تنظيم وإدارة المعلوماتية الحيوية Bioinformatics
في علوم المكتبات والمعلومات

إبراهيم صبرى المتولي

باحث دكتوراه في علوم المكتبات والمعلومات
كلية الآداب – جامعة القاهرة
is_80@hotmail.com

المستخلص:

المعلوماتية الحيوية في أبسط تعريف لها، هي: علم يُعنى بإدارة وتنظيم البيانات الحيوية الجزيئية (DNA, RNA, Proteins)، وكذلك البيانات التي تنتج عن تفاعلاتها الحيوية. وتناقش الدراسة باستخدام المنهج المسحي التحليلي، الأدوات والعمليات الفنية التي يتم إجراؤها على المعلومات الحيوية بمؤسسات المكتبات المعلومات، وترصد الدراسة خطوات علم المكتبات والمعلومات نحو تداخله مع علوم الحاسب في إدارة المعلومات الحيوية، ومدى التشابه والتشابك بينهما في إنشاء، واستخدام، وتطوير أدوات وعمليات التنظيم الفني المتعلقة بـ: (ضبط وتوحيد المصطلحات، وقوائم رؤوس الموضوعات والمكانز الآلية، والفهرسة الوصفية والتسجيلات الحيوية، والتاكسونومي وتصنيف المعلومات الحيوية، والأنطولوجيا الحيوية والعلاقات الدلالية، والبيانات المترابطة، والتنقيب عن البيانات الحيوية، والتنقيب عن الإنتاج الفكري، والمكتبات الرقمية وإدارة مصادر المعلوماتية الحيوية)، كما تناقش الدراسة الأدوار الحالية، والمحتملة التي قد تقوم بها علوم المكتبات والمعلومات، نحو تطوير أدوات تعمل على خزن وتنظيم واسترجاع البيانات من قواعد البيانات، وإدارة مصادر المعلوماتية الحيوية.

الكلمات المفتاحية: المعلوماتية الحيوية؛ تنظيم وإدارة المعلومات؛ علم المكتبات والمعلومات؛ التاكسونومي في المعلوماتية الحيوية؛ الأنطولوجيا الحيوية؛ التنقيب عن البيانات (الحيوية) الضخمة.

تاريخ الاستلام: 2020/02/16 تاريخ القبول: 2020/07/03

التمهيد:

تتشابه الحدود الموضوعية لعلوم المكتبات والمعلومات، والمعلوماتية الحيوية بشكل ملحوظ، فكلا النظامين يتعامل مع المعلومات والمعرفة ومع تخزينها واسترجاعها واستخدامها في خدمة المجتمع، ومع ذلك فإن كل منهما يتأصل في سياقه الاجتماعي والثقافي الفريد الخاص به، في حين أنه لا يمكن إنكار أن تكنولوجيا المعلومات أثرت بشكل كبير في كلا المجالين، إلا أن المبادئ الأساسية لتنظيم المعلومات والمعرفة والتخزين والاسترجاع والاستخدام توفر أساسًا مشتركًا للبحث والممارسة في المجالين (Dalrymple, 2007)

إن النمو الهائل للبيانات البيولوجية في العقد الأخير أدى إلى بعض التحديات المتعلقة بأنشطة جمع المعلومات ومعالجتها، فهذا الفيضان الهائل من المعلومات يهدد تقدم البحث العلمي في هذا المجال، رغم وجود مجموعة كبيرة من البرمجيات التي تساعد على تحليل تلك المعلومات (Kalra, 2005)، فقد سُطرت العديد من المقالات والكتب من قبل المتخصصين في العلوم ذات العلاقة بالمعلومات الحيوية، مثل: علوم الحاسب، والرياضيات، والمعلوماتية الطبية، والتكنولوجيا الحيوية، والبيولوجيا الجزيئية.

ومع ذلك، يُعتقد أن علم المكتبات والمعلومات هو من يمكن أن يلعب دورًا مهما في تنظيم المعلومات لاستخدامها بشكل فعال في المستقبل؛ فعلوم الطب والبيولوجيا الجزيئية تركز عادة على الإجابة على الأسئلة الطبية البيولوجية، أما علم الحاسب الآلي فيهتم ببناء وتحسين الخوارزميات وهياكل المعلومات للمساعدة في التحليل البيولوجي للبيانات الحيوية، في حين إن علم الرياضيات يسعى لتطوير طرق إحصائية ممكنة صالحة لتحليل المعلومات، بينما يركز علم المكتبات والمعلومات على عملية تفاعل المستفيدين مع تلك المعلومات (Li, . Chen & Clintworth, 2013).

ومع التقدم السريع في تكنولوجيات المعلومات، تم توليد كمية كبيرة من البيانات (تسلسل الحمض النووي والبروتين)، وقد أدى ذلك إلى انفجار البيانات البيولوجية، سواء من حيث الحجم أو التعقيد، وهنا تأتي أهمية إدارة المعلومات الحيوية من أجل الاستغلال الكامل للبيانات، والتقنيات الحسابية المتزايدة التعقيد، والوسائل الفعّالة المستخدمة في تخزين البيانات والبحث عنها واسترجاعها، فضلًا عن ضرورة الاستعانة بخوارزميات وأدوات إحصائية قوية. (Geer, 2006)

وتهدف هذه الدراسة إلى التعرف على ماهية إدارة المعلوماتية الحيوية، ودور علم المكتبات والمعلومات في إدارة مصادر هذه المعلومات، دراسة ماهية الأنطولوجيات في مجال المعلومات الحيوية، وأنواعها، ووظائفها، وأوجه الشبه والاختلاف بينها وبين الوسائل الأخرى لتنظيم المعلومات الحيوية (مثل: التاكسونومي)، والتعرف على ماهية تحويل المكانز – التي نُسبت عهودا من الدهر لعلم المكتبات والمعلومات– إلى أنطولوجيات هي الآن

الأداة الأولى في تنظيم المعلومات بقواعد البيانات البيولوجية (المعقدة)، كذلك تهدف الدراسة إلى التعرف عن قرب لبعض نماذج "الأنطولوجيا الحيوية"، إن صح التعبير، وتتوالى عمليات إدارة المعلوماتية الحيوية، حيث يتم إنشاؤها وتجميعها، وحفظها، وإدارتها، والبحث فيها، وأرشفتها، ومشاركة تلك البيانات مع مجتمع المعلومات. (Grimes & Ji, 2014)

1/ **الإطار المنهجي للدراسة**:

1/1 إشكالية الدراسة:

لقد سيطرت حالة من الغيرة الشديدة مع كل تكرارٍ لسماع مصطلح "المعلوماتية الحيوية"، فتم التساؤل عن ماهية هذا النوع من المعلومات، وهل التشابك والتشابه اللفظي لمصطلح المعلومات له ارتباط بعلم المكتبات والمعلومات أم لا؟. لكن توجد فقرة واحدة في الإنتاج الفكري العربي – في مجال المكتبات والمعلومات – تناولت هذا المصطلح بأية درجة من درجات البحث والتحليل، بل زاد الأمر حيرة، في حيرة المتخصصين والباحثين في المجال عند مناقشتهم فكرة البحث، والارتباط المحتمل بين المعلومات الحيوية وعلم المكتبات والمعلومات.

وبالبحث في الإنتاج الفكري الأجنبي، فقد وجد أن الحديث عن علاقة علم المكتبات والمعلومات بالمعلومات الحيوية – علمًا ووظيفة – قد بدأ منذ عَقدين من الزمان، فأُلفت مقالات، وسُطرت دراسات تناقش تطبيقات نظم إدارة المعلومات الحيوية وخدماتها بالمكتبات (الأكاديمية والجامعية والبحثية منها على وجه الخصوص)، في هذه الدراسة محاولة تسريع الخطى للحاق بركب العلم بعد تأخر دام أكثر من عشرين سنة دون أي تحصيل علمي وعملي من جانب الباحثين العرب في هذا المجال.

لذا فإن **إشكالية الدراسة بداية تتلخص في عدم وجود دراسة تأصيلية عربية** في علم المكتبات والمعلومات تُعنى بمناقشة الجوانب العلمية والوظيفية للمعلومات الحيوية. تلك الدراسة التأصيلية التي من المفترض أن تحمل بين طياتها ماهية المعلومات الحيوية، وعوامل ارتباطها بعلم المكتبات والمعلومات، العمليات الفنية التي يتم اجراؤها على المعلومات الحيوية.

2/1 أهمية الدراسة:

تستمد الدراسة أهميتها من ذلك المجال التطبيقي الجديد، فقد زادت أهمية "المعلومات الحيوية" والحاجة إليها مع توافر الكم الهائل من المعلومات التي نتجت عن: (مشروع الجينوم البشري، وجينوم الكائنات الأخرى الحيوانية والنباتية، وتشخيص الفيروسات، وتحديد التتابعات، والتعرف على التراكيب، وإنتاج الأمصال وتصميم الأدوية، وحلول لمشاكل المقاومة للمضادات الحيوية والدواء والمبيدات)، لذا فإن "المعلومات الحيوية" تقدم مجموعة من المهارات والأدوات المتعلقة بجمع تلك المعلومات، والتنقيب عنها، وبناء المعرفة في البحوث

الجارية والتطبيقات الصناعية وغيرها، وذلك باستخدام الحاسب الآلي وتطبيقاته لعمل (أرشيف استنتاجات) من بيانات البيولوجيا الجزئية الحديثة، وعلى ما تقدم فإن الدراسة تستمد أهميتها من أهمية حفظ وتوثيق وتنظيم "المعلومات الحيوية" التي تعمل على إتاحة واسترجاع المعلومات الحيوية.

3/1 أهداف الدراسة:

تهدف الدراسة مجملًا إلى دراسة تقنيات المعلومات الحيوية، وارتباطها بعلم المكتبات والمعلومات، وأهم ما تهدف إليه الدراسة ما يلي:

1. التعرف على ماهية المعلومات الحيوية، وأهميتها في دعم البحوث الجامعية في شتى التخصصات.

2. التعرف على الوسائل والطرق والأدوات والآليات والمعايير التي يتم بها تنظيم المعلومات الحيوية: (وصف وتوثيق وتنظيم وفهرسة وتصنيف وتكشيف وعرض وتمثيل المعلومات الحيوية).

4/1 تساؤلات الدراسة:

من المقترح أن تجيب الدراسة على التساؤلات التالية:

1. كيف يتم تنظيم المعلومات الحيوية وسبل المعالجة الفنية لها، وذلك من حيث تصنيفها، وتكشيفها، وتكويدها، وإدارة تسجيلاتها الحيوية، وأسلوب عرضها؟

2. ما المعايير والأدوات التي يتم اعتمادها واستخدامها في عمليات الوصف والتحليل والتمثيل؟

5/1 منهج الدراسة:

تم الاعتماد في هذه الدراسة على المنهج المسحي التحليلي لدراسة مصادر المعلوماتية الحيوية، والعمليات الفنية التي تتعرض لها المعلومات الحيوية.

6/1 مفاهيم الدراسة:

1/6/1 المعلوماتية الحيوية Bioinformatics:

مصطلح المعلوماتية الحيوية "Bioinformatics" معنى بالتعامل مع البيانات الحيوية بكل أنواعها، (DNA) أو (RNA) أو (Protein) أو (الأنزيمات Enzymes) أو (التمثيل الغذائي Metabolites) كذلك يهتم بالعلاقة بين الجينات وآليات تشغيلها لتحقيق وظيفة بيولوجية معينة، ويهدف علم المعلوماتية الطبية Medical Informatics" إلى التركيز على المعلومات الخاصة بعمليات الرعاية "الطبية المختلفة ومحورها المركزي وهو المريض، بينما يهدف علم المعلوماتية الحيوية إلى التركيز على المعلومات الخاصة بما هو تحت هذا المستوى المرئي من

العلوم الحيوية الأساسية المرتبطة بتطوير الطب والرعاية الصحية كعلم الجينات البشرية والكيمياء الحيوية ووظائف الأعضاء وغيرها.

وترتكز المعلومات الحيوية على المعلومات الجزيئية التي يتم استخراجها من الخلية الحيوية، فالمعلومات الحيوية تقوم بدراسة المعلومات التي لها علاقة بالخلية، والجينات، والجزئ الذي هو يمثل أصغر وحدة أساسية من مركب كيميائي يمكن أن تشارك في التفاعل الكيميائي. فالمعلومات الحيوية هي تكامل بين برمجة الحاسب الآلي، والتحليل الإحصائي، وعلم البيولوجي، وتركز المعلومات الحيوية على البحوث المتعلقة بعلم الجينوم Genomics: (مثل: الجينات Genes)، وعلم دراسة وظائف البروتينات Proteomics: (مثل: البروتينات Proteins الذي هو أساس تكوين الأنسجة الحيوية بجسم الإنسان).

ويعتقد البعض أن المعلومات الحيوية Bioinformatics بمعزل عن المعلومات الصحية Health Informatics أو منفصلة عنها، بينما يؤمن البعض الآخر بالعلاقة القوية بينهما، وعموما فإن على أخصائي المكتبات التسليم بأن المعلومات الحيوية جزء من المعلومات الصحية لثلاثة أسباب رئيسية:

1. المعلومات الحيوية والمعلومات الصحية يتشاركان نفس الطرق والتكنولوجيا.

2. الباحثون في كلا القطاعين يتشاركوا الخبرات ونتائج البحوث، وقنوات اتصالهم بمنتهى السهولة، والدراسون في قاعات البحث يتم تدريبهم على قطاعي المعلومات بنفس القسم.

3. السبب الثالث فلسفي أكثر منه علمي، وهو: إن كلا القطاعين لهما نفس الهدف وهو الرعاية الصحية وتحسينها لأفراد المجتمع، من خلال الحلول التي يتم ابتكارها في مجال تطبيقات المعلومات الصحية.

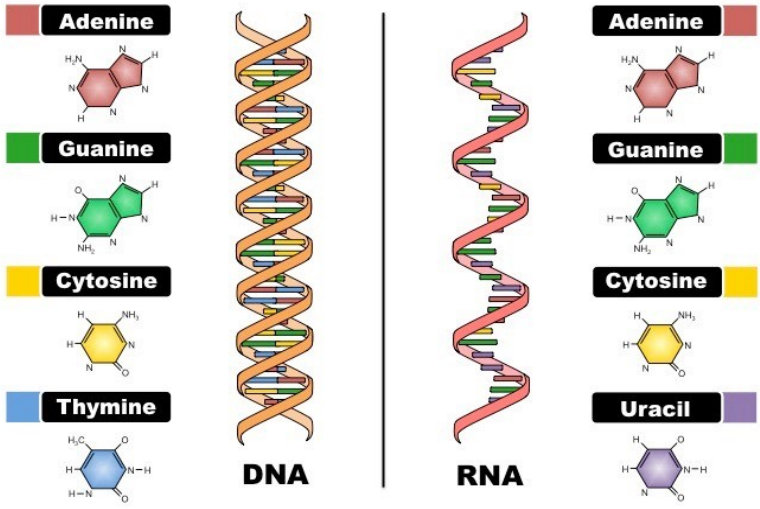
2/ **تمثيل المعلومات الحيوية (أنواع البيانات)**: Data Types

يفترض إلمام القارئ بأساسيات المعلوماتية الحيوية، فلا يَسمَح البحث هنا باستفاضة بشرح المعلوماتية الحيوية وأنواعها وخصائصها، لكن يمكن القول: إن المعلوماتية الحيوية تتباين، ورغم تباينها إلا أن أساسها يكمن في الأنواع الثلاثة الآتية، كما يتضح في الشكلين (1)، (2):

1. تواليات DNA من مكررات لأربعة حروف (A،G،C،T) تمثل القواعد النتروجينية.

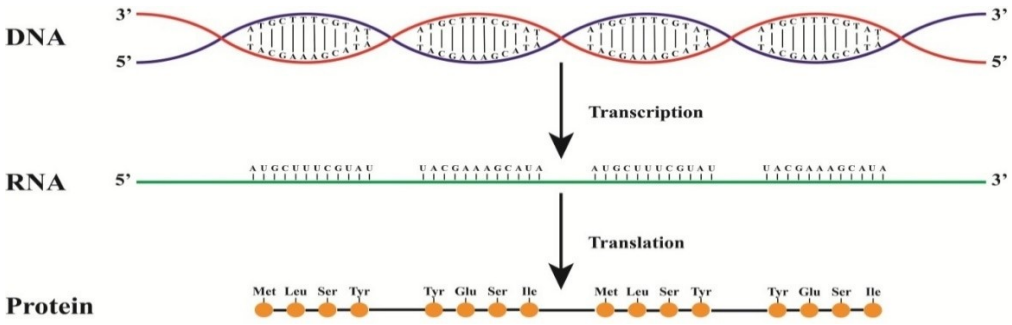
2. تواليات RNA من مكررات لأربعة حروف (A،G،C،U) بعد نسخها من تواليات DNA.

3. تواليات Proteins تتمثل في ثلاثة أحرف من حروف الأحماض الأمينية البالغة 20 حرف.



شكل (1): توالیات DNA، RNA. المصدر: (wikipedia.org)

3/ **تحليل المعلوماتية الحيوية**:



شكل (2): ترجمة RNA إلى بروتين بعد نسخه من شريط DNA. المصدر: (nature.com)

تحليل البيانات هو التطبيق الحقيقي الذي يحدث في مجالات المعلوماتية الحيوية، وفي هذا السياق يجب معرفة أن الأنظمة الحيوية أنظمة معقدة، حيث ترتبط الجزئيات والتراكيب الخلوية فيما بينها بشبكات حيوية Bionetworks معقدة جدا، وتزداد الأمور تعقيدًا بمعرفة أن الأنظمة الحية أو الخلايا لا تكون بمعزل عن المحيط الخارجي؛ أي: إن الخلايا تخضع للقوانين الكيماوية والفيزيائية على حد سواء.

نماذج من عمليات التحليل التي تتم المعلومات الحيوية:

تتعرض المعلوماتية الحيوية لعمليات تحليل تتلخص في الآتي:

1. عمليات التجميع Assembly: لجمع المقروءات الناتجة من عمليات تحديد توالي الجينومات.
2. إيجاد القطيفات Finding Motifs: للتعريف ببعض التواليات الخاصة بالبروتينات.

3. صف التواليات Sequence Alignment: لتحليل الهيكل والوظيفة للتواليات الجديدة.
4. التهميش والتوضيح Annotation: لمعرفة الكثير عن العنصر الذي تطاله المعلوماتية الحيوية.
5. إيجاد العلاقات التطورية (تشابه السلالات) Phylogeny: لفهم العلاقات الوراثية بين الأحياء.
6. تحليل وإيجاد الجين Gene Analysis: لتحديد فيما إذا كان التوالي جين أو قطعة من جين.
7. تصميم البوادئ Primer Design: لإجراء عمليات إكثار DNA بطرق PCR.
8. التنبؤ بتراكيب ومواقع DNA & RNA: للبحث عن وجود المكررات والمناطق غير المشفرة.

4/ **قنوات نشر وتوزيع المعلومات الحيوية**:

قواعد البيانات الحيوية المتاحة على شبكة الإنترنت هي القنوات الرئيسية لنشر المعلومات الحيوية وإتاحتها للباحثين والمهتمين بالمعلوماتية الحيوية، وتتراوح قواعد المعلوماتية الحيوية ما بين المجانية، والمجانية لأعضاء المؤسسات مثل: الجامعات والمراكز البحثية، والمعلوماتية الحيوية بمقابل، ويرجع تباين التقسيم الموضوعي لمصادر المعلوماتية الحيوية، إلى اختلاف وتعدد وجهات النظر عند إعداد الهيكل التقسيمي لتلك المصادر، إلا أن النهايات الصغرى من تلك التقسيمات بعد رصدها تشترك في تقسيم مصادر المعلومات الحيوية إلى الفئات التالية التي سوف نتناول بالتفصيل في دراسة مستقلة بما تشمله من قواعد للبيانات الحيوية Databases، وخدمات على الإنترنت Webservices، وبرمجيات Software.

1. قواعد بيانات الإنتاج الفكري Literature
2. قواعد بيانات علم الوراثة والصحة العامة Genetics & Health
3. قواعد بيانات الجينوم Genomes
4. قواعد بيانات الجينات Genes (DNA, RNA)
5. قواعد بيانات البروتينات Proteins
6. قواعد البيانات الكميائية Chemicals
7. قواعد بيانات البناء Structure

5/ **الفهرسة الوصفية للمعلومات الحيوية**: Cataloging of Bioinformatics

أغلب قواعد البيانات الخاصة بعلوم الحياة تحوي بيانات أولية، غير ذات دلالة بمفردها، وتكون إما تواليات لـ DNA أو RNA أو البروتينات، لذلك كانت الحاجة إلى بيانات إضافية Metadata تعمل على وصف وتمييز التواليات (التسلسلات)، لكن مع ضرورة التفريق الدائم بين البيانات الأولية (التواليات) والبيانات

الإضافية (الوصفية)، لأسباب تتعلق بديناميكية البيانات الأولية، وثبات البيانات الوصفية. (Wu & Stvilia, 2012)

1) **البيانات الأولية**: إما تواليات أحماض نووية أو تواليات الأحماض الأمينية، ففي الحالة الأولى تكون البيانات بمثابة مكررات من 4 حروف (A,G,C,T) لتواليات (DNA)، مكررات من 4 حروف (A,G,C,U) لتواليات (RNA)، أما في حالة البروتينات فتتكون من 21 حمضا أمينيا، كل حمض مكون من ثلاثة حروف (Met, Leu, Ser, Tyr,...etc) وجعل (X) لأي حمض غير معروف الهوية، وبذلك تكون البيانات الأصلية محددة وثابتة (انظر الشكل رقم 2).

2) **البيانات الوصفية**: عناصر البيانات الإضافية (الوصفية) غير محددة، وتختلف من قاعدة بيانات إلى أخرى وغالبًا ما تكون تلك البيانات الوصفية هي الهدف المطلوب من البحث عنها في قواعد المعلومات الحيوية، دون الحاجة إلى البيانات الأصلية، حتى لو كان الهدف الحصول على بيانات أولية سيكون من خلال إحدى البيانات الوصفية، مثل: (اسم الجين، واسم الكائن، واسم الباحث، واسم المرض،...إلخ) البيانات الوصفية المرتبطة بالبيانات الأولية في قواعد المعلوماتية الحيوية، ويمكن تقسيم البيانات الوصفية إلى: (Wroe, Stevens & Goble, 2003)

* **البيانات الوصفية الأولية (البدائية)**: والتي تسمى أيضا الحقائق الأولية، وهي التي لها علاقة بالتوالي (التسلسل) نفسه، أو بالطريقة التي أدت إلى الحصول عليه ومنها: ذكر الكائن الذي اشتق منه التوالي والتوزيع الهندسي للجزئ (حلقية أو مفتوحة)، وكذلك معلومات عن تركيب الكروموسوم، أو البلازميد، أو الفيروس، وطرق تحديد تواليها في المختبر، فضلا عن احتوائها على الصفات الفسيولوجية للكائن، والموقع الجغرافي الذي يوجد فيه، والأعراض المرضية عند إصابة أحد الأحياء به، وغيرها من المعلومات.

* **البيانات الوصفية المشتقة Derived Metadata**: ويتم الحصول عليها أثناء قيام الحاسب الآلي بعمليات التحليل التي تجري على البيانات، وتشمل أيضا التهميش اليدوي Manual annotation الذي يقوم الباحث بإضافته يدويًا، ومن أهم البيانات الوصفية المشتقة هي: أطالس الجينومات Genome Atlases التي تعمل على رسم خريطة ملونة وظيفية للجينات.

وتتعاظم أهمية البيانات الوصفية، وبدونها يصبح من غير الممكن استخدام التواليات للجزيئات الحيوية وذلك للأسباب التالية: (Charles & John, 2007)

- البيانات الوصفية ضرورية لعملية الفهرسة Indexing وهذا يسهل الوصول إلى مجاميع خاصة من المراجع المهتمة بتوالي معين.

– يمكن استخدام البيانات الوصفية للوصول إلى تواليات محددة مثل: البحث عن التواليات الحاوية على AT بنسب محددة فإن ذلك يسهل استخدام البيانات الإضافية.

– المساعدة في معرفة أصل التوالي المستعلم عنه، مثلا: توالي مشتق من كروموسوم أو بلازميد.

– تحديد الكائن أو المجموعة التصنيفية التي ينتمي إليها، والبيئة التي أخذ منها الكائن.

ويمكن الاستنتاج أن البيانات الإضافية ستحسن أو على الأقل تقلل من اللبس الحاصل حول تفسير النتائج، وذلك لكونها تتضمن بالإضافة إلى ما ذكر أعلاه من المعلومات، وصفًا للتقنيات المستعملة والتعريف بوحدات القياس المستعملة، ووصف لظروف أخذ النماذج وطرق إجراء التجارب، وعليه فإن غياب البيانات الوصفية يمكن أن تجعل البيانات المخزونة في قواعد البيانات دون فائدة، لذا كان للبيانات الوصفية دورا كبيرًا في المعلوماتية الحيوية، ويعكس هذا كله الدور الفعلي أو المتوقع لأخصائي المكتبات والمعلومات وما قد يسهم به في تطوير الفهرسة الوصفية للمعلوماتية الحيوية. (Ram & Rao, 2014)

*** التسجيلة الحيوية (الوصفية):**

تحتوي قواعد المعلومات الحيوية على التسجيلات الحيوية، فكما هو موضح في الشكل (3)، احتواء كل تسجيلة على عدد من حقول الوصف، مثل: اسم الكائن الحي (الاسم العلمي)، والاسم العام، تواليات DNA، تواليات البروتينيات، والبيئة التي عزل منها، والمراجع التي استقت منها المعلومات الحيوية: (المؤلف، عنوان المقال، عنوان الدورية...إلخ) مصادر الوصف كما هو متبع في الوصف الببليوغرافي لمصادر المعلومات التقليدية (المتيني & رياض، 2006)، وتختلف طرق الوصف من مركز لآخر – رغم تشابه النهايات الصغرى في عملية الوصف والتوثيق– طبقًا للنظام المعمول به في تلك المؤسسة، وإن حدث وتشابهت حقول الوصف، فإن ترتيب عرض هذه الحقول قد يكون هو اختلاف آخر، وكما سيتضح لاحقًا فإن مؤسسات المعلوماتية الحيوية تواجه تحديًا كبيرًا يتعلق بتوحيد قواعد وصف حصول المعلومات الحيوية داخل التسجيلة الحيوية، حيث تجتهد كل مؤسسة (خاصة المحلية منها) في توصيف التسجيلات الحيوية مما يشكل عائقًا أمام تبادل البيانات الحيوية، ورغم أن العديد من المراكز الحيوية تستخدم صيغة (XML) من أجل أغراض تبادل البيانات العلمية الحيوية، إلا أن الحديث هنا أعمق في مسألة توحيد بيانات الوصف ليشمل كل جوانب الوصف دون إغفال أي حقل من حقول البيانات، وبخاصة حقول الوصف غير المباشرة كتلك التي تتعلق بالجوانب البيئية الكائن الحي.

LOCUS NC_006213 30741 bp **RNA** linear VRL 21-FEB-2019 DEFINITION Human coronavirus OC43 strain ATCC VR-759, complete genome.

ACCESSION NC_006213 **الرقم التعريفي الموحد للتسجيلة الحيوية**
VERSION NC_006213.1
DBLINK BioProject: PRJNA485481 **قاعدة البيانات ذات العلاقة(المشروعات)**
KEYWORDS RefSeq.
SOURCE Human coronavirus OC43 (HCoV-OC43)
ORGANISM Human coronavirus OC43
Viruses; Riboviria; Nidovirales; Cornidovirineae; Coronaviridae;
Orthocoronavirinae; Betacoronavirus; Embecovirus.
REFERENCE 1 (bases 1 to 30741) **البيانات الببليوغرافية**
AUTHORS St-Jean,J.R., Jacomy,H., Desforges,M., Vabret,A.,
Freymuth,F. and Talbot,P.J.
TITLE Human Respiratory Coronavirus OC43: Genetic Stability and
Neuroinvasion
JOURNAL J. Virol. 78 (16), 8824-8834 (2004)
PUBMED 15280490 **رابط المقال في قاعدة البيانات الببليوغرافية**
REFERENCE 2 (bases 1 to 30741)
CONSRTM NCBI Genome Project
TITLE Direct Submission
JOURNAL Submitted (21-FEB-2019) National Center for
Biotechnology
Information, NIH, Bethesda, MD 20894, USA
REFERENCE 3 (bases 1 to 30741)
AUTHORS St-Jean,J.R. and Talbot,P.J.
TITLE Direct Submission
JOURNAL Submitted (29-MAR-2004) INRS-Institut Armand-Frappier,
531 boul.
des Prairies, Laval, Qc H7V 1B7, Canada **تعليق على التسلسل**
COMMENT PROVISIONAL REFSEQ: This record has not yet been
subject to final
NCBI review. The reference sequence is identical to AY585228.
COMPLETENESS: full length.

FEATURES Location/Qualifiers
source 1..30741
/organism="Human coronavirus OC43"
/mol_type="genomic RNA"
/strain="ATCC VR-759"
/serotype="OC43"
/db_xref="ATCC:VR-759"

/db_xref="taxon:31631"
/country="USA"
gene 210..21496
FEATURES Location/Qualifiers **توالى البروتين (الأحماض الأمينية)**

/translation="MSKINKYGLELHWAPEFPWMFEDAEEKLDN
PSSSEVDMICSTTA
QKLETDGICPENHVMVDCRRLLKQECCVQSSLIREIVMNA
SPYDLEVLLQDALQSREA...
ORIGIN **توالى دي ان ايه**
1 attgtgagcg atttgcgtgc gtgcatcccg cttcactgat ctcttgttag
بداية تتابع الجين
61 aatctaaact ttataaaaac atccactccc tgtaatctat gcttgtgggc
121 catagtggtg tttatattca tttctgctgt taacagcttt cagccaggga
181 cctaggcagt ggcccgccca taggtcacaa tgtcgaagat caacaaatac
…
30661 acccagtaat tagtaaatga atgaagttaa ttatggccaa ttggaagaat
30721 aaaaaaaaaa aaaaaaaaaa
a **نهاية تتابع الجين**

شكل (3): جانب من تسجيلة المعلومات الحيوية - جينوم فيروس كورونا وعملية الربط بين المعلوماتية الحيوية وقواعد البيانات الببليوغرافية.

المصدر: Nucleotide Database – NCBI

حقول الفهرسة الوصفية في التسجيلات الحيوية:

التسجيلة الحيوية شأنها شأن التسجيلة الببليوغرافية، تحتوي على حقول بيانات وصفية تستخدم في البحث والاسترجاع، ويتم الربط بين البيانات الحيوية والببليوغرافية من خلال حقل المصدر الببليوغرافي، كما سيتضح من عرضنا التالي لبعض حقول التسجيلة الحيوية:

– رقم التسجيل (Accession number) لتوالي النيوكلوتيد أو البروتين أو الجينوم، من أهم الحقول في التسجيلة الحيوية، وهو الرقم الذي يعطي للجزيئة الحيوية عند إيداعها في قواعد البيانات الحيوية، ويبقى الرقم مرافقا للتوالي بشكل دائم، فهو يعد Primary key الذي يستعمل عند بناء قواعد البيانات، ويكون مشتركا مع المراكز الحيوية الثلاثة: الأمريكي NCBI والياباني DDBI والمركز الأوربي EMBL، فهو أشبه بالترقيم الموحد للكتاب.

– التعريف Definition يتضمن ملخص المعلومات عن سجل التوالي ويشمل الاسم العلمي للكائن المشتق منه التوالي فيما إذا كان معروفا وكذلك فيما إذا كان التوالي كاملا أو جزئيا.

– التاريخ Date الذي أعلن فيه التوالي للاستخدام العام بقاعدة البيانات، إذ إن إيداع التوالي لا يعلن عنه مباشرة وإنما يحتاج إلى مدة للتأكد منه ومقارنته بباقي التواليات.

– المصدر (الكائن الحي) الذي أخذت منه التواليات، تصنيفه أو النسيج أو العضو الذي اشتقت منه.

– الكلمات المفتاحية التي يمكن أن تساعد في عملية البحث.

– المصدر الببليوغرافي الذي ورد بها التوالي: (المؤلف، عنوان المقالة،...إلخ)، وفي حال كانت المعلومات تجريبية وغير منشورة في مصدر ببليوغرافي، فإن هذه الحقول تكون فارغة.

– معلومات حول الجين ومنتجاته وموقعه من التوالي، وكذلك توالي الحوامض الأمينية.

– تواليات الجينوم (الأحماض الأمينية)، وتواليات الأحماض النووية.

*** التوحيد القياسي للبيانات الوصفية:**

وعلى غرار ما يحدث في مجال المكتبات، وفي إطار الحديث عن أنشطة التوحيد القياسي للمعايير التي تحكم إنشاء وتوثيق ووصف وترتيب هذه التسجيلات الحيوية، للتعرف على كيفية توثيق المعلومات الحيوية، فإن "اتحاد معايير الجينوم (GSC)" الذي تأسس عام 2005، يقع على عاتقه استقبال المقترحات بهذا الصدد، ومراجعتها، والتوصية بها بعد نشرها. (Genomic Standards Consortium)

وعلى الرغم من تعاون المؤسسات الدولية مع الاتحاد، إلا أنه لا يملك سلطة إجبار المؤسسات العاملة في مجال المعلوماتية الحيوية على استخدام هذه المعايير، مما جعل العديد من المؤسسات تعمل بشكل فردي في توصيف بياناتها، وهذا تحدي تواجه المعلومات الحيوية نحو توحيد وتكامل البيانات الجينومية من خلال المعايير التي تقدمها لمؤسسات المجتمع الدولي. (Knaup, & Haux, 2010) ويعقد الاتحاد مؤتمرًا سنويًا لمناقشة وبحث آخر التطورات في عملية توصيف البيانات الحيوية، والتوصية بما تراه فعالًا، وقد تم إلى الآن إصدار حوالي 18 معيارا في إطار توحيد البيانات الوصفية والمصطلحات، ومن أمثلة تلك المعايير التي تم إصدارها في عام 2008 معيار"الحد الأدنى من المعلومات الوصفية عن تسلسل الجينوم" = "Minimum Information MIGS About a Genome Sequence"

*** رؤية مستقبلية لوصف التسجيلات الحيوية:**

يُخطط في هذا الجانب مستقبلًا – لأول مرة – بتقديم مقترح علمي لمجتمع المكتبات الدولي، والمعلوماتية الحيوية حول إمكانية استخدام (مارك MARC) و"وصف وإتاحة المصادر RDA" في وصف مصادر المعلومات الحيوية، واستغلال الحقول غير المستخدمة في القيام بوصف المعلوماتية الحيوية، تمهيدًا لدمج التسجيلات الببليوغرافية مع التسجيلات الحيوية في قواعد بيانات واحدة، ولتيسير تبادل التسجيلات بين المكتبات ومراكز المعلوماتية الحيوية.

6/ **التصنيف في مجال المعلوماتية الحيوية** Classification Bioinformatics:

تعتمد المكتبات الطبية في تصنيف مجموعاتها الحيوية الورقية والرقمية البسيطة على التصانيف التقليدية، وقد تنوعت تلك التصانيف ما بين العامة والمتخصصة، ولسنا هنا بصدد التعريف بأنواع التصانيف التقليدية في المكتبات، والتي قتلت بحثًا مثل تصنيف ديوي العشري. إلا أنه وبعيدًا عن مناقشة أسس كل منهما، أي: أنواع التصانيف، فإن المكتبات الطبية والأكاديمية تعتمد على تلك التصانيف في معالجة مجموعاتها في أشكالها التقليدية، لكن بالحديث عن المعلوماتية الحيوية الضخمة فإن المنحنى يذهب مباشرة إلى التصانيف الحديثة التي تستوعب الكميات الهائلة من المعلومات الحيوية (مثل تصنيف التاكسونومي).

نماذج من أدوات التصنيف المستخدمة في المعلوماتية الحيوية:

1- "تصنيف المكتبة الوطنية الطبية الأمريكية" NLM:

يعتبر الأساس في تصنيف المجموعات الورقية والرقمية بالمكتبة الوطنية الطبية الأمريكية، ومركز المعلوماتية الحيوية NCBI. ويبدأ تكويد التصنيف الهجائي الرقمي من الرمز (QS) وحتى (QZ)، ثم من (WA) وحتى (WZ). ويتم تحديث التصنيف بشكل مستمر، والتصنيف بكامله متاح على موقع المكتبة الوطنية الأمريكية، وهناك بعض المكتبات (الطبية) حول العالم تستخدم التصنيف في تنظيم مجموعاتها الورقية والرقمية.

W 26.5	Informatics. Health informatics
	Classify works on use for special subjects by subject.
W 26.55	Special topics, A-Z
W 26.55.A7	Artificial intelligence. Expert systems
W 26.55.A9	Automatic data processing
W 26.55.C7	Computers. Computer systems
W 26.55.D2	Decision making, Computer-assisted
	Classify works on computer-assisted diagnosis (general) in WB 141; on computer-assisted therapy (general) in WB 300.
W 26.55.I4	Information systems. Information storage and retrieval
W 26.55.S6	Software
	Museums, exhibitions, etc.
W 27	Collective
W 28	Individual (Cutter from name of institution or exhibit)

شكل (4): المعلوماتية في خطة تصنيف المكتبة الوطنية الأمريكية

المصدر: تصنيف المكتبة الوطنية الطبية الأمريكية – النسخة الرقمية على شبكة الإنترنت:

7/ **تصنيف التاكسونومي في المعلومات الحيوية**: Taxonomy in Bioinformatics

التاكسونومي "علم التصنيف الذي يتضمن المبادئ العامة التي تقسم على أساسها الأشياء والظواهر إلى أقسام، والتي تنقسم بدورها إلى أقسام فرعية ثم إلى تفريعات من الأقسام الفرعية وهكذا"، وقد عرفه البعض الآخر بأنه: "نظام بنية للمعرفة مشابه لنظام التصنيف والمكانز وتكشيف المصطلحات". ويشير أيضًا إلى "تنظيم المعلومات على الإنترنت، وإن كان يعتمد على نفس الفلسفة الهرمية" (خليل، 2019، 196-198)، ويتم تعريف التاكسونومي أيضا بأنه: "مجموعة من المفاهيم مرتبة على هيئة شجرة مقلوبة، جذعها لأعلى، وأوراقها لأسفل، أي انها مجموعة من المفاهيم مرتبة مع بعض بصورة هرمية". (حسين، 2010) .

وعلم المعلومات الحيوية من المجالات الخصبة التي يتم فيها استخدام التاكسونومي في إدارة وتصنيف المعرفة الحيوية، فالقدرة على تصنيف المعلومات الحيوية وتكشيفها موضوعيا هو السر الكامن وراء الاستخدام الأمثل لتلك المعلومات بعد تحليلها، وقد تم اعتماد نظم تصانيف خاصة تستطيع التعامل مع الكم الهائل من المعلومات المتخصصة في مجال المعلومات الطبية والحيوية، فالمصطلحات الصحية كثيرا ما تكون غير دقيقة وتتغير باستمرار، فالعشرات من المفردات التي تم ضبطها في تطور مستمر، وهو التحدي الحقيقي في هذا الجانب، وتهدف الدراسة في هذا الجانب إلى التعرف على ماهية التاكسونومي في مجال المعلومات الحيوية، وأشكالها، وأنواعها، وكيفية بنائها، وطرق الاستفادة المتبادلة من هذه الأداة في مجال المكتبات والمعلومات سواء من حيث الشكل أو البناء، مع عرض لبعض نماذج التاكسونومي المستخدمة في معالجة المعلومات الحيوية.

وقد عدد أحد الباحثين (Mcknight, 2005) بعض التحديات التي تجعل من الصعب تقنين أدوات تعمل على تصنيف وضبط المصطلحات الطبية الحيوية، هي:

1. تعدد الأنشطة والمهام وتنوع المستفيدين الذي يتوقع أن تشملهم وتخدمهم تلك الأدوات.
2. القدرة على التوافق بين احتياجات المستفيدين ومتطلبات البرمجيات التي يتم تطويرها بدقة.
3. الجوانب الفنية المرتبطة بتقنين مصطلحات يمكن التعامل معها وإدخالها، والبحث عنها واستعراضها، واسترجاعها – واختيار واختبار البرمجيات القادرة على التعامل معها.
4. عدم قدرة اللغة بشكل دائم على تمثيل المفاهيم الطبية بشكل كاف، لعدم توافق المصطلحات الطبية (العملية) مع النماذج والمفاهيم اللغوية (النظرية).
5. صعوبة تحقيق قدر مناسب من الإجماع اللغوي حول المصطلحات الطبية الحيوية من أجل التكيف المحلي لجميع المستفيدين.

6. البنية الخاصة بأنظمة الترميز والتصنيف التقليدية لا تتناسب مع بنية التصانيف الحديثة.

7. القدرة على إدارة التغيير المستمر الذي يطرأ على المصطلحات دون الإخلال بالتسجيلات الحيوية المسجلة بقواعد البيانات الحيوية.

نماذج من نظم تصنيف التاكسونومي:

1- NCBI Taxonomy.

من أشهر نماذج التاكسونومي تلك التي يتم استخدامها بقاعدة بيانات NCBI حيث تحتوي على تصنيف مسبق لأسماء وأنساب الكائنات الحية التي لديها بيانات جزيئية (تواليات DNA-RNA-Proteins) في قواعد البيانات الأخرى بمركز (NCBI)، ويتم تحديث وإضافة العديد من الكائنات الحية بشكل دوري، وقد تم تقسيم المستوى الأعلى من التصنيف إلى سبعة أقسام: (البيكتريا القديمة أو الأصلية، والبيكتريا وحيدة النواة، والكائنات حقيقة النواة مثل: الإنسان، وأشباه الفيروسات، والفيروسات، وغير مصنف، وأخرى)، ثم يأتي تحت كل نوع من هذه التصانيف الكائنات التي تنتمي اليها وهكذا، ورغم ضخامة قاعدة بيانات التصنيف إلا أنها تمثل 10٪ من الأنواع الموجودة على كوكب الكرة الأرضية، ويبلغ عدد (التسجيلات) الكائنات بأنواعها المختلفة داخل قاعدة بيانات التصنيف طبقًا لآخر إحصائية حوالي 589 ألف كائن. (NCBI Taxonomy)

ويعد تصنيف الأحياء أحد الجوانب المهمة في دراسات علوم الحياة، حيث يوفر شجرة للعلاقات التطورية بين الأحياء بالاعتماد على تواليات RNA للوحدات الصغيرة للرايبوزوم، ويتم البحث في مجال التصنيف اعتمادًا على الاسم العلمي للكائن الحي، أو ما يسمى المعرف التصنيفي Taxonomy Identifier ويكون استعمال الأخير هو الأفضل نظرًا للتغير الذي يطرأ على المسمى من وقت لآخر، فاستعمال TaxId يمكن أن يوصل إلى الأسماء الحديثة مع ذكر التسميات القديمة للأحياء. (Rocco & Critchlow, 2003)

Drosophila melanogaster

Taxonomy ID: 7227 (for references in articles please use NCBI:txid7227)
Scientific name: ***Drosophila melanogaster*** Meigen, 1830
homotypic synonym: ***Sophophora melanogaster*** (Meigen, 1830)
Genbank common name: **fruit fly**
Inherited blast name: **flies**
Rank: **species**
Genetic code: Translation table 1 (Standard)

شكل (5): تسجيلة "ذبابة الفاكهة سوداء البطن". المصدر: Taxonomy Database - NCBI

2- Integrated Taxonomic Information System (ITIS)

النظام المعلوماتي المتكامل لتصنيف التاكسونومي (ITIS) الذي تم وضعه عام 1996 هو مشروع شراكة بين العديد من المؤسسات الأمريكية والكندية والمكسيكية، بإدارة المتحف القومي الأمريكي للتاريخ الطبيعي، والذي يهدف إلى توفير معلومات متسقة وموثوق بها لتصنيف الكائنات الحية، حيث يحتوي التصنيف على حوالي (887) ألفًا من الأسماء العلمية للكائنات الحية، منها (132 ألفًا) من الأسماء العلمية الشائعة (أرضًا وبحرًا، من جميع الممالك البيولوجية: الحيوانات والنباتات والفطريات والميكروبات...إلخ). وتعرض التسجيلة التصنيفية للتاكسونومي الأسماء في تصنيفها القياسي، وأهم الحقول: الترقيم الفريد المميز لكل كائن حي Taxonomic Serial Number (TSN)، ومكتشف الكائن الحي، تاريخ الاكتشاف أو الإضافة، والمعلومات الببليوغرافية المتعلقة بالاسم.

***Drosophila melanogaster* Meigen, 1830**
Taxonomic Serial No.: 146290

Download data (Download Help) ***Drosophila melanogaster*** TSN 146290

Taxonomy and Nomenclature

Kingdom:	Animalia
Taxonomic Rank:	Species
Synonym(s):	Drosophila ampelophila Loew, 1862
Common Name(s):	
Taxonomic Status:	
Current Standing:	valid
Data Quality Indicators:	
Record Credibility Rating:	unverified

Taxonomic Hierarchy

Kingdom	Animalia – Animal, animaux, animals
Subkingdom	Bilateria
Infrakingdom	Protostomia
Superphylum	Ecdysozoa
Phylum	Arthropoda – Artrópode, arthropodes, arthropods
Subphylum	Hexapoda – hexapods
Class	Insecta – insects, hexapoda, inseto, insectes
Subclass	Pterygota – insects ailés, winged insects
Infraclass	Neoptera – modern, wing-folding insects
Superorder	Holometabola
Order	Diptera – mosca, mosquito, gnats, mosquitoes, true flies
Suborder	Brachycera – circular-seamed flies, muscoid flies, short-horned flies, mouches muscoïdes
Infraorder	Muscomorpha
Family	Drosophilidae – pomace flies, small fruit flies, vinegar flies, drosophiles, moucherons, mouches du vinaigre
Subfamily	Drosophilinae
Genus	Drosophila
Species	Drosophila melanogaster Meigen, 1830

شكل (6): تسجيلة "ذبابة الفاكهة سوداء البطن". المصدر: ITIS Taxonomy

ويتميز موقع نظام التاكسونومي (ITIS) بالآتي:

– إمكانية البحث أو المقارنة بين قائمة من أسماء الكائنات الحية داخل قاعدة بيانات نظام التصنيف، وليس فقط البحث بكائن حي واحد.

– إمكانية تنزل التصنيف كاملًا، على جهاز الحاسب الشخصي، وإعادة استخدامه بما يخدم مقتضى البحث العلمي، مع إمكانية تنزيل كل مملكة من الممالك على حدة.

تاكسونومي المعلوماتية الحيوية ونظم المعلومات الجغرافية:

البحث في تصنيف (قاعدة بيانات التاكسونومي) ليست هي الطريقة المثلى لوضع (تصنيف) الكائن مع مجاميع الأحياء الأخرى التي تتشارك معه في ظروف بيئية معينة، مثل: الظروف البيئية الحارة أو الباردة أو المعتدلة أو الجافة أو البحرية أو الصحراوية...إلخ الظروف البيئية التي تؤثر على توالیات الجين وتطوره. لذلك يتم استخدام نظم المعلومات الجغرافية GIS – خاصة في عالم الأحياء المجهرية – لتحديد خارطة وجود الأحياء المجهرية في بيئة ما؛ لذلك يمثل GIS أحد مظاهر قوة قواعد البيانات في الكشف عن العلاقات البيئية، وحيث إنه لا تواجد روابط بين نظم المعلومات الجغرافية GIS وقواعد البيانات الجينية، فإن قواعد البيانات الأخيرة تعجز عن الكشف عن الأمراض بشكل دقيق التي يتم تحديد توالياتها وذلك لتداخل البيئة في إحداث الأمراض، وعدم القدرة على التفرقة في تأثير البيئات المختلفة على الجين الواحد، وبخاصة وأن التطور الجيني ليس معتمدًا فقط على التوالیات بشكل مطلق، بل على الظروف البيئية التي ينشأ ويتطور فيها. (Dolan, Holden & Beard, 2006).

وعلى مستوى الجينوم، يتم استخدام نظم المعلومات الجغرافية في الكشف عن الاتجاهات الوراثية للجينات وأنماطها وعواملها البيئية وتحليلها الجيني والبروتيني وتوزيعها؛ وذلك من أجل تصميم إستراتيجيات لمنع المرض من الانتشار من منطقة إلى أخرى، فيمكن استغلال أدوات رسم الخرائط المتطورة لعرض بيانات الجينوم واستكشافها والتفاعل معها.

مستقبل التاكسونومي في المعلوماتية الحيوية:

إن التصنيف يشكل الأساس لكل الأبحاث البيولوجية، لما يترتب عليه من تأثيرات على العديد من المجالات العلمية والتطبيقية الأساسية، وفي هذا الصدد يواجه التصنيف الحيواني والنباتي أزمة كبيرة، ذلك لأن عددًا قليلًا من المتخصصين في هذا السياق من له خبرة تحديد التصنيفات الكثيرة بدقة شديدة، وهنا نستكشف إمكانية استخدام المنهجية القائمة على "التصنيف باستخدام الحمض النووي" للتغلب على هذه المشاكل، إن فائدة تسلسل الحمض النووي للأغراض التصنيفية راسخة جيدًا، ومع ذلك، فإن جميع النهج التصنيفية الحالية تهدف إلى استخدام الحمض النووي، في أحسن الأحوال، كمعيار مساعد لتحديد نوع أو تصنيف الكائن الحي.

ويعكس ذلك مدى أهمية أن يكون بعض متخصصي علم المكتبات والمعلومات على علم أو لهم خلفية علمية، حتى يكونوا شركاء في إنشاء أدوات العمل التي بدا أنها تنحى منحنًا آخر في بنائها، وذلك باعتمادها على المختبرات العلمية، وليس فقط على مبادئ علوم المكتبات والمعلومات.

8/ **الويب الدلالية في المعلوماتية الحيوية**: Semantic Web

الويب الدلالية ثورة جديدة في مجال محركات البحث وتطور أدوات استرجاع المعلومات من الشبكة العالمية الإنترنت، ويهدف في أبسط صوره، إمكانية إكساب البيانات المتاحة على شبكة الإنترنت فرص إعادة استخدامها مرة أخرى من قبل التطبيقات والبرمجيات المختلفة، حيث يعتمد هذا النهج في بنيته على المحددات الفريدة للمصادر، ولغة النص الفائق الموسعة، والأنطولوجيات، ومعيار وصف المصادر، وفي سبيل ذلك حدد العلماء دلالات للعلاقات بين المصطلحات والمفاهيم التي تعمل بدورها على ربط تلك المصطلحات لإعادة تدويرها واستخدامها على شبكة الإنترنت (النشرتي، 2018). والويب الدلالية لها أهمية كبيرة في المعلوماتية الحيوية، حيث تسهم في حل المشكلة الأكبر في مجال المعلوماتية الحيوية، والمرتبطة بعملية استرجاع المعلومات الضخمة المخزنة في قواعد المعلوماتية الحيوية، بدقة وفاعلية، وتقليل اللبس من جراء تشابه تلك المعلومات الحيوية، وذلك من خلال الربط بين المفاهيم الحيوية المختلفة ذات الصلة في أدوات فنية تعمل على ذلك كما سيتضح الآن عند العرض لبعض أدوات الويب الدلالية في مجال المعلوماتية الحيوية.

نماذج من أدوات الويب الدلالية في مجال المعلوماتية الحيوية الطبية:

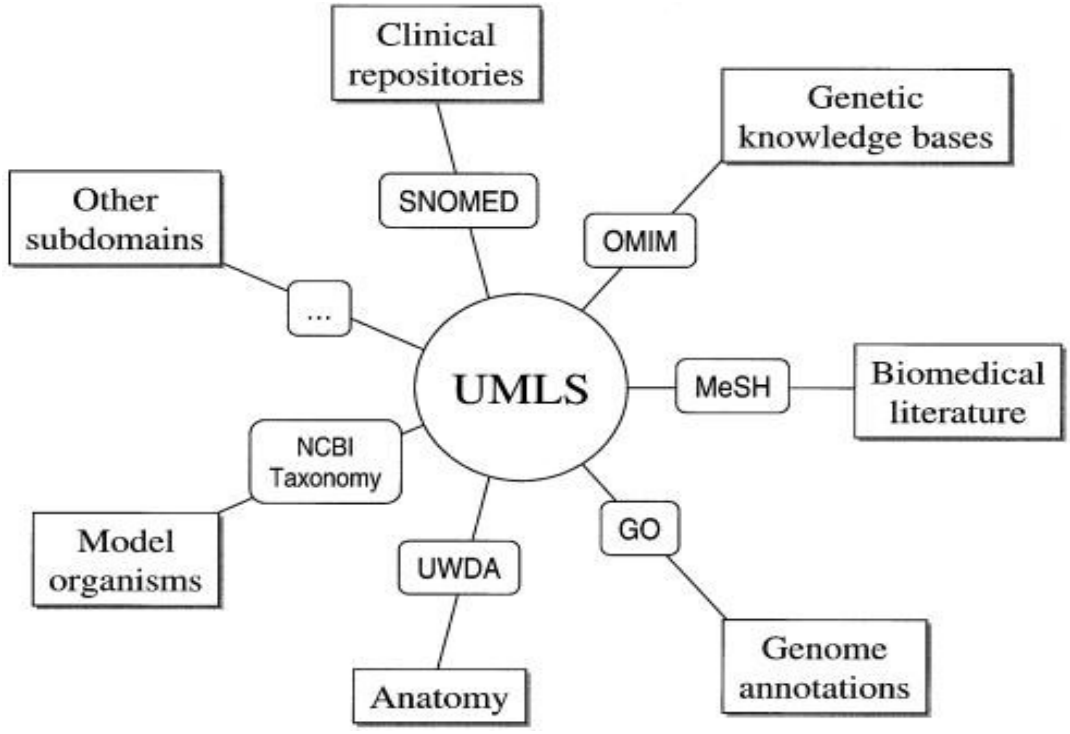
1- Unified Medical Language System. النظام الموحد للغة الطبية

من أكبر وأشهر قواعد البيانات للمصطلحات الطبية والحيوية، والذي سهل بشكل كبير التعامل مع أغلب المصطلحات الطبية بجميع اللغاتK ففي عام 1986 أطلقت المكتبة الوطنية الطبية الأمريكية هذا المشروع الذي هدف إلى إنشاء علاقات مفاهمية للمصطلحات الطبية الحيوية؛ مما يسمح برفع كل العوائق لتطبيق علم الحاسوب في مجال الطب، إن عدد المصادر الطبية الحيوية التي تخرج للباحثين من نتائج البحث كبير جدا، وغالبا ما يشكل هذا عائق بسبب الكم الهائل من نتائج البحث المسترجعة عند إجراء بحث في قواعد بيانات الإنتاج الفكري الطبي. (Michon, 2007)

ومما يحسن عملية الوصول إلى الإنتاج الفكري الأجنبي، تمكين أنظمة الحاسب المتطورة من فهم اللغة الطبيعية الطبية الحيوية، فأهم خاصية يتميز بها هذا النظام هو توفير روابط بين مختلف المصطلحات الطبية الحيوية تم جمعها من مصادر مختلفة، ومن هذا يتبين أن من أهم أهداف الـ(UMLS) هو توفير منصة تسمح بجمع ما

أمكن من مكانز وقواميس ووضع علاقات بين المصطلحات الطبية، حيث يتم تحديث النظام مرتين سنويًا في مايو ونوفمبر، ويستخدم نظام (UMLS) في المهام المتعلقة بربط المصطلحات الطبية والدوائية، وإجراء لغة حوار مشترك بين التخصصات الطبية، تنسيق عملية رعاية المرضى بين عدة أقسام داخل المستشفى، معالجة النصوص لاستخراج المفاهيم والعلاقات، أو المعرفة، كذلك تسهيل عمليات التخطيط والربط بين المصطلحات الطبية الحيوية، وأخيرًا تطوير نظام استرجاع المعلومات الطبية الحيوية. (Bodenreider, 2004)

وبالنظر إلى الرسم التوضيحي التالي يتضح التقسيم الرئيسي والمكونات الفرعية لمصادر المعلومات الفرعية لـ (UMLS)، فعلى سبيل المثال، وكما يستضح آنفًا، أن قائمة رؤوس الموضوعات الطبية (MeSH) تستخدم في تكشيف الإنتاج الفكري في مجال المعلوماتية الطبية الحيوية، وأنطولوجيا الجين (GO) تستخدم في توحيد تمثيل سمات الجين بجميع أنواعه، وأن تصنيف التاكسونومي (NCBI Taxonomy) تستخدم في تصنيف الكائنات الحية بقواعد بيانات المعلوماتية الحيوية، وفيما يلي إيجاز للمكونات الثلاث الفرعية لنظام (UMLS) المستخدم في ضبط المصطلحات للمعلوماتية الطبية الحيوية. (Post, Roos & Marshall, 2007)



شكل (7): المكونات الفرعية للنظام الموحد للغة الطبية. المصدر: النظام الموحد للغة الطبية UMLS

أ) المكنز الوصفي: Metathesaurus

يُعتبر المكنز الوصفي لنظام (UMLS) أكبر قاعدة بيانات للإصطلاحات الطبية في العالم، فهو يمثل قاعدة البيانات للمفاهيم الطبية الموحدة، حيث يتكون المكنز الوصفي من مرادفات، مجموعة من المفاهيم المقترنة، أول نسخة من المكنز الوصفي احتوت على 30000 مفهوم، تحتوي على 60000 كلمة، 100000 علاقة. أما الآن فالنسخة الأخيرة تحوي على أكثر من 160 مصدرًا للمصطلحات (مكانز، قواميس..الخ) و7 ملايين كلمة و13 مليون علاقة، وأهم ما يميز المكنز الوصفي لنظام (UMLS) هو جمعه للمصطلحات من مصادر مختلفة وبلغات مختلفة (Lord, Bechhofer & Wilkinson, 2004)، وقد شكل المكنز الوصفي Metathesaurus حجر الأساس لنظام (UMLS)، حيث تم تصميم القاموس من خلال ربط كل مفهوم بصفاته الخاصة التي تُعَرف معناه، وارتباطاته، وعلاقاته بالمفاهيم الأخرى المتعلقة به في مصادر المفردات المختلفة (مكانز أو قواميس لغوية متعددة)، فالعلاقات المتعددة بين المفاهيم تتدرج هرميا داخل المكنز الوصفي، فإذا استخدمت مفردات مختلفة لنفس المفهوم أو إذا استخدم نفس المصطلح لمفاهيم مختلفة، يتواجد ذلك بشكل موثوق في (Metathesaurus)، ويمكن أن ترتبط المفاهيم في المكنز الوصفي (Metathesaurus) بمصادر خارج قاعدة البيانات من أجل رصد مصادر المعلوماتية الحيوية. (Cannata, Schroder & Marangoni, 2008)

ب) الشبكة الدلالية: Semantic Network (UMLS)

تتكون الشبكة الدلالية لنظام (UMLS) من مجموعة من العلاقات الدلالية، التي توفر تصنيفا متسقا لجميع المفاهيم في المكنز الوصفي Metathesaurus، الشبكة الدلالية هي واحدة من العناصر الثلاثة التي تكون نظام (UMLS)، تتضمن الشبكة الدلالية لنظام (UMLS) 133 نوعا دلاليا و54 علاقة دلالية، بحيث كل مفهوم بالمكنز الوصفي (Metathesaurus) يرتبط بمثيله عبر علاقات دلالات الألفاظ داخل الشبكة الدلالية (Semantic Network). وتتضمن المعلومات عن أنماط دلالات الألفاظ:(المعرف به – التعريف – الأمثلة – المعلومات حول دلالات الألفاظ المحيط بالمعنى والعلاقات المرافقة)، حيث إن الرابط الأساسي بين دلالات الألفاظ هو الرابط (is_a) المؤسس للأنماط التسلسلية الهرمية الخمسة، والتي تُشكل فيما بعد أنماط العلاقات الدلالية، وهذه العلاقات الخمس هي: (Sheth, York & Thomas, 2004)

(1) متعلق جسديا بـ (Physically related to) – (2) متعلق مكانيا بـ(Spatially related to)

(3) متعلق زمانيا بـ (Temporally related to) – (4) متعلق وظيفيا بـ(Functionally related to)

(5) متعلق كمفهوم بـ (Conceptually related to)

ج) القاموس المتخصص: Specialist Lexicon

القاموس المتخصص Specialist Lexicon هو المكون الثالث في نظام (UMLS)، حيث يقدم المعلومات اللازمة لنظام معالجة اللغة الطبيعية (NLP)، فهو يشتمل على المفردات الطبية الحيوية الإنجليزية، فيتكون من مجموعة من المدخلات اللغوية (مدخل واحد لكل هجاء)، بعد ما يتم التأكيد على أن المصطلح هو مدخل تمت تسميته واعتماده بسبب وجوده كمصطلح في القواميس الإنجليزية العامة أو الطبية، أو المكانز مثل قائمة رؤوس الموضوعات الطبية MeSH. كما يدعم القاموس Lexicon مؤخرًا التوسعات في الاختصارات المستخدمة في المجال الطبي الحيوي. (Cudré-Mauroux & Gaugaz, 2005)

ويعتمد القاموس على الاستخدام الفعلي في قواعد بيانات الإنتاج الفكري مثل (MEDLINE)، وقواميس اللغة الإنجليزية العامة، والقواميس الطبية، والمصادر المرجعية مثل: قاموس لونغمان للغة الإنجليزية المعاصرة، وقاموس دورلاند الطبي المصور، وقاموس كولينز كوبويلاند، وقاموس أكسفورد للدارس المتقدم، وقاموس ويبستر الطبي، حيث يتم إدخال محتوى كل هذه القواميس وأكثر في القاموس المتخصص Specialist Lexicon، ومن ثم الربط بين أجزائه في قاموس واحد الذي يحوي الآن قرابة 1.6 مليون مفهوم، وحوالي 5.5 مصطلحات مفردة تم جمعها من المصادر آنفة الذكر.

*** اللغة العربية في النظام الموحد للغة الطبية:**

في هذا السياق لابد أن نتساءل أين اللغة العربية من هذا الزحم من اللغات؟ وأين المصطلحات الطبية المترجمة باللغة العربية؟، إن هناك بعض المصادر المترجمة إلى العربية مثل: "التصنيف الدولي للأمراض ICD" الذي تمت ترجمته عام 2006، من قبل القسم العربي بمنظمة الصحة العالمية، لكن غياب التنظيم الذي يسير الترجمات وينظمها ويعمل على إصدار النسخ المتتابعة منها، يجعل الترجمات لا تُعتمد، ولا يمكن الاعتماد عليها داخل أي نظام وهذا منطقي نظرا لطبيعة علم الطب الحيوي والذي تتغير فيه المصطلحات شهريًا، ولأن المشكل ليسست الترجمة فقط بل تتبع المحذوفات وتغيير العلاقات وترجمتها في حال كانت هناك إضافات (Khayat, 2006). ويبقى السؤال: متى سنرى مؤسسات وفرق بحثية جامعية عربية تتبنى الترجمة الطبية، وتعمل على تنظيم عملية الترجمة للمصادرالطبية؟ حتى نستطيع إدراج اللغة العربية – على الدوام– في قواعد بيانات المصطلحات العالمية، مثل نظام (UMLS).

تحديات الويب الدلالية في مجال المعلوماتية الحيوية:

– التعرف على المصطلحات واستنباط العلاقات بينهم تلقائيًا، واكتشاف التقنيات والطرق التي تدعم إدراج دلالات الألفاظ بطريقة تلقائية على صفحات الويب.

– كيف يمكن لأدوات الويب الدلالية (مثل: الأنطولوجيا) دعم مهمة اختيار المصطلحات بشكل فعال أثناء التحضير لإنشاء الأنطولوجيا واكتشاف المصطلحات في قواعد بيانات علوم الحياة؟.

9/ **الأنطولوجيا وضبط المصطلحات في مجال المعلوماتية الحيوية**:

عرفها فيكرى عالم المعلومات بأنها: "خطة تتضمن استخدام الفئات الدلالية للمفاهيم المهمة في ميدان معين، مع تعريف أو تبصرة توضيحية لكل مفهوم، وآليات لعرض علاقاته بالمفاهيم الأخرى". (عبد الهادي، يوليو 2010). ويشير مصطلح الأنطولوجيا الحيوية المفتوحة OBO) إلى خلق المعاجم المنضبطة، والمتاحة للاستخدام المشترك، ففي خضم الكم الهائل من البيانات الحيوية، فإن الأنطولوجيا هي أداة مهمة من أجل فك الاشتباكات الدلالية بين المصطلحات داخل قواعد البيانات البيولوجية؛ من أجل تعزيز تكاملها، ودعم عمليات تبادل البيانات بين تلك القواعد، حيث تبرز أهمية الأنطولوجيا في قيامها بشرح البيانات البيولوجية، وتحليلها، وتوحيدها، فنظرًا لضخامة حجم البيانات الحيوية، تم التفكير في إجراء عمليات التوحيد للمصطلحات والمسميات، حيث يمكن من خلالها تقييم معاني المصطلحات المستخدمة في البيانات الوصفية، وتنظيمها وفقًا لعلاقتها الدلالية (Carpendale, Chen & Evanko, 2014). ونظرا لتعقد البيانات الحيوية نشأت مشكلات: فهم المعنى، وتصنيف الأفكار، واسترجاع المعلومات وتصنيفها وعمل روابط لها، كل ذلك أدى إلى الحاجة إلى نشوء تخصص علمي جديد هو Ontology، الذي يستخدم من قبل الحاسب الآلي في عمليات الوصف والبحث.

وقد تنوعت الأنطولوجيا في مجال المعلومات الحيوية، ما بين الاعتماد على مكانز سابقة بعد تجهيزها وتحويلها لأنطولوجيات كما حدث في مكنز (MESH)، أو بناء أنطولوجيات جديدة من أجل استيعاب الكم الهائل من البيانات البيولوجية، وتنوعت أيضا ما بين الأنطولوجيا المقيدة التي يتم تطويرها وإضافة مصطلحاتها بواسطة خبراء المعلومات مثل: (Gene Ontology)، وما بين المفتوحة لاقتراح مصطلحات جديدة، مثل: (Open Biomedical Ontology). (Baker, Goble & Bechhofer, 1999). وتتكون الأنطولوجيا من مجموعة من الفئات Classes، وكل فئة تحوي مجموعة من (المصطلحات Terms أو المفاهيم Concepts) التي ترتبط فيما بينها بعلاقات مفاهيمية Relations، يتم اختيارها طبقًا لطبيعة البيانات والمجال المستخدمة فيه.

نماذج من الأنطولوجيا المستخدمة في المعلوماتية الحيوية:

1) Gene Ontology (GO)

أنطولوجيا الجينات (GO) هي مبادرة أمريكية بدأت عام 1998 لتوحيد تمثيل سمات الجين، فقد هدف المشروع إلى: تطوير لغة مقيدة لوصف سمات ووظائف الجين، حيث تشتمل الأنطولوجيا على مجموعة من

المفردات المنظمة المستخدمة لوصف وتصنيف الجينات بناءً على وظيفتها الجزيئية، وموقعها الحيوي في الخلية، ويتم تنقيح مصطلحات الأنطولوجيا أسبوعيًا، لتغير المفاهيم البيولوجية. (Consortium, 2020).

وفيما يخص العلاقات بين المصطلحات، فإن الأنطولوجيا في عرضها للمعلومات تستخدم العلاقات المعروفة في علم المكانز (الأعلى Broad، والأضيق Narrow، والمترابط Related)، حتى إن إحالة "انظر أيضًا" تستخدم من خلال مصطلح (Exact) الذي يشير إلى ترادف المصطلحين، وهذا يوضح أن الأساس الذي بنيت عليه تلك الأنطولوجيات المتخصصة، هو أساس علم المكتبات والمعلومات في تطويره لأدوات ضبط المصطلحات، إلا أن العلاقات الضمنية التي بنيت عليها الأنطولوجيا تستخدم العديد من العلاقات في ربطها للمصطلحات، مثل: (is_a، part_of، has_ part، regulates) والأخيرة بمعنى: "مصطلح ما "ينظم" مصطلح آخر".

وبالنظر إلى بناء أنطولوجيا الجينات (GO)، فإنها تحوي تقسيمًا فرعيًا يشمل ثلاث أنطولوجيات فرعية، لا علاقة بين مصطلحات كل أنطولوجيا بالأخرى، وهذا يقودنا إلى النظر في البناء الحديث لأدوات ضبط المصطلحات، واختلاف الأنطولوجيا عن الأدوات الأخرى التي اعتدنا عليها في قيامها برط كل مصطلحات الأداة الموضوعية في عمل واحد، والذي يستتبعه إنشاء العلاقات بين جميع مصطلحات أداة العمل، وتنقسم الأنطولوجيا إلى ثلاث أنطولوجيات فرعية، تحوي في مجملها (44,700) مصطلح، للتعريف بوظائف (1,351,824) جين، تم جمعها من (4,591) نوع من أنواع الكائنات الحية، وتقدم تعليقات علمية Annotations موثقة من مصادر ببليوغرافية حول عمل الجين تصل في مجملها الى (7,437,710) تعليق توضيحي. (Gene Ontology Resource, 2020)

وفيما يلي موجز للأنطولوجيات الثلاث: (Dessimoz, & Nive, 2017)

1. العمليات الحيوية Biological Process: تحوي مصطلحات العمليات البيولوجية التي تحدث على المستوى الأعلى التي يتم إنجازها بواسطة أنشطة جزيئية متعددة، وليس المستوى الجزيئي كما في فئة "الوظائف الجزيئية"، وتحوى هذه الانطولوجيا (29,393) مصطلح.

2. المكونات الحيوية (الخلوية) Cellular Component: تشير مصطلحاتها إلى المواقع المرتبطة بالهياكل الخلوية التي يؤدي فيها الجين وظيفة ما، ولا تشير هذه الفئة إلى العمليات والوظائف التي يحدثها الجين، بل إلى التشريح الخلوي، وتحوى هذه الأنطولوجيا (4,184) مصطلح.

3. الوظائف الجزيئية Molecular Function: تحوي مصطلحات تصف الأنشطة التي تحدث على المستوى الجزيئي (DNA, RNA, Proteins). وتحوي هذه الأنطولوجيا (11,123) مصطلح.

The gene product "cytochrome c" can be described by the **molecular function** *oxidoreductase activity*, the **biological process** *oxidative phosphorylation*, and the **cellular component** *mitochondrial matrix*
يمكن وصف الجين المسمى بـ "cytochrome c": بأن (**وظيفته الجزيئية**) "احداث نشاط أكسدة"، وأن (**العملية البيولوجية**) التي يقوم بها "مؤكسس فسفوري"، وأن (**مكانه الخلوي**) "مصفوفة الميتوكوندريا"

شكل (8): مثال يوضح فئات ومفاهيم وعلاقات انطولوجيا الجين. المصدر: (Gene Ontology Resource)

2) Medical Subject Headings (MeSH) قائمة رؤوس الموضوعات الطبية

هذه القائمة من الأدوات التي تثبت وتبرهن على الاتصال الوثيق بين علوم المكتبات والمعلومات وبين المعلوماتية الحيوية، فقد بدأت رحلة قائمة رؤوس الموضوعات الطبية كأداة فنيه (ورقية) يتم استخدامها في عملية الفهرسة الموضوعية التي يتم من خلالها ربط الإنتاج الفكري بكلمات مفتاحية تسهل البحث عنها فيما بعد، ومع تطور تقنيات الحاسب الآلي في إدارة المعلومات، بدأ التفكير في تحويل هذه الأداة لأنطولوجيا تعمل على التكشيف التلقائي لمحتويات قاعدة البيانات الطبية الحيوي PubMed. حيث يعد مكنز "قائمة رؤوس الموضوعات الطبية" MeSH: Medical Subject Headings الآن من أهم المكانز الطبية التي تحوي مصطلحات متنوعة في المجالات الطبية، حيث بلغ عدد المصطلحات فيه حوالي (280) ألف مصطلح.

3) EMBL-EBI Ontology

أنطولوجيا تعمل على تكشيف الإنتاج الفكري لمحتويات قاعدة بيانات النص الكامل "Europe PMC" الذي يديرها المركز الأوربي للمعلوماتية الحيوية EMBL-EBI، وتحوي الأنطولوجيا على 243 أنطولوجيا فرعية، وبها (5,737,532) مصطلح.

تحديات تواجه متخصصي علوم المكتبات والمعلومات في مجال الأنطولوجيا الحيوية:

– ضرورة إعادة النظر في الأدوات المستخدمة في المعالجة الفنية لمجموعات المكتبات ومراكز المعلومات، والتفكير في تطوير تلك الأدوات، والعمل على نقلها نقلًا نوعيًا، من استخدامها كأداة تقليدية إلى أنطولوجيا، بعد الانتهاء من دراسات الجدوى من هذا النقل؛ للاختيار ما بين إنشاء أدوات معرفية جديدة أو تطوير أدوات قديمة.

- إنشاء "الأنطولوجيا" الطبية الحيوية العربية"، لمواجهة التحديات المتعلقة بتعامل الحاسب الآلي مع المصطلحات الطبية العربية.

10/ **التنقيب عن البيانات الضخمة في المعلومات الحيوية**: Big Data Mining

تكمن أساسيات المعلوماتية الحيوية في إدارة التحليل والتعامل مع المجاميع الكبيرة من البينات الحيوية، لتكون قادرة على: (Odebode, 2011)

❋ تهيئة قواعد بيانات تسمح بخزن وإدارة وإضافة وتجديد البيانات الحيوية.

❋ تطوير الخوارزميات والبرمجيات التي تعمل على تحديد، وتحليل وإدارة البيانات.

❋ تفسير الأنماط التي تنتج من عملية تحليل الأنواع المختلفة من البيانات الحيوية (تواليات DNA وتواليات RNA والبروتينات).

والتنقيب عن البيانات هي عملية استخراج معلومات مفيدة من بيانات كبيرة غير منظمة، قد يكون مصدر البيانات أوراق أو ملخصات أو صفحات على شبكة الإنترنت أو رسائل بريد إلكتروني...إلخ. وبما أن البيانات غير منظمة أساسا، فإنه يتعين تمثيلها في هيكل من نوع ما، وإيجاد أنماط في النص قبل تطبيق تقنيات استخراج النصوص، والمعلوماتية الحيوية أحد الحقول الرئيسية التي يتم استخدامها بشكل أساسي لاكتشاف المعرفة في النصوص (KDT)، وتتمثل صعوبة التنقيب عن البيانات في المعلوماتية الحيوية والمصطلحات في المجال الطبي ليست متسقة في العديد من المصادر، والاختصارات الغامضة تؤدي إلى صعوبة في تنفيذ استخراج النصوص في مجال المعلوماتية الحيوية، ويتم حل مشكلة الاستخدام الفعال للبيانات عن طريق استخدام تقنيات استخراج النصوص من الإنتاج الفكري المتاح في المكتبات الرقمية (مثل: PubMed، Europe-PMC)، حيث يتم تحليل البيانات غير الهيكلية واستخراج معلومات مفيدة منها (Peng, Mamitsuka & Zhu, 2018). وهذا من شأنه أن يمكن لاكتشافات جديدة في وقت أقل ومساعدة الخبراء في التوصل إلى تشخيص واقعي بمساعدة المعلومات المستخرجة.

التنقيب عن البيانات بشكل عام يحتوي على ثلاث خطوات: التجهيز المسبق، واستخراج النصوص، والمعالجة اللاحقة، ويتم تطبيق مرحلة ما قبل المعالجة في التخلص من العديد من الكلمات غير المرغوب فيها، وإزالة كلمات التوقف، وتقطيعه وتحويله إلى جمل منظمة، ليبقى في النهاية الجزء الرئيسي من النص المستخدم في عملية التنقيب عن البيانات باستخدام نماذج تعلم الآلة أو الشبكات العصبية، وتتضمن خطوة ما بعد التجهيز تفسير وتصور نتائج عمليات استخراج النصوص.

*** مهام التنقيب عن المعلوماتية الحيوية، ودور أخصائي المكتبات والمعلومات:**

هناك العديد من مهام استخراج النصوص في مجال المعلوماتية الحيوية (Cohen & Hunter, 2013) نوجزها في الاتي:

1) استرجاع المعلومات (Information Retrieval): تعني البحث عن المستندات ذات الصلة بالاستعلام، يعد PubMed/MEDLINE مثالًا على نظام استرجاع المقالات العلمية في مجال الطب الحيوي؛ وGoogle عبارة عن نظام استرجاع معلومات لصفحات الويب.

2) تصنيف الوثائق (Document Classification): تصنيف الوثائق هي مهمة تصنيف المستندات إلى فئات موضوعية، فمثلًا قد يتطلب القائمون على قاعدة بيانات استخراج المعلومات المصنفة لكائن معني، أو تصنيف مقالات الدوريات حول ما إذا كانت تتعلق بجنين معين أم لا؟. عادةً ما يستخدم تصنيف المستندات مجموعات ميزات بسيطة للغاية، مثل: وجود أو عدم وجود الكلمات، تُعرف باسم "حقيبة الكلمات Bag of Words". (Caporaso, Baumgartner & Cohen, 2005)

3) التعرف على الكيانات الحيوية (Named Entity Recognition): تعني البحث عن فئات ذات دلالات محددة في النص، مثل: الأشخاص، والأماكن والمنظمات، وأسماء الجينات، والبروتينات...إلخ اللائحة التي يتم تحديدها مسبقًا، تشمل الكيانات التي يتم البحث عنها والتعرف عليها، والتي تسمى "MetaMap" وأشهر مثال على ذلك "Metathesaurus" UMLS.الا أن هناك مشكلات وتحديات تواجه هذا النهج تتعلق بأسماء الجينات الجديدة التي يتم صياغتها باستمرار، وأن العديد من الجينات لها أسماء أو مرادفات مختلفة، ومع ذلك، فإن هذا النهج القائم على استخدام القوائم المسبقة أو المكانز (القوامسيس) يمكن أن يحقق نجاحًا إذا تعرضت تلك الأدوات إلى معالجة مسبقة واسعة النطاق.

4) تطبيع الكيانات الحيوية (Name Entity Normalization): تطبيع الكيان هي عملية البحث عن كيان محدد في قواعد البيانات الحيوية، وإرجاع معرف يشير إليه في المجال البيولوجي، ويتم إجراء تلك المهام على نطاق واسع على الجينات والبروتينات، إلا أن هناك مشكلتين رئيسيتين في تطبيع الجينات: (Armstrong, 2010)

- المشكلة الأولى: العديد من أنواع الجينات لها نفس الاسم، مثل: جين "BRCA1" الموجود في عدد كبير من الحيوانات، وبالتالي، فإن العثور على الجين المناسب يتطلب معرفة الأنواع قيد المناقشة، وهي المشكلة البحثية القائمة بحد ذاتها، والتي يرغب الباحث في التعرف عليها.

- المشكلة الأخرى: الجين الواحد قد يكون له أسماء متعددة، مثل: جين (TRP-1) الذي يتواجد لدى البشر منه خمس جينات بنفس الاسم.

ويبقى التحدي في قيام القاموس بتحديد المصطلح المطلوب، فيستخدم هذا النهج السياق الذي يذكر فيه الجين لتحديد ماهية الجين المطلوب، من خلال المقارنة بين السياق في النص وبين حقول البيانات: (العنوان، المستخلص، النص...الخ) لتسجيلات أخرى مرتبط ذكرها بهذا الجين، أو الربط بين الجين المطلوب البحث عنه وعلاقته بجين آخر دائم الذكر معه.

5) استخراج المعلومات والعلاقات (Relation & Information Extraction): تعني عملية استخراج أنواع محددة من الحقائق، فعلى سبيل المثال: استخراج المعلومات الجينومية، وتفاعلات البروتين، والعلاقات بين الأمراض وعلاجاته، وعلاقات نقل البروتين، وموقع بداية البروتين المنقول، والوجه،. وتستخدم تلك المهمة الأساليب المستندة بناء الجمل النموذجية، فقد تتكون من حروف نصية أو قد تتضمن تحليلات نحوية، لكن الأساليب الحديثة نجحت في استخدام الميزات المأخوذة من التحليل النحوي، وبخاصة تحليل التبعية والعلاقة بين المعلومات. (Hunter, Lu & Firby, 2008)

6) الإجابة على الاستفسارات (Question-Answering): افترضت الأساليب المبتكرة أن العديد من الإجابات على الأسئلة في مجال المعلوماتية الحيوية، موجود في مصادر المعلومات الحيوية (قواعد البيانات النصية). تختلف مهمة إجابة الأسئلة عن استرجاع المعلومات، حيث إن الهدف هو إرجاع إجابة محددة، وليس مستندًا يحتوي الإجابة، والإجابة على الاستفسارات يهدف إلى السماح باستفسارات خاصة مفتوحة، بينما يركز استخراج المعلومات على احتياجات معلومات محددة للغاية، وقد تمت دراسة (Yu, Lee & Kaufman, 2007) للتنقيب عن البيانات باستخدام مهام الأسئلة في المجال الطبي الحيوي، حيث وجد أن درجات الأسئلة تتفاوت من حيث الصعوبة، حيث يتم تحقيق أفضل النتائج مع الأسئلة "الواقعية Factoid"، مثل :"أين"، في حين أن "لماذا" من الأسئلة الصعبة للغاية لأنها تحتاج استنباط من النص، (Zweigenbaum, 2003) والأسئلة الزمانية "متى" قد تتطلب أوقاتًا لإجابتها.

7) التلخيص (Summarization): يعني أخذ مستند أو مجموعة من المستندات كمدخلات وإرجاع نص أقصر، هناك حاجة كبيرة لهذه الإمكانية في المجال الطبي الحيوي، وقد تم تطبيق التلخيص على مقالات الدوريات ومجموعة متنوعة من أنواع المدخلات الأخرى. ومن أمثلة تلك الأنظمة التي تقوم بذلك نظام (MITRE'S MiTAP) (Damianos, Zarrella & Hirschman, 2013) الذي يقوم بتلخيص المستندات، والتقارير الطبية، والأخبار الرقمية والتليفزيونية والإذاعية للكشف عن الأوبئة وتفشي الأمرا،. وفي مجال الجينوم، هناك العديد من النماذج المستخدمة في تلخيص الإنتاج الفكري الحيوي بما ينتج عنه معلومات تقوم بمقام تحليل المعلومات الحيوية والتفاعل فيما بينها:

* أولها: نموذج (GeneRIFs) الذي يعمل على استخراج فقرة نصية قصيرة يبلغ طولها أقل من 255 حرفًا، تعمل على إدخال وظيفة الجين في قاعدة بيانات PubMed.

* ثانيها: نموذج (Edmundsonian) وهو نهج يعمل على إعطاء الجمل الملخصة نقاطًا (وزنًا) وفقًا لمجموعة من المعايير النسبية، مثل: موضع الجملة في الوثيقة، فالعنوان يختلف عن المستخلص وعن النص الأصلي، كذلك نوع الجمل ما بين "الكلمات الجدلية Cue Words" التي تشير إلى ارتباط الجملة بالنص، والكلمات غير ذات العلاقة "Stigma Words" التي تشير إلى أنه من غير المرجح أن تكون الجملة ملخصا جيدا.

ويبقى التحدي هنا في إيجاد أفضل جملة تعبر عن النص، كذلك الجمل التي تبرز بعض الحقائق حول عملية تحليل المعلوماتية الحيوية، مثل: "تأكيد تفاعل بروتين معين أم لا" من داخل النص"، ويتم التغلب على ذلك بفكرة BioCreative المتمثلة في اشتراك مجموعة من المقالات في إثبات التساؤل ذاته، لتعطي أفضل دليل على حدوث هذا التحليل (التفاعل)، وذلك بالنظر والبحث عن الإشارات والأشكال أو الجداول المشتركة باستخدام خوارزميات معنية بذلك.

نماذج من تطبيقات التنقيب عن البيانات الحيوية:

1) Biodoop - Bioinformatics on Hadoop:

أباتشي هادوب هو برنامج أو منصة برمجية مفتوحة المصدر مكتوبة بلغة ألجافا لتخزين ومعالجة البيانات الضخمة بشكل موزع، أي: تخزين البيانات الضخمة على عدة أجهزة ومن ثم توزيع عملية المعالجة على هذه الأجهزة لتسريع نتيجة المعالج،. وبرنامج "Biodoop" هو مجموعة من الأدوات لعلم الأحياء الحسابي (المعلوماتية الحيوية) التي تركز على التنفيذ الفعال للمهام ذات البيانات الحيوية الضخمة، وهو يتألف من مكونات خاصة وتطبيقات متعلقة بعمليات تحليل المعلوماتية الحيوية، مثل: محاذاة التسلسل، وتعمل التطبيقات بشكل عام على واجهة برمجة تطبيقات Pydoop لـ Hadoop وهي مصممة لتتناسب بشكل جيد مع الحوسبة المتوفرة وكمية البيانات المطلوب معالجتها، مما يجعلها مناسبة بشكل خاص لمعالجة مجموعات البيانات الكبيرة. (Leo, Santoni & Zanetti, 2009)

2) Litminer:

من أمثلة التطبيقات التي تقوم بهذه الوظائف تطبيق تحليل النصوص LitMiner الذي أعد لتمكين البيولوجيين من استكشاف العلاقات بين الجينات الواردة في الأبحاث المنشورة بقواعد بيانات النصوص الكاملة، فيتميز التطبيق بمجموعة من الواجهات التي تمكن المستخدمين من البحث في قواعد البيانات المحلية،

وقواعد بيانات PubMed وتحليل واستكشاف العلاقة بين الجينات الواردة في تلك القواعد، فيوفر التطبيق مجموعة من الأدوات التى تمكن المستفيدين من التركيز على المعلومات (وصفا وتحليلا) في تطبيق واحد بدلا من التبديل بين التطبيقات المتعددة، ويتميز التطبيق بالاتي: (Demaine, Martin & Wei, 2006)

- إمكانية استيراد الملفات من قواعد البياناات المتعددة، والتحكم في النتائج.
- تجميع نتائج البحث في شكل هرمي بناء على التقسيم الموضوعي للكلمات المفتاحية.
- التعرف على الجينات الواردة في أدبيات الموضوع والمقارنة بينها، وتصوير العلاقات بين تلك الجينات – الواردة في قواعد بيانات النصوص الكاملة – باستخدام الرسوم البيانية.
- إمكانية التعديل في مكنز المصطلحات بناء على ما يراه البحث مناسبا عند الإشارة إلى الجينات الواردة في البحوث الحيوية.

11/ **الحوسبة السحابية في مجال المعلومات الحيوية**: Cloud Computing

يشير هذا المصطلح إلى إمكانية حصول الشركات والأفراد على بعض الخدمات كالتخزين وقواعد البيانات والبرمجيات وغيرها من الخدمات المختلفة من خلال شبكة الإنترنت، بعبارات بسيطة، يُقصد بالحوسبة السحابية تأجير تكنولوجيا المعلومات بدلًا من شرائها، فبدلًا من الاستثمار الكامل في قواعد البيانات والبرامج والمعدات، تختار الشركات الوصول إلى قوة الحوسبة عبر الإنترنت وتدفع مقابل استخدامه. وتنقسم أنواع خدمات الحوسبة السحابية التي تقدمها إلى الأنواع الثلاث التالية: (Google Cloud)

1. خدمات البنية الأساسية (Infrastructure as a service) والذي يشار إليه بـ (IaaS) وتُقدم خدمات التخزين عبر الإنترنت من خلال هذا النوع من الخدمات السحابية، وتقدر أسعار الخدمة بناء على حجم المساحة التخزينية المطلوبة، وسرعة نقل البيانات.

2. خدمات المنصة (Platform as a service) والتي يشار إليه بـ (PaaS) وتتألف بشكل أساسي من مكتبات، وبرامج وسيطة، وتحديثات وأدوات يحتاجها المطورون في تطوير التطبيقات والبرمجيات كخدمة.

3. خدمات البرمجيات (Software as a service) والتي يشار إليه بـ (SaaS) ويُعنى هذا النوع من الخدمات السحابية بتقديم خدمات البرمجيات المختلفة عبر الإنترنت، المتعلقة بالمستخدم النهائي، حيث تتيح العديد من الحلول البرمجية في مختلف المجالات.

وتتميز خدمات الحوسبة السحابية بتوافرها كل الوقت ومن أي مكان، وقلة إمكانية تواجد الخطأ، وحل المشكلات تكون بشكل أسهل وأسرع، إضافة للمرونة في إمكانية التوسع دون انتظار أو تأخير، ضمان جودة عالية لخدمة المستخدمين من حيث أداء الأجهزة، ومعدل نقل البيانات، وسعة الذاكرة، ولا تتطلب خدمات

الحوسبة السحابية تمويلًا كبيرًا، فلا حاجة لوجود نفقات رأسمالية، حيث يمكن للمستخدمين دفع ثمن الخدمات والسعة حسب الحاجة.

تطبيقات الحوسبة السحابية في المعلوماتية الحيوية:

يزخر مجال المعلوماتية الحيوية بأمثلة على تطبيقات الحوسبة السحابية في مجال المعلوماتية الحيوية، نتيجة لانخفاض تكاليف الأجهزة، وعملية الحوسبة السحابية، والقدرة على إجراء المعالجة الموازية، والتقدم الخوارزمي، وما صاحبه من انخفاض تكاليف ووقت عمليات التحليل الحيوي، وفيما يلي عرض لنماذج من موردي خدمات الحوسبة السحابية في المعلوماتية الحيوية:

1) Google Cloud Life Sciences:

شركة Google العملاقة لا غنى عن التعريف عنها وعن حجم الحلول التي تقدمها لمجتمع المعلومات، وتمكن علوم الحياة السحابية (Cloud Life Sciences) مجتمع علوم الحياة من معالجة بيانات الطب الحيوي على نطاق واسع، حيث تتيح Cloud Life Sciences التركيز على تحليل البيانات ، وإمكانية إعادة التحليلات أثناء استخدام الحوسبة السحابية Google Cloud Platform، وذلك بفضل التكلفة المناسبة التي تقدمها شركة Google، والتي تناسب جميع المستويات مهما كان حجم التحليلات الحيوية المطلوبة.

2) Aws Amazon – Bioinformatics & Genomes:

طورت شركة أمازون خدمة الحوسبة السحابية للمعلوماتية الحيوية، حيث تم بناء أدوات تعمل على تخزين وتحليل البيانات الحيوية، وتقديم حلول البنية التحتية القائمة على الحوسبة السحابية في مجال المعلوماتية الحيوية، إضافة إلى توفير إمكانيات دمج التحليلات الحيوية، وإمكانيات تعلم الآلة، وإمكانيات تعمل على دمج نتائج التحليلات الحيوية التي يتم التوصل إليها في الممارسة الطبية، والذي ينعكس ذلك بدوره على تقدم الرعاية الصحية لمجتمع المستفيدين. (Genomics Cloud Computing)

تحديات الحوسبة السحابية في مجال المعلوماتية الحيوية:

لا يزال هناك عدد من التحديات، والتي تُعالج حاليًا من قبل الباحثين والممارسين، وفيما يلي عرض لهذه التحديات بإيجاز:

– الأداء (Performance): بسبب طبيعة البيانات الحيوية، فإن المشكلة الكبرى تتمثل في الأداء لبعض التطبيقات ذات البيانات الكثيفة، وفيها، قد تفتقر الحوسبة السحابية للأداء المناسب. والمستخدمون البعيدون عن مزودي السحابة ربما يواجهون ا في نقل البيانات.

– الأمن والخصوصية (Security and Privacy): بسبب حساسية المعلومات الحيوية وبخاصة المعلومات الوراثية للإنسان، فلا تزال الشركات تشعر بالقلق إزاء الأمن عند استخدام الحوسبة السحابية، فالمستفيدون قلقون بخصوص التعرض لهجمات عندما تكون المعلومات المهمة ومصادر تقنية المعلومات خارج جدار الحماية.

– التحكم (Control): تشعر بعض الشركات بالقلق لأن مقدمي الحوسبة السحابية يمتلكون السيطرة الكاملة على المنصات، ومقدمو الحوسبة السحابية لا يقومون عادة بتصميم منصات لشركات محددة وممارساتها التجارية.

– تكاليف معدل نقل البيانات (Data Transfer Rate Costs): مع الحوسبة السحابية، يمكن للشركات توفير الأموال المهدرة على الأجهزة والبرمجيات، إلا أنها يمكن أن تتحمل ارتفاع رسوم معدل نقل البيانات للشبكة، وقد تكون تكلفة معدل نقل البيانات منخفضة بالنسبة لتطبيقات الإنترنت الصغيرة، والتي لا تُعد كثيفة البيانات، ولكنها يمكن أن ترتفع بشكل كبير بالنسبة لتطبيقات البيانات الحيوية الكثيفة، وبخاصة وأن هذه إحدى السمات الرئيسية للبيانات الحيوية.

– الدقة والموثوقية (Accuracy and Reliability): لا تزال الحوسبة السحابية لا تقدم موثوقية دائمة على مدار الساعة، لأن هناك بعض الحالات التي تعاني فيها خدمات الحوسبة السحابية من انقطاع التيار لعدد قليل من الساعات، ولكن في المستقبل، نتوقع أن نرى المزيد من مزودي الحوسبة السحابية، بمعايير معمول بها، وممارسات أفضل.

12/ **استرجاع المعلوماتية الحيوية**: Bioinformatics Retrieval Systems

توفر المعلوماتية الحيوية الكم الهائل من البيانات المخزونة في قواعد البيانات ويكون في كثير من الأحيان استرجاع هذه البيانات لأغراض شتى، وبخاصة استرجاع التواليات والتعرف على التواليات من حيث موقعها في الجين أو وظيفتها أو أية معلومات موجودة ضمن عملية التهميش المرافقة للتواليات، ويعد استرجاع البيانات من المهام الرئيسة المنوطة بها المعلوماتية الحيوية سواء البيانات الأولية (التواليات) أو البيانات الوصفية، ولعملية الاسترجاع في حقل المعلوماتية الحيوية مشاكل كثيرة تعيق عمليات البحث نتيجة من عدة أسباب، يمكن تلخيص بعضها فيما يلي:

– عدم إيفاء البيانات الوصفية للتسلسلات بجميع المعلومات التي يحتاجها الباحث، فالنهايات الصغرى من البيانات الوصفية تختلف من مؤسسة لأخرى.

– اللغة المستعملة في البحث ليست قياسية، وحتى عند استعمال اللغة الإنكليزية كلغة تواصل فهناك اختلافات في الإملاء، فالإنكليزية الأمريكية تكتب (اللون Color)، في حين أن الإنكليزية البريطانية يكتبها (Colour).

– تفسير الاختصارت اللغوية أحيانًا بشكل خاطئ، فمثلا:(Media) تعني الإعلان ووسائله في العموم، في حين تعني: "الوسط الزرعي لتنمية الأحياء المجهرية" للعاملين في مجال الأحياء المجهرية، لذلك لابد من وجود بيانات وصفية إضافية تمكن الحاسب من إرجاع الكلمات المستعملة ضمن النطاق المطلوب.

تسهل أنظمة الاسترجاع الدخول إلى عدة قواعد بيانات للحصول على نتائج البحث، وعند استخدام استعلامات معقدة في قاعدة ما، لابد من استخدام العلاقات الجبرية Boolean operators لتضيق عمليات البحث، حيث إن أغلب قواعد البيانات البيولوجية تستعمل بعض هذه العلاقات الجبرية.

نماذج من أنظمة البحث واسترجاع البيانات بقواعد المعلومات الحيوية:

1) Entrez Retrieval System (NCBI)

نظام Entrez لاسترجاع البيانات يمثل نقطة البداية في البحث في قواعد معلومات NCBI، حيث يعمل النظام على البحث في جميع قواعد بيانات NCBI البالغ عددها حوالي (38) قاعدة بيانات، والتي تتنوع ما بين قواعد بيانات الأحماض النووية (DNA, RNA)، والأحماض الأمينية (Proteins)، فضلا عن قواعد بيانات الإنتاج الفكري مثل PubMed...إلخ ، لذلك لا يحتاج المستخدم عند البحث الولوج إلى قاعد بيانات بعينها، بل البحث فيها جميعًا، ومن ثم عرض النتائج مصنفة حسب قواعد البيانات.

2) EMBLNEW retrieval system SRS (EBI)

نظام استرجاع تابع للمركز الأوربي وهو أكثر شمولية من نظام Entrez. وتوفر واجهته إمكانيات كبيرة للحصول على المعلومات مثل: الدخول إلى أكثر من 80 من قواعد البيانات التي تتمثل في قواعد بيانات التواليات، والجينومات الكاملة (المحدد تواليها)، والخرائط الوراثية والطفرات ومواقعها، فضلا عن إمكانيات للتنبؤ بالتركيب الثلاثي للبروتينات...إلخ. وبطبيعة الحال يتيح النظام إمكانية الربط التشعبي لنتائج البحث من خارج النظام الأوربي EBI. ويتم تجديد القواعد بشكل دوري كل ثلاثة أشهر، وتحتوي النتائج على التوالي المستعلم عنه والتهميش الخاص به وروابط إلى المراجع الخاصة، وينتج نظام الاسترجاع هذا صيغ ملفات تعرف

بـ EMIBL format وبخاصة بالمركز الأوربي تحوي معلومات تخص التواليات المستعلم عنها، وهناك العديد من الصيغ التي يمكن أن تخص كل قاعدة بيانات أو برنامج سيتم الإشارة اليها لاحقًا. (The European Bioinformatics Institute, 2020)

الإرشادات التي تساعد في عمليات البحث بقواعد المعلوماتية الحيوية:

- ترجمة تواليات DNA إلى تواليات البروتين ثم البحث عن الأخيرة في قواعد البروتينات، بدلًا من البحث مباشرة بتواليات DNA، لأن الترجمة تعطي نتائج أفضل وأدق.
- البحث في أكثر من قاعدة بيانات أو برنامج للوصول إلى نتائج معينة.
- استخدام وسائل البحث الموجهة والسريعة، ثم الانتقال إلى وسائل البحث الدقيقة.
- عند وجود الإمكانية يفضل استعمال تواليات البروتينات بدلا عما يقابلها من توالياDNA .
- يفضل البحث باستعمال تواليات استعلام طويلة مثلا لـ DNA تكون أكبر من 1000 من القواعد النتروجينية وللبروتينات بحدود 200 حامض أميني.
- في حالة احتواء توالي الاستعلام على المكررات Repeats فيجب إزالتها ولهذا توجد برامج خاصة، ثم معاودة البحث في قواعد البيانات.

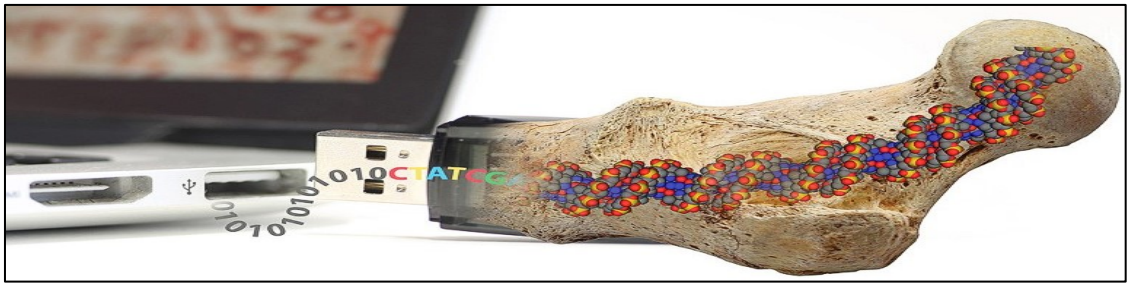
13/ **المادة الوراثية ومستقبل وسائط التخزين**: DNA and the future of data storage

"الاكتفاء بتخزين البيانات في الأقراص الصلبة يعني إلقاء جميع بياناتنا في ثقب أسود مجهول دون أن ندري." كانت تلك مقولة لفينت سيرف الباحث في مجال الإنترنت معلقا ومقارنا بين وسائط التخزين المتعددة، وقد تساءل المختصون في علم المعلومات عن المدة التي قد تصمد فيها وسائل تخزين البيانات المستخدمة حاليا قبل أن تتلف، وفي هذا الصدد قامت شركة "باكبلاز Backblaze" الأمريكية بإجراء دراسات حول معدل استنزاف وسائط التخزين الحالية وقدرتها على التكيف مع الأوضاع البيئية والفيزيقية، وقد خلصت تلك الدراسات إلى: إذا تم تشغيل تلك الوسائط لمدة خمس سنوات، فإن معدل استنزاف تلك الوسائط قد يصل إلى 22٪ كل أربع سنوات، مما يعني أن خُمس وسائط التخزين معرضة للتلف بما تحمله من بيانات كل أربع سنوات، وهذا يعني أن أكثر الوسائط حظا لن تدوم أكثر من عشرين عامًا، وأن الذي بات مؤكا أن أيا من تلك الوسائط لن تدوم إلى الأبد.

لهذا يعكف المعنيون بنظم المعلومات على البحث عن بديل يسمح بتخزين البيانات لمدد أطول تمتد لمئات القرون، ولقد أوحت فكرة المادة الوراثية بالخلية DNA (الحمض النووي) وما تقوم به من دور في حفظ البيانات

الوراثية لملايين السنين الحل الأمثل لحفظ البيانات الأخرى غير الوراثية، وبخاصة أنها تتصف بقدرات هائلة على حفظ المعلومات في مساحة صغيرة للغاية والذي أسهم بدوره في تحفيز الباحثين لتخزين البيانات عليها بدلا من الأقراص الصلبةK إن DNA يُعتبر حجر الأساس فى الحياة، لكن من المحتمل فى المستقبل القريب أن يصبح أيضًا مستودعا ومخزنا لمختلف البيانات الرقمية، فقد كشفت أحدث التجارب التي تم عرضها في المؤتمر السنوي للجمعية الكيمائية الأمريكية عن اتصاف الـ DNA بميزة أخرى هي: استمرار فترة احتفاظها بالبيانات المخزنة عليها لفترة تصل إلى ألفي عام دون أن تتعرض للتلف، بينما تواصل الأقراص الصلبة تخزين البيانات لفترة زمنية لا تتجاوز العقود القليلة. (Stephen & Wong , 2001)

لقد قام باحثون بالمعهد الفيدرالي السويسري للتكنولوجيا ETH بمدينة زيورخ بتجريب نظام تخزين البيانات المبنية بنظام (1،0) داخل الحمض النووي DNA، وتم التوصل إلى: كل (1) جرام من المادة الوراثية DNA قادر على تخزين 455 إكسا بايت Exabyte ، حيث إن كل (1) اكسا بايت يعادل مليار جيجا بايت. مما يعكس الكثافة التخزينية الهائلة Large Storage Density للمادة الوراثية، لكن في الوقت ذاته تبقى مشكلة التحلل السريع Degrades Quickly لتلك المادة الوراثية أمر قائم، لكن يعتقد العلماء أن الحفاظ على استقرار المادة الوراثية من خلال طرق الحفظ التقليدية الموجودة: (التبريد، التجفيف، التغليف،...إلخ)، أو طرق الحفظ المجهرية الزجاجية يمكن أن يحافظ على المعلومات في خيوط DNA سليمة لآلاف السنين، لهذا فإن الجمع بين ميزة الكثافة التخزينية للبيانات وبين معالجة أمر التحلل السريع لتلك المادة الوراثية هو التحدي القائم في الوقت الراهن.



شكل (9): صورة تخيلية لتحويل البيانات من ترميز الحاسب الى ترميز DNA. المصدر: (CNN.com)

ترجمة البيانات من لغة الحاسب إلى ترميز DNA بالخلية:

هناك العديد من المحاولات التي قامت بمحاولات لإتمام هذا الأمر:

فريق بجامعة هارفارد، قام بترميز 5.2 مليون بت تقريبًا، ولكن الفريق استخدم طريقة ترميز بسيطة (شفرة بسيطة)، حيث مثلت قواعد الحمض النووي (DNA): الأدينين (A) والسيتوزين (C) أصفارًا (0)، ومثلت قواعد الجوانين (G) والثايمين(T) الرقم واحد (1). وأدّى هذا في بعض الأحيان إلى استطالة الحروف نفسها، بطريقة يصعب قراءة التسلسل الوراثي لها، وأدت في النهاية إلى وقوع بعض الأخطاء. (Merelli, Pérez-Sánchez & Gesing, 2014).

– وفي محاولة أخرى، استطاع العلماء بنجاح تحويل ٧٣٩ كيلوبايت من البيانات الرقمية إلى أكواد ورموز جينية واستعادتها مرة أخرى بدقّة ١٠٠%، حيث بدأ العلماء تحويل مجموعة ملفات بمختلف الصيغ والأنواع: بعض قصائد شكسبير، وتسجيل صوتي، وأوراق بحثية نصية. وكما نعلم فإن تلك البيانات يتم تخزينها على القرص الصلب فى الحاسوب على شكل سلسلة أكواد من رقمين: الصّفر والواحد، لكن قام الباحثون بعمل نظام لترجمة هذا النظام الثنائي إلى نظام آخر ذي أربعة أحرف (AGCT) وهو الكود الذي يتكوّن منه الحمض النّووى DNA وقد استخدموا هذا الكود الوراثى الجينى لتوليف خيوط DNA الفعليّة مع البيانات الجديدة المُدخلة فى هيكلها، لكن في حقيقة الأمر كانت النّتائج بداية غير مثيرة للانبهار، مجرد بضع ذرّات من أشياء بالكاد مرئية وموجودة فى أسفل أنبوب الاختبار، لكنهم حقّقوا النّجاح الباهر عندما تم عكس العمليّة، حيث قام العلماء باحضار الجينات الخاصّة بالمعلومات المُخزّنة وأعادوا ترجمتها إلى نظامها الأصلى الثنائي، وكانت النتيجة هى استعادة المعلومات كاملة دون أدنى خطأ، وكما رأينا، فلم تك المرة الأولى التى يتم فيها تخزين البيانات على شريط DNA، لكن المحاولة الأخيرة كانت أكثر كفاءة ودقة بكثير من سابقتها.

إذًا، ماذا يميز تخزين البيانات فى DNA عن أنظمة التخزين الأخرى؟

– أولا: يتميز الحمض النووى بكثافة تخزين عالية جدًا.

– ثانيًا: يستطيع DNA التواجد فى نطاق من الظروف المختلفة، حيث إنه لا يعانى من هشاشة أو حساسية تجاه أشياء معينة مثل أنظمة التّخزين الموجودة حالياّ.

– ثالثًا: لدى DNA سمعة جيدة فى التخزين الآمن للمعلومات، حيث إنّه يختزن فى داخله تاريخ الحياة منذ نشأتها على الأرض.

– رابعًا: هذا النّظام قد يكون مفيدًا فى مجال الأرشفة أو المعلومات التى لا تتطلّب الدخول المتكرر، بحيث تكون كنسخة احتياطيّة عند الطوارئ أو شئ من هذا القبيل.

– العديد من الاكتشافات التكنولوجية المهمة في عصرنا الحالي، من الذكاء الاصطناعي حتى الأبحاث الطبية، تحتاج إلى قواعد بيانات عملاقة من الآحاد والأصفار، وأي اختراع يجعل هذه المعطيات أسهل في التخزين قد يحسن من الإنتاجية والمردود.

ولكن، ما قد يسبب التراجع لهذه التقنية ما يلي:

– أولًا: التكلفة، حيث إن تغيير وتوليف DNA على المعلومات الجديدة يُعتبر عملية مكلفة، لكن يؤكد العلماء بأنها كباقي التقنيات الجديدة قد بدأت بتكلفة عالية ثم ما لبثت أن تقل تكلفتها مع تطور تكنولوجيا المعلومات.

– ثانيًا: عدم التوصل بعد لنظام لبحث وأرشفة المعلومات الحيوية.

خاتمة البحث:

من هذا المنطلق تبرز أهمية تلك الدراسة، فيحاول البحث في هذا الصدد إشراك المعنيين من خبراء المكتبات والمعلومات في عملية تنظيم وإدارة المعلومات الحيوية؛ لفتح آفاق قد تقلب موازين علم المكتبات والأرشيف والمعلومات، وكما أسلفنا القول، فإن اهتمام علماء المكتبات والمعلومات بهذا الأمر في الأساس هو من أجل البحث عن الإمكانيات الهائلة في إدارة وأرشفة المعلومات الحيوية.

وقد خلُصت الدراسة للنتائج الآتية:

1) علم المكتبات والمعلومات يمكن أن يلعب دورًا مهما في تنظيم المعلوماتية الحيوية لاستخدامها بشكل فعال في المستقبل، حيث يركز علم المكتبات والمعلومات على عملية تفاعل المستفيدين مع تلك المعلومات، كذلك يهتم علم المكتبات والمعلومات بإدارة مصادر المعلوماتية الحيوية، بما تشمله من: خدمات على الإنترنت Webservices، وبرمجيات Software، وقواعد البيانات الحيوية Databases، مثل: (قواعد بيانات الإنتاج الفكري Literature، علم الوراثة والصحة العامة Genetics & Health، الجينوم Genome، الجينات Genes (DNA, RNA)، البروتينات Proteins، قواعد البيانات الكميائية Chemicals، قواعد بيانات البناء Structure).

2) غياب التوحيد القياسي فيما يتعلق بوصف المعلومات الحيوية للتسلسلات الأولية، وقيام العديد من المؤسسات العاملة في مجال المعلومات الحيوية بتوصيف بياناتها بشكل فردي.

3) يواجه التاكسونومي (التصنيف الحيواني والنباتي) أزمة كبيرة، ذلك لأن عددًا قليلًا من المتخصصين في هذا السياق من له خبرًة تحديد التصنيفات الكثيرة بدقة شديدة، وهنا نستكشف إمكانية استخدام المنهجية القائمة على "التصنيف باستخدام الحمض النووي" للتغلب على هذه المشاكل.

4) اعتماد بناء أدوات ضبط المصطلحات على أساليب جديدة غير التي تعودنا عليها في بنائنا لأدوات ضبط المصطلحات في تخصص المكتبات، فالأنطولوجيا الواحدة قد تحتوى على العشرات والمئات من الأنطولوجيات الفرعية، من غير أن يكون هناك علاقة إحداها بأخرى طبقًا للاحتياجات الموضوعية والوظيفية، وقد يفسر ذلك بالتمدد الطبيعي لمصطلحات الأنطولوجيا، الذي يكون من السهل إعادة ترتيب وعلاقات المصطلحات في إحداها دون المساس بالأنطولوجيات الأخرى.

5) الويب الدلالية لها أهمية كبيرة في المعلوماتية الحيوية، حيث تسهم في حل المشكلة الأكبر في مجال المعلوماتية الحيوية، والمرتبطة بعملية استرجاع المعلومات الضخمة المخزنة في قواعد المعلوماتية الحيوية، بدقة وفاعلية، وتقليل اللبس من جراء تشابه تلك المعلومات الحيوية.

وأوصت الدراسة بما هو آت:

1) ضرورة الاهتمام بالتعليم والتدريب في المعلوماتية الحيوية من خلال المقررات التعلمية بالجامعة، والبرامج التدريبية بمؤسسات المكتبات والمعلومات، وعدم وجود خلفية بيولوجية قوية لن يكن عائقًا أمام نجاح الطلاب في فهم واستيعاب المعلوماتية الحيوية.

2) الدراسة الرأسية التفصيلية لعمليات وأدوات إدارة المعلوماتية الحيوية، والتعرف على الملامح البنائية والوظيفية لها، حيث إن هذه الدراسة حاولت على استحياء التعرض الأفقي للموضوعات المشتركة بين تخصص المكتبات والمعلومات والمعلوماتية الحيوية.

3) إصدار معايير موحدة تعمل على التوحيد القياسي لوصف المعلومات الحيوية، والذي من شأنه أن يسهل عملية تبادل البيانات، كما هو الحال في معيار (مارك MARC) في المكتبات.

4) حث الباحثين والدارسين والأكاديميين في مجال المكتبات والمعلومات على تصنيف مؤلفاتهم في الجوانب العديدة التي تتعلق بنظم المعلومات والتخصصات الأخرى، من أجل إثراء العلم بإنتاج فكري متنوع في مجال تكنولوجيا المعلومات.

5) دخول علم المكتبات والمعلومات كشركاء في إنشاء أدوات العمل التي بدا أنها تنحى منحنًى آخر في بنائها، وذلك باعتمادها على المختبرات العلمية، وليس فقط على مبادئ علوم المكتبات والمعلومات.

قائمة المصادر:

أولا المصادر العربية:

حسين، حسين علي. (2010). بنية التصنيف على شبكة الإنترنت [الفوكسونومي - التاكسونومي - التصنيف الوجهي]. (ط.1). القاهرة: ايبيس كوم.

خليل، سميرة خليل محمد. (2019). التاكسونومي ودوره في تنظيم المعلومات في البيئة الرقمية العربية. المجلة العلمية للمكتبات والوثائق والمعلومات. 1(1)، 195-199.

عبد الهادي، محمد فتحي. (يوليو 2010). الانطولوجيات ودورها في تنظيم المعلومات في بيئة الويب 2.0. في الملتقى الملتقى العربي الرابع لتكنولوجيا المكتبات والمعلومات بعنوان: الجيل الثاني للويب وتأثيره على قطاع المكتبات والمعلومات، القاهرة.

المتيني، أحمد يوسف & رياض، سناء أحمد & عبد العزيز، آمال أحمد. (2006). الجينوميكس والمعلوماتية الحيوية. الإسكندرية: مكتبة بستان المعرفة.

النشرتي، مؤمن سيد. (2018). دور العلاقات الببليوغرافية في التكامل المعرفي. (رسالة دكتوراة غير منشورة). جامعة القاهرة، الجيزة. ص. 9.

ثانيا المصادر الأجنبية:

Armstrong, C. (2010). Aliases and Ambiguity: A case study of gene aliases, and implications for information curation and AI. Ideals.illinois.edu. Retrieved 12 January 2018, from https://www.ideals.illinois.edu/handle/2142/15063

Baker, P. G., Goble, C. A., Bechhofer, S., Paton, N. W., Stevens, R., & Brass, A. (1999). An ontology for bioinformatics applications. Bioinformatics, 15(6), 510–520. doi:10.1093/bioinformatics/15.6.510.

Bodenreider, O. (2004). The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Research, 32(90001). doi: 10.1093/nar/gkh061.

Cannata, N., Schröder, M., Marangoni, R., & Romano, P. (2008). A Semantic Web for bioinformatics: goals, tools, systems, applications. BMC Bioinformatics, 9(Suppl 4), S1. doi:10.1186/1471-2105-9-s4-s1.

Caporaso JG, Baumgartner WA Jr, Cohen KB, Johnson HL, Paquette J, et al.. (2005). Concept recognition and the TREC Genomics tasks. In: The Fourteenth Text REtrieval Conference (TREC 2005) Proceedings.

Carpendale, S., Chen, M., Evanko, D., Gehlenborg, N., Gorg, C., Hunter, L., … Strobelt, H. (2014).

Ontologies in Biological Data Visualization. IEEE Computer Graphics and Applications, 34(2), 8–15. doi: 10.1109/mcg.2014.33.

Charles F. Priore, Jr. and John L. Giannini (March/April 2007). Integrating Bibliographic Software, Database Searching, and Molecular Modeling in an Introductory Biology Course: A Collaboration Between the Biology Department and the Science Librarian. Journal of College Science Teaching. 36(5), 21-23.

Cloud Life Sciences | Google Cloud. (2020). Retrieved 29 January 2020, from https://cloud.google.com/life-sciences/

Cohen KB, Hunter LE (2013). Text Mining for Translational Bioinformatics. PLoS Comput Biol 9(4): e1003044. doi.org/10.1371/journal.pcbi.1003044.

Consortium, G. (2020). AmiGO 2: Drill-down Browser. Retrieved 14 January 2020, from http://amigo.geneontology.org/

Cudré-Mauroux, P., Gaugaz, J., Budura, A., & Aberer, K. (2005). Analyzing semantic interoperability in bioinformatic database networks. W8: Semantic Network Analysis, 82.

Dalrymple, P.W. & Roderer, N.K. (2007). Library and information science and biomedical informatics: converging disciplines. Information Research, 12(4).

Damianos, L., Zarrella, G., & Hirschman, D. (2013). The MiTAP System for Monitoring Reports of Disease Outbreak. Retrieved 14 February 2020, from https://www.mitre.org/publications/technical-papers/the-mitap-system-for-monitoring-reports-of-disease-outbreak

Demaine, J., Martin, J., Wei, L., & Bruijn, B. D. (2006). LitMiner: integration of library services within a bio-informatics application. Biomedical Digital Libraries, 3(1). doi: 10.1186/1742-5581-3-11.

Dessimoz, C., & Škunca Nives. (2017). The gene ontology handbook. New York: Springer.

Dolan, M., Holden, C., Beard, M., & Bult, C. (2006). Genomes as geography: using GIS technology to build interactive genome feature maps. BMC Bioinformatics, 7(1). doi:10.1186/1471-2105-7-416.

Geer, R.C. (2006). Broad issues to consider for library involvement in bioinformatics. *Journal of the Medical Library Association : JMLA, 94 3*, 286-98.

Gene Ontology Resource. (2020). Retrieved 14 January 2020, from

http://geneontology.org/stats.html

Genomic Standards Consortium. (n.d.). Retrieved February 27, 2020, from https://press3.mcs.anl.gov/gensc

Google Cloud (Mar, 2018). Google Genomics Documentation. Google Cloud. (2019). Retrieved 13 December 2019, from: https://cloud.google.com/genomics

Grimes, S. M., & Ji, H. P. (2014). MendeLIMS: a web-based laboratory information management system for clinical genome sequencing. BMC Bioinformatics, 15(1). doi: 10.1186/1471-2105-15-290.

Human coronavirus OC43 strain ATCC VR-759, complete genome - Nucleotide - NCBI. (n.d.). Retrieved February 27, 2020, from https://www.ncbi.nlm.nih.gov/nuccore/1578871709

Hunter, L., Lu, Z., Firby, J., Baumgartner, W. A., Johnson, H. L., Ogren, P. V., & Cohen, K. B. (2008). OpenDMAP: An open source, ontology-driven concept analysis engine, with applications to capturing knowledge regarding protein transport, protein interactions and cell-type-specific gene expression. BMC Bioinformatics, 9(1). doi: 10.1186/1471-2105-9-78.

Integrated Taxonomic Information System. (n.d.). Retrieved February 27, 2020, from https://www.itis.gov/

Kalra, H. P. (2005). Bioinformatics and the Library and Information Science Community. IFLA Journal, 31(4), 333–341. doi:10.1177/0340035205061398.

Khayat, M. H. (Trans.). (2006). Icnp version 1: international classification for nursing practice [version 1]. Retrieved from http://applications.emro.who.int/dsaf/dsa740.pdf

Knaup, P., & Haux, R. (2010). Recommendations of the International Medical Informatics Association (IMIA) on Education in Health and Medical Informatics. Methods of Information in Medicine, 49(03), 105–120. doi: 10.1055/s-0038-1634340.

Leo, S., Santoni, F., & Zanetti, G. (2009). Biodoop: Bioinformatics on Hadoop. In *International Conference on Parallel Processing Workshops*, 415-422.

Li, M., Chen, Y.-B., & Clintworth, W. A. (2013). Expanding roles in a library-based bioinformatics service program: a case study. Journal of the Medical Library Association: JMLA, 101(4), 303–309. doi: 10.3163/1536-5050.101.4.012.

Lord, P., Bechhofer, S., Wilkinson, M. D., Schiltz, G., Gessler, D., Hull, D., … Stein, L. (2004). Applying Semantic Web Services to Bioinformatics: Experiences Gained, Lessons Learnt. The

Semantic Web – ISWC 2004 Lecture Notes in Computer Science, 350–364. doi: 10.1007/978-3-540-30475-3_25.

Mcknight, M. (2005). Librarians, Informaticists, Informationists, and Other Information Professionals in Biomedicine and the Health Sciences. Journal of Hospital Librarianship, 5(1), 13–29. doi: 10.1300/j186v05n01_02.

Michon, J. (2007). Biomedicine and the Semantic Web: A Knowledge Model for Visual Phenotype. Cataloging & Classification Quarterly, 43(3-4), 149–160. doi: 10.1300/j104v43n03_08

NCBI Taxonomy Homepage. (n.d.). Ncbi.nlm.nih.gov. Retrieved 7 April 2019, from: https://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html

Odebode, I. E. (2011). The application of information retrieval techniques to the mining of bioinformatics data (Unpublished Doctoral dissertation), Morgan State University, Maryland.

Peng, S., Mamitsuka, H., & Zhu, S. (2018). MeSHLabeler and DeepMeSH: Recent Progress in Large-Scale MeSH Indexing. Methods in Molecular Biology Data Mining for Systems Biology, 203–209. doi: 10.1007/978-1-4939-8561-6_15.

Post, L. J. G., Roos, M., Marshall, M. S., Driel, R. V., & Breit, T. M. (2007). A semantic web approach applied to integrative bioinformatics experimentation: a biological use case with genomics data. Bioinformatics, 23(22), 3080–3087. doi: 10.1093/bioinformatics/btm461.

Ram, S., & Rao, N. L. (2014). Metadata Description Framework for Integration of Bioinformatics Information Resources: A Case of iBIRA. DESIDOC Journal of Library & Information Technology, 34(5), 384–392. doi:10.14429/djlit.34.6254

Rocco, D., & Critchlow, T. (2003). Automatic discovery and classification of bioinformatics Web sources. Bioinformatics, 19(15), 1927–1933. doi: 10.1093/bioinformatics/btg353.

Sheth, A., York, W., Thomas, C., Nagarajan, M., Miller, J. A., Kochut, K.,... & Yi, X. (2004, October). Semantic Web technology in support of Bioinformatics for Glycan Expression. In W3C Workshop on Semantic Web for Life Sciences. 27-28.

The UMLS Semantic Network. (n.d.). Retrieved February 27, 2020, from: https://semanticnetwork.nlm.nih.gov

Wroe, C., Stevens, R., Goble, C., Roberts, A., & Greenwood, M. (2003). A Suite Of Daml Oil Ontologies To Describe Bioinformatics Web Services And Data. International Journal of

Cooperative Information Systems, 12(02), 197–224. doi: 10.1142/s0218843003000711.

Wu, S., Stvilia, B., & Lee, D. J. (2012). Authority Control for Scientific Data: The Case of Molecular Biology. Journal of Library Metadata, 12(2-3), 61–82.

Yu, H., Lee, M., Kaufman, D., Ely, J., Osheroff, J. A., Hripcsak, G., & Cimino, J. (2007). Development, implementation, and a cognitive evaluation of a definitional question answering system for physicians. Journal of Biomedical Informatics, 40(3), 236–251. doi: 10.1016/j.jbi.2007.03.002.

Zweigenbaum P. (2003). Question answering in biomedicine. In: Proceedings of the workshop on natural language processing for question answering. 1-4.

●●●

Bioinformatics Management Applied
In Library and Information Science

Ibraheem Sabry El-Metwally
Ph.D. Student
Libraries and Information Depart.
Faculty of Arts - Cairo University
is_80@hotmail.com

Bioinformatics in its simplest definition, the science of the management of molecular data (DNA, RNA, Proteins), as well as data that results from their interactions. Using the survey method, the study discusses tools and processes that are performed on Bioinformatics in library institutions. The study monitors the steps of library and information science toward its interrelationship with computer science in the management of bioinformatics, how similar and intertwined they are in the creation, use, and development of organizational tools and processes related to control and standardize terms, metadata, indexing, taxonomy and classification of bioinformatics, semantic relationships, data mining, digital libraries, and management of bioinformatics resource. The study also discusses the current and potential roles that libraries, information, and computer science may play toward developing tools that store, organize, and retrieve data from bioinformatics Databases.

key words: **Bioinformatics; Information Management; Library and Information Sceince; Taxonomy in Bioinformatics; Bio-Ontology; Big Data Mining.**