أطار عمل عالي الجودة لجمع البيانات لأنظمة البيانات الضخمة

إعداد

الاء محمد حماد

المشرف

د. محمد عبدالله

جامعة الزيتونة الأردنية، 2021

الملخص

أدت الزيادة السريعة في تكنولوجيا المعلومات و الصناعات و المواقع الاخبارية و التعليمية وغيرها إلى ظاهرة البيانات الضخمة, ،حيث تم تجميع البيانات الضخمة باستخدام عدة طرق من ضمنها اجهزة الاستشعار ،اطار عمل ،انظمة و الويب ،وبعد ذلك ظهرت مجموعة من المعاير و التشريعات مثل GDPR وISO /IEC 25010 لتحسين نوعية البيانات و لكن لم يتم تنفيذها بشكل كامل في جميع المنظمات وذلك بسبب تنوع الطبيعة التشغيلية للمؤسسات ,و كما انها لم تغطي انواع مختلفة من البيانات ،لذلك تم تطوير اطار عمل لجمع البيانات ذات الجودة من الانظمة الضخمة حيث يتكون من سبع عوامل رئيسية (موثوقية مصدر البيانات ، والجدارة بالثقة ، وملائمة البيانات ، وحفظ البيانات ، وسالمة البيانات ، وجمع البيانات السريع ، وإمكانية إعادة استخدام البيانات) و عشرون عامل فرعي ، تم تطوير اطار العمل عن طريق دمج مجموعة من تشريعات GDPR و عناصر ISO /IEC 25010 مع إضافة مجموعة من العوامل لتحسين كفاءة عملية جمع البيانات والحصول على بيانات عالية الجودة, تم التحقق من صحة عملية جمع البيانات عالية الجودة وتحليلها باستخدام المرشحات المناسبة في أداة Weka وباستخدام ابعاد جودة البيانات (DQD) مثل حساب البيانات المفقودة ، وتم تقليل كمية البيانات الضائعة ، وما إلى ذلك ، ولكن المرشحات لم تغطي عاملين فرعيين مصدر موثوق وبيانات قانونية لذلك تم تضمينها في العقد المبرم بين مزود البيانات ومتلقي البيانات ..