استخدام أدوات اكتشاف السرقات العلمية ببحوث المكتبات والمعلومات: دراسة تحليلية للتخطيط لأداة اكتشاف النص العربي

د. وليد محمد هيكل
مدرس بقسم المكتبات والمعلومات
كلية الآداب – جامعة حلوان

مستخلص :

تعد السرقة العلمية خيانة للأمانة الأكاديمية وأخلاقيات البحث العلمي؛ لذا يتناول هذا البحث ما يتعلق بالانتحال وأشكاله المتعددة، وكذلك الأوجه المختلفة للكشف عن السرقات العلمية باستخدام الأدوات التكنولوجية، ومدى استعانة لجان الفحص الرسمية في مجال المكتبات والمعلومات بها. كذلك يتناول البحث التحديات التي تواجه هذه الأدوات بغرض إرساء خطوات فعلية في إنشاء أداة لكشف السرقات العلمية لبحوث المكتبات والمعلومات باللغة العربية.

الكلمات المفتاحية :

اكتشاف السرقات، الانتحال، الغش، السرقة العلمية، السرقة الأدبية، بحوث المكتبات والمعلومات.

أ - تمهيد :

يعتبر الانتحال ظاهرة ليست بالحديثة، حتى إن "نيوتن" "Newton"، و"ليبنيز" "Leibniz" قد شُك في أصالة ابتكاراتهما، بل اُتهم "شكسبير" "Shakespeare" بأنه نسخ من "مارلو" "Marlowe" (Heather, 2010, p.647). وفي السنوات الأخيرة اجتذب موضوع الانتحال تغطية إعلامية كبيرة؛ نتيجة للنمو المطّرد في الانتحال بالمؤسسات الأكاديمية. بالإضافة إلى ذلك فقد شهد الاتجاه العام وصف الانتحال بأنه "وباء" كعَرَض وسبب؛ نتيجة التراجع في المعايير الأكاديمية، فضلاً عن تقديم الكثير من الأدلة التي تدين الإنترنت كعامل مؤثر سلباً على المجتمع. وعلى الجانب الآخر ،كان للإنترنت أثر في تسهيل تنفيذ الانتحال والسرقات العلمية، كما أنه سهَّل الكشف عن هذا بوسائل مختلفة. ونتيجة لتاريخ هذه الظاهرة برزت البحوث المبكرة، التي اعتمدت

في كشف الانتحال على برامج الحاسب؛ حيث برزت في الآونة الأخيرة الأدوات التكنولوجية المتطورة التي تُنفذ تلقائياً عملية الكشف عن الانتحال مستندة إلى النص (McKeever, 2006, p.155,156). يعني هذا أن التقنية سلاح ذو حدين، قد سهلت بالفعل حياة الإنسان، لكن البعض أساء استخدامها؛ مما أسهم في عولمة الجريمة المعلوماتية، الأمر الذي أثر سلباً على أخلاقيات مجتمع المعلومات، حيث أصبحت تلك الأخلاقيات تُنتهك من قبل المتعاملين مع المعلومات. وبرغم أهمية قضية السرقات العلمية وحيويتها؛ بل وخطورتها لم تنل ما تستحقه من البحث والدراسة (السالم، ٢٠١٠، ص٢).

وقد تم تقسيم هذا البحث إلى: الشِّق المنهجي لهذه الدراسة، ثم الجانب النظري المتعلق بتحديد معالم الانتحال والسرقات العلمية من خلال أنواعه وأشكاله والأدوات المستخدمة في هذا المجال، فضلاً عن التقنيات والخوارزميات المطبقة في هذه الأدوات، والمراحل التي تمر بها الوثيقة لتحديد أصالة ودقة الأوراق والبحوث، ومن ثمَّ تناول المزايا والعيوب والتحديات التي تواجه أدوات اكتشاف الانتحال والسرقات العلمية. وتناول الشِّق الأخير من الدراسة الإطار التحليلي المتمثل في توجهات لجان الحكم والفحص المتخصصة في مجال المكتبات، فضلاً عن تحليل الأدوات والخدمات في كشف الانتحال، وقد تم وضع مواصفات أو سمات أساسية في تصميم أداة للكشف عن الانتحال والسرقات العلمية بالنص بالعربي في بحوث المكتبات والمعلومات وغيرها من البحوث.

ب- مشكلة الدراسة:

من خلال استكشاف الأدوات المتاحة على الإنترنت، لم تلاحَظ أي أداة لكشف النص العربي. وكشفت إحدى الدراسات أن معظم الأدوات باللغة الإنجليزية، فضلاً عن تكييف بعضها للتعامل مع اللغة الصينية، ولا توجد أداة لاكتشاف السرقة في النص العربي؛ لمكافحة السرقات العلمية للبحوث بالنص العربي (Alzahrani, Salim & Alsofyani, 2009). مما ينتج عنه ضعف الاعتماد على مثل هذه الأدوات في اكتشاف السرقات العلمية في البحوث والنصوص باللغة العربية، فضلاً عن أن المشكلة الحقيقة تتمثل في أن عدم مواجهة ومكافحة هذه الظاهرة ومنع اكتشاف الانتحال يهدم الإبداع الإنساني ولا يحفظ للمبدعين حقوقهم في البيئة الرقمية.

جـ- أهمية الدراسة:

يُعد الانتحال والسرقة العلمية جريمة في حقوق الملكية الفكرية؛ حيث تُعد بمثابة جريمة جنائية بسبب خطورتها على الأمن الفكري والاقتصاد الوطني. وقد ازدادت السرقات العلمية مع تزايد استخدام الإنترنت؛ لذا وجب التصدي لهذه الظاهرة بالاستعانة بالتقنيات الحديثة المتمثلة في البرامج ومحركات البحث ومواقع الويب المتخصصة في الكشف عن الانتحال والسرقات العلمية. وتتضح أهمية الدراسة أكثر في الاهتمام بالوسائل التي تتبعها لجان التحكيم العلمية بالمؤسسات الأكاديمية والدوريات العربية المحكمة في مجال المكتبات والمعلومات، وذلك بسبب عدم

توافر مواقع وأدوات عربية متخصصة في اكتشاف وتتبع السرقات العلمية، على العكس تماماً في اللغات الأجنبية. ومن ثم الاهتمام بدراسة سمات الأدوات المتاحة في كشف الانتحال والسرقات العلمية بغرض وضع أسس لأداة كشف النص باللغة العربية.

د - تساؤلات الدراسة:

١- ما هو المقصود بالانتحال والسرقات العلمية؟

٢- ما هي الأساليب المتبعة في الأدوات التكنولوجية لتتبع السرقات العلمية؟

٣- ما مدى استخدام أدوات اكتشاف الانتحال والسرقات العلمية في تحكيم بحوث المكتبات والمعلومات؟

٤- ما هي الخصائص الوظيفية والتقنية لأدوات اكتشاف الانتحال والسرقات العلمية؟

هـ- أهداف الدراسة:

١- التعرف على اتجاهات المسئولين عن نشر وتحكيم البحوث في المكتبات والمعلومات في اكتشاف السرقات العلمية.

٢- تحديد الجوانب الوظيفية والتقنية لأدوات اكتشاف الانتحال والسرقات العلمية.

٣- الإسهام في وضع أسس أداة عربية لكشف الانتحال، للحد من هذه الظاهرة.

و - مجالات الدراسة وحدودها:

الحدود الموضوعية: تناولت الدراسة ظاهرة السرقات العلمية والانتحال وكيفية مواجهتها بواسطة الأدوات التكنولوجية الناتجة عن التطورات التي حدثت في مجال الحاسب الآلي.

الحدود الزمنية: تغطي الدراسة أدوات اكتشاف الانتحال والسرقات العلمية منذ بداية ظهورها وحتى عام ٢٠١٤.

الحدود اللغوية: تم دراسة الأدوات المتاحة باللغة الإنجليزية واللغة العربية إذا توافرت.

ز- مصطلحات الدراسة:

القرصنة (سرقة المؤلفات) Piracy:

هي "نشر عمل أدبي بدون الحصول على تصريح من المؤلف أو دفع أية مبالغ له، وهو ما يُعد مخالفاً لقانون الملكية الفكرية". (الشامي، ٢٠١١).

القرصنة الإلكترونية Electronic Piracy:

تعني "الاستخدام غير المشروع لمحتويات الإنترنت من برامج ومعلومات وغيرها، وذلك من خلال انتهاك حقوق النشر والتأليف" (السالم، ٢٠١٠، ص٥).

السرقة الأدبية literary theft:

يُقصد بها "السطو على أفكار الآخرين المنشورة من بحوث ومقالات ودراسات ذات قيمة علمية، بما في ذلك الانتحال Plagiarism، والغش Cheating، والقرصنة، والسطو على المادة العلمية بمختلف اتجاهاتها الأدبية والفكرية"(السالم، ٢٠١٠، ص٥).

الانتحال (سرقة مؤلفات الغير) Plagiarism:

ورد في لسان العرب: "النِّحْلَةُ: الدَّعْوَى.

وانْتَحَل فلانٌ شِعْرَ فلانٍ. أَو قالَ فلانٍ إِذا ادّعاه أَنه قائلُه.

وتَنَحَّلَه: ادَّعاه وهو لغيره.

ونَحَلْتُه القولَ أَنْحَلُه نَحْلاً، بالفتح: إِذا أَضَفْت إِليه قولاً قاله غيره وادّعيتَه عليه.

وفلان يَنْتَحِلُ مذهبَ كذا وقبيلةَ كذا إِذا انتسب إِليه". (ابن منظور)

ويرجح قاموس أكسفورد (Oxford) (٢٠١٣) "أن بداية استخدام هذا المصطلح كانت في أوائل القرن ١٧؛ حيث يعود أصل الكلمة إلى الكلمة اللاتينية (plagiarius)، والتي هي بمعنى (الخاطف "kidnapper") أي المدعي كذباً بالتأليف على أنه تأليف شخصي. ويتفق قاموس المكتبات والمعلومات المباشر (Online Dictionary for Library and Information Science ODLIS) مع ذلك؛ حيث يعتبر الخاطف هو الذي ينسخ أو يزيف عن كثب أعمال المؤلفين، دون الحصول على إذن مع العزم على تمرير النتائج وكأنها العمل الأصلي. وفي مجال النشر وقوانين حقوق التأليف والنشر تُجَرَّم السرقة الأدبية كجريمة جنائية. ويعتبر الانتحال في كثير من الكليات والجامعات قضية أخلاقية، ويفرض أعضاء هيئة التدريس عقوبات على الطلاب الذين ينخرطون في ذلك. ويمكن تجنب الانتحال من خلال التعبير عن الفكرة أو المفهوم بتسجيلها بكلمات واضحة من المرء نفسه. ولابد من توثيق المصادر في الحواشي أو في نهاية العمل، ونفس الطريقة عند الاقتباس المباشر" (Reitz, 2007).

ونتيجة ازدياد استخدام الإنترنت "فقد سهُل الاستيلاء على الأفكار أو التعبيرات من الآخرين، وهو ما أطلق عليه (الانتحال السيبراني) (Cybercheating). وذلك أن القدرة على القص واللصق من معظم برامج معالجة النصوص وبرامج استعراض صفحات الويب سهلت كثيراً من الانتحال الرقمي. ويعد تقديم مقالة أو ورقة عمل تم شراؤها من الإنترنت وإعادة كتابتها مرة أخرى من أكثر أشكال الانتحال الصارخة. وهناك الكثير من أعضاء هيئة التدريس المخضرمين يواكبون استخدام أحدث التقنيات للكشف عن عمليات الانتحال وأشكال الغش المختلفة المشابهة لذلك" (Reitz, 2007).

وهناك الكثير من المصطلحات المرادفة للانتحال، والتي تستخدم في المجال العلمي لمعالجة الانتحال، مثل: السرقة stealing OR theft، والغش cheating، وسوء السلوك misconduct، والتضليل dishonest (Pardeshi, Rajput, Chaudhary & Patil, 2012, p.29).

برمجيات اكتشاف الانتحال
plagiarism detection software

هو "برنامج تم تصميمه للتحقق من الأوراق البحثية والمقالات الطلابية بقواعد البيانات؛ لغرض تحديد المحتوى غير الأصلي"(Reitz, 2007).

ح - منهج الدراسة وأدواتها:

في ضوء ما حدده البحث من أهداف، فقد اعتمد على المنهج الوصفي من أجل تحليل أدوات كشف الانتحال والسرقات العلمية لتحديد سماتها ووظائفها، فضلاً عن تحديد نهج دوريات المكتبات والمعلومات المحكمة ولجنة فحص الإنتاج العلمي لشغل وظائف الأساتذة والأساتذة المساعدين في الكشف عن السرقات العلمية. وقد تم الاستناد إلى التحليل الوثائقي بغرض استكمال ما قام به السابقون من جهود تجريبية على بعض الأدوات التقنية لتحليل واكتشاف النص العربي.

ط/١. أدوات جمع البيانات:

يتعين على هذا البحث بعد تحديد الأهداف الموضوعة، وتحديد منهج الدراسة، تحديد أدوات جمع البيانات، التي نبينها على النحو الآتي:

الاستبيان(*): يتوجه هذا الاستبيان إلى رؤساء تحرير دوريات المكتبات والمعلومات المحكمة (المتاحة على الإنترنت أو المطبوعة)، وإلى أعضاء لجنة فحص الانتاج العلمي لشغل وظائف الأساتذة والأساتذة المساعدين؛ للتعرف على الأساليب والطرق المتبعة في الكشف عن الانتحال والسرقات العلمية.

قائمة المراجعة(**): تستخدم قائمة المراجعة للكشف عن سمات أدوات كشف الانتحال والسرقات العلمية للتخطيط لأداة عربية.

ط/٢. مجتمع الدراسة:

يتوزع مجتمع الدراسة بين الأدوات المختلفة للكشف عن الانتحال والسرقات العلمية، وكذلك ما تتبعه هيئات التحكيم في دوريات المكتبات والمعلومات المحكمة، وأعضاء اللجنة العلمية لتحكيم البحوث الأكاديمية لترقية الباحثين(***)؛ لذا سيتم أخذ كل مجتمع الدراسة لصغر حجمه، ولكن سيتعين تحديد عينة أدوات اكتشاف السرقات العلمية باستخدام أسلوب العينة العمدية. وفيما يتعلق باللجنة العلمية، كانت استجابة أعضاء اللجنة للاستبيان بواقع ستة أعضاء من أصل تسعة أعضاء، أي بنسبة ٦٦,٧%.

(*) قد تم تحكيم الاستبيان من قبل الأساتذة المحكمين:

١- أ.د. أسامة السيد محمود أستاذ المكتبات والمعلومات بكلية الآداب- جامعة القاهرة.

٢- أ.د. شريف كامل شاهين أستاذ ورئيس قسم المكتبات والمعلومات بكلية الآداب- جامعة القاهرة.

٣- أ.د. إيناس حسين صادق أستاذ المكتبات والمعلومات المساعد بقسم المكتبات والمعلومات بكلية الآداب- جامعة حلوان.

(**) قد تم تحكيم قائمة المراجعة من قبل الأساتذة المحكمين:

١- أ.د. أسامة السيد محمود أستاذ المكتبات والمعلومات بكلية الآداب- جامعة القاهرة.

٢- أ.د. زين الدين عبدالهادي أستاذ ورئيس قسم المكتبات والمعلومات بكلية الآداب- جامعة حلوان.

(***) اللجنة العلمية لفحص الإنتاج العلمي لشغل وظائف الأساتذة والأساتذة المساعدين، الدورة الحادية عشر (يناير ٢٠١٣- ديسمبر ٢٠١٥)، تخصص المكتبات والمعلومات، الآداب، المجلس الأعلى للجامعات، وزارة التعليم العالي، جمهورية مصر العربية.

قائمة بالدوريات العربية المحكمة:

تصدر في مجال المكتبات والمعلومات ما يقرب من ١٢ دورية محكمة على المستوى المحلي والإقليمي، بينما تتكون القائمة التالية (جدول رقم ١) من تسعة عناوين؛ ومن ثمَّ تمكن الباحث من الحصول على وسائل الاتصال بها لإرسال الاستبيان، وهذا قد تم إرسال الاستبيان لهذه الجهات ثلاث مرات على مدار ثلاثة أشهر.

جدول رقم (١)

عينة الدوريات المحكمة في تخصص المكتبات والمعلومات

م	عنوان الدورية	التوزيع الجغرافي	الاستجابة
١	مجلة المكتبات والمعلومات العربية	إقليمي	✓
٢	العربية ٣٠٠٠	إقليمي	×
٣	اعلم	إقليمي	×
٤	الاتجاهات الحديثة في المكتبات والمعلومات	مصر	✓
٥	دراسات عربية في المكتبات وعلم المعلومات	مصر	✓
٦	بحوث في علم المكتبات والمعلومات	مصر	✓
٧	مجلة مكتبة الملك فهد الوطنية	السعودية	×
٨	دراسات المعلومات	السعودية	×
٩	مجلة جمعية المكتبات والمعلومات المصرية	مصر	✓
الإجمالي		٥ = ٥٥,٦%	

عينة أدوات اكتشاف السرقات العلمية:

الشروط والمتغيرات:

- الأدوات التي تعمل في بيئة تشغيل ويندوز والأخرى المتاحة على الإنترنت.
- أن تتوافر الواجهة باللغة الإنجليزية.
- عدم تكرار نفس التقنية المستخدمة في بعض الخدمات المباشرة على الإنترنت.
- إمكانية استخدام الخدمات المقدمة بمقابل من خلال نسخة تجريبية.
- أن تكون الاستجابة الأولية للكشف عن سرقات هي بالفعل نتيجة صحيحة.
- أن تتيح خدمة الكشف عن النصوص وليس الكشف فقط عن محتوى مواقع الويب من خلال "معرف الموارد المحدد" URL.

ونتيجة لما سبق، فقد تم حصر عدد ٣٣ خدمة وعدد ٩ برامج، وبعد تطبيق الشروط وصل عدد الأدوات إلى ٢٤ أداة، مقسمين إلى: ١٧ خدمة على الإنترنت بنسبة ٥١,٥%، وعدد ٧ برامج بنسبة ٧٧,٨%، أي بنسبة كلية تساوي ٥٧,١%. وقد تم حصر أدوات الإصدار الكامل أو الإصدار التجريبي وتوزيعها على فئتين:

الأولى: البرامج المستخدمة في بيئة تشغيل الويندوز، وهي كالآتي في الجدول رقم (٢).

جدول رقم (٢)

الأدوات المتاحة في بيئة تشغيل ويندوز

م	اسم الأداة	الشركة	الترخيص
١	Plagiarism-Finder	Mediaphor Software Entertainment AG	تجريبي
٢	AntiPlagiarist	ACNP Software	تجريبي
٣	Plagiarism Checker X	Plagiarism Checker X, LLC.	تجريبي
٤	AntiPlagiarism.Net	Ined-Trade.Ltd	تجريبي
٥	Plagiarism Detector	SkyLine LLC	تجريبي
٦	Desktop Plagiarism Checker	plagiarisma.net	تجريبي
٧	Viper Plagiarism Scanner	All Answers Limited	تجريبي

الثانية: الخدمات المقدمة على الإنترنت، كما في الجدول رقم (٣).

جدول رقم (٣)

الأدوات المتاحة في بيئة تشغيل ويندوز

م	اسم الخدمة	الشركة	الترخيص
١	Academic Plagiarism	Academicplagiarism Inc	تجريبي
٢	Advanced Plagiarism Checker	Searchengine Reports	مجاني
٣	ChimpSky	Michael Palmer	مجاني
٤	Plagiarism Checker.com	PlagiarismChecker.com	مجاني
٥	Doc Cop	Mark McCrohon	تجريبي
٦	Write Check	iParadigms, LLC	تجريبي
٧	Glatt Plagiarism Self-Detection Program	Glatt Plagiarism Services, Inc	تجريبي
٨	PlagScan	PlagScan	تجريبي
٩	Free plagiarism Finder	plagiarismFinder.com	مجاني
١٠	Plagiarisma.Net	Plagiarisma.Net	مجاني
١١	checkforplagiarism.net	Academic Paradigms, LLC	تجريبي
١٢	PlagTreacker	plagtreacker.com	تجريبي
١٣	PD PplagiarismDetect	PlagiarismDetect.com	تجريبي
١٤	PaperRater	PaperRater.com	مجاني
١٥	Small SEO Tools	SmallSEOTools.com	مجاني
١٦	Plagium	Septet Systems Inc.	تجريبي
١٧	PlagiarismCheck.org	PlagiarismCheck.org	مجاني

والجدير بالذكر، أنه قد تم التركيز في هذا البحث على الأدوات فقط، وتم استبعاد المواقع التي تنشر معلومات عن البحوث التي تم انتحالها، مثل:

- موقع نادي لصوص الكلمة: موقع عربي يرصد السرقة العلمية.
- ملتقى الأدباء والمبدعين العرب: تم تخصيص أحد أقسامه للسرقات الأدبية.

ط - الدراسات السابقة:

من أبرز الدراسات والبحوث التي تناولت موضوع أدوات كشف الانتحال والسرقات العلمية:

دراسة "ماكيفر" "McKeever" (٢٠٠٦)، وقد قامت على خدمات كشف الانتحال على الإنترنت، من حيث التقنيات المستخدمة والأساليب المتبعة في كشف الانتحال والسرقات العلمية، وكذلك مميزات وعيوب هذه الخدمات.

بينما قام كلٌّ من "الزهراني وآخ" "Alzahrani et al." (٢٠٠٩) بإعداد دراسة هدفت إلى اقتراح أداة فعالة لاكتشاف السرقة في النص العربي يتم دمجها مع نظم التعليم الإلكتروني لتقييم واجبات وأبحاث الطلبة، لكن هذه الدراسة لم تقدم التصور الكافي لأداة تكشف الانتحال في النص العربي.

كما حدد "الجوة، بلغيث وحمادو" (٢٠١١) طريقة للكشف الآلي عن عمليات الانتحال في النصوص المكتوبة باللغة العربية، وتعتمد هذه الطريقة على استخراج الكلمات الهامة في النص المراد كشفه؛ ومن ثمَّ تكوين سلاسل لغوية حسب تواترها في النص، ثم في مرحلة ثانية يتم استخراج النصوص التي تندرج في نفس إطار البحث الأولي باستعمال محرك البحث Google، فضلاً عن القيام باستعمال الترجمة الآلية للسلاسل اللغوية بغرض استخراج السلاسل (مجموعة من الحروف المترابطة) المرادفة باللغة الإنجليزية، ومن ثم استعمال محركات البحث لاستخراج النصوص المدوَّنة باللغة الإنجليزية والتي تتطابق مع السلاسل بعد ترجمتها باستخدام برنامج Babylon translator.

وفي دراسة أخرى أجرى كلٌّ من (Jadalla & Elnagar, 2012) تجربة أداة للكشف على النص العربي بالاعتماد على محرك بحث، وقد أطلق على هذه الأداة "اقتباس" "Iqtebas1.0"، وتعتمد على عنصرين: التكشيف لمجموعة من الوثائق الأصلية، والبحث الذي يحتسب القيمة الأصيلة للوثائق المشتبه فيها، وقد تم فيها عرض الجوانب التقنية فقط للتعامل مع اللغة العربية.

كذلك هناك دراسة أولية تجريبية لكشف الانتحال في الوثائق العربية، من خلال إجراء مجموعة من التجارب للتعرف عن قرب على تأثير بعض الملامح الأسلوبية الشائعة على اللغة العربية. ولم تركز الدراسة على كشف التشابه بين الجمل المسروقة، ولكن الكشف كان عن الجمل غير

الطبيعية فيما يتعلق بأسلوب الكتابة العامة للوثيقة (متوسط طول : الكلمة، والجملة). وتمت التجارب على وثائق باللغة العربية تم نقلها حرفياً (النقل الصوتي للحروف العربية transliterate) إلى الإنجليزية؛ لأن الأداة Stylysis المستخدمة في الكشف عن الانتحال لا تدعم اللغة العربية (Rosso, Bensalem & Chikhi, 2012).

وقد حاولت دراسات أخرى تقديم أداة لكشف الانتحال للمقارنة بين الوثائق باللغة العربية لتحديد أوجه التشابه المحتملة، وتستند الأداة على خوارزمية مقارنة تستخدم طرقاً استدلالية لمقارنة الوثائق المشتبه بها في مستويات هرمية مختلفة لتجنب المقارنات التي لا داعي لها، وتم تطبيقها على وثائق باللغة العربية لإظهار قدرتها على تحديد النسخ المباشر، وإعادة ترتيب الجمل واستبدال المترادفات، وأبرزت أهم الأساليب المتبعة لتحليل المحتوى، وركزت على أحد المداخل الإحصائية والمعروف باسم Stylometry، والذي يعتمد على تحليل أسلوب الكتابة باستخدام عوامل من نفس الوثيقة، أو المقارنة بين وثيقتين لنفس المؤلف (Menai, 2012; Menai & Bagais, 2011, August).

وفي سياق آخر، اتجه كلٌّ من "عماد عيسى" و "أماني السيد" (٢٠١٢) إلى دراسة دور المكتبات الأكاديمية في اكتشاف الانتحال لما لها من دور فعال في دعم استراتيجيات الجامعات للحد من السرقات العلمية؛ حيث يتمثل دورها في الحد من الانتحال من خلال زيادة الوعي بمفهوم الانتحال وآثاره السلبية وأساليب تجنبه؛ ودورها أيضاً في كشف الانتحال من خلال الاشتراك في خدمات كشف الانتحال وفي خدمات برمجيات الاستشهادات المرجعية، هذا بجانب استعراض بعض الأدوات المستخدمة في كشف الانتحال.

١- ما هو الانتحال:

الانتحال هو عمل من أعمال نسخ ما كُتب من قبل شخص آخر، سواء كان ذلك مجرد نص أو محادثة أو عمل فني أو حتى فكرة، ويتم ذلك بأي وسيلة تقليدية أو إلكترونية. وعلى كل حال، يعتبر الانتحال قضية غير أخلاقية، وعدم النظر لها بعين الاعتبار سيشجع المزيد من الانتحال؛ مما سيؤدي إلى تدني مستوى جودة البحوث والمنشورات العلمية. ولذلك يُعد التغاضي عن الانتحال أيضاً سلوكاً غير أخلاقي؛ مما يشجع على انتهاج هذا السلوك من قبل المنتحلين. وإذا سمح الأكاديميون لأنفسهم بالانتحال فكيف يمكن معاقبة الطلاب على نفس الفعل، لذلك تعد هذه الظاهرة قضية أخلاقية خطيرة وبحاجة إلى تحجيم (Shahabuddin, 2009, p.355). ومن ثمَّ فإن الانتحال والسرقات العلمية تُعد انتهاكاً صريحاً لحقوق الإنسان ويجب أن تعتبر كأحد قضايا حقوق الإنسان المعاصرة.

ويمكن تقسيم الانتحال إلى فئتين:

١- انتحال النصوص من الوثائق: وتنحصر معظم الأعمال في سرقة الوثائق ذات الأغراض الأكاديمية. وتعتبر مهمة اكتشاف الانتحال ذات أهمية للطلاب وخصوصاً لطلاب الدراسات العليا.

٢- انتحال التعليمات (الأكواد) البرمجية: لقد تم وضع نهج مختلف لكشف سرقة التعليمات البرمجية، وصممت للمقارنة بين أكواد لغات البرمجة المختلفة، والتي تستخدم في الغالب تقنية الترميز (قطع دلالية) tokenization وسلسلة من الخوارزميات algorithm لمطابقة وقياس درجة التشابه.

وجدير بالذكر، أنه من السهل نسبياً اكتشاف التعليمات البرمجية دون شك أو تداخل مع تعليمات أخرى، ولكن يصعب في كثير من الأحوال اكتشاف النص باللغة الطبيعية (Menai & Bagais, 2011, August, p.1379).

٢ - أنواع وأشكال السرقة العلمية :

يُعد الانتحال مشكلة عالمية تحدث في كثير من المجالات في حياتنا. وهناك العديد من أشكال الانتحال المختلفة، حيث يُعد الانتحال في المدارس والكليات من أبرز الأشكال وله جوانب سلبية كثيرة، ويعتبر عاملاً مثبطاً لهمم المعلمين وأيضاً للطلاب والباحثين. ومن أشكال الانتحال:

أ – الانتحال من الآخرين:

(١) انتحال الأفكار وإعادة استخدامها: يعني إدراج الأفكار أو الطرق من أعمال مؤلفين آخرين دون الاعتراف أو الإشارة إلى المؤلف الأصلي (Masic, 2012, p.111).

(٢) الإخفاق في وضع النص المقتبس بين علامات الاقتباس.

(٣) نسخ الكثير من الكلمات من مصدر ما ويشكل ذلك الغالبية العظمى من العمل، سواء أعطى المؤلف الثاني إشارة مرجعية أم لا (Stowers & Hummel, 2011, p.165).

(٤) اقتباس النص فيما يزيد عن الحد المتعارف عليه.

(٥) ويتمثل أكثر الأنواع شيوعاً في نسخ الكثير من الوثيقة أو بعض أجزاء منها مع إعادة هيكلة الجملة (أي تغيير الجملة من مجهول إلى معلوم)، أو إعادة صياغة المحتوى نفسه بكلمات مختلفة (مترادفات، ومتضادات)، وذلك بالاعتماد على أفكار الآخرين.

(٦) الإشارة إلى العمل بشكل غير صحيح أو ذكر مصدر وهمي غير موجود في الأساس كالإحالة إلى مصادر الويب التي تم إزالتها.

(٧) انتحال الأعمال التي يتم ترجمتها؛ حيث يتم ترجمة المحتوى واستخدامه دون الإشارة إلى العمل الأصلي، حتى ولو كان ذلك بنسب قليلة من النص الأصلي.

(٨) انتحال عمل شخص آخر وتقديمه في وسائل مختلفة ووسائط متعددة (والنتيجة النهائية يمكن أن تتضمن نصوصاً، أو صوراً أو فيديو) دون الذكر المناسب للعمل الأصلي (Alzahrani, et al., 2009, p.4).

(٩) شراء الأعمال البحثية من إحدى المواقع التجارية، وتقديمها على أنها من عمل الباحث أو الطالب.

(١٠) انتحال قطعة من عمل جماعي على أنه جهد فردي (Atkinson, & Yeoh, 2008, p.225).

(١١) الانتحال من المصادر الثانوية في أحد الأعمال دون الرجوع إلى المصدر الأصلي، ويتم الحصول على هذه المعلومات وتسجيلها بالبحث دون الإشارة إلى أنها منقولة عن شخص ما (Clough, 2003, p.2).

(١٢) قيام الباحثين الكبار بطمس حق وجهود الباحثين الذين يعملون معهم دون إثبات هذه الحقوق.

ب – الانتحال الشخصي:

١- الانتحال الذاتي self-plagiarism، ويحدث عند إعادة استخدام المحتوى سواء كان جزئياً أو كلياً دون الإشارة إلى أنه تم نشره من قبل. ويعتبر التداخل الكبير بين البحوث المنشورة هو السمة الرئيسة في جميع أشكال الانتحال الذاتي، والأهم هو عدم وجود إشارة توضح العلاقة بين التكرارات المختلفة والبحوث ذات الصلة (Roig, 2010, p.297). بمعنى إعادة تدوير النص من عدة أعمال مع إضافة بعض التحليلات.

٢- النشر المتكرر duplicate publishing: يعني إعادة نشر نفس الأبحاث في دوريات مختلفة أو في محافل علمية مختلفة (Shahabuddin, 2009, p.353). أو التداخل مع أعمال أخرى من حيث الفرضيات والنتائج دون الاعتراف بما سبقها. ومن الصور الأخرى أيضاً تقسيم العمل الأكبر المنشور إلى أجزاء صغيرة وإعادة نشرها مستقلة (Hoseinpourfard, Bahadori, Izadi & 2012, p.170)، ويطلق عليه النشر المجزأ segmented publication. وهناك شكل آخر وهو النشر الزائد عن الحاجة redundant publication، ويعد الشكل الأكثر شيوعاً هو النشر المتكرر.

بينما تنقسم أنواع الانتحال إلى نوعين رئيسين، هما على النحو الآتي:

أ- الانتحال المتعمد/المقصود: هو ذلك النوع الذي يتم فيه استخدام أعمال كاتب آخر دون أن يشير الناقل إلى المصدر ويعكس للآخرين كما لو كان هذا النص من بنات أفكاره، وذلك عمداً لمصلحته. ويعد هذا النوع أسوء جريمة في مجال البحث العلمي.

ب- الانتحال غير المتعمد: هو نسخ عَرَضي، ويمكن ارتكابه من قبل طلاب الدراسات العليا، والباحثين، والأكاديميين.

ج- الانتحال من التأليف: يحدث عندما يدعي شخص ما أنه قام بتأليف عمل كامل، وهذا العمل في الأصل كُتِب من قبل شخص آخر، وأفضل مثال لهذا النوع عندما يقدم أحد العلماء ورقة بحثية للنشر وقد سبق نشرها لعالم آخر (Pardeshi et al, 2013, p.29).

٣- أنواع أدوات اكتشاف السرقة العلمية :

هناك العديد من أدوات كشف الانتحال المتاحة للاستخدام إما مجانية أو بمقابل، وتنقسم إلى ٣ فئات رئيسة:

١- أدوات تعتمد على الإنترنت web-enabled systems: يتم تصميم وتطوير تلك الأدوات للعمل على إحدى منصات العمل في بيئة الإنترنت، لما لها من قدرات تقنية للتغلب على مشاكل كشف الانتحال؛ فضلاً عن سهولة توافر تلك الأدوات للكثير من المستخدمين ولما تتمتع به من توسيع البحث عن المصادر المسروقة على الإنترنت بيسر (Bin-Habtoor, Zaher, 2012, p.185). وهذه الأدوات تنقسم إلى فئتين:

أ- خدمات الكشف على الإنترنت اعتماداً على قواعد البيانات Online Detection Services: ويتم تقديم خدمات الكشف عن الانتحال عبر الإنترنت من خلال إنشاء قواعد بيانات محدثة باستمرار بالبيانات الوصفية للأعمال الأكاديمية وبأدوات التكشيف اللازمة. ويتم تحديث قواعد البيانات بوثائق الأساتذة والطلاب، كذلك يتم تأمينها وحمايتها بسبب تسجيل الأعمال الأصلية بها؛ ومن ثمَّ تقديم خدمات فحص الأعمال المشتبه فيها (Vernon, Bigna & Smith, 2001, p.195).

ب- خدمات الكشف على الإنترنت مباشرةً: لا يوجد لها تعريف في أدبيات الإنتاج الفكري، ولكنها عبارة عن مواقع إنترنت تقدم خدمة الكشف عن الانتحال والسرقات العلمية سواء كانت بالمجان أو بمقابل، من خلال نسخ ولصق النصوص إلى مربع البحث أو من خلال رفع الملفات النصية إلى الموقع، ويتم مقارنتها بالوثائق المتاحة على الإنترنت وتحليلها؛ ومن ثمَّ ظهور نتيجة الكشف ونسبة التشابه بين النصوص.

٢- أدوات تستند على نظم التشغيل stand-alone systems: تم تطوير هذه الأدوات ليتم تثبيتها على أجهزة الحاسب، ويطلق عليها تطبيقات سطح المكتب، مع شرط الاتصال بالإنترنت لتفعيل مهمة البحث عن النصوص المتشابهة (Bin-Habtoor, Zaher, 2012, p.185). تُعد هذه الأداة أسهل في الكشف عن الانتحال؛ نظراً لكونها تركز على المحددات البحثية بدلاً من البحث في النص الكامل فقط، على الرغم من التشابه الكبير بين البرامج المتاحة في الأسواق (Clough, 2003). ويمكن تقسيم هذه الأدوات إلى:

أ- برامج تجارية مغلقة المصدر.

ب- برامج مجانية مفتوحة المصدر.

٣- محركات البحث: تعد من الأشكال الأكثر استخداماً على نطاق واسع لكشف الانتحال

على الإنترنت، وذلك باستخدام محرك بحث أو محرك بحث متعدد؛ لمحاولة إيجاد مصدر الجملة بشكل محرف أو مشكوك فيه. وتعد هذه التقنية سريعة وبسيطة، ومع ذلك هناك تناقضات في وسائل تغطية محرك البحث؛ حيث تعتمد على نهج النجاح والإخفاق، وذلك بوضع العلامات المحددة والعبارات المراد البحث عنها على الإنترنت فقط دون الاعتماد على قواعد البيانات للبحوث، بدلاً من العمل التلقائي في برامج كشف الانتحال (McKeever, 2006, p.157). أي على المسئول عن اكتشاف الانتحال القيام بعمليات النسخ واللصق ثم البحث ومضاهاة النتائج يدوياً مما يرهق القائم على ذلك.

٤ - تقنيات أدوات السرقة العلمية Plagiarism Techniques

تتبع أدوات اكتشاف الانتحال والسرقات العلمية العديد من التقنيات لتتبع مثل تلك التعديات، ومنها:

١ - تقنيات الكشف عن التشابه (التشابه التناظري)

يُعد من أكثر الأمثلة شيوعاً في هذا المجال ما يسمى بتقنيات فرز السمات. وتنشأ هذه التقنية "بصمة" خاصة للمقارنة بين الملفات، بما في ذلك المقاييس مثل متوسط طول الخط، وحجم الملف، ومتوسط عدد الفواصل في كل سطر، أي ما يعرف بسمات الملفات الإلكترونية. ولكن يعتبر هذا الأسلوب تقليدياً ولايمكن الاعتماد عليه، ونادراً ما يستخدم في الوقت الحاضر؛ حيث تعتمد أدوات كشف الانتحال في العادة على تقنيات مضاهاة المحتوى. وتشمل التقنيات الأكثر شيوعاً سلسلة من المسارات؛ لإيجاد التغطية المشتركة بين اثنين من الملفات وتحليل شجري مقارن (Bin-Habtoor & Zaher, 2012, p.186).

وتعمل تقنيات سرعة كشف الانتحال Fast Plagiarism Detection technique (FPDS) على تحسين أداء خوارزميات اكتشاف الانتحال عن طريق استخدام بنية خاصة بتكشيف البيانات لتخزين المدخلات على مجموع الملفات. كما يعتبر الترميز (القطع الدلالية) تقنية شائعة الاستخدام لمواجهة إعادة تسمية المتغيرات variables وحلقات التكرار المرتبطة بعداد loop في برامج الحاسب. وخوارزميات الترميز في أبسط صورها على سبيل المثال تتمثل في استبدال عناصر من أكواد البرنامج برمز واحد single tokens؛ حيث يمكن أن يتم استبدال كل المعرفات identifiers بالرمز "IDT"، وكل القيم values بالرمز "VALUES"، وستكون الجملة كالتالي: a line a = b + 45؛ وسيحل محلها عند سرقتها: <IDT>=<IDT>+<VALUE>؛ وبالتالي فإن إعادة تسمية المتغيرات لن تساعد مستكشف الانتحالات والسرقات العلمية (Bin-Habtoor & Zaher, 2012, p.186).

٢ - خوارزميات الانتحال (لكشف سرقة أكواد البرمجة):

هناك عدد من خوارزميات كشف الانتحال؛ حيث تعتمد أبسط تلك الخوارزميات على سلسلة مقارنات، كالآتي:

- إزالة جميع التعليقات.
- تجاهل كل الفراغات وأسطر الشروح (جمل يكتبها المبرمج لشرح الكود)، إلا عند الحاجة كمحددات delimiters.
- إجراء سلسلة أحرف character string للمقارنة بين ملفين.
- الحفاظ على عدد من النسب المئوية المرتبطة بالأحرف، ويتم تشغيل تلك الخوارزمية لجميع أزواج البرامج الممكنة (Bin-Habtoor & Zaher, 2012, p.186).

٣ – التحليل اللغوي:

يستخدم نفس الأساليب المتبعة في اللغويات القضائية (لتحديد صحة أقوال الشهود)؛ حيث تعمل على مبدأ أن لكل شخص أسلوبه في الكتابة والتعبير. ويتم إرسال النص بشكل إلكتروني بصورة متعددة للوثائق التي يتم فحصها لإجراء التعديلات المطلوبة لتفادي التشابه مع نصوص أخرى؛ ومن ثمَّ يتم حذف خامس كلمة من الجملة ويتم إرجاع النص إلى الكاتب مرة أخرى، حيث يطلب من الكاتب ملء الفراغات بنفسه للتأكد من عدم سرقتها. ويتم احتساب نسبة الانتحال استناداً إلى الكلمات الصحيحة التي تم تعبئتها، والوقت الذي يستغرقه من أجل استكمال العملية (McKeever, 2006, p.157). وتستعمل هذه الطرق معايير عديدة منها عدد مرات ظهور الأسماء وتعقيد المفردات وكلمات الربط ومقاطع الكلمات، وهناك محاولات لعدد من الباحثين لترتيب هذه المعايير بينت أهمية تعقيد الصفات والكلمات ومدى تعقيد الأفعال المستخدمة كمقياس لاكتشاف الانتحال، وجميع هذه الطرق تركز على تحليل الأسلوب اللغوي لفقرات النص؛ ومن ثمَّ الكشف عن التغيرات التي تطرأ على الأسلوب والتي تدل على إمكانيات الاحتيال (الجوة وآخ، ٢٠١١، ص4). ولكن بعد إجراء مسح عن برامج وخدمات كشف الانتحال والسرقات العلمية لم تتوافر هذه التقنية إلا في عدد محدود من البرامج؛ إذاً لم تعد تستخدم هذه التقنية على نطاق واسع من البرامج.

٥ – مميزات وعيوب أدوات اكتشاف السرقات العلمية:

١ – مميزات أدوات اكتشاف السرقات العلمية:

- توفير الوقت والجهد الشاق المبذول في الكشف بالطرق التقليدية، رغم خبرة الأساتذة والمحكمين، وأيضاً بسبب تعدد الباحثين واختلاف أماكن النشر، فضلاً عن اختلاف أساليب الكتابة.
- الانفجار المعرفي الهائل بسبب انتشار استخدام الإنترنت، أدى إلى اتجاه الباحثين للنشر الإلكتروني بدلاً من المطبوع مما يسهل عملية الكشف (McKeever, 2006, p.163).
- تعطي هذه الأدوات إمكانية لتعقب المنتحل.
- إمكانية معاقبة المنتحلين من خلال إقامة حظر على أي أعمال يتم إرسالها للنشر

في المستقبل من قبلهم (Shahabuddin, 2009, p.353).

٢ – عيوب أدوات اكتشاف السرقات العلمية:

- محدودية النطاق:

تسلط الكثير من الخدمات الضوء على متواليات النص أي النسخ الحرفي؛ ومن ثمَّ فمن الممكن على السارق التلاعب أو تغيير الترتيب لخداع النظام وإخفاء الانتحال. ومن هنا يلاحظ مدى جمود وتعقيد تلك المبادئ، حيث إن تحديد درجة الانتحال تتعلق بالسماح للقياس الكمي للكلمات ونسبة ذلك لإجمالي النص. كما تقتصر خدمات كشف الانتحال من حيث التغطية على المحتوى الرقمي بالإنترنت، ولكن نسبة كبيرة من هذا المحتوى لم تجد طريقها إلى نتائج البحث بسبب توافر عوامل الحماية والأمان في بعض قواعد بيانات النص الكامل والتي من المحتمل احتوائها على مصدر قد تم السرقة منه (McKeever, 2006, p.160).

- القيود التقنية:

على الرغم من التطور فإن بعض الخدمات لاتزال تفرض بعض القيود التقنية؛ ومن ثمَّ يواجه بعض المستفيدين صعوبة في التعامل معها، وذلك على سبيل المثال بسبب الاختلاف في قبول صيغ الملفات في خدمات كشف الانتحال، فضلاً عن اشتراط بعض الخدمات على المستفيدين القيام ببعض الخطوات التي تتطلب كفاءة تقنية من جانب المستفيد. أما من حيث منصة العمل فهي تنصب على اختيار المستفيد النهائي، ما بين الخدمات المستندة على الويب web-based التي ليست بحاجة إلى تنصيب أو تحميل أي أدوات، والأختيار الآخر للأدوات التي تستند على نظم التشغيل (McKeever, 2006, p.160).

- عدم استقرار الأدوات:

نتيجة للتطور التكنولوجي والاقتصادي، فإن قطاع كشف الانتحال متقلب للغاية، وهذا يجعل قرار الشراء للمؤسسات الأكاديمية أكثر خطورة؛ حيث تظهر بعض خدمات الاكتشاف وتختفي أو تتغير تغيراً جذرياً. كذلك فإن الدعم عبر الإنترنت غير متطور بشكل جيد (McKeever, 2006, p.161).

- القضايا الإدارية والتكاليف:

تعد حماية البيانات من العقبات الإدارية الرئيسة التي يتعين التغلب عليها في خدمات كشف الانتحال. أما التكلفة فتختلف التراخيص من خدمة إلى أخرى، فالبعض يضع اعتباراً للمؤسسات الأكاديمية ويقدم الخدمة مجاناً، والبعض الآخر يقدم رسوماً ثابتة وكبيرة بغرض تقديم الخدمة بثبات ولمواصلة تحسين وتطوير الخدمات (McKeever, 2006, p.162).

٦- مراحل اكتشاف السرقات العلمية:

يعتبر "أوتينستاين" "Ottenstein" (1976) أول من اقترح نظام آلي لكشف الانتحال والسرقات العلمية باستخدام أنظمة الحاسب والبرمجيات؛ من خلال الاعتماد على لغة برمجة فورترن "FORTRAN".

وترتكز أدوات كشف الانتحال والسرقات العلمية على اثنين من العناصر الحاسمة؛ أولهما إرسال المدخلات وثانيهما تجميع الوثائق في ذخائر نصية corpus للبحث فيها. هذا بجانب أربعة شروط: أولاً التحقق من البحوث خلال أدوات كشف الانتحال. ثانياً: تحتاج الخطوة الأولى إلى تحديد صيغة الملفات. ثالثاً: إذا تم العثور على أجزاء متشابهة يظهر تقرير بالمصادر المشتبه بها ويتم تغيير لون النص المتشابه. وأخيراً: يتم الحكم على ما إذا كان هذا النص الذي تم تحديده رجع إلى المصدر الأصلي بشكل صحيح أم لا (Alzahrani, et al., 2009, p.9).

وتتفق أغلب الأدوات على طريقة العمل التالية:

- استخراج النص من الوثيقة (إما في صيغة Word، أو PDF).
- ثم البحث على الإنترنت، أو في قاعدة بيانات خاصة، أو الاثنين معاً.
- تحديد تشابه النص المستخرج مع مصادر أخرى، ثم عرض النتائج (Heather, 2010, p.649).

ويمكن توضيح المراحل بالتفصيل كما بالشكل التالي رقم (١)؛ حيث تشمل المرحلة الأولى تجميع الوثائق في ذخائر نصية، وهي إما أن تكون داخلياة Intra-corpus أو مشتركة Inter-corpus؛ حيث إن الوثائق في البيئة الداخلية عبارة عن وثائق لم تتوزع عبر الشبكات، أي يتم البحث والعثور عليها في نفس مكان التخزين مثل المكتبات الرقمية. أما على الجانب الآخر، فإن معظم التقنيات الحالية تهدف إلى البحث عن الوثائق المتشابهة في الإنترنت؛ وبالتالي فإن تلك الأدوات تدعم وثائق الإثبات المنتشرة عبر الشبكات.

أما المرحلة الثانية فتتمثل في تحليل الوثائق التي تم جمعها؛ لإيجاد أنماط متماثلة مع وثائق الاستعلام. وخلال مرحلة التحليل يعطى لكل زوج من الوثائق درجة أو نسبة تشابه استناداً إلى منهج محدد. أما عن المرحلة الثالثة فتتمثل في "الإثبات" والتأكيد؛ لتجنب ما يسمى "بالإيجابية الكاذبة"، التي تحدث عندما يتم إعطاء زوج من الوثائق نقاط تشابه عالية على الرغم من أنها تحتوي على القليل من التشابه، أو تعطي "سلبية كاذبة" للوثائق المتشابهة ولكنها لا تكتسب نقاط تشابه عالية. أما المرحلة الأخيرة فتتمثل في "التحقيق"، وهي تحتاج إلى تدخل الإنسان؛ لضمان ما إذا كانت الأجزاء الموضحة مسروقة حقاً أو مستشهد بها أو مقتبسة أو تم الإشارة إليها بشكل صحيح (Alzahrani, et al.,2009,p.10).

شكل رقم (١)

مراحل اكتشاف الانتحال

٧- التحديات التي تواجه أدوات السرقة العلمية:

١- هناك العديد من قواعد البيانات المتاحة على الإنترنت تتطلب التسجيل أو الاشتراك للاطلاع على محتواها؛ وبالتالي يصعب

الوصول إلى المحتوى من خلال أدوات اكتشاف السرقة العلمية، ومن الممكن أن تكون في متناول المنتحل (Atkinson & Yeoh, 2008).

٢- صعوبة اكتشاف معظم السرقات العلمية، حيث إن معظم أدوات كشف الانتحال غير فعالة على الوثائق المنشورة في شكل ورقي ولا تغطيها قواعد بيانات هذه الأدوات.

٣- كما تواجه هذه الأدوات الفشل المؤكد في اكتشاف الانتحال في حالة السرقة من وثائق مكتوبة بلغات أخرى (Maurer, Kappe & Zaka, 2006).

٤- عدم قدرة الأدوات على مقارنة الصور؛ وبالتالي صعوبة اكتشاف انتحال الصور.

٥- استخدام بعض الأحرف باللون الأبيض مع الخلفية البيضاء وإظهارها على أنها مساحات فارغة؛ مما يصعب رؤيتها بالعين المجردة، رغم اكتشافها من قبل الأدوات.

٦- تكشف الأدوات التكنولوجية الجمل والعبارات المسروقة، ولكن لا تكتشف الفكر أو الأفكار المسروقة (Pardeshi et al, 2013, p.29)؛ حيث إن صعوبة الكشف عن سرقات الأفكار تتطلب تقنيات دقيقة للغاية لتصوير وتمثيل الأفكار وتطوير طريقة يمكن الاعتماد عليها لاستخلاصها من البناء النصي (Mozgovoy, Kakkonen& Cosma, 2010, p.515).

٧- كل ما ينشر على الإنترنت لا يتم تكشيفه، أو تتأخر عملية التكشيف من جانب محركات البحث أو برامج كشف الانتحال.

٨- وبعد إجراء بعض المحاولات، وجد أن إعادة صياغة الجملة وإعادة ترتيب بنيتها يؤدي إلى خداع أدوات كشف الانتحال.

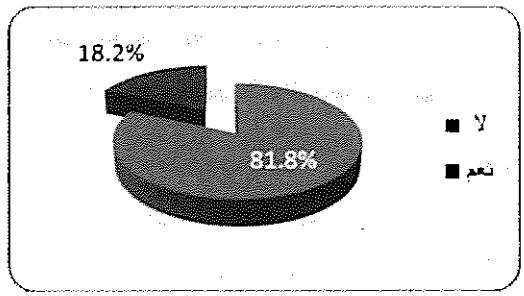
٨ - الجهات المسؤولة عن تحكيم ومراجعة البحوث المتخصصة في المكتبات والمعلومات: المشهد العام

تنحصر تلك الجهات كما سبق ذكر ذلك، في الدوريات المتخصصة المحكمة في مجال المكتبات والمعلومات، بالإضافة إلى لجنة فحص الإنتاج العلمي. وقد اتفق الجميع على عدم توافر أي سياسة واضحة ومحددة تنص على أوجه وأبعاد السرقات العلمية وكيفية معاقبة المنتحلين والتصرف معهم، وذلك بنسبة ٨١,٨%، كما في الشكل رقم (٢). ولكن يرى ١٨,٢% -وهذه النسبة من أعضاء اللجنة العلمية لفحص الإنتاج العلمي- أن لائحة المجلس الأعلى للجامعات الخاصة بشغل وظائف الأساتذة والأساتذة المساعدين تنص في بعض موادها على معاقبة المنتحلين، وذكروا أيضاً أنها توصي إذا تبين وجود حالة ثُبت عليها عدم الالتزام بالأمانة العلمية، فعلى اللجنة أن توقف إجراءات الفحص، وأن تقدم تقريراً مدعوماً بالدلائل القاطعة إلى المجلس الأعلى للجامعات؛ ومن ثمَّ يُتخذ اللازم ويتم مراسلة عميد كلية العضو المنتحل، وبعدها يُعرض على مجلس تأديب. وبكل تأكيد هذه المواد لا بأس بها ولكن هي في حد ذاتها بنود أو مواد ضمن لائحة

المجلس الأعلى للجامعات، وليست سياسة كاملة وواضحة توضح أبعاد وأشكال الانتحال والسرقات العلمية، وتعدد درجات الجزاء والعقاب والتأديب الذي يمكن أن يقع على الباحثين جراء فعل هذا التعدي على حقوق الآخرين.

شكل رقم (٢)

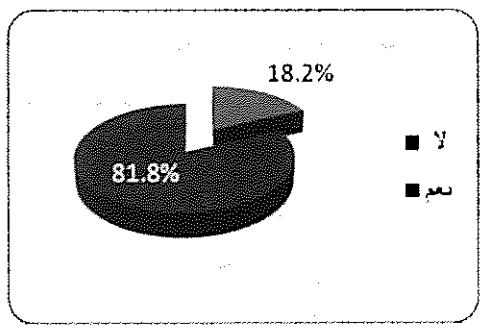
نسب توافر سياسات للسرقات العلمية



يتضح من خلال الشكل رقم (٣)، أن أغلب الجهات المسؤولة عن تحكيم البحوث في مجال المكتبات والمعلومات تعتمد اعتماداً كلياً على اكتشاف السرقات العلمية وحالات الانتحال من خلال الاعتماد على الوسائل التقليدية، وذلك بنسبة ٨١,٨%، وعدم الاعتماد بشكل أو بآخر على أحد أدوات اكتشاف السرقات العلمية سواء المجانية أو التجارية.

شكل رقم (٣)

اكتشاف السرقات العلمية بالاعتماد على الوسائل التقليدية

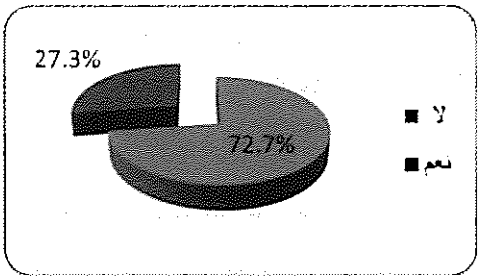


وبالتعرف على توجهات البعض نحو الاستعانة ببعض الأدوات التكنولوجية من باب اليقين من أن البحث المفحوص خالٍ من التشابهات والانتحال، أظهرت النتائج كما في الشكل رقم (٤) أن ما يقرب من ٧٢,٧% لا يعتمد على المستوى الشخصي بأي أداة قبل كتابة التقرير عن البحث.

شكل رقم (٤)

مراجعة الأدوات التكنولوجية

قبل الحكم على البحث بشكل فردي

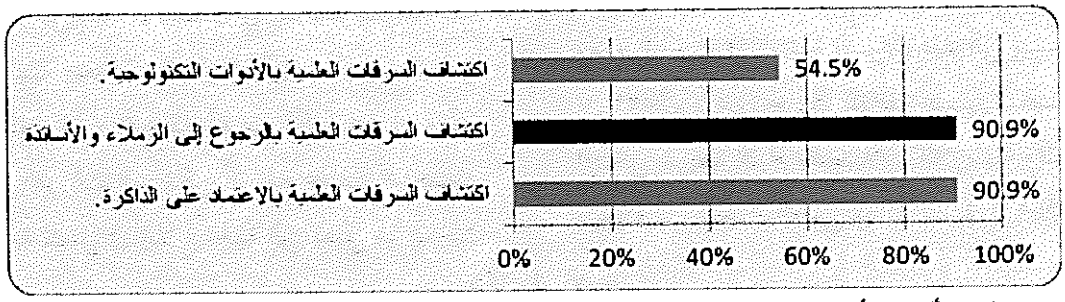


وتعكس المؤشرات السابقة، أن جميع المشتركين في العينة يعتمدون بشكل كبير على الذاكرة في اكتشاف السرقات العلمية، دون غيرها من الأدوات التكنولوجية المتاحة، ولم تتباين الردود ووجهات النظر كثيراً بشأن أفضلية الأساليب والطرق التي يمكن الاعتماد عليها في اكتشاف السرقات العلمية، فكانت الإجابة متماثلة بشكل كبير-كما بالشكل رقم (٥)، حيث إن نسبة ٩٠,٩% تفضل اكتشاف السرقات بالاعتماد على الذاكرة وأيضاً استشارة الزملاء والأساتذة، أي الاعتماد الكلي على الأدوات والأساليب التقليدية البحتة لبحث وجود سرقات علمية، ونسبة ٥٤,٤% على الأقل تفضل الاعتماد على الأدوات المخصصة لاكتشاف الانتحال والسرقات

العلمية. وهذا ينعكس بكل تأكيد على عدم الاعتماد بشكل رسمي على أيٍّ من هذه الجهات في الحكم على أصالة البحوث بأحد أدوات اكتشاف السرقات العلمية، وذلك بنسبة ١٠٠%.

شكل رقم (٥)

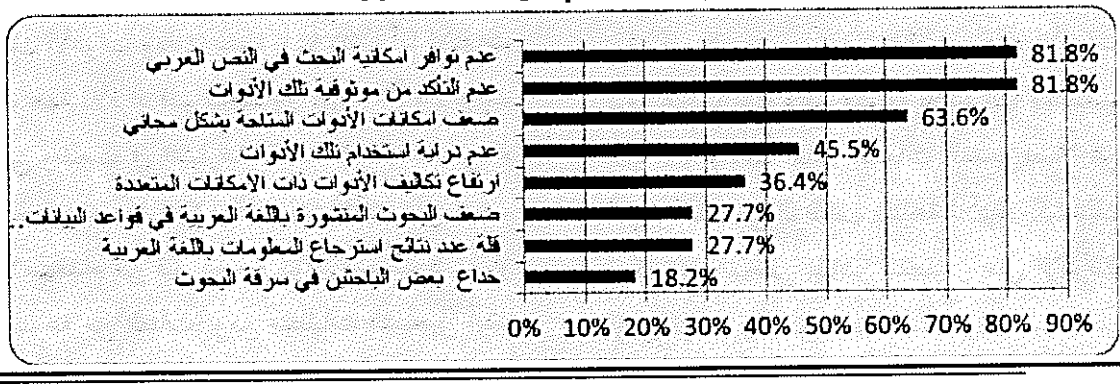
أفضلية اكتشاف السرقات العلمية



وقد كان أكثر الأسباب وراء عدم اعتماد اللجنة العلمية على مثل هذه الأدوات متمثلاً في وعدم توفر أدوات متخصصة في الكشف عن الانتحال والسرقات العلمية تدعم اللغة العربية بشكل كامل، وعدم التأكد من موثوقية وشمولية هذه الأدوات، فضلاً عن التخوف من استخدام الأدوات المجانية بسبب ضعف نتائجها وقدرتها المحددودة وإمكاناتها الضعيفة، ثم يلي ذلك أيضاً عدم الدراية باستخدام أي أداة، كما أن ارتفاع التكاليف يلعب دوراً في عدم الاعتماد على مثل تلك الأدوات؛ بسبب قلة الدعم المالي وضعف الميزانيات، كذلك يري البعض أن ضعف البحوث المنشورة باللغة العربية على الإنترنت بشكل عام وقواعد البيانات والمستودعات الرقمية المتخصصة بشكل خاص له أثر كبير في عدم الاعتماد على مثل هذه الأدوات، ويتبع ذلك بكل تأكيد وبنفس النسب المئوية قلة نسبة النتائج المسترجعة باللغة العربية الناتج عن ضعف النشر الرقمي، وأخيراً يمكن القول بأن من بين هذه الأسباب أيضاً خداع بعض الباحثين لمثل هذه الأدوات من خلال التعرف على نقاط الضعف في عملية الفحص والتحليل والمقارنة بين النصوص وغيرها من الأساليب المبتعة في أغلب البرامج والأدوات الشهيرة في هذا المجال، وتتضح هذه النسب على التوالي كما بالشكل التالي رقم (٦).

شكل رقم (٦)

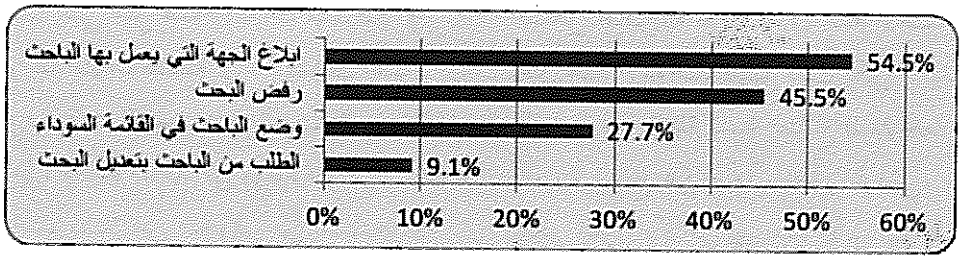
أسباب عدم الاعتماد الرسمي على الأدوات التكنولوجية



كما تتعامل جهات فحص البحوث في مجال المكتبات والمعلومات مع الباحثين المنتحلين في حال ثبوت مثل هذه الأفعال في البحوث المقدمة بأساليب متعددة، وتتفاوت النسب بينهم كما في الشكل رقم (٧)، فقد أعرب الكثير عن أن من أكثر القرارات التي يمكن أن تقع على الباحث صرامةً تتمثل في إبلاغ الجهة التي يعمل بها الباحث وذلك بنسبة ٥٤,٥% حتى يتسنى لها أن تتخذ معه الإجراءات اللازمة للتحقيق في مثل هذه الواقعة، وثانيها رفض البحث وذلك بنسبة ٤٥,٥%، أما وضع الباحث في القائمة السوداء فقد أتى بنسبة ٢٧,٧%، وأخيراً جاءت الموافقة على البحث المنتحل شريطة القيام بالتعديلات اللازمة وفقاً لأخلاقيات البحث العلمي وذلك بنسبة ٩,١%.

شكل رقم (٧)

اتجاهات لجان التحكيم نحو الباحثين المنتحلين



٩- سمات أداة كشف الانتحال في النص العربي:

١/٩ التسجيل ومتطلباته:

يمثل التسجيل النسبة الأعلى للأدوات التجارية التي يمكن الوصول إليها من خلال الإصدارات التجريبية وذلك للكشف عن السرقات العلمية، وتصل نسبتها إلى ٦٦,٧%. ولم تحدد أغلبية الأدوات الجمهور المستهدف من استخدام تلك الأدوات وذلك بنسبة ٧٩,١%، أما الأدوات الموجهة للأفراد فلم تتجاوز ٢٠,٨% والأدوات التي وضعت في اعتبارها الاستخدام المؤسسي لم تتجاوز نسبتها ١٢,٥%، ويدل ذلك على تعميم الاستخدام من قبل الجميع دون تحديد فئة أو مجموعة من الفئات. ومن متطلبات التسجيل أيضاً تحديد رسوم الاشتراك فيما يخضع لسياسة الترخيص التجارية، وقد كانت أعلى النسب لفئة الرسوم تتراوح ما بين ٢٠:٨٠ دولار شهرياً بنسبة ٣٧,٥%. وكان لابد من التعرض للعروض المالية للاشتراكات، حيث يرتبط التسجيل أحياناً في تلك الأدوات بدفع رسوم مقابل حزمة من الخدمات، وقد تنوعت تلك الخطط في عدد من البرامج بنسبة ٤٥,٨% ولكن نسبة ٣٣,٤% لم توفر أي شيء، بينما ٢٠,٨% من هذه الأدوات توفر خططاً ثابتة وغير متنوعة للاشتراك؛ ويتسم هذا التنوع وفقاً لعدد الخدمات المقدمة فضلاً عن اختيار عدد من الوظائف المتاحة للأداة، كذلك إمكانيات الفحص المباشر للمصادر على الإنترنت أو في قواعد بيانات محددة سلفاً وغيرها من

السمات المتنوعة التي تضفي على الخطط تنوعاً من الممكن أن يجذب المستفيد النهائي لاختيار أي منها وفقاً لرغباته واحتياجاته ومتطلباته الفنية والمادية. وبما أن مهمة هذه الأدوات تنصب على الفحص كان لابد من التعرف على أكثر التقنيات المرتبطة بطبيعة الاشتراك كتحديد عدد الكلمات في الوثيقة التي يتم ارسالها ووجد أن ٦٢,٥% لم تحدد عدداً معيناً من الكلمات وفقاً لقيمة الاشتراك، وكذلك أظهر الفحص أن ٧٠,٨% من الأدوات لم تربط قيمة الاشتراك بعدد الملفات التي يتم إرسالها. ويدل هذا على أن النسبة الأعلى من هذه الأدوات لا تهتم بمحاسبة المستفيد على عدد الكلمات أو الملفات التي يرغب في فحصها أو حتى عدد مرات استخدام الأداة على مدار فترة الاشتراك. ويتضح أن أكثر البرامج انتشاراً من فئة البرامج التجارية ذات الوظائف والخيارات المتعددة والواجهات الرسومية.

أما نسبة وجوب التسجيل سواء الأدوات المجانية أو التجارية فقد تعدت ٨٣%، ولكن أداة واحدة فقط هي التي أتاحت إمكانية تعدد الاشتراك لحساب واحد فقط (أي حساب واحد للعديد من المستخدمين، كأن يكون لعضو هيئة التدريس حساب رئيسي، والطلاب الذين يدرسون تحت إشرافه يحصلون على اشتراك فرعي يتيح لهم بعض الميزات والخيارات) ويساوي ذلك ٤,١% فقط، ورغم أن هذه الميزة تعد قيمة مضافة للبرنامج وتعد ميزة تنافسية لم تتوفر في غيره من البرامج فإنه من المستحسن إضافتها لما لها من فوائد على الطلاب وهيئات التدريس. أما بالنسبة لطبيعة الإتاحة للأدوات والحصول على خدماتها، فتبين أن ٧٩,١% اهتمت بإتاحة خدماتها في بيئة الإنترنت دون البرمجيات التي يتم تنصيبها وتشغيلها على نظم التشغيل.

٢/٩ واجهة الاستخدام:

لم تتوافر واجهة رسومية لأكثر من نصف العينة، وذلك بنسبة ٥٤,٢%، وللتعرف على أكثر تصنيفات الواجهات وفقاً لطبيعة الاستخدام، ظهر أن الفئة الأكثر شيوعاً هي الواجهات الشاملة لجميع الوظائف وذلك بنسبة ٩١,٧%، يليها الواجهات الخاصة بأعضاء هيئة التدريس والواجهات الخاصة بالطلاب بنسبة ١٢,٥% لكل منهم، أما الواجهات الخاصة بالكُتّاب والمؤلفين فكانت بنسبة ٤,٢%، ونفس النسبة لواجهات مديري النظم. وهذا يدل على أن الاتجاه الأكثر شيوعاً هو توافر واجهة موحدة سواء كانت بسيطة أو شاملة تتوافر فيها كل الوظائف والخدمات دون تصميم واجهات مختلفة لكل فئة من فئات الجمهور المستهدف. وقد كانت نسبة الواجهات أحادية اللغة ٧٥%، مما يعني اكتفاء الأدوات بلغة واحدة فقط للواجهة، وفي الغالب وبنسبة كبيرة تنصب في ناحية اللغة الإنجليزية.

٣/٩ التجهيزات المادية والبرمجية:

من الأساليب المتبعة عند شراء أو الاشتراك في أي برنامج أن تُذكر المتطلبات المادية والبرمجية، وقد أظهرت عينة الدراسة في مجملها عن عدم الاكتراث بوصف وتحديد تلك المتطلبات وذلك بنسب متفاوتة. فإن عدد الأدوات التي لم تذكر

نظام التشغيل المتوافق مع الأداة ٦٢,٥%، وأغلب الأدوات تعمل في بيئة نظم تشغيل ويندوز بنسبة ٤١,٧%. ولم يتم ذكر التطبيقات المساعدة على التشغيل بنسبة ٥٨,٣%، وما تم ذكره من تلك التطبيقات تركز على ذكر المتصفحات المناسبة وإصدارة الجافا الملائمة وأخيراً برنامج تشغيل الفلاش وبرامج أخرى غير شائعة وذلك بنفس الترتيب ٢٩,٢%، ١٢,٥%، ٤,٢%، ١٢,٥%، وفيما يتعلق بذكر معلومات عن الذاكرة الدائمة أو المؤقتة ظهرت النتيجة متساوية بنسبة ٢٠,٨%. أي أن الكثير من الأدوات لم تُخبر المستخدم بالمتطلبات التقنية الواجب توافرها قبل تشغيل أدوات كشف الانتحال للتعرف على مدى توافقها مع بيئة التشغيل وتطبيقاتها المساعدة.

٤/٩ الإمكانات الوظيفية والتقنية:

١/٤/٩ تجهيز الملفات:

أ- إعداد إحصائية عن النص أو الملف الذي تم رفعه كعدد الكلمات واسم الملف ووقت الفحص المنقضي ...إلخ.

ب- معالجة جميع صيغ الملفات بما في ذلك (صيغ Pdf,Doc)، فضلاً عن تحويل الملفات إلى ملف نصي، وقبل التحويل يفضل فحص الميتاداتا.

ج- إزالة العلامات المميزة من الملفات (Chuda & Navrat, 2010).

٢/٤/٩ أساليب وخوارزميات الكشف في النص العربي(*)

أ – المعالجة الأولية للنصوص:

- تحويل الصور النصية إلى نص قابل للكشف: حتى لا يتم تجاهل النصوص التي تم أخذ صورة منها عن طريق الماسحات الضوئية، ويتم تحويلها بالاعتماد على نظم التعرف الضوئي على الحروف Optical Characters Recognition OCR (Patel, Bakhtiyari & Taghav, 2011).
- تجهيز (تهيئة) النص:
  - إزالة التشكيل: لابد من إزالة تشكيل الكلمات؛ لمنع حدوث سلبيات كاذبة، بسبب أخطاء الإدخال للنصوص العربية.
  - قائمة الاستبعاد Stop list: هي قائمة للمفردات التي لا تحمل معاني اصطلاحية أو مفاهيم مهمة لأغراض استرجاع المعلومات (الصوينع، ٢٠٠٨، ص٨٣). وتشمل على أدوات التعريف وحروف الجر، ويتم استبعادها قبل معالجة النصوص، لتحسين البحث.

(*) استكمالاً لجهود الباحثين في مجال معالجة اللغة العربية، قد استفاد الباحث من تجارب البحوث السابقة، بذكر أهم العناصر التي يمكن الاستفادة منها في تصميم أداة كشف الانتحال باللغة العربية، كما يمكن الرجوع إلى الوثائق والبحوث والدراسات المعدة في هذا الجزء للتعرف أكثر عن الأساليب المتبعة في معالجة اللغة العربية، حرصاً على ألا تتكرر الجهود.

- توحيد أشكال الأحرف: أي توحيد أشكال الحروف ذات الأشكال المختلفة.
- تغيير لون النص أو تغيير الخلفية للكشف عن علامات الترقيم والكلمات غير المرئية (Patel, et al., 2011).
- التصحيح اللغوي والإملائي: حيث يتم التنويه إلى توافر اقتراحات بشأن تركيب الجملة أو بعض الأخطاء اللغوية، وليس بالضرورة أن يتم تغيير كل الاقتراحات. ومن خلال الكشف عن مدى توافر خاصية التصحيح اللغوي في الأدوات المتاحة وجد أنها توافرت بنسبة ٥٤,٢%، وأن خاصية التصحيح الإملائي توافرت بنسبة ٦٦,٧%.

- الاستعانة بالترجمة المتعددة من اللغة العربية إلى العديد من اللغات الأجنبية؛ ثم العودة مرة أخرى إلى اللغة العربية، بحيث يمكن استخدامها للحصول على نتائج متطورة (Patel, et al., 2011). ولكن يتوافر في بعض برامج عينة الدراسة التغطية لأكثر من لغة لتوسيع نطاق الفحص وليس لغرض الترجمة من لغة إلى أخرى ولكن قبول النصوص بلغات مختلفة، وكانت حصيلة اللغة الإنجليزية هي الأكبر بنسبة ١٠٠%، تليها اللغة الفرنسية والألمانية والإسبانية بنسبة ٣٧,٥%، ثم الإيطالية بنسبة ٢٩,٢%، ثم الروسية بنسبة ٢٠,٨%، ثم اليابانية بنسبة ١٦,٧%، وأخيراً الهولندية واللغة العربية بنسبة ١٢,٥%.

- تجزئة النص Segmentation: هي عملية تقسيم النص إلى فقرات أو جمل متماسكة، ويذكر أن متوسط طول الجملة العربية ٣٧ كلمة، وهو أعلى من أي لغة أخرى. وليس من السهل تحديد نهاية الجملة وهي مشكلة في طرق كشف الانتحال، ويمكن حلها عن طريق تحديد علامات الترقيم المتعارف عليها، والتي تعمل على تقليل عدد كلمات الجملة (Jadalla & Elnagar, 2012, p.148).
- تجزئة الجمل: وهي عملية تقطيع الجمل العربية إلى كلمات لتسهيل كشف الانتحال، مثل: "يمضي محمد إلى المدرسة" ويتم تقطيعها كالتالي: "يمضي،محمد،المدرسة" (Menai & Bagais, 2011, August, p.1381).
- تجذيع الكلمة (التشذيب) Stemming: هي عملية حسابية لتجريد الكلمات من السوابق (ال،بال،و،ك) واللواحق (ون،ات،ها) وبعض الحروف غير الأصلية؛ لإنتاج كلمات لها نفس الجذر أو الأصل (Jadalla & Elnagar, 2012, p.148). بغرض تسهيل المقارنة بين الكلمات والمساعدة في التعرف على مدى التطابق. ويمكن بناء مشذب عربي Stemmer فضلاً عن إتاحة الكثير من أدوات التشذيب العربية. ومشكلة هذا الأسلوب تتمثل في عدم فهم الحالات الشاذة في اللغة العربية كجمع التكسير، ولابد من وضع بعض القيود حتى لا يتم حذف أصل الكلمات. وعلى الرغم من ذلك، فقد تبين من خلال أدوات كشف الانتحال (ما يعالج منها اللغات الأجنبية واللغة

العربية) محل الدراسة أن الغالبية العظمى منها لم يستخدم مثل هذه التقنية وذلك بنسبة ٨٣,٣%.

- المعالجة الصرفية: يتم مقارنة ما استخرج من التجذيع بقائمة من هياكل الجذور، ولكي تتم عملية المعالجة بشكل صحيح لابد من القيام ببعض العمليات منها: إرجاع الفعل المضارع إلى الماضي وفك التضعيف (قدّم-قددم)، وفي حالة عدم وجود تطابق بين الكلمة وقائمة الهياكل، فهذا يعني إما أن هناك خطأ في الكلمة أو أن هناك حاجة إلى معالجة إضافية في المرحلة التالية.
- المعالجة الإضافية: وهذه المرحلة للكلمات التي بحاجة إلى معالجة إضافية مثل جمع التكسير (قلم-أقلام)، أو الإبدال (وصف-اتصف)، أو الإعلال (قاضي-قاضٍ) (الحوراني).
- استخدام تقنية N-gram: رغم أن هذه التقنية لم تثبت نجاحها في عمليات استرجاع المعلومات باللغة العربية بسبب الطبيعة العالية للاشتقاق، فإنه يمكن الاستفادة منها في كشف الانتحال حيث تعمل هذه التقنية على قياس درجة التشابه بين الكلمات. وترتكز على تقصي عدد المرات التي تتكرر فيها مجموعة ذات عدد محدد (س) من الأحرف في كل كلمة، ومن ثم يتم احتساب التشابة بين الكلمتين كدالة للعدد (س) من الأحرف التي ترد في كلتا الكلمتين (الربيعة و السلمان، ٢٠١٠، ص٣).
- استخدام المكانز: تعد أداة فعالة لمقارنة النصوص، ومفيدة في مكافحة استبدال الكلمات المرادفة في النصوص. فمن الضروري استخدام المكانز جنباً إلى جنب مع النماذج التوضيحية لمعنى الكلمة من أجل التأكيد على أن مجموعة المرادفات التي يتم استخراجها دقيقة ومقبولة (Mozgovoy, et al., 2010, p.522,524). وبهذا الأسلوب يعتبر المستوى الأدنى للمعنى الدلالي؛ حيث يتم استخراج مجموعة من الكلمات المتصلة والمرادفة لجمل النص الأصلي. ويمكن أن تتوافر في هذه السلاسل النصية إمكانية إضافة معلومات لغوية أخرى بخلاف المرادفات كالعلاقات الشمولية. ولا يتعدى عدد الكلمات المكونة لتلك السلاسل ست كلمات (الجوة وآخ، ٢٠١١، ص٥).

أما ما يتعلق بتقنيات الجذر، وتجزئة النص والجمل، فضلاً عن استخدام المكانز، فهذه التقنيات غرضها تبسيط البحث عن الكلمات ومترادفاتها، ولكن بعد إجراء دراسة وفحص أدوات الكشف المختارة وجد أن ما نسبته ٨٣,٣% لا تستخدم مثل هذه التقنية، ولكنها تبحث بطريقة مطابقة الجملة exact sentence match.

ب - **بصمة الكاتب**: من الممكن إنشاء بصمة لكاتب معين على أساس أن له مفردات خاصة، ونمط في الكتابة وفي تركيب الجمل. ويمكن استخدام هذه السمات لتحديد مؤلف النص. وهذه الأساليب قادرة على

كشف المؤلفين خلال مجموعة مقيدة ومعرفة مسبقاً من المؤلفين فقط (Mozgovoy, et al., 2010, p.522,524). وهذه البصمات مشابهة لبصمات الأصابع، وهناك طريقتان: الأولى إعطاء بصمة لكل الوثيقة تتكون من مجموعة من كل السلاسل النصية الفرعية substrings المتسلسلة من الطول a في الكلمة، وهناك (N−α+1) للسلاسل الفرعية، حيث إن N هو طول الوثيقة في الكلمات، وهذه البصمة تختار تطابق السلاسل النصية الفرعية. فعند المقارنة بين وثيقة X إلى وثيقة Y حيث إن حجم وثيقة X هو |X| وإذا كان N هو عدد السلاسل النصية الفرعية المشتركة لكلتا الوثيقتين فبعد ذلك n/|X| وهو مقياس مدى توافر نص وثيقة X في وثيقة Y. الطريقة الثانية: يطلق عليها البصمات الانتقائية، وأبسط نوع منها يتمثل في اختيار جميع السلاسل النصية الفرعية، ولكن لا يتم اختيار السلاسل النصية الفرعية المتطابقة. وبدلاً من ذلك، يتم اختيار كل السلاسل النصية الفرعية غير المتطابقة من الحجم (a في الكلمة) داخل الوثيقة، على سبيل المثال: إذا كان (a) هو يعبر عن ٣، فتنفذ الاستراتيجية باختيار ثالث كلمة من بداية الجملة (HaCohen-Kerner, Tayeb & Ben-Dror, p.422). ويتم تطبيق هذه التقنية على مستوى شجري منطقي (الوثيقة، الفقرة، الجملة). والجدير بالذكر أن تقنية هي البصمة أحد تقنيات الذكاء الاصطناعي.

ويمكن مقارنة فقرتين باستعمال معايير كموضع الفقرات وتكرار الكلمات، وقد قام بعض الباحثين بدمج معايير طول التواقيع ومواضعها من خلال عدة معادلات لمقارنة التواقيع كالتشابه التناظري، كالتالي: التشابه التناظري: م(ف١)∩م(ف٢)/ (م(ف١)+م(ف٢))، م(ف١) يمثل مجموعة التواقيع للفقرة١. (الجوة وآخ، ٢٠١١، ص٤).

جـ – **مقاييس التشابه:** هناك العديد من هذه المقاييس مستخدمة في أدوات كشف الانتحال، والخوارزمية الأكثر ملائمة لاكتشاف أفضل درجة تشابه تعرف بـ Longest Common Substring (LCS). (Menai & Bagais, 2011, August, p.1381)

د – **مرحلة تحليل التشابه:** تصمم الأدوات للكشف عن التشابه بين النصوص وذلك بحساب التشابه بين محتوى الوثائق المحددة، وذلك بعد إزالة كلمات التوقف، وإزالة السوابق واللواحق (Hariharan, 2012, p.324).

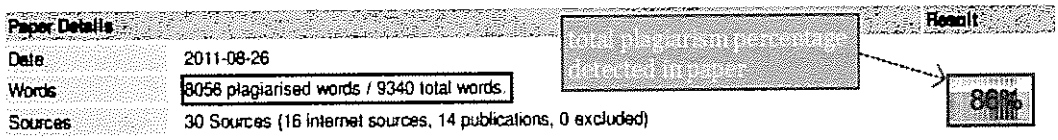
هـ – **التكشيف والبحث :** يشبه هنا بنية محرك البحث من حيث أنه يحافظ على بناء الكشاف المقلوب، وهو عبارة عن هياكل البيانات للكلمات الدالة مع الإشارة إلى مواقعها في مجموعة من الوثائق. أما البحث فيتم من خلال نوعين رئيسين في نظم الاسترجاع: الاستعلامات المنطقية

Boolean، والاستعلامات التصنيفية Ranked، أو من نوع مختلط ويسمى الاستعلامات المنطقية التصنيفية (Jadalla & Elnagar, 2012, p.149).

و – **مرحلة ما بعد المعالجة:** ضرورية لتحديد النتائج وحساب قيمة دقة الأصالة لجمل الاستعلام؛ ومن ثمّ فحص كل الوثيقة (Jadalla & Elnagar, 2012, p.150). وفي هذه المرحلة على الأداة أن تحدد النسبة المئوية للتشابه بدقة من خلال التعرف على حذف التكرار بين مصادر البحث المتشابه، وهذا ما تغفله الكثير من الأدوات المتاحة، حيث من الأفضل حساب النسبة المئوية على أساس تحديد الكلمات المتشابهة بالوثيقة مع الوثائق الأخرى؛ ومن ثمّ حساب نسبة هذه الكلمات المتشابهة بالأخرى غير المتشابهة في نفس نص الوثيقة، كما في الشكل التالي رقم (٨).

شكل رقم (٨)

جزء من تقرير خدمة Check for Plagiarism.net

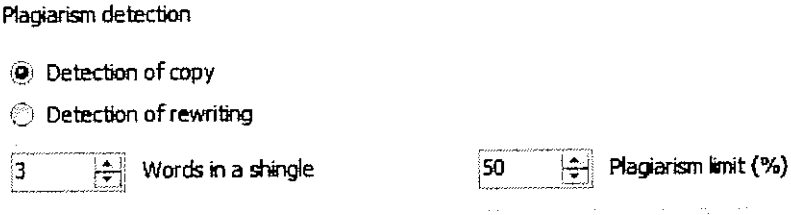


٣/٤/٩ مستوى الكشف:

قد تم التعرف على مستوى الكشف في الأدوات، إذا ما كان من خلال تحديد عدد الجمل والعبارات أو عدد محدد من الكلمات، والأخيرة كانت هي النسبة الأكبر؛ حيث تشكل نسبة ٩٥,٨%. ويتم تحديد عدد الكلمات غالباً من قبل المستخدم كما في الشكل رقم (٩).

شكل رقم (٩)

تحديد عدد الكلمات (برنامج AntiPlagiarism.Net)



٤/٤/٩ البنية التقنية:

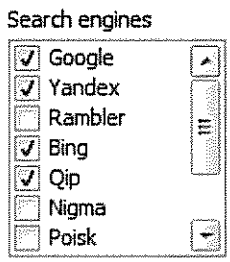
تجاوزت نسبة الأدوات، التي تقدم خدمة الفحص في مجموعات من الوثائق الداخلية كمقالات وبحوث الطلاب، نصف العينة بحوالي ٥٤,٢%. بينما تغطي أغلب الأدوات الفحص المباشر للمصادر المتاحة على الإنترنت وذلك بنسبة ٩١,٧%. واتضح أن نسبة البحث أو الفحص في أحد قواعد بيانات النصوص الكاملة للدوريات أو المجلات أو الكتب تساوي ٢٩,٢%، ويتضح أن

أغلب هذه الأدوات تقدم خدماتها على الإنترنت وذلك بنسبة ٧١,٤%. كما اتضح أن ٧٥% من الأدوات لم تعتمد على فحص المصادر المطبوعة – أي المصادر التي يتم تحويلها فيما بعد إلى شكل إلكتروني. وقد كانت ٧٥% من الأدوات لا تسمح بالاستعانة بأحد محركات البحث في عمليات الفحص والكشف عن الانتحال والسرقات العلمية، كما يبين الشكل الآتي (رقم ١٠) بعض البرامج التي تسمح بالاختيار من عدة محركات بحث يمكن أن تساند عملية الفحص.

وبعد ظهور نتيجة الفحص بالمصادر المتشابهة تُقدم أغلب الأدوات إمكانية حفظ هذه النتائج بنسبة ٧٠,٨%، هذا بخلاف الحفظ والتحميل الهابط للتقارير، وتعتبر هذه الخاصية غايةً في الأهمية حيث يمكن الاستعانة بها كدليل إثبات على السرقات العلمية في لجان تحكيم البحوث. وفيما يتعلق بجدولة الفحص زمنياً بشكل تلقائي لم تضف هذه الخاصية في الأدوات عينة الدراسة.

شكل رقم (١٠)

اختيار من متعدد لمحركات البحث (برنامج (AntiPlagiarism.Net



٥/٤/٩ التحقق من الاستشهادات المرجعية:

أما عن لتحقق من الاستشهادات المرجعية سواء من حيث دقة صياغتها أو حقيقة وجودها أو الإخفاق في عدم كتابتها، وهو من العناصر الفاصلة للتأكد من صحة نقل المعلومات والبيانات من المصادر الأصلية، فوجد أن نسبة الأدوات التي قامت بتصميم خوارزميات الكشف عن الاستشهادات المرجعية لم تزد عن ١٩,٨%. مما يعني أن أغلب الأدوات لم تُعنَ بفحص دقة وصحة الاستشهادات المرجعية؛ مما يؤدي في بعض الأحوال إلى الخروج بنسب انتحال غير دقيقة، ويتطلب هذا معه مجهوداً بشرياً أكبر في عملية التحقق من دقة الاستشهادات المرجعية.

٦/٤/٩ الدقة:

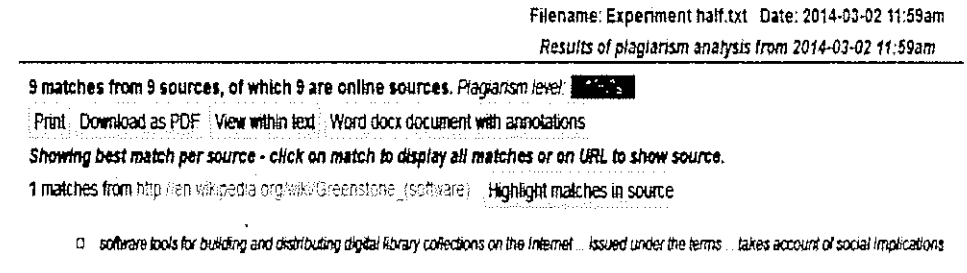
في بعض الحالات تعمل بعض الأدوات على التحقق من صحة ما كتبه المؤلف وذلك بإرسال الوثيقة مرة أخرى وبها مساحات فارغة بالكلمات وعليه أن يكملها كما في الوثيقة الأصلية تماماً، وهذه التقنية لم تطبق بنسبة ٧٩,٢%. وقد تم الإشارة فيما سبق فيما يتعلق بالتصحيح اللغوي والإملائي.

٧/٤/٩ التقارير:

في غالب الأدوات يتم إصدار تقارير بالنتيجة النهائية لعملية الفحص، ولكن تصدر هذه التقارير في شكل واحد فقط دون تنوع في صيغ الملفات بنسبة ٧٠,٨%، وتتراوح بين ملفات Html أو PDF أو Word كما في الشكل رقم (١١). ويأتي التقرير مفصلاً وواضحاً بنسبة ٥٤,٢%. وبعد ظهور نتيجة الفحص يوضح التقرير المصادر الأصلية التي تم الانتحال منها، وهذا يتوافر بنسبة ٧٩,٢%، كما في الشكل رقم (١٢).

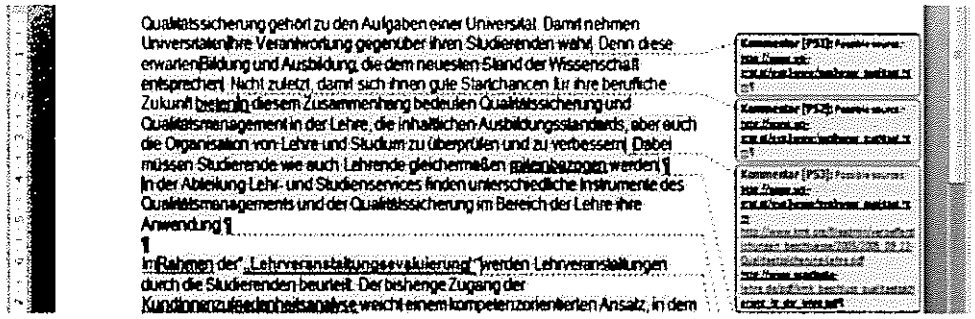
شكل رقم (١١)

نتيجة الفحص (خدمة **PlagScan**)



شكل رقم (١٢)

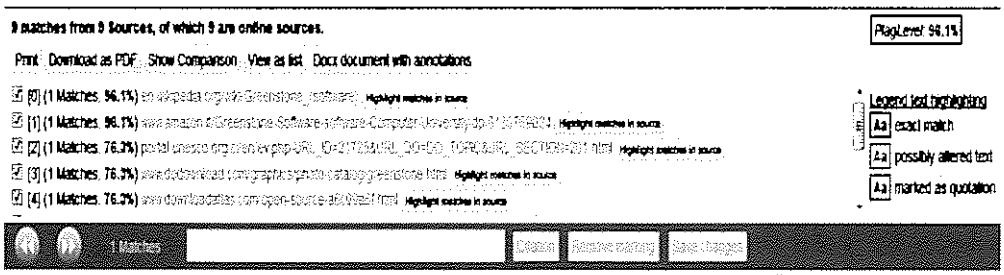
توضيح المصادر الأصلية بأحد ملفات برنامج معالجة النصوص (**MSWord**)



وتكاد تخرج أغلب التقارير في الأدوات بنسب مئوية للتعرف على مدى الانتحال من المصادر الأصلية، وذلك كان بنسبة ٩١,٧%، ويبين الشكل الآتي رقم (١٣) تقريراً من أحد البرامج، ويظهر به نسب مئوية على كل مصدر أصلي تم الأخذ منه، وبعد ذلك يتم حساب إجمالي النسبة المئوية للنص المنتحَل أو المقتبس منه.

شكل رقم (١٣)

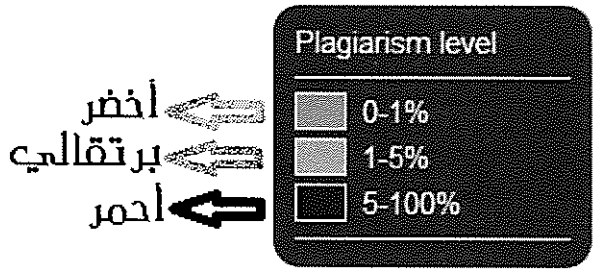
تقرير بنسب المصادر التي تم الانتحال منها (خدمة **PlagScan**)



وفي بعض الأحوال تقوم بعض الأدوات بتمييز ألوان النصوص التي تم انتحالها أو سرقتها من النصوص والمصادر الأصلية بألوان مختلفة تدل على مستوى الانتحال مقابل النسبة المئوية، وذلك كما في الشكل رقم (١٤)، الذي يوضح اللون المناسب للنسبة المئوية لتسهيل التعرف على مستوى السرقة والانتحال في البحث.

شكل رقم (١٤)

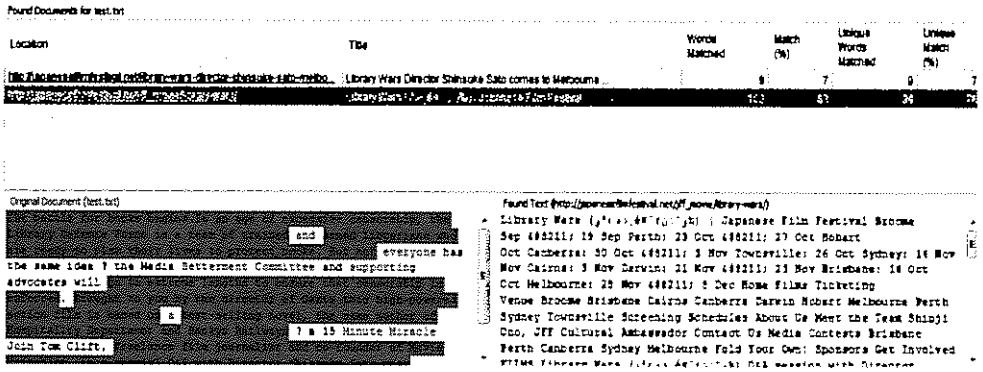
تحديد ألوان نسب السرقات العلمية (خدمة PlagScan)



وقد لوحظ أن استخدام الألوان عامل لا غنى عنه في مثل هذه الأدوات للتعرف على النص الذي تم اقتباسه أو سرقته كما هو من المصادر الأصلية، وذلك بنسبة ٦٦,٧%، ويتضح ذلك في بعض البرامج كما في الشكل رقم (١٥)، الذي يُظهر نصاً تم اقتباسه كما هو بلون أحمر على اليسار من النص الأصلي الذي يظهر في المربع على اليمين لتسهيل عملية المقارنة، كما يظهر في أعلى هذا المربع الموقع الإلكتروني المتاح على الإنترنت لسهولة التأكد والانتقال إليه في الحال.

شكل رقم (١٥)

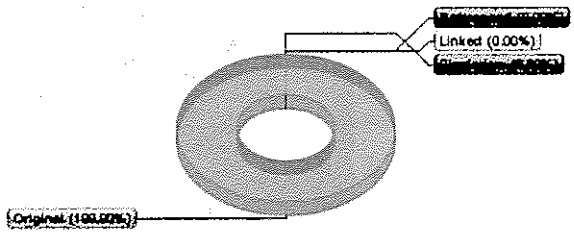
النصوص المتشابهة للنص الأصلي (برنامج Viper)



وقد خرجت أغلب التقارير ونتائج الفحص للسرقات والانتحال دون رسوم بيانية توضح وتبين مدى الانتحال في البحث المراد فحصه وذلك بنسبة ٩١,٧%، أي أن ٨,٣% فقط من الأدوات اهتمت بتزويد نتيجة الفحص برسم بياني كما في الشكل رقم (١٦)؛ مما يوضح أن المقالة أو البحث خالٍ بنسبة ١٠٠% من سرقات أو انتحال، كما يكشف الشكل عن نسبة المراجع والروابط التي يمكن أن تثبت ذلك.

شكل رقم (١٦)

رسم بياني عن نتيجة الفحص (برنامج Plagiarism Detector)



Referenced 0% / Linked 0%
Original - 100% / 0% - Plagiarism

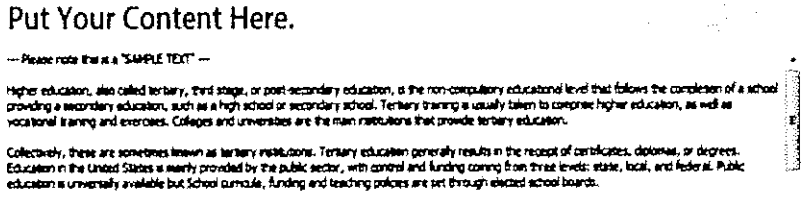
وتزود العديد من الأدوات بإمكانات طباعة أو تحميل التقرير بنسبة ٦٢,٥%، ولكن نسبة الأدوات التي تسمح بمشاركة التقارير مع حسابات أخرى أو حسابات فرعية للحساب الرئيس لا تتعدى ١٣%، وذلك يمكن أن يكون بسبب جوهري يتمثل في أن أغلب البرامج لا تتيح إمكانية الاشتراك بحساب رئيس ويشترك معها في نفس الوقت مجموعة الحسابات الفرعية للعديد من الأشخاص أو الطلاب أو غير ذلك.

وفيما يتعلق بتوافر إمكانية إرسال تقارير الفحص عبر البريد الإلكتروني، قد تبين أن ٦٢,٥% من الأدوات لم تدعم هذه الإمكانية، حيث تقتصر على حفظ التقارير أو طباعتها.

٩/٥ طريقة عمل الأداة:

شكل رقم (١٧)

إمكانية نسخ ولصق النص (برنامج Plagiarism Checker)

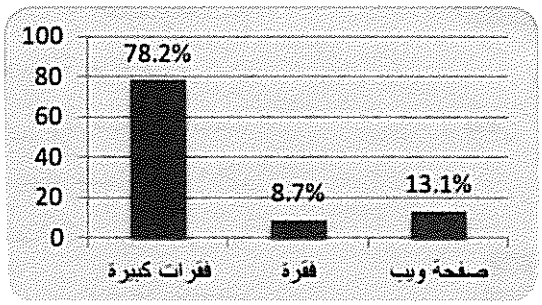


تتوفر خاصية النسخ واللصق لإجراء علمية الفحص في حالة التعذر لرفع الملفات إلى الأداة، أو إذا كانت صيغة الملف لا تدعمها الأداة، أو في حالة رغبة المستخدم فعل ذلك لأي سبب آخر، وقد وصلت النسبة إلى ٧٩,٢% من الأدوات توفر مثل هذه الإمكانية، كما هو واضح في الشكل رقم (١٧)، وتتفاوت هذه النسبة بين أدوات تقبل فقرات كبيرة وأخرى تقبل فقرة

واحدة فقط، والبعض يوفر إمكانية نسخ ولصق صفحة ويب ليتم فصحها، ويبين الشكل رقم (١٨) النسب الموزعة بين هذه السمات الثلاثة، وهناك بالتأكيد إمكانية لصق فقرات كبيرة في مربع الفحص للكشف عن أصالتها.

شكل رقم (١٨)

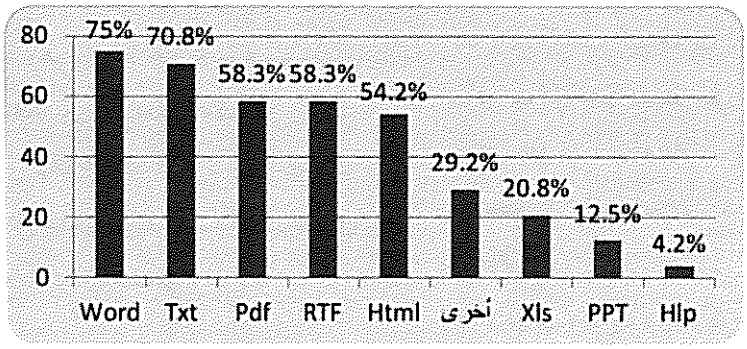
حجم النص المقبول خلال عملية اللصق



وفي مقابل النسخ واللصق، توفر بعض الأدوات إمكانية رفع الملفات إلى الأداة لتسهيل عملية الفحص، واتضح أن ٧٥% من الأدوات توفر مثل هذه الإمكانية، وكان لابد من التعرف على أكثر الملفات التي تقبلها تلك الأدوات وتتوافق مع برمجتها كما في الشكل رقم (١٩)، وكانت صيغة ملفات برامج معالجة النصوص (MS Word) بنسبة ٧٥%، ثم صيغة (Txt) بنسبة ٧٠,٨%، وتليها صيغة (PDF) و (Rtf) بنفس النسبة ٥٨,٣%، وكذلك صيغة (Html) بنسبة ٥٤,٢%. بينما لم تدعم العديد من الأدوات بنسبة ٧٩,٢% إمكانية تحويل الملفات النصية إلى صيغ نصية أخرى تقبلها الأداة أو أن يتم تحويلها إلى نص مرئي في مربع يظهر به ما هو مكتوب في هذا الملف لإجراء وإتمام عملية الفحص بشكل مباشر.

شكل رقم (١٩)

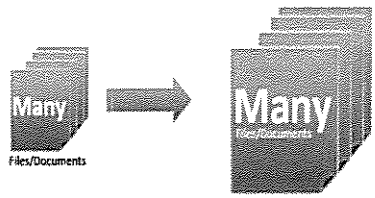
أكثر صيغ الملفات التي تقبلها أدوات كشف الانتحال



وعند قياس قدرة الأداة على مدى التوسع في عملية الفحص عن مدى التشابه بين النصوص، يكون ذلك من خلال التعرف على الطرق المتبعة في المضاهاة بين النصوص والملفات. وقد يتوافر بالأداة الكثير من هذه السمات أو سمة واحدة فقط، الحالة الأولى وهي المضاهاة بين ملف وبين ما هو متاح من موارد على الإنترنت، وكانت هذه النسبة هي الأعلى حيث تمثل ٩٥,٨%، والحالة الثانية والمتمثلة في المضاهاة بين مجموعة ملفات مقابل مجموعة أخرى من الملفات كانت نسبتها ٤٥,٨% وتم تمثيلها في الشكل رقم (٢٠).

شكل رقم (٢٠)

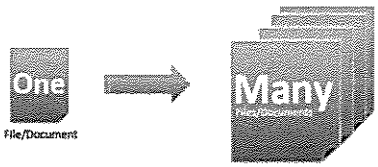
مجموعة ملفات مقابل مجموعة من الملفات



والحالة الثالثة وهي المضاهاة بين ملف واحد وعدة ملفات يتم تحديدها من قبل المستخدم وكانت النسبة تساوي ٣٣,٣%، وقد تم تمثيلها في الشكل رقم (٢١)، وهذه الحالة تستخدم حين يتم مقارنة ملف مع عدة ملفات ذات علاقة به. أما الحالة الأخيرة وهي عبارة عن مجموعة من الملفات يتم المضاهاة بين كل ملف مقابل كل ملف على حدة وكأنها محاكاة لمستودع واحد به كل الملفات، وتمثل هذه الطريقة نسبة تساوي ٢١,٧% وتتضح أكثر في الشكل رقم (٢٢).

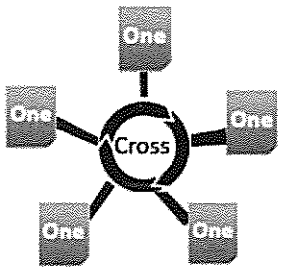
شكل رقم (٢١)

مقارنة ملف مقابل عدة ملفات



شكل رقم (٢٢)

فحص كل ملف مقابل كل ملف



وجاءت النتيجة ٠% فيما يتعلق بمدى توفر إمكانية اختيار أكثر من لغة بغرض توسيع نطاق الفحص وما إذا كان قد تم ترجمته إلى لغات أخرى.

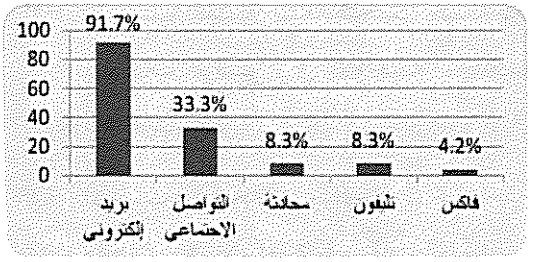
٦/٩ المساعدة والدعم الفني:

وفي محاولة للتعرف على مدى اهتمام الأدوات محل الدراسة بإعلام وإحاطة المستخدم وإمداده بمعلومات حول موضوع "السرقات والانتحال للنصوص" من خلال ما تنشره من نشرات، كانت النتيجة أن نسبة ٦٢,٥% من الأدوات لم تهتم بتقديم مثل هذه المعلومات. وفيما يتعلق بإذا ما كان هناك أسئلة متكررة حول كيفية استخدام الأداة والتعريف بها جاءت النتيجة بنسبة ٥٠%. بينما تتوفر نشرات التعريف بالأداة بنسبة ٨٣,٣%، ولم يتوفر بالأدوات المحددة أي وسائط متعددة يعتمد عليها المستخدم في التعرف على

كيفيه تشغيل الأداة وما تضمه من سمات ووظائف بها وكانت النسبة ٥٨,٣%؛ ولتقديم خدمة الدعم الفني بالوسائل الإلكترونية كما في الشكل رقم (٢٣) وجد أن أكثر الوسائل استخداماً في تقديمه تمثلت في البريد الإلكتروني بنسبة ٩١,٧%، بينما جاءت مواقع التواصل الاجتماعي في المرتبة الثانية وتمثلت بنسبة ٣٣,٣%، وتليها إمكانية إجراء دعم فني عبر برامج المحادثة الفورية بنسبة ٨,٣%، ونفس النسبة كانت من نصيب الهواتف، وأخيراً الفاكس بنسبة ٤,٢%، وفيما يتعلق بتوقيت تقديم الخدمة إذا كان يقدم بشكل متواصل دون انقطاع خلال اليوم وطوال أيام الأسبوع، وجد أن نسبة ٧٩,٢% من الأدوات لم تقدم تلك الميزة ضمن باقاتها، بينما على العكس ظهر أن نسبة ٢٠,٨% من إجمالي الأدوات أكدت تقديم خدمات الدعم الفني بشكل مستمر دون انقطاع.

شكل رقم (٢٣)

أبرز وسائل الاتصال للدعم الفني



١٠- النتائج:

١- عدم وجود سياسات واضحة وصارمة للانتحال والسرقات العلمية.

٢- تعتبر الوسائل والطرق التقليدية هي التي تجد سبيلها في كشف الانتحال والسرقات العلمية دون الأدوات الحديثة، كما أن لها الأفضلية عنها.

٣- ضعف نسب الاستفادة من الأدوات الحديثة في كشف الانتحال والسرقات العلمية بشكل فردي، وانعدام الاعتماد عليها في تحكيم البحوث في مجال المكتبات والمعلومات.

٤- يعد من أهم الأسباب وراء العزوف عن استخدام أدوات كشف الانتحال والسرقات العلمية عدم دعم وتمثيل اللغة العربية في أدوات كشف الانتحال والسرقات العلمية على عكس اللغات الأخرى؛ وبالتالي كان العنصر الثاني في الترتيب هو عدم الموثوقية في نتائج هذه الأدوات.

٥- لم تتعدَ العقوبات التي يمكن أن تقع على الباحثين المنتحلين عقوبة إبلاغ الجهة أو رفض البحث أو وضع الباحث في القائمة السوداء، وربما تعديل البحث بما يتوافق مع شروط البحث العلمي.

٦- تعتبر نسبة تمثيل اللغة العربية في أدوات كشف الانتحال ضعيفة للغاية وغير مجدية في اكتشاف السرقات العلمية.

١١- التوصيات:

١- ضرورة تضافر الجهود من المؤسسات البحثية المهتمة بتكنولوجيا المعلومات لإنشاء أداة تدعم اللغة العربية بصورة كبيرة لتقليل نسبة السرقات العلمية في البحوث باللغة العربية، فضلاً عن أنها يمكن أن تحسن بشكل غير

مباشر من سلوك الباحثين في الانتباه لمخاطر السرقة العلمية.

٢- لابد من توفير برامج معتمدة في كشف الانتحال والسرقات العلمية في المؤسسات البحثية والأكاديمية، تدعم اللغة العربية وبالتأكيد اللغة الإنجليزية.

٣- صياغة وتنفيذ سياسة واضحة المعالم والأركان تتضمن كافة القضايا والأوجه المختلفة للسرقات العلمية.

٤- من المستحسن لأعضاء اللجان العلمية لفحص الإنتاج الفكري الاعتماد على أي برامج تكشف الانتحال والسرقات العلمية سواء بشكل رسمي أو فردي، فضلاً عن استخدامها من قبل أعضاء هيئة التدريس لمتابعة بحوث الطلاب في المرحلة الجامعية الأولى وخاصةً طلاب الدراسات العليا.

٥- لابد من إدماج خدمات كشف الانتحال مع برامج التعليم الإلكتروني.

٦- يجب على الباحثين الاهتمام بالأرشفة الذاتية لحماية حقوقهم الفكرية، وتزويد المستودعات الرقمية دوماً بالبحوث العلمية بغرض زيادة إنتاجية المحتوى الرقمي العربي.

٧- التوعية للطلاب والباحثين وأعضاء هيئة التدريس وهيئات التحكيم بأدوات كشف الانتحال والسرقات العلمية.

٨- اتباع أحد الأساليب المعيارية للتوثيق العلمي للوثائق الورقية والرقمية.

٩- يمكن الاستعانة بأحد محركات البحث الداعمة للغة العربية عند تصميم البنية الهيكلية لأداة كشف الانتحال والسرقات العلمية.

١٢- الدراسات المستقبلية

١- دراسة أفضل الأدوات المستخدمة في الكشف عن الانتحال والسرقات العلمية كعناصر مرجعية لمقارنتها بأداة للنص العربي.

٢- دراسة اتجاهات أعضاء هيئة التدريس والطلاب نحو استخدامات أدوات كشف الانتحال والسرقات العلمية.

٣- الأساليب العلمية المتبعة لتجنيب الباحثين والطلاب الوقوع في الأوجه المختلفة للسرقات العلمية.

٤- الأسباب والمشكلات وراء سلوك الباحثين والطلاب في الاستهانة بالسرقات العلمية.

المراجع

(١) ابن منظور (٢٠٠٣). لسان العرب. بيروت: دار صادر.

(٢) الجوة، ماهر، الجوة، فاطمة، بلغيث، ليساء و حمادو، عبدالمجيد (٢٠١١). كشف حالات الانتحال في النصوص المدونة باللغة العربية بالاعتماد على السلاسل اللغوية. الجمعية العربية للحاسبات والاتصالات، ٤(٢).

(٣) الحوراني، خالد (د.ت). المحلل المصرفي للغة العربية: المصرف. ٥ص.

(٤) الربيعة، مها سليمان و السلمان، عبدالملك (٢٠١٠). التطورات والاتجاهات الحديثة في استرجاع المعلومات العربية. في: ورشة عمل إثراء المحتوى الرقمي العربي، أكتوبر، دمشق. تاريخ الإطلاع ٢١ نوفمبر، ٢٠١٣. استرجعت من http://mahaalrabiah.files.wordpress.com/2011/01.

(٥) السالم، سالم بم محمد (٢٠١٠، ٦-٧ أبريل). السرقات العلمية في البيئة الإلكترونية: دراسات للتحديات والتشريعات المعنية بحماية حقوق التأليف، في: مؤتمر البيئة المعلوماتية الآمنة: المفاهيم والتشريعات والتطبيقات، الرياض، ٢٠١٠.

(٦) الشامي، أحمد محمد (٢٠١١). مصطلحات المكتبات والمعلومات والأرشيف. تاريخ الإطلاع ١٠ أكتوبر، ٢٠١٣. استرجعت من http://www.elshami.com.

(٧) الصوينع، علي بن سليمان (٢٠٠٨). استرجاع المعلومات في اللغة العربية. الرياض: مكتبة الملك فهد الوطنية.

(٨) عيسى، عماد و السيد، أماني (٢٠١٢، ٩-١١ يوليو). دور المكتبات الأكاديمية في منع السرقات العلمية واكتشافها: دراسة استكشافية لخدمات المكتبات وبرمجيات كشف الانتحال. في: المؤتمر الدولي للتعلم الإلكتروني في الوطن العربي: مشكلاته وآفاق تطويره، القاهرة، ٢٠١٢. تم استرجاعه ٠٣ نوفمبر، ٢٠١٣، من http://www.slideshare.net/esalh/ss-13988012.

References

(1) Alzahrani, S. M., Salim, N., & Alsofyani, M. M. (2009, April). Work in progress: Developing Arabic plagiarism detection tool for e-learning systems. In Computer Science and Information Technology-Spring Conference, 2009. IACSITSC'09. International Association of (pp. 105-109). IEEE.

(2) Atkinson, D., & Yeoh, S. (2008). Student and staff perceptions of the effectiveness of plagiarism detection software. Australasian Journal of Educational Technology, 24(2), 222-240. Retrieved 03 November, 2013, from http://ascilite.org.au/ajet/ajet24/atkinson.html.

(3) Bahadori, M., Izadi, M., & Hoseinpourfard, M. (2012). Plagiarism: Concepts, Factors and Solutions. MilMed Journal, 14(3), 168-177.

(4) Bensalem, I., Rosso, P., & Chikhi, S. (2012). Intrinsic Plagiarism Detection in Arabic Text: Preliminary Experiments. In II Spanish Conference on Information Retrieval (CERI'12). Retrieved 22 November, 2013, from http://users.dsic.upv.es/grupos/nle/ceri/papers/ceri2012_bensalem.pdf.

(5) Bin-Habtoor, A. S., & Zaher, M. A. (2012). A Survey on Plagiarism Detection Systems. International Journal of Computer Theory and Enginee-ring, 4(2), 185-188.

(6) Chuda,D., Navrat, P. (2010). Support for checking plagiarism in e-learning. Procedia Social and Behavioral Sciences, 2, 3170-3144. Retrieved 22 December, 2013, from

(7) Clough, P. (2003). Old and new challenges in automatic plagiarism detection. In National Plagiarism Advisory Service, Retrieved 29 October, 2013, from http://ir.shef.ac.uk/cloughie/index.html.

(8) HaCohen-Kerner, Y., Tayeb, A., & Ben-Dror, N. (2010, August). Detection of simple plagiarism in computer science papers. In Proceedings of the 23rd International Conference on Computational Linguistics (pp. 421-429). Association for Computational Linguistics. Retrieved 23 November, 2013, from, https://www.aclweb.org/anthology-new/C/C10/C10-1048.pdf.

(9) Hariharan, S. (2012). Automatic plagiarism detection using similarity analysis. Int. Arab J. Inf. Technol., 9(4), 322-326.

(10) Heather, J. (2010). Turnitoff: Identifying and fixing a hole in current plagiarism detection software. Assessment & Evaluation in Higher Education, 35(6), 647-660. Retrieved 03 November, 2013, from http://epubs.surrey.ac.uk/107387/2/turnitoff-named.pdf.

(11) Jadalla, A., & Elnagar, A. (2012). A plagiarism detection system for arabic text-based documents. In Intelligence and Security Informatics (pp. 145-153). Springer Berlin Heidelberg. doi: 10.1007/978-3-642-30428-6_12

(12) Masic, I. (2012). Plagiarism in scientific Publishing. Acta Informatica Medica, 20(4), 208.

(13) Maurer, H., Kappe, F., & Zaka, B. (2006). Plagiarism: A survey. Journal of Universal Computer Science,12(8), 1050-1084. Retrieved 05 November, 2013, from http://citeseerx.ist.psu.edu/viewdoc/downlo ad?doi=10.1.1.102.5629&rep=rep1&type=pdf.

(14) McKeever, L. (2006). Online plagiarism detection services—saviour or scourge?. Assessment & Evaluation in Higher Education, 31(2), 155-165.

(15) Menai, M. E. (2012). Detection of Plagiarism in Arabic Documents. International Journal of Information Technology and Computer Science (IJITCS), 4(10), 80-89. Retrieved 24 November, 2013, from http://www.mecs-press.org/ijitcs/ijitcs-v4-n10/IJITCS-V4-N10-10.pdf.

(16) Menai, M. E., & Bagais, M. (2011, August). APlag: A plagiarism checker for Arabic texts. In Computer Science & Education (ICCSE), 2011 6th International Conference on (pp. 1379-1383). IEEE.

(17) Mozgovoy, M., Kakkonen, T., & Cosma, G. (2010). Automatic student plagiarism detection: future perspectives. Journal of Educational Computing Research, 43(4), 511-531. Retrieved 22 November, 2013, from, http://www4.ncsu.edu/~brad_m/research/plagiarism_files/Mozgovoy_automaticPlagiarism_JECR10.pdf.

(18) Ottenstein, K. J. (1976). An algorithmic approach to the detection and prevention of plagiarism. ACM Sigcse Bulletin, 8(4), 30-41.

(19) Pardeshi, C. V., Rajput, P. V., Chaudhary, K. S., & Patil, G. B. (2012). Plagiarism in scientific research: needs lock-up to unlock the ethical publications. Current Trends in Biotechnology and Pharmacy, 6(1), 28-34. Retrieved 16 November, 2013, from http://www.abap.co.in/files/CTBP_6_1.pdf#page=34.

(20) Patel, A., Bakhtiyari, K., & Taghavi, M. (2011). Evaluation of cheating detection methods in academic writings. Library Hi Tech, 29(4), 623-640. Retrieved 18 December, 2013, from http://www.researchgate.net/publication/216345619_Evaluation_of_cheating_detection_methods_in_academic_writings/file/79e41511ce513b4eba.pdf.

(21) Reitz, J. M. (2007). ODLIS: Online Dictionary for Library and Information Science. Retrieved 30 September, 2013, from http://www.abc-clio.com/ODLIS/odlis.aspx.

(22) Roig, M. (2010). Plagiarism and self-plagiarism: What every author should know. Biochemia Medica, 20(3), 295-300. Retrieved 07 November, 2013, from http://hrcak.srce.hr/file/89432.

(23) Shahabuddin, S. (2009). Plagiarism in Academia'. International Journal of Teaching & Learning in Higher Education, 21(3), 353-359. Retrieved 05 November, 2013, from http://www.isetl.org/ijtlhe/pdf/IJTLHE641.pdf.

(24) Stowers, R. H., & Hummel, J. Y. (2011). The Use of Technology to Combat Plagiarism in Business Communication Classes. Business Communication Quarterly, 74(2), 164-169. Retrieved 07 November, 2013, from http://bcq.sagepub.com/content/74/2/164.

(25) Vernon, R. F., Bigna, S., & Smith, M. L. (2001). Tech Notes: Plagiarism and the Web. Journal of Social Work Education, 37(1), 193-196.

(26) Oxford University (2013). Oxford Dictionary. Retrieved 10 October, from http://www.oxforddictionaries.com.