آموزش داده می‌شود. تولید محتواها و کلیپ‌های آموزشی نقش بسیار موثری در آموزش صحیح پیشگیری از سرقت علمی ایفا می‌کنند. از آنجا که کلیپ‌های آموزشی به علت وضوح و روشنی بیشتر، اطلاعات زیاد و مفیدی را در طول مدت زمان کوتاهی به آموزش گیرنده انتقال می‌دهند، اثر گذاری آن بیشتر از یک فایل متنی طولانی است. بنابراین اکتفا کردن به گذاردن یک لینک به فایل متنی بدون ساختاری که مطالعه آن خارج از حوصله دانشجویان است، نمی‌تواند در امر آموزش پیشگیری از سرقت علمی موجه باشد. پس از ارائه همه آموزش‌های لازم، چنانچه دانشجویی ناقض قوانین دانشگاه شود و مرتکب سرقت علمی شود با ارزیابی‌های لازم، مجازاتی طبق میزان تخلف صورت گرفته در حق وی اعمال می‌شود. مجازات‌ها معمولاً به صورت پلکانی از اخطار به دانشجوی، کم شدن نمره درس، گذاردن مجدد واحد درسی و در نهایت اخراج از دانشگاه در اکثر دانشگاه‌های معتبر دنیا اعمال می‌گردد. سیاست سرقت علمی دانشگاه لیسیستر^۱ و دانشگاه کوئینزلند^۲ دو نمونه از سیاست‌های دانشگاهی است که به ارائه طبقه بندی مشخصی از نوع سرقت علمی، ویژگی آن و مجازات قابل اعمال اشاره دارند.

1. <http://www2.le.ac.uk/offices/sas2/assessments/plagiarism/penalties>
2. <http://www.uq.edu.au/economics/plagiarism-policy-151827>

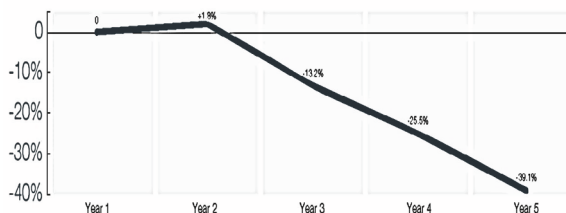
تشخیص سرقت علمی

همانطور که آموزش مباحث اخلاق انتشار علمی می‌تواند در کاهش نرخ ارتکاب سرقت علمی موثر باشد، طراحی و توسعه روش‌های ماشینی تشخیص سرقت علمی نیز حائز اهمیت است. وجود نرم‌افزارهای تشخیص سرقت علمی، افراد خاطی را آگاه می‌سازد که در صورت تخطی از موارد اخلاقی و ارتکاب سرقت علمی، سازوکارهایی برای تشخیص آن در منابع اطلاعاتی بکار برده می‌شود که در نهایت از دید داوران و افراد مسئول پنهان نمی‌ماند (علیزاده، ۱۳۹۵).

از آنجایی که مقالات علمی به عنوان واسطه‌هایی به شمار می‌روند که دانش نورا به دانش پیشین مرتبط می‌سازند، کیفیت دانش تا حد زیادی به کیفیت کار داوران مجلات بستگی پیدا می‌کند (فتاحی، ۱۳۹۴). طراحی و توسعه نرم‌افزارهای تشخیص سرقت علمی می‌تواند کمک شایانی به داوران و ویراستاران در کنترل کیفیت و سرعت بخشی در کار آنها کند. استفاده از ابزار تشخیص سرقت علمی در نشریات علمی باعث حفظ اعتبار آثار علمی منتشر شده می‌شود. ویژگی‌های جدید محمل‌های اطلاعاتی و قابلیت کپی و ارسال سریع آنها نیاز به روش‌های تشخیص ماشینی سرقت علمی را بیش از پیش نمایان می‌سازند. تعدد این نرم‌افزارها نشان از نیاز روزافزون به یک نرم‌افزار جامع و قابل اعتماد در زمینه تشخیص علمی می‌باشد.

با توجه به حجم روزافزون انتشار اطلاعات، شناسایی و ردیابی سرقت علمی از عهده داوران انسانی به تنهایی بر نمی‌آید. از این رو نرم‌افزارهای زیادی جهت تشخیص و ردیابی سرقت علمی بوجود آمده‌اند. هر یک از این نرم‌افزارها با توجه به روش‌های گوناگون شناسایی، عملکردهای تقریباً متفاوتی دارند. به طور مثال نرم‌افزار ترنیتین^۱ یکی از مشهورترین نرم‌افزارهای شناسایی سرقت علمی است که توسط اکثر دانشگاه‌های معتبر دنیا جهت ارزیابی تکالیف درسی و مقالات دانشجویی بکار می‌رود و تنها در آمریکا بیش از ۳۵۰۰ موسسه آموزش عالی از این نرم‌افزار استفاده می‌کنند. این نرم‌افزار می‌تواند مدارک را در ۱۸ زبان شناسایی کند (ترنیتین اف‌ای کیو).

ترنیتین علاوه بر اینکه به کشف سرقت علمی می‌پردازد سیستم اتوماسیون نمره دهی را نیز تسهیل کرده است. صرفه جویی در زمان با استفاده از این سیستم نمره دهی در مقایسه با روش قلم و کاغذ سنتی، تبادل نظر بهینه‌تر بین دانشجو و استاد از فواید این سیستم است. طی تحقیقی با تحلیل مقالات موجود در یک دوره پنج ساله در بیش از ۱۰۰۰ دانشگاه مشخص شد که میزان ارتکاب سرقت علمی با بکارگیری این نرم‌افزار کاهش چشمگیری را داشته است. شکل ۴.



شکل ۴- کاهش چشمگیر نرخ ارتکاب سرقت علمی در آموزش عالی آمریکا در یک دوره پنج ساله (گزارش ترنیتین، ۲۰۱۴)

علاوه بر ترنیتین، نرم افزارهای زیادی چه به صورت کلاینت و یا مبتنی بر سرور وجود دارند که در این زمینه به جامعه علمی باری رسانده اند؛ روند توسعه این نرم افزارها حاکی از اهمیت این مسئله دارد. مطالعات در زمینه بکارگیری روش های آماری و محاسباتی در پردازش زبان طبیعی و تشخیص سرقت علمی از سال ۱۹۹۰ آغاز گردید (الزهرانی، سلیم، آبراهام، ۲۰۱۲). این روش ها امروزه مورد توجه محققان زیادی در رشته های بازیابی اطلاعات، بازیابی اطلاعات بین زبانی، زبانشناسی رایانه ای، هوش مصنوعی و زبانشناسی جنایی قرار گرفته است. تعدد نرم افزارهای کشف سرقت علمی نیز نشان دهنده نیاز به تولید و توسعه این گونه نرم افزارها با قابلیت های گوناگون می باشد.

محققانی که در زمینه توسعه و پیشرفت روش های تشخیص سرقت علمی فعالیت دارند علل عمده نیاز به توسعه نرم افزارهای تشخیص سرقت علمی را موارد زیر معرفی کرده اند:

- ویژگی های محمل های جدید اطلاعات: افزایش حجم عظیم منابع اطلاعاتی، سهولت در تکثیر و ارسال اطلاعات، وجود نرم افزارهای تبدیل تصویر به متن، زمینه را برای ارتکاب سرقت علمی مناسب می کند (استاماناتوس، ۲۰۱۱؛ پتاست، بارون، استین، روسو، ۲۰۱۱).
- نشریات علمی: انتشار روزافزون نشریات و بالا رفتن تقاضای چاپ مقالات و گزارشات علمی در پی افزایش تعداد دانشجویان تحصیلات تکمیلی و پژوهشگران منجر به طولانی شدن فرایند داوری نشریات و هزینه های زیاد زمانی و مالی آن، نیاز به کمک گرفتن از نرم افزارها در تشخیص اصالت اثر را بیشتر نشان می دهد (استاماناتوس، ۲۰۱۱).
- بازار فروش مدارک علمی: معضل فروش پایان نامه ها و مقالات علمی هم در ایران و هم در خارج از کشور شیوع زیادی پیدا کرده است (پاتل، بختیاری و تقوی، ۲۰۱۱). این مشکل متولیان امر در وزارت علوم، تحقیقات و فناوری ایران را به تشکیل کارگروه هایی جهت برون رفت از آن وا داشته است. بنابراین شناخت و توسعه روش های تشخیص سرقت علمی می تواند گامی موثر در راستای حل این مسئله بر دارد.

روش‌های تشخیص سرقت علمی

همانطور که اشاره شد، با توجه به حجم عظیم انتشارات و منابع علمی، کار تشخیص سرقت علمی بدون کمک گرفتن از ماشین امکان پذیر نیست و توسعه نرم‌افزارهای متعدد در این زمینه نشان دهنده این امر است. باید اشاره کرد که تفاوت بین نرم‌افزارهای موجود مربوط به نوع الگوریتم‌های بکار رفته در آنها می‌باشد. همانطور که در شکل ۱ مشاهده شد روش‌های تشخیص سرقت علمی در حوزه هوش مصنوعی و بازیابی اطلاعات هوشمند مورد بحث قرار می‌گیرند. این بحث در حوزه بازیابی اطلاعات به میزانی مورد استقبال بوده است که همانند دیگر کنفرانس‌های بازیابی اطلاعات مثل TREC و CLEF که به بحث و تبادل نظر در رابطه با بهبود روش‌های بازیابی اطلاعات می‌پردازند، مورد بحث قرار می‌گیرد. PAN قسمتی از کنفرانس CLEF (بازیابی اطلاعات بین زبانی) می‌باشد که به ارزیابی الگوریتم‌های تشخیص سرقت علمی می‌پردازد. در این مقاله تلاش شده است تا با مرور منابع به ارائه روش‌های کلی تشخیص سرقت علمی پرداخته شود.

تشخیص سرقت علمی دستی و ماشینی

هر چند که تشخیص سرقت علمی به صورت دستی می‌تواند بدین معنا باشد که شخص با مطالعه مدرک، ارتکاب سرقت علمی و یا عدم آن را تشخیص دهد؛ اما این روش علاوه بر صرف وقت بسیار نیاز به تجربه و دانش فراوان شخص در زمینه موضوعی مدرک دارد که عملاً آن را بسیار سخت و غیر ممکن می‌نمایاند. تفسیر دیگری نیز از روش تشخیص به صورت دستی آمده است؛ انتخاب پاراگراف شاخصی از متن و جستجوی آن از طریق موتورهای جستجوی قوی مانند گوگل را یک نوع روش دستی در تشخیص سرقت علمی معرفی کرده اند (مورر، کاپه و زاکا، ۲۰۰۶)، روش‌های ماشینی و دستی را با عنوان روش‌های تحت اللفظی^۱ و هوشمند^۲ نیز تقسیم بندی کرده اند (الزهرانی، سلیم، آبراهام، ۲۰۱۲).

همه روش‌های تشخیص سرقت علمی را می‌توان از طریق پیاده‌سازی با الگوریتم‌های یادگیری ماشینی به صورت خودکار به اجرا درآورد، بنابراین همه در دسته روش‌های ماشینی نیز قرار می‌گیرند. روش‌های ماشینی تشخیص سرقت علمی نیز خود به تشخیص سرقت علمی در کدهای برنامه‌نویسی و تشخیص سرقت علمی در متون تقسیم بندی می‌شوند (علیزاده، ۱۳۹۵). فنونی که در تشخیص سرقت کدهای برنامه‌نویسی استفاده می‌شود با فنون تشخیص سرقت علمی در متون متفاوت است. در سرقت کدهای برنامه‌نویسی، شخص بدون استناد دادن به کدهای نوشته شده توسط شخصی دیگر، با کپی برداری از کل کد و یا قسمتی از آن و یا دستکاری و تغییر، از آنها در برنامه خود استفاده می‌کند. روش‌های ماشینی در تشخیص سرقت علمی متون مانند همه روش‌های پردازش زبان طبیعی می‌توانند مستقل از زبان و یا وابسته به زبان عمل کنند. بدین معنا که تشخیص سرقت علمی را در زبانی خاص یا در تمامی زبانها بتوانند مشخص کنند.

1. Literal Plagiarism
2. Intelligent Plagiarism

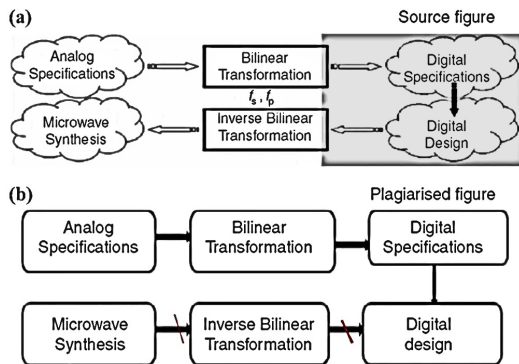
تشخیص سرقت علمی درونی^۱ و بیرونی^۲

روش‌های تشخیص سرقت علمی را از جنبه مرجع مورد مقایسه می‌توان به دو دسته درونی و بیرونی تقسیم کرد. در روش بیرونی، مدرک مشکوک با مجموعه‌ای از سایر مدارک مورد مقایسه قرار می‌گیرد. هر چه تعداد مدارک موجود در این مجموعه بیشتر باشد کار مقایسه با دقت و اطمینان بالاتری صورت می‌گیرد (طاهرعلی، ۲۰۱۱). تشخیص شباهت با روش‌های تشخیص بیرونی به دو نوع مکانی و سراسری تقسیم می‌شود:

تشخیص شباهت مکانی: از پرکاربردترین روش‌های تشخیص سرقت علمی روش انگشت نگاری^۳ مدرک می‌باشد. در این روش هر مدرک به زیر رشته‌هایی^۴ تقسیم می‌شود که به عنوان اثر انگشت عمل می‌کنند. زیر رشته‌های مدرک مشکوک با زیر رشته‌های مدارک موجود مقایسه می‌شوند و تشخیص متن مشکوک صورت می‌گیرد. این زیر رشته‌ها به نوعی مدرک را تحلیل و خلاصه می‌کنند و مشابه نمایه معکوب در بازیابی مدارک عمل می‌کنند. این روش سند را به صورت منحصر به فرد (مانند اثر انگشت برای انسان) شناسایی می‌کند. عملکرد روش مذکور بستگی به واحد انتخاب شده از متن دارد که می‌تواند در سطح حرف، کلمه و یا جمله باشد. تشخیص شباهت سراسری: در روش قبلی (تشخیص شباهت مکانی)، تشابهات قسمت‌های محدودی از متن بازیابی می‌شدند اما در تشخیص شباهت سراسری، قسمت‌های بزرگتری از متن و یا کل مدرک مورد تحلیل قرار می‌گیرند. روش‌هایی که بر اساس تحلیل میزان رخداد کلمه در کل مدرک عمل می‌کنند، مانند مدل فضای برداری، در این دسته قرار می‌گیرند. (راکیان، اصفهانی، رستگاری، ۱۳۹۲، گیپ، ۲۰۱۳: ۲۰) در تحلیل سراسری، حروف اضافه در نظر گرفته نمی‌شوند و فقط ریشه اصلی کلمات در نظر گرفته می‌شود تا بردارهای کلمات متن را تشکیل دهند. اما در تحلیل شباهت مکانی توالی تطابق کلمات مدرک مشکوک با مدرک اصلی که حداقل تعداد کاراکترها و کلمات را دارند مشخص می‌شود.

سرقت ایده نیز از مصادیق سرقت علمی است که در زمره روش‌های تشخیص بیرونی است. همانطور که در شکل ۵ مشاهده می‌شود، شکل (b) ایده مدل خود را از شکل (a) سرقت کرده است و فقط نوع نمادها و جهت آن‌ها را تغییر داده است. تشخیص تصاویر و نمودارها نیز در تشخیص سرقت علمی کاربرد دارد.

1. Intrinsic Plagiarism
2. Extrinsic Plagiarism
3. Fingerprinting
4. Substring



شکل ۵- نمونه‌ای از سرقت ایده (ایسا، سلیم، الزهرانی و گورمن، ۲۰۱۵)

از آنجا که گردآوری همه نوشتجات نیز عملاً غیر ممکن است، روش‌های تشخیص درونی می‌توانند با هوشمندی بالاتر جایگزین مناسبی باشند. در روش درونی، مدرک مشکوک با سایر مدارک مقایسه نمی‌شود بلکه با استخراج ویژگی‌هایی^۱ از خود متن، ادعای مبتنی بر ارتکاب سرقت علمی و یا عدم ارتکاب آن بررسی می‌شود. تشخیص سبک نگارش یا استایلومتری^۲ از جمله روش‌های تشخیص سرقت علمی درونی است و مبنای آن بر این است که هر نویسنده سبک نگارشی منحصر به فردی دارد. در این روش قسمت‌های مختلف متن از لحاظ سبک نگارشی مورد مقایسه قرار می‌گیرند. قسمت‌هایی از متن که از نظر سبک نگارش با سایر قسمت‌ها، ناهماهنگی دارند به عنوان قسمت‌های مشکوک و سرقت علمی تشخیص داده می‌شوند. در این روش سبک نگارش فرد در یک متن با متن مقالات قبلی وی مورد مقایسه قرار می‌گیرد. شکل ۶ در زیر نشان دهنده یک مورد ساده در این روش می‌باشد.

“Our goal is to identify files that came from the same source or contain parts that came from the same source. We say that two files are similar if they contain a significant number of common substrings that are not too small. We would like to find enough common substrings to rule out chance, without requiring too many so that we can detect similarity even if significant parts of the files are different. However, my interest in plagiarism lies within academic institutions, so the document domain will be local research articles. The limited scope of domain will make it easier to determine if it is same source or not.”

شکل ۶- نمونه‌ای از سرقت علمی با تشخیص سبک، (مورر، کاپیه و زاکا، ۲۰۰۶)

1. Feature Selection
2. Stylometry

اگر با دقت به متن شکل ۶ بنگرید متوجه تناقضات زیر خواهید شد:

۱. تغییر در ضمائر از **our/we** به **my**

۲. نویسنده در خطوط اول متن از حرف اضافه **the** قبل از کلمات **same source** استفاده می‌کند اما در موارد بعدی حرف **the** را حذف کرده است.

کلماتی که با حروف پررنگ نمایش داده شده اند نشان دهنده تناقض و احتمال وجود سرقت علمی هستند. در نرم‌افزار تشخیص سرقت علمی **Glatt** از یک آزمون^۱ تشخیص سبک نگارش استفاده می‌شود که فرض آن بر این است که هر شخصی خود بهتر می‌تواند کلمات مورد استفاده در نوشتار خود را یادآوری کند. قبل از پذیرش نهایی مقاله، این آزمون از نویسنده گرفته می‌شود تا از عدم وجود هر گونه سرقت علمی اطمینان یابند. بعد از انتخاب یک فراز از متن مشکوک، پنجمین کلمه هر سطر حذف می‌شود و سپس از نویسنده خواسته می‌شود تا جاهای خالی را پر کند (شکل ۷). برنامه با محاسبه تعداد جواب‌های درست و همچنین زمان پاسخگویی، امتیاز کسب شده توسط نویسنده را محاسبه می‌کند و به نمایش در می‌آورد (مورر، کاپیه و زاکا، ۲۰۰۶). امتیازی که به سرعت در پاسخگویی تعلق می‌گیرد موجب می‌شود نویسنده زمان کافی برای مراجعه به متن اصلی به قصد فریفتن نرم‌افزار را نداشته باشد.

Your job is to fill in the blanks with the EXACT word you think you used.

Use your cursor to move from one blank to the next blank; DO NOT USE THE TAB KEY.

Do not look at your original paper or the test results will be invalid. Each blank represents ONE word.

Type the word that you think belongs in each blank. Continue until the end of the text. Remember, you can always go back and make any changes to your answers. When you are satisfied, push the submit button.

Remember, do NOT consult your paper or the test results will be INVALID.

The proposed framework is a very effective approach to deal with information available to any individual. It provides precise and selected news and information with a very high degree of convenience due to its capabilities of natural interactions with the system. The proposed user modelling and information domain ontology offers a very useful tool for

Text:

Submit Test

Reset Form

شکل ۷- نمونه ورود اطلاعات در نرم‌افزار **Glatt** (مورر، کاپیه و زاکا، ۲۰۰۶)

1. Wilson Taylor's cloze procedure

از جمله ویژگی‌های متنی که در روش تشخیص سبک مورد استفاده قرار می‌گیرند می‌توان به موارد زیر اشاره کرد:

- آمارهای متنی که در سطح حروف ارائه می‌شوند.
 - بررسی نقش کلمات^۱ در متن و اندازه‌گیری میزان استفاده از هر کدام از آنها.
 - میزان استفاده از واژه‌های خاص و میانگین فراوانی انواع کلمات.
 - ویژگی‌های نحوی که سبک نگارش را در سطح جمله اندازه‌گیری می‌کنند (زو ایسن، استین، ۲۰۰۶).
- تشخیص سبک علاوه بر نشان دادن سرقت علمی در موارد حقوقی نیز کاربرد دارد. تشخیص مولف یکی از شاخه‌های زبان شناسی حقوقی یا زبان شناسی قانونی به شمار می‌رود که در آن سعی می‌شود نویسنده واقعی متن مشخص شود. (محمدیان، ۱۳۸۹، مومنی و عزیزی، ۱۳۹۱) مثلاً در مورد نامه‌های تهدید آمیزی که به صورت ناشناس برای افراد ارسال می‌شوند، تشخیص نویسنده و یا نویسندگان مظنون به حل دعاوی قانونی و حقوقی کمک شایانی می‌کند. (استاجی، ۱۳۸۷) همچنین سبک نویسندگی زنان و مردان قابل تشخیص است و زنان و مردان در استفاده از ضمائر، صفات و قیدها متفاوت عمل می‌کنند. این امر هم در مورد متون داستانی و هم غیرداستانی صادق است. (ژنگ، ۲۰۰۶، مومنی و عزیزی، ۱۳۹۱) در سال ۱۷۸۵ و پلموت^۲ نویسنده واقعی نمایش نامه‌های شکسپیر را بیکن دانست (استاجی، ۱۳۸۷).

از جمله روش‌های تشخیص سرقت علمی درونی می‌توان به مباحث مطرح در مجموعه رقابت‌های PAN اشاره کرد:

- تشخیص نگارنده متن^۳: در مورد مدرکی که نویسنده ندارد، با مقایسه مدرک مشکوک با مجموعه‌ای از مدارک با سبک‌های نگارشی مختلف، صورت می‌پذیرد و نویسنده مدرک مشخص می‌شود.
- سنجش درستی یا نادرستی متون منسوب به نویسنده‌ای خاص^۴: به این مسئله می‌پردازد که آیا می‌توان با در دست داشتن نمونه‌هایی از متون نوشته شده توسط نویسنده A، حدس زد که آیا مدرک مشکوک متعلق به نویسنده A می‌باشد یا خیر. در فرایند این مقایسه، تعلق مدرک به نویسنده تایید یا رد می‌شود (الزهرانی، سلیم، آبراهام، ۲۰۱۲).
- تعیین هویت نگارنده^۵: از طریق استخراج ویژگی‌هایی از متن صورت می‌گیرد و به تشخیص سن، جنس و ویژگی‌های اخلاقی نویسنده می‌پردازد. این مبحث در زمینه مسائل جنایی برای تشخیص ویژگی‌های مظنون و همچنین در زمینه بازاریابی هوشمند^۶ با تحلیل وبلاگ‌ها، نقدها و نظرات افراد در مورد رضایت یا

1. POS: Part-of-speech
2. J. Wilmot
3. Authorship attribution
4. Authorship verification
5. Author profiling
6. Intelligent Marketing

عدم رضایت از کالا و خدمات در وب سایتهای تجاری برای تشخیص علاقمندی‌های عموم مردم استفاده می‌شوند (استاماتانوس، پتاست، رانگل، روسو، استین، ۲۰۱۵).

هرساله رقابت برای دستیابی به برترین الگوریتمها در زمینه تشخیص سرقت علمی در سه مبحث اشاره شده برگزار می‌گردد و در نهایت الگوریتم برنده اعلام می‌شود.

تحلیل خطاهای گرامری و نحوی^۱ متون: از طریق تحلیل و پردازش این خطاها می‌توان ویژگی‌هایی برای نویسندگان مقالات انتخاب کرد. به طور مثال ممکن است نویسنده‌ای در نگارش خود زمانی که قصد استفاده از حرف اضافه at را داشته باشد، به اشتباه از حرف اضافه on استفاده می‌کند؛ یا اینکه کلمه‌ای را معمولاً غلط تایپ می‌کند، مثلاً به جای across کلمه اشتباه across را تایپ کند و این خطاها در دفعات زیادی صورت می‌گیرد. تحلیل محاسباتی و آماری این گونه خطاها اطلاعات مفیدی را در تشخیص شخص نویسنده ارائه می‌دهد.

عمدتاً تحقیقات در زمینه روش‌های درونی تشخیص سرقت علمی در زمره تحقیقات حوزه بازیابی اطلاعات هوشمند قرار می‌گیرند. مدل بولی، فضای برداری، نمایه‌سازی پنهان، الگوریتمهای یادگیری ماشینی با ناظر و بدون ناظر، بازیابی اطلاعات فازی، بازیابی اطلاعات مبتنی بر هستی‌شناسی و تکنیک‌های مبتنی بر گراف از جمله روش‌های بازیابی اطلاعات هستند که در تشخیص سرقت علمی مورد استفاده قرار می‌گیرند (راکیان، صافی اصفهانی، رستگاری، ۱۳۹۲).

روش‌های مبتنی بر قانون^۲ و مبتنی بر پیکره^۳

روش‌های مبتنی بر قانون، با بکارگیری قواعد و قوانینی مانند گرامرها و قاعده‌های if.then به تشخیص می‌پردازند. از نمونه روش‌های مبتنی بر قانون، تشخیص سرقت علمی مبتنی بر استناد می‌باشد. بررسی و ارزیابی الگوهای استنادی^۴ موجود در مقالات می‌تواند تشابهاتی را بین مدارک بازیابی و مشخص کند. ترتیب قرار گرفتن ارجاعات مقاله اگر در دو مدرک مشابه بود می‌تواند احتمال ارتکاب سرقت علمی توسط نویسنده یا نویسندگان را نشان دهد. همچنین بررسی و تحلیل ترتیب استنادات درون متنی نیز در این زمینه اطلاعات مفیدی را ارائه می‌دهد (گیپ، ۲۰۱۳: ۲۳). چنانچه ترتیب قرار گرفتن ارجاعات و یا استنادهای درون متنی دو مدرک شباهت بسیار زیادی به هم داشته باشند، احتمال ارتکاب سرقت علمی وجود دارد. نمونه‌ای از الگوی استنادی در مورد تشابهات بین ارجاعات دو مدرک در شکل ۸ نشان داده شده است. متن سمت چپ مرتکب سرقت علمی شده است و از طریق یافتن مشابهت در ترتیب استنادها این مسئله مشخص شده است. همچنین در استنادهای درون متنی نیز یافتن چنین تشابهاتی می‌تواند به مشخص شدن سرقت علمی کمک کند.

1. Error Analysis
2. Rule based
3. Corpus based
4. Citation patterns

| Plagiat Hehn 2007 S. 145-146 | Quelle Unser 1999 S. 156-158 |
|------------------------------------|---|
| Dörner (1990) | Dörner, D. (1990) |
| Dörner (1986) | Forkel, M. (1995) |
| Forkel (1995) | Hastie / Pennington (1995) |
| Hastie and Pennington (1995) | Kirsch, W. (1988) |
| Kluwe (1990) | Schwarz, N. (1982) |
| Kluwe (1995) | Dörner, D. (1986) |
| Kirsch (1988) | Kluwe, R. H. (1990) |
| Newell and Simon (1972) | Kluwe, R. H. (1995) |
| Seel (1991) | Newell, Simon (1972) |
| Schwarz (1982) | Seel, N. M. (1991) |
| Stemberg (1996a) | Sternberg, R. J. (1996a) |
| Banyard et al. (1995) | Banyard, P. (1995) |
| Simon (1979a) | Simon, H. A. (1979a) |
| Slovic et al. (1977a) | Slovic, P./Fischhoff, Tergan, S.-O. (1986) |
| Tergan (1986) | Zimbardo, P. G. (1992) |
| Zimbardo (1992) | Pervin, L. A. (1987) |
| Pervin (1987) | Oldenburger, H.-A. (1981) |
| Oldenburger (1981) | Wessels, M. G. (1984) |
| Furstenau (1994) | Dörner, D. (1987) |
| Richter (1996) | Dörner, D. (1988) |
| Tergan (1986) | Kluwe, R. (1979) |
| Wessels (1984) | Zimbardo, P. G. (1992) |
| Dörner (1987) | Furstenau, B. (1994) |
| Dörner (1988) | Richter, A. (1996) |
| Kluwe (1979) | Tergan, S.-O. (1986) |
| Zimbardo (1992) | Furstenau, B. (1994) |
| Furstenau (1994) | Dutke, S. (1994) |
| Dutke (1994) | Kluwe, R. H. (1990) |
| Kluwe (1990) | Kluwe, R. H. (1995) |
| Kluwe (1995) | Opwis, K. (1985) |
| Opwis (1985) | Reason, J. (1990) |
| Reason (1990) | Seel, N. M. (1991) |
| Seel (1991) | Simon, H. A. (1991) |
| Simon (1991) | Johnson-Laird, P. N. (1983) |
| Johnson-Laird (1983) | Johnson-Laird, P. N. (1995) |
| Johnson-Laird (1995) | Gentner, D./Stevens, A. L. (1983) |
| Gentner und Stevens (1983) | McCain, R. A. (1992) |
| McCain (1992) | |

شکل ۸- تشخیص سرقت علمی در الگوی ارجاعات (گیپ، ۲۰۱۳: ۱۳۱)

اما روش‌های مبتنی بر پیکره^۱، بدون کمک گرفتن از هرگونه قاعده و قانونی کار می‌کنند آنها روش‌های آماری هستند که از طریق بدست آوردن فراوانی‌هایی مانند رخداد کلمات به تشخیص سرقت علمی می‌پردازند. علاوه بر رخداد کلمات کلیدی و شاخص متن، در بعضی موارد حتی رخداد ایست و آژه‌ها^۲ یا کلمات بسیار عام نیز می‌تواند اطلاعات ارزشمندی در تشخیص سرقت علمی در اختیار بگذارند. روش ان-گرام ایست و آژه‌ها^۳ برخلاف روش‌های دیگر که این کلمات حذف می‌شوند، با ننگه داشتن آنها، بقیه کلمات متن حذف می‌شوند. این روش می‌تواند تشابه نحوی را شناسایی کند. استاماتاتوس^۴ (۲۰۱۱) معتقد است وقتی شخصی مرتکب سرقت علمی می‌شود، اندیشه‌های اصلی را با کلمات دیگری بیان می‌کند و معمولاً ایست و آژه‌ها ثابت می‌مانند و کمتر تغییر می‌کند. Corpus- پیکره یا دادگان به مجموعه‌ای نوشتاری یا گفتاری آوانویسی شده می‌گویند که می‌توان آن را به عنوان مبنایی برای تحلیل و توصیف زبانی به کار برد. محقق با کمک پیکره می‌تواند به بررسی‌های آوایی، نحوی، اجتماعی و دیگر زمینه‌های زبان بپردازد. (عاصی، ۱۳۸۳)

2. StopWords
3. stopword n-grams
4. Stamatatos

می‌کنند. این روش بر اساس فهرستی از ایست واژه‌ها بوجود آمده است. ان-گرام^۱ این کلمات، اطلاعات مهمی را برای ردیابی سرقت علمی آشکار می‌کنند، زیرا آنها قادرند شباهت‌های نحوی بین مدرک اصلی و مدرک مشکوک را تشخیص دهند. آنها می‌توانند برای ردیابی دقیق متونی که در آنها سرقت علمی صورت گرفته، استفاده شوند. این روش در مواردی که متن، بسیار تغییر کرده و یا بسیاری از کلمات و اصطلاحات با واژه‌های مترادف جایگزین شده اند مفید است (استاماناتوس، ۲۰۱۱). در این روش وقتی مدرکی وارد نظام اطلاعاتی می‌شود ابتدا همه کلمات و علائم به جز ایست واژه‌ها حذف می‌شوند. شکل ۹ نمونه‌ای از مشخص شدن ایست واژه‌های یک متن و شکل ۱۰ بازنمون متن پس از حذف سایر کلمات را نشان می‌دهد.

These savage birds are very common in Maine, where they make great havoc among the flocks of wild-ducks and Canada grouse, and will even, when driven by hunger, venture an attack on the fowls of the farm-yard.

شکل ۹. نمونه متن جهت تشخیص ایست واژه‌ها

are in they the of and and will by an on the of the

شکل ۱۰. باقی ماندن ایست واژه‌ها پس از حذف سایر کلمات

با مقایسه دو متن زیر در شکل ۱۱ و ۱۲ می‌توان این روش را بهتر درک کرد. همانطور که مشاهده می‌شود ان گرام ایست واژه‌های این دو متن کاملاً مشابه یکدیگر است، بنابراین متن شکل شماره ۱۱ مشکوک به سرقت علمی می‌باشد.

This came into existence likely from the deviance in the time-period of the particular billet. As the premier is to be nominated for not more than a period of four years, it can infrequently happen that an ample wage, fixed at the embarkation of that period, will not endure to be such to its end.

شکل ۱۱. متنی که سرقت علمی در آن صورت گرفته است

This probably arose from the difference in the duration of the respective offices. As the President is to be elected for no more than four years, it can rarely happen that an adequate salary, fixed at the commencement of that period, will not continue to be such to its end.

شکل ۱۲. متن اصلی که مورد دستبرد علمی واقع شده است

۱. ان-گرام: این واحد جهت تطبیق پرس و جوها با مدرک استفاده می‌شود. هدف، بدست آوردن تطبیق نادقیقی است که نسبت به اشتباه‌های املایی یا خطاهای تشخیص حروف نوری انعطاف پذیر خواهد بود. ان گرامها مستقل از زبان هستند و مدل زبانی را نشان می‌دهند.

روش‌های تشخیص بین زبانی

گسترش ترجمه ماشینی و سهولت در استفاده از آن زمینه را برای ارتکاب سرقت علمی هموار می‌سازد. تشخیص سرقت علمی بین زبانی^۱ یا سرقت علمی ترجمه^۲ از مباحثی است که در آن از الگوریتم‌های بازیابی اطلاعات هوشمند استفاده می‌شود (پتاست، بارون، استین، روسو، ۲۰۱۱). به طور کلی روش‌هایی که مستقل از زبان هستند می‌توانند سرقت علمی را در همه زبان‌ها تشخیص دهند. روش‌های بین زبانی به تشخیص سرقت علمی در زبان‌های مختلف می‌پردازند. با ساخت دیکشنری دو زبانه در زبان‌های مختلف و مترجم‌های ماشینی که قادر به تبدیل زبان مبدأ به زبان مقصد باشند، تشخیص سرقت علمی در زبانهای مختلف صورت می‌پذیرد. سوء استفاده از ساختارهای متفاوت در زبانهای غربی و شرقی زمینه‌ای است که افراد را ترغیب به ارتکاب سرقت علمی می‌کند. بطور مثال با ترجمه متنی از انگلیسی به اسپانیایی و سپس فارسی و بعد از آن ترجمه به مالایایی، سپس به فرانسه و چینی و در نهایت وقتی متن بدست آمده به انگلیسی ترجمه شود متنی متفاوت با متن اولیه بدست می‌آید که اشخاص خاصی به راحتی از آن استفاده می‌کنند (پاتل، بختیاری و تقوی، ۲۰۱۱). در تشخیص سرقت علمی بین زبانی، معمولاً به پیکره‌های هم تراز شده^۳ که در ترجمه ماشینی نیز کاربرد دارند، نیاز است. این پیکره‌ها معمولاً به دو شکل موازی^۴ و تطبیقی^۵ می‌باشند. در پیکره‌های موازی متون کلمه به کلمه از زبان مبدأ به زبان مقصد ترجمه شده‌اند، مثل قوانین اتحادیه اروپا که در آن قوانین به تمام زبان‌های رسمی اروپا عیناً ترجمه شده است. اما در پیکره‌های تطبیقی، متون دقیقاً ترجمه یکدیگر نیستند بلکه به لحاظ موضوعی با هم مرتبط هستند، مثلاً اخبار انتشار یافته در مورد وقایع مشترک توسط کشورهای مختلف با زبان‌های متفاوت (برادران هاشمی، ۱۳۸۹). مدل‌های تشخیص سرقت علمی بین زبانی بر اساس بکارگیری هر کدام از این دو نوع پیکره‌ها به مدل‌های مبتنی بر پیکره‌های موازی و یا مبتنی بر پیکره‌های تطبیقی دسته بندی می‌شوند (پتاست، بارون، استین، روسو، ۲۰۱۱).

همانطور که اشاره شد ضرورت دستیابی به نرم‌افزارهای هوشمند تشخیص سرقت علمی امروزه در جامعه علمی بسیار نمایان است. داوری دقیق مقالات علمی، ارزیابی دقیق تکالیف دانشجویی، سنجش اصالت پایان نامه‌های تحصیلات تکمیلی از دغدغه‌های جامعه علمی امروز است که امید می‌رود با پیشرفت فناوری‌های اطلاعاتی بتوان به نتایج پربراری رسید.

تشخیص سرقت علمی در زبان فارسی

اکثریت پیشرفت‌های صورت گرفته در زمینه ساخت نرم‌افزارهای سرقت علمی در حوزه زبان‌های غیر

1. cross-language plagiarism detection (CLPD)
2. Translation Plagiarism
3. Aligned Corpus
4. Parallel Corpora
5. Comparable Corpora

فارسی بوده است. با اینکه بسیاری از ابزارهای پردازش زبانی می‌توانند به صورت مستقل از زبان عمل کنند، اما توسعه این ابزارها با توجه به ویژگی‌ها و مختصات زبان فارسی بسیار ضروری است. در تشخیص سرقت علمی دسترسی به پیکره‌های مناسب که حاوی مقالاتی متشکل از نمونه‌های واقعی سرقت علمی باشند بسیار مثر است (استاماناتوس، ۲۰۱۱). در زمینه ساخت نرم‌افزارهای تشخیص سرقت علمی در زبان فارسی می‌توان به تلاش‌های صورت گرفته توسط مرکز تحقیقات کامپیوتری اسلامی در راه اندازی سامانه مشابه یاب روایات و مشابهت یاب مقالات یا سمیم نور اشاره کرد. پژوهشگاه علوم و فناوری اطلاعات نیز سامانه‌ای آزمایشی را به نام همانندجو راه اندازی کرده است. نگارندگان بعد از کار با این سامانه‌ها پی بردند که با وجود مثبت بودن تلاش‌های صورت گرفته، این سامانه‌ها هنوز در ابتدای راه هستند و نیاز به تحقیق و توسعه بیشتری دارند.

نتیجه‌گیری

سرقت علمی از جمله سوء رفتارهای پژوهشی است که از دیرباز مورد نگرانی جامعه علمی بوده است. عواملی همچون افزایش تعداد پژوهشگران، شعار هر چه پر تولید تر، علمی تر با ایجاد فشار بر جامعه پژوهشگران، زمینه را برای ارتکاب به این امر نكوهیده مهیاتر می‌سازد. از طرفی محمل‌های نوین اطلاعاتی با قابلیت‌های فراوانی چون تکثیر فراوان، ارسال سریع، آسان و ارزان نیز بدین امر دامن زده‌اند. از آنجا که علاج واقعه را قبل از وقوع باید کرد، لزوم تدوین برنامه مدون آموزشی در جهت پیشگیری از سرقت علمی اهمیت زیادی دارد. تولید محتواهای آموزشی مناسب، پر رنگ تر شدن نقش کتابداران و کتابخانه‌ها در آموزش، ورود مبحث پیشگیری از سرقت علمی در سرفصل دروس دانشگاهی از جمله روش‌های جلوگیری از این امر مذموم می‌باشد. تدوین برنامه مدون اعمال مجازات نیز از ضروریات امر آموزش به شمار می‌رود؛ از طرفی رشد و توسعه نرم‌افزارهای تشخیص سرقت علمی می‌تواند به کنترل رفتار افراد خاطی کمک کند. ضرورت دستیابی به نرم‌افزارهای هوشمند تشخیص سرقت علمی تا بدانجاست که کنفرانس‌های بازیابی اطلاعات نیز بدین امر پرداخته‌اند و با برگزاری رقابت‌های علمی سعی در دستیابی به الگوریتم‌های قوی تر در تشخیص سرقت علمی دارند.

References

1. Estagi Azam(1387), "Detecting Literary and Legal Texts : A Discussion in Forensic Linguistics. Language and Language Study",
2. Baradarn Hashemi(1389), "Using Adaptive Configurations for Inter-language Information Retrieval in Persian- English", Computer Engineering Master's Thesis, Tehran University.
3. Betahi Zahra(1390), "Plagiarism: Definition, Examples and Prevention Methods, Prevention and Detection", Karagah Quarterly, 2nd Period,4th Year, No.16, p. 18-6.
4. Bahadori Mohammad Karim(1391), " Plagiarism: Concepts, Factors and Solutions", Military Medicine Magazine. 14(3), p. 168-177.
5. Hori Abbas(1390), "Scientific Publications Ethics", Shiraz: Islamic World Sciences Documentation base.
6. Rakian Shima, Safi Esfahani, Faramarz, Rastegari Hamid(1392), "Investigating Outer Plagiarism Detection Methods in monolingual environment, First Iranian National Conference on Modern Approaches in Computer Engineering and Information Retrieval", Islamic Azad University, Roudsar Branch.
7. Asi Mostafa(1383), "Persian Grammar Processing with Computer", Persian Academy Letter, No.1, p.52-29.
8. Alizadeh Hamid(1395), "Familiarization with all Types of Plagiarism Detection Software (workshop). Retrieved in <https://ricest.ac.ir>.
9. Fatahi Rahmatollah(1394), " Judging Research and Scientific Essays: Approaches, Criteria and Challenges, Workshop held in Fars Informatics and Librarianship Society", Regional Center for Informatics of Science and Technology, Azar 26th, 1394.
10. Falahati Fomani Mohammad Reza(1394), "Writing List of References in software-assisted APA Method (Workshop), Retrieved in <https://ricest.ac.ir>.
11. Mohammadian Amir(1389), "Detecting Personal Style of Writers based on Kiosam Analytical Method and Reviewing its Efficacy in the Persian Language", Kavosh Nameh Research-Scientific Quarterly, 11th Year, No.21, p. 112-89.
12. Mansourian Yazdan(1393), " Research Method in information and Scientology", Tehran: Samt.
13. Mousavi Doost, Sara and Fenoudi Hannaneh(1389), " Violations and Plagiarism", Nesha Elm Quarterly, 1(1), p. 21-29.
14. Momeni Negar, Azizi Cyrus(1391), "Plagiarism and the Role of Linguistic Tools

- in Identifying and Analyzing it, Karagah Quarterly, 2nd Period, 5th Year, No. 5, 69-44.
15. Alzahrani, S.M. Salim, N., Abraham, A. (2012) Understanding Plagiarism Linguistic Patterns, Textual Features, and Detection Methods. IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS. PART C: APPLICATIONS AND REVIEWS. 42(2).pp.133-149.
16. APA (2013). Research Misconduct. Retrieved November 15, 2013 from <http://www.apa.org/research/responsible/misconduct/>
17. Eisa, T. A., Salim, N., Alzahrani, S., & Gorman, G. (2015). Existing plagiarism detection techniques: A systematic mapping of the scholarly literature. *Online Information Review*, 39(3).
18. George, S., Costigan, A. and O'Hara, M. (2013). Placing the library at the heart of plagiarism prevention: The University of Bradford experience. *New Review of Academic Librarianship*. Vol. 19, No. 2, pp.141-160.
19. Gipp, Bela. (2013). Citation based plagiarism detection. Berlin: Springer
20. iThenticate (2012). True Costs of Research Misconduct. Retrieved November 15, 2013 from <http://www.ethics.elsevier.com/pdf/ithenticate-MisconductReport-2012.pdf>
21. Maurer, H. A., Kappe, F., & Zaka, B. (2006). Plagiarism-A Survey. *J. UCS*, 12(8), 1050-1084.
22. Patel, A., Bakhtiyari, K., & Taghavi, M. (2011). Evaluation of cheating detection methods in academic writings. *Library Hi Tech*, 29(4), 623-640.
23. Potthast, M., Barrón-Cedeño, A., Stein, B., & Rosso, P. (2011). Cross-language plagiarism detection. *Language Resources and Evaluation*, 45(1), 45-62.
24. Roig, M. (2006) Avoiding plagiarism, self-plagiarism, and other questionable writing practices: A guide to ethical writing. <http://www.cse.msu.edu/~alexliu/plagiarism.pdf>
25. Sindhu, L., Thomas, B. B., & Idicula, S. M. (2011). A Study of Plagiarism Detection Tools and Technologies. *International Journal of Advanced Research In Technology*, 1, 64-70.
26. Stamatatos, E. (2011). Plagiarism detection using stopword n-grams. *Journal of the American Society for Information Science and Technology*. 62(12).P.2512.
27. Stamatatos, E., Potthast, M., Rangel, F., Rosso, P., & Stein, B. (2015). Overview of the PAN/CLEF 2015 Evaluation Lab *Experimental IR Meets Multilinguality, Multimodality, and Interaction* (pp. 518-538): Springer.
28. Stein, B., Lipka, N., Prettenhofer, P. (2011). Intrinsic plagiarism analysis. *Lang Resources & Evaluation*, 45:63-82. DOI 10.1007/s10579-010-9115-y.

29. Tahir Ali, A. M. et. (2011). Overview and Comparison of Plagiarism Detection Tools in V. Snasel, J. Pokorny, K. Richta (Eds.): Dateso2011 . PP. 161-172. <http://ceur-ws.org/Vol-706/poster22.pdf>
30. Turnitin, (2014) Turnitin Effectiveness in U.S. Colleges and Universities. Retrieved from <http://go.turnitin.com/paper/effectiveness-turnitin-higher-education> at 26.9.2015
31. Youmans, R. J. (2011) Does the adoption of plagiarism-detection software in higher education reduce plagiarism? *Studies in Higher Education*, 36(7), 749-761.
32. Zheng, R., Li, J., Chen, H., & Huang, Z. (2006). A framework for authorship identification of onlinemessages: Writing-style features and classification techniques. *Journal of the American Society for Information Science and Technology*, 57(3), 378–393. DOI: 10.1002/asi.20316.
33. Zu Eissen, S. M., & Stein, B. (2006). Intrinsic plagiarism detection. In *Advances in Information Retrieval* (pp. 565-569). Springer Berlin Heidelberg.
34. http://turnitin.com/en_us/features/faqs
35. <http://forum.wordreference.com/showthread.php?t=1462123>
36. PAN website: <http://www.uni-weimar.de/medien/webis/events/pan-15/pan15-web/>