/10التنقيب عن البيانات الضخمة في المعلومات الحيوية: Big Data Miningتكمن أساسيات المعلوماتية الحيوية ف إدارة التحليل والتعامل مع المجاميع الكبيرة من البينات الحيوية،* تهيئة قواعد بيانات تسمح بخزن وإدارة وإضافة وتجديد البيانات الحيوية.* تطوير الخوارزميات والبرمجيات التي تعمل على تحديد،* تفسير الأنماط التي تنتج من عملية تحليل الأنواع المختلفة من البيانات الحيوية (تواليات DNAوالتنقيب عن البيانات هي عملية استخراج معلومات مفيدة من بيانات كبيرة غير منظمة،البيانات أوراق أو ملخصات أو صفحات على شبكة الإنترنت أو رسائل بريد إلكتروني.غير منظمة أساسا، فإنه يتعين تمثيلها ف هيكل من نوع ما،النصوص،النصوص (،KDTوتتمثل صعوبة التنقيب عن البيانات ف المعلوماتية الحيوية والمصطلحات ف المجال الطبيليست متسقة ف العديد من المصادر،مجال المعلوماتية الحيوية، ويتم حل مشكلة الاستخدام الفعال للبيانات عن طريق استخدام تقنيات استخراجPubMedحيث يتم تحليلPeng, 2018وهذا من شأنهالمستخرجة.والمعالجة اللاحقة،وإزالة كلمات التوقف، وتقطيعه وتحويله إلى جمل منظمة، ليبقى ف النهاية الجزء الرئيسي من النص المستخدم فـعملية التنقيب عن البيانات باستخدام نماذج تعلم الآلة أو الشبكات العصبية، وتتضمن خطوة ما بعد التجهيزتفسير وتصور نتائج عمليات استخراج النصوص مهام التنقيب عن المعلوماتية الحيوية، ودور أخصائي المكتبات والمعلومات: 2013)1استرجاع المعلومات (:)Information Retrievalتعني البحث عن المستندات ذات الصلةبالاستعلام، يعد PubMed/MEDLINEمثالًا على نظام استرجاع المقالات العلمية ف مجال الطب الحيوي؛)2تصنيف الوثائق (:)Document Classificationتصنيف الوثائق هي مهمة تصنيف المستندات إلىفئات موضوعية،تصنيف مقالات الدوريات حول ما إذا كانت تتعلق بجنين معين أم لا؟. عادةً ما يستخدم تصنيف المستنداتCaporaso, وأسماء الجينات،إلخ اللائحةالتي يتم تحديدها مسبقًا، تشمل الكيانات التي يتم البحث عنها والتعرف عليها، والتي تسمى ""MetaMapبأسماء الجينات الجديدة التي يتم صياغتها باستمرار، وأن العديد من الجينات لها أسماء أو مرادفات مختلفة، ومعذلك، فإن هذا النهج القائم على استخدام القوائم المسبقة أو المكانز (القوامسيس) يمكن أن يحقق نجاحًا إذا)4تطبيع الكيانات الحيوية (:)Name Entity Normalizationتطبيع الكيان هي عملية البحث عن كيانمحدد ف قواعد البيانات الحيوية، ويتم إجراء تلك المهام على نطاقواسع على الجينات والبروتينات،ـ المشكلة الأولى: العديد من أنواع الجينات لها نفس الاسم، وبالتالي، فإن العثور على الجين المناسب يتطلب معرفة الأنواع قيد المناقشة، وهي المشكلة البحثيةالقائمة بحد ذاتها، والتي يرغب الباحث ف التعرف عليها.ـ المشكلة الأخرى: الجين الواحد قد يكون له أسماء متعددة، مثل: جين ( )1–TRPالذي يتواجد لدىالجين لتحديد ماهية الجين المطلوب، من خلال المقارنة بين السياق ف النص وبين حقول البيانات: (العنوان،المستخلص، النص. أو الربط بين الجين المطلوب البحث عنه)5استخراج المعلومات والعلاقات (:)Relation & Information Extractionتعني عملية استخراج فعلى سبيل المثال: استخراج المعلومات الجينومية، وتفاعلات البروتين، وعلاقات نقل البروتين، وموقع بداية البروتين المنقول، وتستخدم تلك المهمةالأساليب المستندة بناء الجمل النموذجية، لكن وبخاصة تحليل التبعية والعلاقةHunter, Lu & Firby,بين المعلومات)6الإجابة على الاستفسارات (:)Question–Answeringافترضت الأساليب المبتكرة أن العديد منالنصية. حيث إن الهدف هو إرجاع إجابة محددة،مستندًا يحتوي الإجابة، بينما يركزاستخراج المعلومات على احتياجات معلومات محددة للغاية، Lee & Kaufman, حيث وجد أن درجات الأسئلة تتفاوت حيث يتم تحقيق أفضل النتائج مع الأسئلة "الواقعية ،"Factoidمثل :"أين"، ف حين أن) Zweigenbaum, 2003والأسئلة الزمانية"متى" قد تتطلب أوقاتًا لإجابتها.)7التلخيص (:)Summarizationيعني أخذ مستند أو مجموعة من المستندات كمدخلات وإرجاع نصأقصر، هناك حاجة كبيرة لهذه الإمكانية ف المجال الطبي الحيوي، وقد تم تطبيق التلخيص على مقالات ومن أمثلة تلك الأنظمة التي تقوم بذلك نظام 2013))MITRE'S MiTAP(والتقارير الطبية، وف مجال هناك العديد من النماذج المستخدمة ف تلخيص الإنتاج الفكري الحيوي بما ينتج عنه معلومات تقومحرفًا، تعمل على إدخال وظيفة الجين ف قاعدة بيانات .* ثانيها: نموذج ( )Edmundsonianوهو نهج يعمل على إعطاء الجمل الملخصة نقاطًا (وزنًا) وفقًالمجموعة من المعايير النسبية، مثل: موضع الجملة ف الوثيقة، فالعنوان يختلف عن المستخلص وعن النص كذلك نوع الجمل ما بين "الكلمات الجدلية "Cue Wordsالتي تشير إلى ارتباط الجملة بالنص،ويبقى التحدي هنا ف إيجاد أفضل جملة تعبر عن النص، كذلك الجمل التي تبرز بعض الحقائق حول عملية مثل: "تأكيد تفاعل بروتين معين أم لا" من داخل النص"، ويتم التغلب على ذلكبفكرة BioCreativeالمتمثلة ف اشتراك مجموعة من المقالات ف إثبات التساؤل ذاته، لتعطي أفضل دليل علىحدوث هذا التحليل (التفاعل،باستخدام خوارزميات معنية بذلك.نماذج من تطبيقات التنقيب عن البيانات الحيوية:1) Biodoop – Bioinformatics on Hadoop:أباتشي هادوب هو برنامج

lakhasly.com © تم تلخيص النص بواسطة موقع لخصلي

أو منصة برمجية مفتوحة المصدر مكتوبة بلغة ألجافا لتخزين ومعالجة البيانات أي: تخزين البيانات الضخمة على عدة أجهزة ومن ثم توزيع عملية المعالجة على هذهالمعلوماتية الحيوية) التي تركز على التنفيذ الفعال للمهام ذات البيانات الحيوية الضخمة، وهو يتألف منمكونات خاصة وتطبيقات متعلقة بعمليات تحليل المعلوماتية الحيوية، مما يجعلها مناسبة بشكل خاص لمعالجة مجموعات البيانات الكبيرة.Leo, Santoni & Zanetti, 2009)من أمثلة التطبيقات التي تقوم بهذه الوظائف تطبيق تحليل النصوص LitMinerالذي أعد لتمكينالكاملة، فيوفر التطبيقالتبديل بين التطبيقات المتعددة، ويتميز التطبيق بالاتي:)(Demaine, Martin & Wei, 2006ـ إمكانية استيراد الملفات من قواعد البياناانات المتعددة،

lakhasly.com © تم تلخيص النص بواسطة موقع لخصلي