تكمن أساسيات المعلوماتية الحيوية ف إدارةBig Data Mining :التنقيب عن البيانات الضخمة في المعلومات الحيوية10/ 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 والتنقيب عن البيانات هي عملية استخراج معلومات مفيدة من بيانات كبيرة غير منظمة،البياناتDNA البيانات الحيوية (تواليات أوراق أو ملخصات أو صفحات على شبكة الإنترنت أو رسائل بريد إلكتروني.غير منظمة أساسا، فإنه يتعين تمثيلها ف هيكل من وتتمثل صعوبة التنقيب عن البيانات ف المعلوماتية الحيوية والمصطلحات ف المجالKDT،) نوع ما،النصوص،النصوص الطبيليست متسقة ف العديد من المصادر،مجال المعلوماتية الحيوية، ويتم حل مشكلة الاستخدام الفعال للبيانات عن طريق وهذا من شأنهالمستخرجة.والمعالجة اللاحقة،وإزالة كلمات2018 ,Pengحيث يتم تحليلPubMedاستخدام تقنيات استخراج التوقف، وتقطيعه وتحويله إلى جمل منظمة، ليبقى ف النهاية الجزء الرئيسي من النص المستخدم فعملية التنقيب عن البيانات باستخدام نماذج تعلم الآلة أو الشبكات العصبية، وتتضمن خطوة ما بعد التجهيزتفسير وتصور نتائج عمليات استخراج النصوص Information(:) مهام التنقيب عن المعلوماتية الحيوية، ودور أخصائي المكتبات والمعلومات: 2013)1استرجاع المعلومات مثاً لا على نظام استرجاع المقالاتPubMed/MEDLINE تعني البحث عن المستندات ذات الصلةبالاستعلام، يعدRetrieval تصنيف الوثائق هي مهمة تصنيفDocument Classification(:) العلمية ف مجال الطب الحيوي؛)2تصنيف الوثائق المستندات إلىفئات موضوعية،تصنيف مقالات الدوريات حول ما إذا كانت تتعلق بجنين معين أم لا؟. عادةً ما يستخدم تصنيف وأسماء الجينات،إلخ اللائحةالتي يتم تحديدها مسبقاً، تشمل الكيانات التي يتم البحث عنها والتعرف ,Caporasoالمستندات بأسماء الجينات الجديدة التي يتم صياغتها باستمرار، وأن العديد من الجينات لها أسماء أوMetaMap"" عليها، والتي تسمى مرادفات مختلفة، ومعذلك، فإن هذا النهج القائم على استخدام القوائم المسبقة أو المكانز (القوامسيس) يمكن أن يحقق نجاحاً إذا تطبيع الكيان هي عملية البحث عن كيانمحدد ف قواعد البياناتName Entity Normalization(:))4تطبيع الكيانات الحيوية 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 الذي يتواجدTRP-1( ) يرغب الباحث ف التعرف عليها.ـ المشكلة الأخرى: الجين الواحد قد يكون له أسماء متعددة، مثل: جين لدىالجين لتحديد ماهية الجين المطلوب، من خلال المقارنة بين السياق ف النص وبين حقول البيانات: (العنوان،المستخلص، Relation & Information(:) النص. أو الربط بين الجين المطلوب البحث عنه)5استخراج المعلومات والعلاقات تعني عملية استخراج فعلى سبيل المثال: استخراج المعلومات الجينومية، وتفاعلات البروتين، وعلاقات نقل البروتين،Extraction وموقع بداية البروتين المنقول، وتستخدم تلك المهمةالأساليب المستندة بناء الجمل النموذجية، لكن وبخاصة تحليل التبعية افترضت الأساليبQuestion-Answering(:) بين المعلومات)6الإجابة على الاستفسارات,Hunter, Lu & Firbyوالعلاقة المبتكرة أن العديد منالنصية. حيث إن الهدف هو إرجاع إجابة محددة،مستنداً يحتوي الإجابة، بينما يركزاستخراج المعلومات حيث وجد أن درجات الأسئلة تتفاوت حيث يتم تحقيق أفضل ,Lee & Kaufman ،على احتياجات معلومات محددة للغاية والأسئلة الزمانية"متى" قد تتطلب أوقاتا2003 ,Zweigenbaum) (مثل :"أين"، ف حين أنFactoid"، النتائج مع الأسئلة "الواقعية يعني أخذ مستند أو مجموعة من المستندات كمدخلات وإرجاع نصأقصر، هناكSummarization(:) لإجابتها.)7التلخيص حاجة كبيرة لهذه الإمكانية ف المجال الطبي الحيوي، وقد تم تطبيق التلخيص على مقالات ومن أمثلة تلك الأنظمة التي تقوم بذلك والتقارير الطبية، وف مجال هناك العديد من النماذج المستخدمة ف تلخيص الإنتاج الفكري)MITRE'S MiTAP((نظام 2013) الحيوي بما ينتج عنه معلومات تقومحرفاً، تعمل على إدخال وظيفة الجين ف قاعدة بيانات .* ثانيها: نموذج وهو نهج يعمل على إعطاء الجمل الملخصة نقاطاً (وزناً) وفقاًالمجموعة من المعايير النسبية، مثل: موضعEdmundsonian(التيCue Words" الجملة ف الوثيقة، فالعنوان يختلف عن المستخلص وعن النص كذلك نوع الجمل ما بين "الكلمات الجدلية تشير إلى ارتباط الجملة بالنص،ويبقى التحدي هنا ف إيجاد أفضل جملة تعبر عن النص، كذلك الجمل التي تبرز بعض الحقائق المتمثلة ف اشتراكBioCreative حول عملية مثل: "تأكيد تفاعل بروتين معين أم لا" من داخل النص"، ويتم التغلب على ذلكبفكرة مجموعة من المقالات ف إثبات التساؤل ذاته، لتعطي أفضل دليل علىحدوث هذا التحليل (التفاعل،باستخدام خوارزميات معنية أباتشي هادوب هو برنامج:Biodoop – Bioinformatics on Hadoop (بذلك.نماذج من تطبيقات التنقيب عن البيانات الحيوية:1

Summarized by © lakhasly.com

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 من أمثلة التطبيقات التي تقوم بهذه الوظائف تطبيق تحليل(Leo, Santoni & Zanetti, 2009.مجموعات البيانات الكبيرة (Demaine, :الذي أعد لتمكينالكاملة، فيوفر التطبيقالتبديل بين التطبيقات المتعددة، ويتميز التطبيق بالاتيLitMiner النصوص إمكانية استيراد الملفات من قواعد البياناانات المتعددة ـMartin & Wei, 2006،

Summarized by © lakhasly.com