

دراسة بحثية في كلية تكنولوجيا المعلومات تبين كشف الاستلال بناءً على التحليل النصي و الدلالي

بينت دراسة بحثية مشتركة في كلية تكنولوجيا المعلومات اعنتها الباحثة هديل قاسم الخفاجي مع الدكتورة ايمان صالح الشمري والموسومة بـ:
" كشف الاستلال بناءً على التحليل النصي و الدلالي "
وبينت الباحثة
ان ابسط وصف للاستلال هو اما ان يكون نسخ ولصق نص حتى وان تمت الاشارة الى المصدر او تغيير بعض الكلمات بأخذ المعنى بدون الاشارة الى المصدر , وحيث ان تحديد المعنى هو الاصعب والاعقد .
يمكن ان ينظر الى الاستلال باعتباره واحدا من الجرائم الإلكترونية مثل (فيروسات الكمبيوتر، قرصنة الكمبيوتر، البريد الإلكتروني غير المرغوب وانتهاك حقوق التأليف والنشر)، لذلك اصبح هذا الموضوع مثيرا للاهتمام لأنه جزء مهم من أخلاقيات البحث العلمي.
تم تحقيق ثلاث اهداف في هذا العمل ، الاول هو بناء قاعدة بيانات ذات هيكل خاص لتخزين الوثائق واستخدامها للمطابقة مع المستند المصدر (الوثيقة المدخلة) للكشف عن الاستلال .
الثاني هو الكشف عن الاستلال النصي، حيث ان خوارزمية shingle ومعامل Jaccard تم توظيفهما لهذا الغرض. ان خوارزمية Shingle التي تستخدم كخاصية لقياس التشابه النصي في الوثائق هي وسيلة فعالة لمقارنة مجموعات من النصوص وحيث انها ستعمل مع معامل Jaccard الذي يقيس التشابه بين المجاميع . أيضا في هذه المرحلة ، تم تحديد مواقع النصوص المستلة من حيث وجودها في الخلاصة ، النتائج والاستنتاجات. وقد أعطيت هذه المواقع أولوية بفرض أنه يجب أن تكون خالية من الاستلال.
الثالث هو الكشف عن الاستلال الدلالي في الخوارزمية او النظام المقترح المكتوب في البحوث الاستلال الدلالي مازال مشكلة كبيرة حيث ان المنتحل يقوم بتغيير كلمات أو جمل إلى كلمات ذات معاني مماثلة للتغلب على هذه المشكلة
تم استخدام WordNet حيث يعطي مرادفات للكلمات الموجودة في الوثيقة. في هذه المرحلة ، تم توسيع WordNet بكلمات اقترحت أن تكون مرادفات لكلمات اخرى شائعة الاستخدام عند كتابة البحوث.
هذه الطرق تم اثبات فعاليتها في مجالات الاستلال النصي والدلالي. تحديد مواقع النصوص المستلة أثبتت وجود الاستلال النصي في الاجزاء المهمة من البحث ، وتوسيع WordNet قد اعطى نتائج جيدة في كشف الاستلال الدلالي .
وأخيرا تم تنفيذ النظام المقترح باستخدام لغة Java NetBeans IDE 8.0.2 وتم بناء قاعدة البيانات بواسطة برنامج MySQL Workbench