جمهورية العراق
وزارة التعليم العالي والبحث العلمي
جامعة الموصل/ كلية الآداب
مجلة آداب الرافدين

مَجَلَّةُ

آدَابِ الرَّافِدَيْنِ

مجلة فصليَّة علميَّة مُحكَّمة

تصدر عن كلية الآداب – جامعة الموصل

ملحق

العدد الواحد والتسعين/ السنة الثانية والخمسون

جمادى الثانية – ١٤٤٤هـ / كانون الأوَّل ٢٠٢٢/١٢/٢٩م

رقم إيداع المجلة في المكتبة الوطنية ببغداد : ١٤ لسنة ١٩٩٢

ISSN 0378- 2867
E ISSN 2664-2506

للتواصل:

radab.mosuljournals@gmail.com

URL: https://radab.mosuljournals.com

آداب الرافدين

مجلة محكّمة تعنى بنشر البحوث العلميَّة الموثّقة في الآداب والعلوم الإنسانيَّة

باللغة العربيَّة واللغات الأجنبيَّة

ملحق العدد الواحد والتسعين السنة: الثانية والخمسون/ جمادى الثانية – ١٤٤٤هـ / كانون الأوَّل ٢٠٢٢م

رئيس التحرير: الأُستاذ الدكتور عمار عبداللطيف زين العابدين (المعلومات والمكتبات) كلية الآداب/ جامعة الموصل/ العراق

مدير التحرير: الأُستاذ المساعد الدكتور شيبان أديب رمضان الشيبانيّ (اللغة العربية) كلية الآداب/ جامعة الموصل/ العراق

أعضاء هيئة التحرير :

الأُستاذ الدكتور حارث حازم أيوب (علم الاجتماع) كلية الآداب/جامعة الموصل/العراق

الأُستاذ الدكتورة وفاء عبداللطيف عبد العالي (اللغة الإنكليزية) كلية الآداب/ جامعة الموصل/ العراق

الأُستاذ الدكتور مقداد خليل قاسم الخاتوني (اللغة العربية) كلية الآداب/ جامعة الموصل/ العراق

الأُستاذ الدكتور علاء الدين أحمد الغرايبة (اللغة العربية) كلية الآداب/جامعة الزيتونة/الأُردن

الأُستاذ الدكتور قيس حاتم هاني (التاريخ) كلية التربية/جامعة بابل/العراق

الأُستاذ الدكتور مصطفى علي الدويدار (التاريخ) كلية العلوم والآداب/جامعة طيبة/ السعودية

الأُستاذ الدكتورة سوزان يوسف أحمد (الإعلام) كلية الآداب/جامعة عين شمس/مصر

الأُستاذ الدكتورة عائشة كول جلب أوغلو (اللغة التركية وآدابها) كلية التربية/جامعة حاجت تبه/ تركيا

الأُستاذ الدكتورة غادة عبدالمنعم محمد موسى (المعلومات والمكتبات) كلية الآداب/جامعة الإسكندريّة

الأُستاذ الدكتور كلود فينثز (اللغة الفرنسية وآدابها) جامعة كرنوبل آلب/فرنسا

الأُستاذ المساعد الدكتور أرثر جيمز روز (الأدب الإنكليزي) جامعة درهام/ المملكة المتحدة

الأُستاذ المساعد الدكتور سامي محمود إبراهيم (الفلسفة) كلية الآداب/ جامعة الموصل/ العراق

سكرتارية التحرير:

التقويم اللغوي: م.د.خالد حازم عيدان – مقوّم لغوي/ اللغة العربيَّة

م.م.عمَّار أحمـــــد محمود – مقوّم لغوي/ اللغة الإنكليزيَّة

المتابعـــــــة:

مترجم.إيمان جرجيس أمين – إدارة المتابعـــــة

مترجم.نجلاء أحمد حسين – إدارة المتابعـــــة

قواعد تعليمات النشر

١- على الباحث الراغب بالنشر التسجيل في منصة المجلة على الرابط الآتي:
https://radab.mosuljournals.com/contacts?_action=signup .

٢- بعد التسجيل ستُرسل المنصة إلى بريد الباحث الذي سجل فيه رسالة مفادها أنَّه سجَّل فيها، وسيجد كلمة المرور الخاصة به ليستعملها في الدخول إلى المجلة بكتابة البريد الإلكتروني الذي استعمله مع كلمة المرور التي وصلت إليه على الرابط الآتي:
https://radab.mosuljournals.com/contacts?_action=login .

٣- ستمنح المنصة (الموقع) صفة الباحث لمن قام بالتسجيل؛ ليستطيع بهذه الصفة إدخال بحثه بمجموعة من الخطوات تبدأ بملء بيانات تتعلَّق به وببحثه ويمكنه الاطِّلاع عليها عند تحميل بحثه .

٤- يجب صياغة البحث على وفق تعليمات الطباعة للنشر في المجلة، وعلى النحو الآتي :

- تكون الطباعة القياسية على وفق المنظومة الآتية: (العنوان: بحرف ١٦/ المتن: بحرف ١٤/ الهوامش: بحرف ١١)، ويكون عدد السطور في الصفحة الواحدة: (٢٧) سطرًا، وحين تزيد عدد الصفحات في الطبعة الأخيرة عند النشر داخل المجلة على (٢٥) صفحة للبحوث الخالية من المصورات والخرائط والجداول وأعمال الترجمة، وتحقيق النصوص، و (٣٠) صفحة للبحوث المتضمنة للأَشياء المشار إليها يدفع الباحث أجور الصفحات الزائدة فوق حدّ ما ذُكر آنفًا .
- تُرتَّب الهوامش أَرقامًا لكل صفحة، ويُعرَّف بالمصدر والمرجع في مسرد الهوامش لدى وورد ذكره أول مرة، ويلغى ثبت (المصادر والمراجع) اكتفاءً بالتعريف في موضع الذكر الأول ، في حالة تكرار اقتباس المصدر يذكر (مصدر سابق).
- يُحال البحث إلى خبيرين يرشِّحانه للنشر بعد تدقيق رصانته العلمية، وتأكيد سلامته من النقل غير المشروع، ويُحال – إن اختلف الخبيران – إلى (مُحكِّم) للفحص الأَخير، وترجيح جهة القبول أَو الرفض، فضلًا عن إحالة البحث إلى خبير الاستلال العلمي ليحدد نسبة الاستلال من المصادر الإلكترونيَّة ويُقبل البحث إذا لم تتجاوز نسبة استلاله ٢٠% .

٥- يجب أن يلتزم الباحث (المؤلِّف) بتوفير المعلومات الآتية عن البحث، وهي :

- يجب أن لا يضمّ البحث المرسل للتقييم إلى المجلة اسم الباحث، أي: يرسل بدون اسم .
- يجب تثبيت عنوان واضح وكامل للباحث (القسم/ الكلية او المعهد/ الجامعة) والبحث باللغتين: العربية والإنكليزية على متن البحث مهما كانت لغة البحث المكتوب بها مع إعطاء عنوان مختصر للبحث باللغتين أيضًا: العربية والإنكليزية يضمّ أَبرز ما في العنوان من مرتكزات علمية .
- يجب على الباحث صياغة مستخلصين علميين للبحث باللغتين: العربية والإنكليزية، لا يقلَّان عن (١٥٠) كلمة ولا يزيدان عن (350)، وتثبيت كلمات مفتاحية باللغتين: العربية والإنكليزية لاتقل عن (٣) كلمات، ولا تزيد عن (٥) يغلب عليهنَّ التمايز في البحث.

٦- يجب على الباحث أن يراعي الشروط العلمية الآتية في كتابة بحثه، فهي الأساس في التقييم، وبخلاف ذلك سيُردّ بحثه ؛ لإكمال الفوات، أمَّا الشروط العلميَّة فكما هو مبيّن على النحو الآتي :

- يجب أن يكون هناك تحديد واضح لمشكلة البحث في فقرة خاصة عنوانها: (مشكلة البحث) أو (إشكاليَّة البحث) .
- يجب أن يراعي الباحث صياغة أسئلة بحثيَّة أو فرضيَّات تعبِّر عن مشكلة البحث ويعمل على تحقيقها وحلِّها أو دحضها علميًّا في متن البحث .
- يعمل الباحث على تحديد أهمية بحثه وأهدافه التي يسعى إلى تحقيقها، وأن يحدِّد الغرض من تطبيقها.
- يجب أن يكون هناك تحديد واضح لحدود البحث ومجتمعه الذي يعمل على دراسته الباحث في بحثه .
- يجب أن يراعي الباحث اختيار المنهج الصحيح الذي يتناسب مع موضوع بحثه، كما يجب أن يراعي أدوات جمع البيانات التي تتناسب مع بحثه ومع المنهج المتَّبع فيه .
- يجب مراعاة تصميم البحث وأُسلوب إخراجه النهائي والتسلسل المنطقي لأفكاره وفقراته.
- يجب على الباحث أن يراعي اختيار مصادر المعلومات التي يعتمد عليها البحث، واختيار ما يتناسب مع بحثه مراعيًا الحداثة فيها، والدقة في تسجيل الاقتباسات والبيانات الببليوغرافية الخاصة بهذه المصادر.
- يجب على الباحث أن يراعي تدوين النتائج التي توصل إليها ، والتأكُّد من موضوعاتها ونسبة ترابطها مع الأسئلة البحثية أو الفرضيات التي وضعها الباحث له في متن بحثه .

٧- يجب على الباحث أن يدرك أنَّ الحُكْمَ على البحث سيكون على وفق استمارة تحكيم تضمّ التفاصيل الواردة آنفًا، ثم تُرسل إلى المُحكِّم وعلى أساسها يُحكَّم البحث ويُعطى أوزانًا لفقراته وعلى وفق ما تقرره تلك الأوزان يُقبل البحث أو يرفض، فيجب على الباحث مراعاة ذلك في إعداد بحثه والعناية به .

تنويه:

تعبِّر جميع الأفكار والآراء الواردة في متون البحوث المنشورة في مجلتنا عن آراء أصحابها بشكل مباشر وتوجهاتهم الفكريَّة ولا تعبِّر بالضرورة عن آراء هيأة التحرير فاقتضى التنويه

رئيس هيئة التحرير

المحتـــويـــات

العنـوان	الصفحة
بحوث اللغة العربية	
صور إضافة الظرف (مع) إلى ضمير المُخاطب ودلالاتها في القرآن الكريم أحمد عبدالستار فاضل و فراس عبدالعزيز عبدالقادر	30-1
الآخر محاربًا في شعر ابن الدَّهان الموصليّ (ت581هـ) عجيل مد الله أَحمد و مقداد خليل قاسم	80 -31
الطَّيفُ في شعرِ ابنِ نُباتةَ المَصريّ فارس ياسين محمد الحمداني	102 -81
اعتراضات نُقْرَه كار (ت776هـ) الصرفيَّة في شرح شافية ابن الحاجب (646هـ) هلال علي محمود	132 -103
الشخصية الإشكالية ومستويات وعيها في عالم (متاهات) برهان شاوي الروائي نورا وريا عزالدين و شادان جميل عباس	172 -133
الزمن السّردي في قصص جابر خليفة جابر يونس جاسم محمد سالم و بسام خلف سليمان	198 -173
الصورة المشهديَّة: الثابتة والمتحرِّكة في شعر حسب الشيخ جعفر ملكة عصام ياسين	219 -199
التوكيد بوصفه عارضًا نحويًّا في الحديث النبوي الشريف حديث: "إِنَّما الأَعمالُ بالنيَّات" أُنموذجًا مصعب إسماعيل عمر و ثامر عبدالجبّار نصيّف	244 -220
أَنماطُ الحالِ ودلالاتُها في معلَّقةِ طرفة بن العبد إسماعيل حميد حمد أَمين و مظفر الدين عثمان حمد صالح	265 -245
دلالة الأَفعال المقيَّدة بحرف الجر في قصيدة مديح الظل العالي لمحمود درويش إسراء غانم أَحمد	301 -266
الارتداد الزمني في رواية ظلال الوأد لـ(منيرة السبيعي) سروى صباح رجب	329 -302
تنوع الدلالات في نماذج مختارة من شعر الهذليين أَحمد يعقوب الجبوري	352 -330
ميميَّة حسَّان بن ثابت ألم تسأل الدار "دراسة تحليليَّة نقديَّة" وضَّاح حسن خضر حسن	372 -353
الصورة بوصفها إدانة للواقع في رواية (العراق سينما) لأَحمد إبراهيم السعد ليث طالب ذنون	386 -373
المعرب على أَكثر من ثلاثة أَوجه من المصدر المعرفة المنصوب المحذوف فعله في كتاب الدر المصون للسمين الحلبي شذى محمد مصطفى رشيد	405 -387
بحوث التاريخ والحضارة الإسلاميَّة	
العلاقات الاقتصاديَّة التركيَّة الليبيَّة 1989-2011م صبا طلال عمر طلال و محمد علي محمد عفين	420 -406
مُعلِّمو السلاطين العثمانيين الشيخ آق شمس الدين أُنموذجًا (1389-1459م) دراسة تاريخيَّة أَمين غانم محمد و عماد عبدالعزيز يوسف	455 -421
انتفاضة علي باشا جان بولاد في ولاية حلب 1605-1607م أَحمد محمد نوري أَحمد العالم	489 -456
كتابُ "تاريخ مدينة السلام" ومؤلِّفه الخطيب البغدادي حميدي خضير جمعة	508 -490

بحوث علم الاجتماع	
منظمات المجتمع المدني ودورها في الاستقرار السياسي دراسة تحليليَّة في علم الاجتماع السياسي إيمان حماديّ رجب و حسن راشد جاسم	509 - 535
عزوف التلاميذ والطلبة عن التسجيل للدراسة التركمانيَّة في محافظة نينوى (الأَسباب- النتائج- الحلول) عدنان حازم عبد أَحمد	536- 570
بحوث الفلسفة	
فلسفة أَفلاطون على نظريَّة الفارابي السياسيَّة (دراسة تحليليَّة مقارنة) ليلى يونس صالح	571- 601
بحوث الشريعة والتربية الإسلاميَّة	
خصائص النبي (ﷺ) في الآيتين (157_158) من سورة الأَعراف -دراسة تحليليَّة تفسيريَّة- نغم قاسم أحمد الأَرمني و رائد سالم شريف	602- 621
بحوث المعلومات وتقنيات المعرفة	
المكتبة العامة المركزية في الموصل: دراسة في واقعها ومقترحات تطويرها وسن سامي سعدالله	622- 649
بناء نظام خبير لتصنيف الرسائل والأَطاريح الجامعيَّة باستخدام خوارزميَّة (Naïve Bayes): دراسة تجريبيَّة إيناس جاسم هادي	650- 689
بحوث علم النفس وطرائق التدريس	
اتجاه طلبة الجامعة نحو التعليم الإلكتروني وعلاقته بالتنظيم الذاتي الأَكاديمي عدي نعمت بطرس عجاج	690- 737
صعوبات تدريس مادة الفيزياء في المرحلة الثانوية من وجهة نظر مدرسي الفيزياء في مدينة الموصل طارق موفق سحري	738- 776
بحوث القانون	
أثر حالة الضرورة لارتكاب الجريمة في المسؤولية في الشريعة الإسلاميَّة والقانون العراقي شيروان عمر رسول و عادل عبدالله حمد	777-817

CC BY

بناء نظام خبير لتصنيف الرسائل والأَطاريح الجامعيَّة باستخدام خوارزميَّة (Naïve Bayes): دراسة تجريبيَّة

إيناس جاسم هادي *

تأريخ التقديم: 2021/11/26 تأريخ القبول: 2022/1/18

المستخلص:

صار تصنيف البيانات من الحقول المعرفية واسعة الانتشار، ولاسيَّما في ظل التطور الهائل لتكنولوجيا المعلومات والنظم الخبيرة، ومن بين أَبرز حقول المعرفة التي تعنى بتصنيف البيانات هو تصنيف البيانات النصية، في بحثنا هذا نقدم نموذجاً لنظام خبير مهمته تصنيف الرسائل الجامعية اعتمادًا على عناوينها العربية وباستخدام خوارزمية (Naïve Bayes). في البحث تم استخدام قاعدة بيانات تم تجميعها باستخدام تقنية (Scrapping) وتتمثل في جمع البيانات من المستودعات المنتشرة على شبكة الإنترنت؛ إذ جرى بناء قاعدة بيانات قوامها (6100) عنوان لرسالة ماجستير وأطروحة دكتوراه، توزعت هذه العناوين على (سبعة) فئات تمثل الاختصاص العام للرسالة – الأطروحة. من بين أَبرز التحديات التي واجهت البحث هو طبيعة اللغة العربية ودرجة تعقيدها. أَثبتت نتائج البحث أَن الأُنموذج كان قادرًا على تصنيف الرسائل بنسبة (86%) وهي نسبة تعد واعدة بشكل كبير. كما جرى أَيضًا تقييم للنموذج المقترح عن طريق توزيع استبانة لعدد من المختصين في علم الحاسبات ونظم المعلومات، اشتمل الاستبانة على قياس أثر متغيرين مستقلين هما دقة النظام الخبير واهميته. اثبتت النتائج ان هنالك تأثيراً ذو دلالة إحصائيَّة للمتغيرين المستقلين على المتغير المعتمد (الفائدة المتوخاة من النظام الخبير). ان نتائج البحث تفتح افاقاً جديدة في عملية تصنيف البيانات النصية وبالأخص اللغة العربية، وتفتح المجال صوب تنفيذ خوارزميات أُخرى ومقارنة النتائج.

* مدرس/ كلية الآداب/الجامعة المستنصريَّة.

الكلمات المفتاحية: تصنيف النصوص العربية، النظم الخبيرة، خوارزمية Naïve Bayes.

أولا: الإطار العام للدراسة

1-1 المقدمة

لقد منّ الله تعالى على البشرية جمعاء من خلال منحها العقل البشري، الذي يميز الانسان عن سائر المخلوقات من خلال القابلية على التعلم والتحليل والتمييز. يتميز الذكاء البشري في عمليات لا حصر لها يتم تنفيذها داخل دماغ الانسان ويمكن ان تكون هذه الأنشطة حركية أَو فكرية أَو غيرها. هنالك عمليات حسابية ومنطقية هائلة تحدث داخل العقل البشري أصبحت وما تزال الشغل الشاغل للعلماء والباحثين. إِذ من الممكن ان يتم تعريف الذكاء البشري بانه القدرة على مواجهة المصاعب والمشكلات المختلفة والتكيف مع الظروف المتغيرة والطارئة.

ان مبدأ الذكاء الطبيعي يعد من أَبرز الأمور التي الهمت العلماء في عصر الثورة التقنية وتطور الحاسوب إِلى بناء نماذج محوسبة تحاكي ذكاء الانسان إِلى حد ما. وفي هذا الصدد نرى نشأة الذكاء الاصطناعي اتسمت بالتطور السريع والمتزايد وخصوصا في العقدين الأخيرين. يتلخص علم الذكاء الاصطناعي في بناء نماذج محوسبة لها القدرة على تسهيل الأمور الحياتية من خلال القدرة على الفهم والادراك والتنبؤ وحل مختلف المشكلات.

وتعد النظم الخبيرة واحدة من أَبرز المجالات والتطبيقات التي ينتجها علم الذكاء الاصطناعي. تم الاستعانة بالنظم الخبيرة في إِجراء العديد من البحوث والدراسات والتجارب العلمية التي تخدم مختلف التخصصات العلمية وجوانب الحياة نظرا لأهميتها والدور البارز الذي تلعبه هذه الأنظمة في مختلف المجالات.

1-1 مشكلة الدراسة

يُسهم التقدم التكنولوجي في اثراء مختلف جوانب الحياة بالآليات والتقنيات التي تساهم إِلى حد كبير في تطور الجوانب العلمية والمعرفية. وتصنيف البيانات غير بعيد عن التقدم الحاصل في التكنولوجيا، إِذ برزت العديد من الطرائق والخوارزميات منذ

وقت ليس بالقليل تعالج موضوع تصنيف البيانات والمعلومات. الأنظمة الخبيرة تعد من أبرز الجوانب التي تستخدم في تصنيف البيانات والمعلومات. يتبنى البحث الحالي تصنيف الرسائل الجامعية اعتماداً على العنوان باللغة العربية، وفي هذا السياق ينبثق تحديين، الأول يتعلق بطول السلسلة النصية التي يتم معالجتها، إذ يعد عنوان الرسالة أو الاطروحة قصيراً نسبياً مما يقدم كمية قليلة من الخصائص التي تستخدم في عملية التصنيف، الأمر الثاني هو العمل مع اللغة العربية، إذ تعد من اكثر اللغات تعقيداً فيما يخص المعالجة الرقمية وذلك نظراً للطبيعة الشكلية للنصوص العربية إذ يعتمد كتابة الكلمة ورسمها على موضع الحرف، فضلًا عن العلامات الاعرابية الَّتي تتحكم بلفظ الكلمة ومعناها. ومن خلال البحث في المستوعبات، يمكن ملاحظة عن هنالك قلة في عدد الدراسات التي تعالج البيانات النصية باللغة العربية مقارنة باللغات الأُخرى كالإنجليزية مثلاً. ويمكن ان تتلخص مشكلة البحث في الإجابة على الأَسئلة التالية، هل بالإمكان بناء نظام خبير لتصنيف الرسائل الجامعية باستخدام العنوان؟، هل يمكن معالجة النصوص العربية بطريقة تمكن من الحصول على أَداء جيد فيما يخص التصنيف.

1-2 اهداف الدراسة

ان الهدف الأساسي من البحث هو بناء أنموذج لنظام خبير لتصنيف الرسائل الجامعية باستخدام العنوان وبالاعتماد على خوارزمية (Naïve Bayes). ويتفرع من الهدف الرئيسي الأهداف الفرعية التالية:

- بناء قاعدة بيانات لعناوين الرسائل الجامعية.
- استخدام طرائق جيدة للمعالجة المبدئية للنصوص العربية.
- فحص وتشغيل النظام الخبير.
- تقييم أَداء النظام الخبير.

1-3 فرضيات البحث

يقدم البحث فرضياته فيما يتعلق بتقييم النظام الخبير المقترح، ويتم وفق ذلك قياس تأثير متغيرين مستقلين (أَهميَّة النظام الخبير، دقة النظام الخبير) على متغير معتمد

وهو (الفائدة المتوخاة من النظام الخبير)، وعلى هذا الأساس يقدم البحث الفرضيات التالية:

1- وجود تأثير معنوي للمتغير المستقل (أَهميَّة النظام الخبير) على المتغير المعتمد (الفائدة المتوخاة من النظام الخبير).

2- وجود تأثير معنوي للمتغير المستقل (دقة النظام الخبير) على المتغير المعتمد (الفائدة المتوخاة من النظام الخبير).

1-4 أَهميَّة الدراسة

يعد تصنيف النصوص واحد من أَبرز الجوانب البحثية التي تتطور بشكل سريع في السنوات الأخيرة، لما تحمله النصوص من أَهميَّة واستخدامات يومية تفوق التوقعات. ولاسيَّما بعد انتشار وسائل التواصل الاجتماعي وأصبح العالم منفتح بشكل غير مسبوق في تبادل الآراء والمقالات، لذا ظهرت الحاجة إِلى تصنيف النصوص وتحليلها. واللغة العربية غير بعيدة عن هذا التطور، إِذ تعد أحد ست لغات عالمية معتمدة من قبل الأمم المتحدة، فضلًا عن ان هنالك 25 دولة تعتمدها لغة رسمية وبعدد من المتكلمين يفوق ثلاثمئة مليون فرد. لذلك من المهم جدا ان يتم معالجة النصوص العربية وتطبيق السياقات البرمجية الذكية بغية انجاز أنظمة وطرائق من شانها الارتقاء باللغة واستخداماتها. إِلى جانب ذلك تساهم الأنظمة الخبيرة إِلى حد كبير في اثراء مختلف الجوانب المعرفية بآليات رصينة من انها ان تزيد من الدقة والإِنتاجية في مختلف التخصصات. وان استخدام النظم الخبيرة في تصنيف الرسائل الجامعية يشكل نقطة تحول في علم المعلومات وبداية لإنجاز العديد من الطرائق والمهام التي تخدم علم المعلومات ومختلف فروعه.

1-5 منهجية الدراسة

تم اعتماد المنهج التجريبي في عملية بناء النظام الخبير، إِذ تم برمجة خوارزمية Naïve Bayes وإدخال البيانات النصية لها. وبغية تقييم النظام، تم اعتماد المنهج الوصفي المسحي من خلال إِجراء استبانة تم توزيعه لعدد من مختصي المعلومات والمكتبات فضلًا عن تكنولوجيا المعلومات.

1-6 حدود الدراسة:

الحدود الموضوعية: الأنظمة الخبيرة، تصنيف النصوص العربية، عناوين الرسائل الجامعية.

1-7 الدراسات السابقة:

بغية إحاطة موضوع الدراسة من مختلف الجوانب، تم اخذ الدراسات السابقة من منظورين، الأوَّل يعرض دور الأنظمة الخبيرة في اثراء مختلف العلوم المعرفية وكيفية تطبيقه وما هي النتائج التي تم تحصيلها نتيجة تبني الأنظمة الخبيرة، أمَّا المنظور الثاني فيعرض عدداً من الدراسات السابقة التي تناولت موضوع تصنيف النصوص العربية باستخدام الخوارزميات الذكية.

أولا: دراسات تناولت دور تطبيق الأنظمة الخبيرة في مختلف التخصصات:

1- دراسة (رشيد، 2013) :

قدَّمت الدراسة نموذجا ذكياً لنظام خبير معتمد على المنطق المضبب في المجال الطبي وذلك لتعيين القيم التقديرية لمرضى الثلاسيميا، تضمنت الدراسة دمج وتهجين المنطق المضبب (Fuzzy Logic) مع الشبكات العصبية (Neural Networks)، تم تطبيق النموذج على عدد من البيانات مع ملاحظة دقة الأَداء له، اثبتت الدراسة أَنَّ النموذج المهجن يقدم أَداء جيداً مع دقة عالية في تقدير المعلمات.

2- دراسة (مهدي، 2014):

تضمنت الدراسة تصميم وتنفيذ نظام خبير مضبب لتشخيص امراض الغدد الصماء باستخدام خوارزمية الطيور. هدفت الدراسة إلى الحصول على الحل الأمثل عن طريق مطابقة النتائج التي يتم الحصول عليها من خلال النظام مع نتائج واقعية عبر محاكاة آلية عمل الخوارزمية. قدمت الدراسة عددا من النتائج الجيدة والمتعلقة في إمكانيَّة تحديد الإصابة من عدمها. مما يثبت فاعلية النظم الخبيرة في تقديم أَداء متميز من شأنه رفد التخصصات الطبية بالعديد من الآليات والإجراءات الفاعلة.

3- دراسة (الحسين، 2015)

تم في هذه الدراسة تصميم نظام خبير هجين يقوم بتشخيص امراض صدأ القمح عن طريق الدمج بين المنطق المضبب والشبكات العصبية إضافة إلى طريقة Bayesian. هدفت الدراسة إلى مراقبة واقتراح التدابير العلاجية اللازمة للسيطرة على المرض، قدمت الدراسة نتائج امتازت بالدقة والأداء الجيد. مما يرجح من استخدام الأنظمة الخبير لرفد القطاع الزراعي بالتقنيات الذكية التي ترفع كفاءة الإنتاج وتقليل مخاطر الأمراض المختلفة التي تصيب النباتات.

4- دراسة (الشهري، 2012):

تحاول هذه الدراسة البحث في الإجابة عن تساؤل معين وهو ما هي إمكانيَّة نظم المعلومات المتكاملة في إنتاج نظام خبير ذكي؟، استخدم الباحث لحل مشكلة البحث المنهج الوصفي التحليلي والمتمثل بدراسة الآليات والمعالجات التي يتم استخدامها في توظيف أنظمة المعلومات لبناء الأنظمة الخبيرة. عرض الباحث كيفية بناء مكونات النظم الخبيرة وماهي الخوارزميات المتبعة في انشاءها، والدور الذي يلعبه الذكاء الاصطناعي في رفد العلوم بشتى الوسائل والطرائق المميزة. وتوصلت الدراسة إلى ان نظم المعلومات عند تكاملها مع الأنظمة الخبيرة، فأنها تقدم نموذجاً متطوراً لا غنى عنه في مختلف الأجهزة الإدارية.

5- دراسة (الصعيدي، 2015):

هدفت الدراسة إلى الكشف عن أثر تصميم نظام خبير تعليمي على تنمية الجوانب المعرفية والادائية المتلقة بمهارات إنتاج المقررات الإلكترونيَّة لطلبة الدراسات العليا في كلية التربية، تتضمن الدراسة تصميم نظام خبير تعليمي للمقررات الدراسية. استخدم المنهج الوصفي التحليلي والمنهج التجريبي لقياس تأثير المتغيرات. تضمن البحث أيضًا إجراء تجربة على عدد من طلبة الدراسات العليا لتقييم النظام الخبير المقترح، اسفرت النتائج عن وجود تأثير معنوي للأنظمة الخبيرة على فهم المقررات الدراسية.

6- دراسة (الجليل، 2017) :

تناولت الدراسة موضوع النظم الخبيرة وعلاقتها بمجال اتخاذ القرار في عملية منح القروض المصرفية، وترمي الدراسة إلى إثبات ان النظم الخبيرة هي ليس مقتصرة على الصناعات والتقنية فقط، ولكن يمكن استخدامها في شتى المجالات. تم توظيف المنطق المضبب في تصميم النظام الخبير فضلًا عن الشبكات العصبية، ومن النتائج التي خرجت بها الدراسة هي تميز أنظمة دعم القرار بالحيوية والديناميكية فيما يخص اتخاذ القرارات التي تعنى بالقروض المصرفية، إذ يمكن للنظام من اقتراح قبول أو رفض القرض بناءً على عدد من المعطيات.

ومن الدراسات آنفًا يتضح أهميَّة الأنظمة الخبيرة في رفد مختلف التخصصات بالآليات والتقنيات المفيدة الَّتي من شانها رفع كفاء الأداء وزيادة جودة المعلومات ومن ثُمَّ تحسين في نوعية الخدمة المقدمة.

ثانياً: دراسات تناولت تصنيف النصوص العربية:

1- دراسة (Ghanem & Ashour, 2012) :

ناقشت الدراسة تأثير تطبيق التقنيات المشتقة على تجميع النصوص، وذلك بتحديد الإجراء الأكثر ديناميكية لمرحلة ما قبل المعالجة .تم استخدام ثلاث طرائق الاشتقاق المستند إلى الجذر، الاشتقاق الخفيف وبدون الاشتقاق للمقارنة .علاوة على ذلك، تم تطبيق خوارزمية التجميع K-mean لعملية التجميع. أثبتت نتائج التجربة أن الاشتقاق الخفيف يفوق التقنيات الأُخرى من حيث الاسترجاع والدقة.

2- دراسة (Sharef, Omar, & Sharef, 2014)

افترضت الدراسة نهجًا لحل وتعزيز أداء نمط تصنيف النصوص العربية عبر تقنية مقترحة جديدة تسمَّى طريقة تراكم نسبة التردد، التي تعتمد على نمط رياضي غير معقد. تم تنفيذ العديد من خطوات المعالجة المسبقة لمهمة معالجة الميزة الَّتي تم التعامل معها كمتطلبات أساسية لمهمة التصنيف. تم استخدام نظرية Bayes لتقييم أداء النموذج المقترح مقارنة مع ثلاث مصنفات تحت خوارزمية (Naïve Bayes NB)، أكدت الاستنتاجات النهائية لهذا النهج أن النموذج حقق 95.1 ٪ نسبة دقة في تصنيف النصوص العربية ومما يثبت كفاءة النموذج.

3- دراسة (Al-Saleem, 2010):

قامت الدراسة بالتقصي حول استخدام طريقة NB وخوارزمية Support Vector Machine (SVM) وتطبيقها على مجموعة من البيانات العربية مختلفة. يتضمن عملية تقييم الأَداء قياسه استخدام أكثر المؤشرات شيوعاً. أظهرت النتائج التجريبية لمجموعات بيانات التصنيف لنصوص عربية مختلفة أن خوارزمية SVM يمكن أن تتفوق على NB فيما يتعلق بالبيانات المستخدمة في هذه الدراسة.

4- دراسة (Elnagar, Al-Debsi, & Einea, 2020):

تم في هذه الدراسة تقديم مجموعات بيانات جديدة غنية وغير متحيزة لمهام تصنيف النص العربي ذات التسمية وهي (SANAD) و (NADiA)، كلا المجموعتين متاحتان مجانًا لمجتمع البحث في اللغويات الحاسوبية العربية. علاوة على ذلك، تقدم الدراسة مقارنة شاملة للعديد من نماذج التعلم العميق (DL) لتصنيف النص العربي من أَجل تقييم فعالية هذه النماذج. من مميزات الدراسة انها لا تتطلب مرحلة ما قبل المعالجة ويعتمد بالكامل على نماذج التعلم العميق. إِلى جانب ذلك، تمت دراسة تأثير استخدام نماذج تضمين word2vec لتحسين أَداء مهام التصنيف، وأَظهرت النتائج التجريبية أَداء قويًا لجميع نماذج مع مجموعة SANAD مع دقة لا تقل عن 91.18٪، أما بالنسبة إِلى NADiA، فقد حققت أعلى دقة إِجمالية قدرها 88.68٪ لمجموعة فرعية من 10 فئات كحد أقصى.

ثانياً: الجانب النظري

2-1 النظم الخبيرة:

بشكل عام يعد النظام الخبير برنامجاً محوسباً يتم تصميمه لحل عدد من المشاكل المعقدة التي يتطلب حلها جهداً بشرياً كبيراً، فيقوم النظام بمحاكاة التفكير البشري وإيجاد أفضل الحلول الممكنة. وتستخدم النظم الخبيرة مبادئ المعرفة لدى الإِنسان لحل مشاكل في الغالب تحتاج إِلى الذكاء البشري عن طريق تمثيل معلمات معارف الخبراء على شكل عدد من القواعد والبيانات الإِلكترونيَّة، ويتم استخدام هذه القواعد لحل مختلف المشكلات المرتبطة بجانب معين. وتعد الأنظمة الخبيرة أحد فروع الذكاء الاصطناعي الَّذي يتكون من عدد من السلوكيات والانماط تحاكي القدرات

الذهنية وطريقة عملها، وامن أَمثلة هذه الأنماط: التعلم والاستنتاج ورد الفعل على أوضاع معينة (عفيفي، 2014). تهدف النظم الخبيرة بشكل عام إِلى اعداد وتطوير برمجيات الكترونية يمكن لها ان تتخذ القرارات وتعالج المشكلات المختلفة والوصول إِلى نفس النتائج التي قد يصل اليها الشخص الخبير (حسن، الحميد، & عفيفي، 2014).

2-2-1 خصائص النظم الخبيرة:

هناك عدد من الخصائص التي تميز النظم الخبيرة أَو التي يجب إتباعها عند بناء وتصميم النظم الخبيرة ومنها: (الفقي، 2012; عيسى، 2016) (القرة غولي، 2005)

1- الخبرة:

تُعدُّ الخبرة من أَبرز الأسس والخصائص التي يجب أن تتوافر في النظم الخبيرة لكي تعمل بكفاءة، بمعنى أن تحقق النظم الخبيرة نفس مستوى الأَداء الذي يحققه الخبراء البشر في مجال معين، فالنظم الخبيرة يمكنها أن تصل إِلى حلول للمشكلات بسرعة وبكفاءة معتمدة في ذلك على الحيل والطرائق المختصرة الناجمة عن الخبرة التي تمتلكها وذلك بنفس أسلوب الخبير البشرى.

2- القدرة على الشرح:

لكي يحاكى النظام الخبير أَداء الخبير البشرى يجب أن تكون لديه القدرة على شرح وتفسير قراراته المتعلقة بطريقة حل مشكلة معينه أَو اتخاذ قرار معين، إذ أن تكلفة أي قرار خطأ يتم اتخاذه قد تكون مرتفعة وكبيرة جداً، ويقصد بالشرح الإِجابة على أَسئلة معينة تتعلق بالطرائق التي تتبعها النظم الخبيرة في الاستنتاج المنطقي أَو إيجاد الحلول للمشكلات واتخاذ القرارات.

3- القدرة على التعلُّم:

فالنظام الخبير يجب أن يكون قادراً على التعلم من تجاربه السابقة فكما هو الحال مع الخبراء البشر فإنَّ القدرة على التعلم تعنى أن النظام الخبير سيكون قادراً على الوصول إِلى حلول مشابهة بسرعة أكبر عندما تعرض عليه نفس المشكلة أكثر من مرة.

4- الاستنتاج الرمزي:

تستخدم النظم الخبيرة عند حلها للمشكلات الاستنتاج الرمزي، بمعنى اختيار مجموعة من الرموز (Symbols) لتمثيل أنواع مختلفة من المعارف مثل الحقائق، المفاهيم، القواعد التي تخص المسألة أَو المشكلة المطلوب إيجاد حل لها، فالنظم الخبيرة يتم بناءها لمعالجة المعارف وذلك على عكس البرامج التقليدية التي تقوم بمعالجة البيانات العددية الرقمية وذلك باستخدام الخوارزميات بمعنى أنها تتبع تسلسلاً من الخطوات المحددة سلفاً لكي تصل إِلى الحل، بينما النظم الخبيرة لا تتبع تسلسلاً من الخطوات المحددة من قبل إذ أنها تسمح بحدوث تفكير غير دقيق، ويمكنها التعامل مع البيانات الناقصة والمشوشة، ويقصد بالرمز "سلسلة من الحروف والأرقام تمثل مفاهيم

العالم الحقيقي.

5- الكفاءة والفاعلية:

يجب أن يتميز النظام الخبير بالعمق أي يعمل بفاعلية وكفاءة في مجال تطبيق دقيق ومحدود ويحتوي على مشاكل تتسم بالصعوبة تمثل تحدياً للنظام الخبير، ومن هنا فإنَّ القواعد التي يحتويها النظام الخبير تكون معقدة بالضرورة سواء كان هذا التعقيد راجعاً إِلى طبيعة تركيبها أَو إِلى عددها، فالنظم الخبيرة تستخدم غالباً في مجالات تتصل بمشاكل واقعية وحقيقية.

6- المعرفة الذاتية:

يجب أن يمتلك النظام الخبير المعرفة التي تحقق له الاستدلال المنطقي فيما يقوم به هو نفسه من عمليات هذا فضلًا عن امتلاكه لبنية أساسية من المعرفة تحقق تبسيط هذا الاستدلال، وذلك بغرض اختبار النظام الخبير لمدى دقة ومعقولية النتيجة التي توصل إليها. علاوة على استنباط الحجج والبراهين التي تشرح الاستدلال المنطقي الذي أدى إِلى النتيجة التي توصل إليها ويطلق على هذه المعرفة اسم المعرفة الذاتية.

2-2-2 مكونات النظام الخبير:

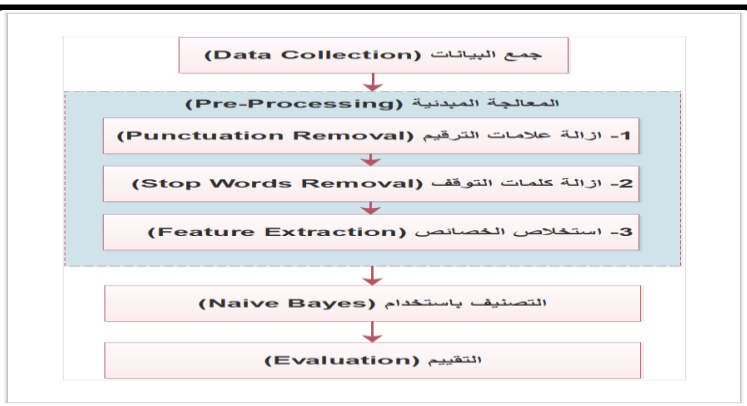
حتى يتسنى للنظام الخبير أن يقوم بالدور المنوط اليه لتحقيقه يجب أن يشتمل على عدد من المكونات الأساسية الآتية: (بسيونى، 2005; عفانة، 2012)

- قاعدة المعرفة Knowledge Base
- محرك الاستدلالInference Engine
- واجهة التفاعل User Interface
- وحدة الشرح Explanation Facility
- وحدة إمكانيَّة التحديث Knowledge Update Facility

ثالثاً: الجانب العملي

3-1 منهجية البحث:

نقدم في هذه الفقرة من البحث جميع الإجراءات التي تم اتباعها من لتحقيق اهداف البحث والإجابة على تساؤلاته. ويضم هذا الفصل الآليات المعتمدة في الحصول على البيانات فضلًا عن جميع الإجراءات الخاصة بمعالجة المبدئية للبيانات وتصفيتها وانتهاءً بتنفيذ عمليات التصنيف باعتماد تقنيات الذكاء الاصطناعي، وأخيراً مرحلة تقييم نتائج النظام المقترح، وكما هو موضح في الشكل رقم (1). ان جميع الإجراءات الخاصة بتنفيذ وبرمجة النظام الخبير المقترح تم تنفيذها باستخدام اللغة البرمجية بايثون (Python). وتعد هذه اللغة من أقوى اللغات البرمجية المستخدمة لأغراض تنفيذ تقنيات تعلم الالة والذكاء الاصطناعي.



شكل رقم (1): مخطط منهجية البحث

وفيما يلي من فقرات سيتم التطرق إلى تفاصيل كل مرحلة من المراحل سابقة الذكر في منهجية البحث.

3-2 جمع البيانات :

كما تم ذكره مسبقاً، أنَّ البحث الحالي يتمحور حول تقديم مقترح لنظام خبير هدفه تصنيف الرسائل الجامعية باستخدام عناوينها، بمعنى ادق، يشمل عمل النظام الخبير على وضع الاطروحة أو الرسالة ضمن التخصص العام لها على وفق العنوان، إضافة لذلك ان العناوين التي سيتم تصنيفها هي باللغة العربية. تم جمع بيانات البحث من المستودع الرقمي العراقي للرسائل والأطاريح [1] الذي يعد من أكبر المستودعات العراقية التي تضم عناوين الرسائل الجامعية، وتم بناء هذا المستودع بدعم من الأمانة العامة للعتبة العباسية المقدسة واشراف مباشر من مركز المعلومات الرقمية التابع لمكتبة العتبة العباسية بالتعاون مع الجامعات العراقية. كما تم استخدام تقنيات (Scrapping) بواسطة لغة بايثون وهي من أشهر لغات البرمجة حالياً الَّتي تستخدم على نطاق واسع في تنفيذ مختلف العمليات البرمجية المختلفة ولاسيَّما ما يتعلق بالخوارزميات والذكاء الاصطناعي، أَمَّا تقنية (Scrapping) فتم الاستعانة بها من أَجل

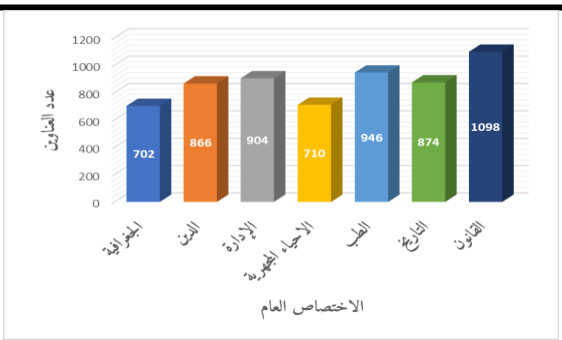
(1) المستودع الرقمي العرقي: يمكن الوصول له من خلال الرابط الالكتروني التالي **(https://iqdr.iq)**

تسريع عملية جمع البيانات. تتضمن تقنية (Scrapping) عدداً من الإجراءات البرمجية الَّتي تأخذ على عاتقها استخلاص البيانات من المواقع الإلكترونيَّة وتحميلها إلى المستخدم، إذ عندما يتم البحث في المستوعبات الإلكترونيَّة وتظهر نتائج البحث، تظهر الحاجة لاستخلاص هذه النتائج بشكل ملف منفصل لغرض إدخاله ضمن الخوارزميات، تقوم تقنية (Scrapping) بهذه المهمة وهي تحويل النتائج التي تم الحصول عليها من المواقع الإلكترونيَّة إلى ملفات جدولية يستخدمها الباحث للمعالجة والتحليل من خلال إدخالها في مختلف الخوارزميات. بلغ عدد العناوين التي تم جمعها (6100) عنواناً لرسالة ماجستير وأُطروحة دكتوراه، كما توزعت هذه العناوين على وفق (7) أصناف تمثل الاختصاص العام وكما هو موضح في الجدول رقم (1).

جدول (1): وصف عام لبيانات البحث

ت	الاختصاص العام	عدد العناوين
1	الجغرافية	702
2	الدين	866
3	الإدارة	904
4	الاحياء المجهرية	710
5	الطب	946
6	التاريخ	874
7	القانون	1098
	المجموع	6100

وكما هو معلوم أَنَّه هناك العديد من الاختصاصات الأُخرى التي كان بالإمكان تضمينها في النظام الخبير، ولكن ولكون الأنظمة الخبيرة تعمل بنحو أفضل مع البيانات الضخمة، لذا تم اختبار النظام على التخصصات السبعة في الجدول آنفًا كونها تضمنت اكثر من 500 عنوان في قاعدة البيانات ضمن المستودع الرقمي، كانت نتائج البحث عن الاختصاصات آنفًا (500) عنوان وهو ما يصب في الجناب الإيجابي لنتائج البحث. والشكل رقم (2) يوضح مخططاً لبيانات البحث وتوزيعها على وفقت الاختصاصات المختلفة.



شكل (2): مخطط توزيع بيانات البحث على وفق التخصصات

والجدول رقم (2) يعرض نماذج للبيانات البحث التي تم جمعها، إذ يتم تقسيم العناوين على وفق الاختصاص العام لها.

جدول (2): نماذج للبيانات التي تم جمعها

ت	الاختصاص	العنوان
1	الجغرافية	دراسة مناخية لتكرار الظواهر الجوية الغيوم والعواصف الرعدية والبرد في العراق
2	الدين	الاصول الدينية للتعايش الانساني في الاديان السماوية
3	الإدارة	الجودة وأثرها على الاستراتيجية التسويقية للخدمة المصرفية دراسة تحليلية على عينة من المصارف التجارية العراقية
4	الاحياء المجهرية	عزل بعض انواع البكتريا الخيطية من الاشخاص المعرضين لفرط التحسس الرئوي وبيئتهم في الموصل
5	الطب	استخدام البوليمرات ذات القابلية على التحلل البايولوجي في تثبيت الكسور
6	التاريخ	الاحوال العامة للمرأة في الجزيرة العربية قبل الاسلام حتى منتصف القرن الاول الهجري
7	القانون	دور الضبط الاداري في حماية البيئة في القانون العراقي دراسة مقارنة

ومن الجدير بالذكر أنَّ البيانات التي تم جمعها تتنوع فيما بيناه من حيث طول العنوان (عدد الحروف والرموز التي يحتويها. واستناداً إلى الجدول رقم (3) الَّذي يمثل خلاصة إحصائيَّة لبيانات، يمكن ملاحظة أنَّ أقصر عنوان تم الحصول عليه يحوي على (49) حرفاً ورمزاً، فيما بلغ أطول عنوان (251) حرفاً ورمزاً. وكما موضح أيضًا أن متوسط عدد الأحرف لجميع العناوين يبلغ (90.55)، ونلاحظ هنا أنَّ البيانات التي تم جمعها تمتاز بقلة عدد الكلمات، ويعود ذلك إلى المعايير الخاصة

بوضع عناوين الرسائل؛ إذ يجب أن تكون وافية ومركزة. أنَّ قلة عدد الكلمات تشكل تحدياً في عملية التصنيف؛ إذ من المعروف أنَّه دقة التصنيف تزداد كلما زادت البيانات المدخلة إلى الخوارزمية.

جدول (3): ملخص احصائي لوصف البيانات التي تم جمعها

ت	الفقرة	الوصف
1	مجموع العناوين	6100
2	الوسط الحسابي لعدد الاحرف والرموز	90.55
3	أقصر عنوان	49
4	أطول عنوان	251
5	عدد الاحرف والرموز لـــ 25% من البيانات	68
6	عدد الاحرف والرموز لـــ 50% من البيانات	83
7	عدد الاحرف والرموز لـــ 75% من البيانات	109

3-3 المعالجة المبدئية للبيانات (Pre-Processing)

بغية معالجة النصوص التي تم جمعها، وتحليل البيانات التي تحملها هذه النصوص، هنالك عدد من الإجراءات والمعالجات الواجب عملها لغرض إتمام العمل على أكمل وجه. تعد البيانات النصية واحدة من البيانات المعقدة التي يتعامل معها الحاسوب، وذلك لتنوع البنية للجمل النصية واختلافها بشكل ملحوظ. كما ان تصنيف عنوان معين ووضعه في اختصاص معين يتطلب من النظام الخبير ان يفهم إلى حد ما عدد من المعطيات الموجود في هذه الجملة النصية. ان تحليل البيانات النصية ومعالجتها ليست بالعملية اليسيرة بل وتتطلب معالجة دقيقة للحفاظ على البيانات النوعية وعدم فقدانها خلال مراحل المعالجة.

أنَّ عملية تحليل وتصنيف النصوص تقدمت بشكل ملحوظ وازدادت البحوث والدراسات التي تعالج هكذا نوع من البيانات، ويأتي ذلك نتيجة لتطور وسائل التواصل الاجتماعي وانتشار البيانات النصية واتساع رقعة استخدامها عالميًا، ولكن على الرغم من هذا التطور الهائل في مجال معالجة وتحليل البيانات النصية، نجد أنَّ هذا التطور اقتصر نوعا ما على اللغة الإنجليزية، وهنالك افتقار واضح فيما يتعلق بتحليل النصوص العربية. ان هذا النقص في تطوير عمل النصوص العربية وتحليلها

حاسوبياً يأتي نتيجة لصعوبة اللغة العربية من حيث بنيتها. فاللغة العربية تمتاز عن باقي اللغات الحية ببنيتها الشكلية (Orthographical)، أي أَنَّ الكلمة العربية تعتمد في رسم حروفها بحسب موقع الحرف، على سبيل المثال حرف الميم من الممكن ان يأخذ اربعة اشكال ضمن موقعه في الكلمة ("مـ "، " ـمـ"، "م"، "ـم")، الأَمر الذي يعدُّه الحاسوب حالة مختلفة كل مرة يأتي بها الحرف على شكل معين؛ إِذ في اللغة الإنجليزية الاختلاف يكون فقط في كون الحرف كبيراً أَو صغيراً بغض النظر عن موقعه في الكلمة، إضافة لذلك هناك أَمر آخر فيما يتعلق ببنية الكلمة في اللغة العربية، إِذ تعتمد الكلمة على العلامات الإِعرابية (Diacritics) وهي (الضمة، الفتحة، الكسرة، السكون، الشدة)، تؤثر هذه العلامات بشكل رئيسي في معنى الكلمة على سبيل المثال لو اخذنا الكلمة (علم)، هذه الكلمة لا يمكن التكهن بمعناها بشكلها المجرد وبدون وجودها في جملة معينة، فمن الممكن ان تأخذ هذه الكلمة الاشكال التالية (عِلْمٌ، عَلَمٌ، عَلِمَ، عَلَّمَ، عُلِمَ، عُلِّمَ)، نلاحظ أَنَّه باختلاف العلامات الاعرابية يختلف المعنى أَيضًا وبشكل ملحوظ. من جهة أُخرى، في اللغة العربية هنالك إِمكانيَّة لإطالة الكلمات لأسباب تتعلق بأسلوب التعبير أَو لأسباب ترتيبية لترتيب شكل نص معين، مثلاُ كلمة ("سعيد") يمكن ان تكتب بشكل ("سعيـــــــــــــــــــــــد")، إِذ بالإمكان إضافة واحد أَو عدة رموز متتابعة من الرمز(ـــ) لإطالة الكلمة وهذه الإِطالة يمكن ان تكون لترتيب الشكل أَو للتعبير عن المبالغة في السعادة. (Oussous, Benjelloun, Lahcen, & Belfkih, 2020)

إِنَّ جميع ما تم ذكره في آنفًا، يوجب توخي الدقة في إِجراء المعالجات البرمجة على النصوص العربية، إِذ من الممكن ان تفقد الكلمات معناها نتيجة لتطبيق الآليات البرمجية لمعالجة النصوص، اضافة لكون العناوين قصيرة ومختصرة، الأَمر الذي يضيف تحدياً آخر لتصنيف النصوص العربية. (Alsayat & Elmitwally, 2020)

واستناداً لمبادئ الذكاء الاصطناعي والنظم الخبيرة، وبغية الشروع بإدخال البيانات النصية ضمن خوارزمية التصنيف، فإنَّ هناك عددًا من الإجراءات والمعالجات المبدئية يجب تطبيقها على النصوص بغية استخلاص البيانات المفيدة منها وكما هو موضح في الفقرات اللاحقة.

3-3-1 إزالة علامات الترقيم (Punctuation Removal):

إنَّ مبدأ عمل الأنظمة الذكية وتحليل البيانات، يحتم التخلص من جميع البيانات غير الضرورية وتصفية البيانات بطريقة تبقي فقط على ما هو ضروري وأَساسي، والغالية من ذلك هو لإثراء خوارزميات التصنيف بالبيانات المؤكدة فقط، فضلًا عن ذلك، استبعاد البيانات غير الضرورية يُسهم في تقليل حجم البيانات التي يتم معالجتها، ويؤدي ذلك إلى رفع كفاءة خوارزميات التصنيف والحصول على أَداء أَعلى ومعالجة أسرع.

إنَّ وجود علامات الترقيم أَمر طبيعي في كل اللغات العالمية، تُسهم علامات الترقيم في إضفاء دلالات معينة ضمن النصوص المختلفة، وتختلف علامات الترقيم باختلاف مدلولاتها على سبيل المثال (!" # $ % & \ ' () * + , - . / : ; < = > ? @ [\ ] ^ _ ` | ~)، نلاحظ أَنَّ هناك عددًا غير قليل من علامات الترقيم الَّتي من الممكن أَن توجد ضمن بيانات البحث ومن أشهرها (الاقتباس، السؤال، الاقواس بأنواعها المختلفة، الفواصل، الاستطالات وغيرها)، إنَّ وجود علامات الترقيم ليس له فائدة من حيث تحليل النص؛ إذ إنَّه لمعرفة عنوان معين وتصنيفه بشكل صحيح، نحتاج إلى تحليل الكلمات الفعلية وليس العلامات الأُخرى. (Zahidi, El Younoussi, & Al-Amrani, 2021) ، (Ibrahim, Alhakeem, & Fadhil, 2021)

ومن الجدير بالذكر أَنَّ علامات الترقيم تكون قليلة في البيانات الخاصة ببحثنا، ولكن قد تكون كثيرة ولاسيَّما عند التعامل من النصوص الطويلة، ولكن بغية اتباع القواعد السليمة لتصنيف النصوص، تم تنفيذ عملية إزالة علامات الترقيم من جميع النصوص الموجودة في بيانات البحث. وبغية تنفيذ هذه الآلية، توفر لغة بايثون البرمجية مكتبات جيدة تمكن من إزالة علامات الترقيم بسهولة.

3-3-2 إزالة كلمات التوقف (Stop Words Removal):

إنَّ مصطلح كلمات التوقف يشمل عددًا من الكلمات التي توجد ضمن الفقرات النصية الَّتي لا تكون مفيدة في عملية تحليل النص؛ لأَنَّ الأساس في تمييز عنوان ومعين وتصنيفه هو إيجاد ما يميزه عن باقي العناوين في قاعدة البيانات. ان علامات الترقيم سابقة الذكر والكلمات التوقف لا تمثل تميزاً في نص دون الآخر؛ إذ تتكرر هذه العلامات والكلمات في الكثير من النصوص دونما تأثير على تفرد نص دون الاخر. ومن الامثلة على كلمات التوقف الموجودة في اللغة العربية ما هو موضح في جدول رقم (4). (Mostafa, 2017)

جدول (4): كلمات التوقف في اللغة العربية

إذ	إنا	بكما	خلا	عند	كيفما	لئن	هذان	إما	هاته	ثم	كلتا	ته
إذا	أنت	بكن	دون	غير	لا	ليت	هذه	إن	هاتي	عدا	لنا	ريث
إذما	أنتم	بل	ذا	فإذا	لاسيما	ليس	هذي	حاشا	هاتين	كليكما	نحو	كلا
إذن	أنتما	بلى	ذات	فإن	لدى	ليسا	هذين	حبذا	هاك	لها	وإذ	لما
أف	أنتن	بما	ذاك	فلا	لست	ليست	هكذا	حتى	اهنا	ها	لو	مهما
أقل	إنما	بماذا	ذان	فمن	لستم	ليستا	هل	حيث	هذا	وإن	إليكن	الَّذي
أكثر	إنه	بمن	ذانك	في	لستما	ليسوا	هلا	حيثما	ولكن	لوما	إليك	لهما
ألا	أنى	بنا	ذلك	فيم	لستن	ما	هم	حين	ولو	إليكما	بس	ولا
التي	آه	به	ذلكم	فيما	لسن	ماذا	هما	عل	وما	بعض	تي	تين
الذي	آها	بها	ذلكما	فيه	لسنا	متى	هن	على	ومن	تينك	سوف	اللواتي
الذين	أو	بهم	ذلكن	فيها	لعل	مذ	هنا	عليك	وهو	شتان	كلاهما	إيه
اللاتي	أولاء	بهما	بعد	قد	لك	مع	هاتان	عليه	يا	كلما	لن	تلكما
اللائي	أولئك	بهن	ذو	كأن	لكم	مما	هنالك	عما	أم	له	نحن	ذينك
اللتان	أوه	بي	ذوا	كأنما	لكما	ممن	هو	عن	بكم	نعم	الَّذين	كل
التيا	آي	بين	ذواتا	كأي	لكن	من	هؤلاء	كم	ثمة	وإذا	لهن	لم
اللتين	أيها	بيد	ذواتي	كأين	لكنما	منه	هي	كما	عسى	لولا	لي	مه
اللذان	أين	تلك	ذي	كذا	لكي	منها	هيا	كي	كليهما	بك	سوى	هيهات
اللذين	أينما	تلكم	ذين	كذلك	لكيلا	منذ	هيت	كيت	لهم	إليكم	إلى	كيف

وكما تم ذكره مسبقاً، ان وجود الكلمات آنفًا في نص معين من عدمه لا يشكل فرقاً جوهرياً فيما يتعلق بخصائص النص، إذ يتم البحث في النصوص عن الكلمات الفريدة الَّتي قد لا تتواجد في نصوص أُخرى، لذلك يتحتم رفع جميع كلمات التوقف من النصوص لغرض الإبقاء فقط على الكلمات والمصطلحات الأساسية، وكذلك تقليل حجم النصوص والاستفادة من هذا الجانب. وكما هو الحال في علامات الترقيم، أَيضًا توفر لغة برمجة بايثون آلية ممتازة لحصر كلمات التوقف وازالتها من النصوص، وتم تطبيق هذه الآلية ولمعرفة ذلك يمكن تصور المثال الآتي:

العنوان الأصلي: *"دور تكنولوجيا المعلومات الاستراتيجية في توزيع الحصص المائية على الجداول النهرية دراسة تطبيقية في مديرية الموارد المائية في محافظة بابل"*

العنوان بعد المعالجة: *"دور تكنولوجيا المعلومات الاستراتيجية توزيع الحصص المائية الجداول النهرية دراسة تطبيقية مديرية الموارد المائية محافظة بابل"*

3-3-4 استخلاص الخصائص (Features Extraction)

يُعدُّ استخلاص الخصائص من أَبرز مراحل بناء النظم الخبيرة؛ إذ يتم في هذه المرحلة استخراج البيانات التي لها قيمة معنوية فقط من مجموع البيانات الداخلة في عملية التصنيف. تهدف هذه المرحلة إلى إنتاج بيانات مشتقة من البيانات الأَصليَّة بهدف تقليل حجمها ومعالجة المفيد منها فقط. ففي أَي نظام خبير أَو أُنموذج ذكاء اصطناعي لا يتم ادراج البيانات كما هي، بل يتم إدخال فقط البيانات ذات الفائدة المعنوية الَّتي تخدم عملية التحليل والتصنيف.

إنَّ البيانات النصية الفعلية (أي الكلمات كما هي) لا يتم إدخالها مباشرة إلى خوارزميات التصنيف ما لم يتم تحويلها إلى مصفوفات رقمية، تحوي هذه المصفوفات على قيم معينة تمثل دلالات على اوزان الكلمات في النصوص، ان تحويل جميع الكلمات إلى قيم رقمية، يرافقه تكوين مصفوفات كبيرة الحجم، فمثلًا إذا أَردنا تحويل الكلمات في نص يحتوي على (1000) كلمة نحتاج إلى مصفوفة تستوعب (1000) قيمة رقمية، وفي الحقيقة ان اغلب خوارزميات معالجة النصوص تعالج عشرات بل ومئات الآلاف من الكلمات مما يعني ان هنالك ضرورة للتعامل مع كم هائل من

البيانات داخل المصفوفات، الأمر الذي يسبب استنزاف موارد الحاسوب واستهلاك وقت طويل للمعالجة.

عليه فإنَّ استخلاص الخصائص، يتضمن الإبقاء فقط على البيانات التي لها تأثير معنوي على مستوى الأداء في خوارزميات التصنيف. وهنا تكون هنالك فائدتين، الأولى: هو تقليل حجم المصفوفات الداخلة في المعالجة والتخلص من جميع البيانات غير المفيدة في التصنيف، والأمر الثاني هو زيادة كفاءة عملية التصنيف بتحسين الأداء وتسريعه. تتلخص عملية استخلاص البيانات المفيدة في انشاء مصفوفة من البيانات تحتوي هذه المصفوفة على مؤشرات معينة لها أوزانها وقيمها. ويتم ذلك في حصر الكلمات التي لها اوزان معنوية ضمن الجمل النصية.

هناك عدد من الطرائق التي يتم استخدامها في استخلاص خصائص النصوص، والطريقة التي نعتمدها في بحثنا هذا هي طريقة (مصطلح تردد الوثيقة العكسي) وهو ما يعرف بالغة الإنجليزية بالعبارة الآتية: (Term Frequency – Inverse Term Frequency) ويتم اختصارها على الشكل (TF-IDF)، هذه الطريقة هي عبارة عن مقياس إحصائي يقيس مدى صلة كلمة معينة في نص محدد مع مجموعة من النصوص في نفس قاعدة البيانات. فمثلا في بحثنا يتم قياس مدى صلة كلمة في عنوان الرسالة أو الاطروحة مع جميع العناوين الموجود في قاعدة البيانات. وتعمل عن طريق زيادة متناسبة مع عدد المرات التي تظهر فيها الكلمة في نص معينة، ولكن يقابلها عدد النصوص الأُخرى التي تحتوي على الكلمة. لذا، فإنَّ الكلمات الشائعة في كل نص على سبيل المثال كلمات التوقف التي تم ذكرها مسبقاً في جدول رقم (4) تأتي في مرتبة منخفضة على الرغم من أنها قد تظهر عدة مرات؛ لأنَّها لا تعني الكثير لهذا النص على وجه الخصوص، أي أنَّها منتشرة في جميع النصوص وليس لها فائدة معنوية في نص دون غيره؛ لذا تكون قيم هذه الكلمات منخفضة ويتم استبعادها.

يتم حساب TF-IDF لكلمة في نص معين بضرب مقياسين مختلفين :

1. تكرار الكلمة في نص. هناك عدة طرائق لحساب هذا التكرار، أبسطها هو عدد الحالات التي تظهر فيها كلمة ما في نص ما. بعد ذلك، توجد طرائق لضبط التكرار، حسب طول النص، أو حسب التردد الأولي للكلمة الأكثر شيوعًا في النص

2. تكرار المستند العكسي للكلمة عبر مجموعة من النصوص. هذا يعني مدى شيوع أو ندرة الكلمة في مجموعة النصوص بأكملها. كلما اقتربنا من الصفر، كلما كانت الكلمة أكثر شيوعًا، ويمكن حساب هذا المقياس بأخذ العدد الإجمالي للنصوص (عناوين الرسائل الجامعية)، وتقسيمه على عدد النصوص التي تحتوي على الكلمة، وحساب اللوغاريتم لها. لذلك، إذا كانت الكلمة شائعة جدًا وظهرت في العديد من النصوص، فسيقترب هذا الرقم من الصفر، وإلا فسيقترب من الواحد. ينتج عن ضرب هذين الرقمين درجة TF–IDF للكلمة في النص. كلما زادت القيمة، زادت صلة هذه الكلمة بهذا النص تحديداً. لوضعها في مصطلحات رياضية أكثر رسمية، يتم حساب درجة TF–IDF للكلمة(t) في النص (d) من مجموعة من النصوص (D) على النحو الآتي:

$$tf - idf = tf(t,d)\ .idf(t,D) \dots\dots (1)$$

$$tf(t,d) = log(1 + freq(t,d)) \dots\dots (2)$$

$$idf(t,D) = log(\frac{N}{count(d \in D : t \in d)}) \dots\dots (3)$$

ولو فرضنا أَنَّنا أَمام احتساب قيم تكرار الكلمة (TF) في نص معين، وعلى فرض النصين الآتيين:

1. أثر استخدام النظم الخبيرة في تصنيف البيانات والمعلومات (دراسة تجريبية في المكتبات العراقية).

2. النظم الخبيرة ودورها في إثراء المكتبات العراقية (دراسة تطبيقية).

نلاحظ من النصين آنفًا أَنَّ هناك عددًا من الكلمات مكررة في النصين وهي (المكتبات، العراقية، نظم، خبيرة، في)، لذلك يتم أولا توليد مصفوفة الكلمات الفريدة في النصين مع استبعاد تكرار الكلمات المكررة، وكما هو موضح في الجدول رقم

(5). بعدها يتم احتساب قيمة (TF) لكل كلمة من الكلمات وذلك بقياس صلتها بالنصوص الموجودة.

جدول (5): مثال على احتساب قيمة (TF)

ت	مصفوفة الكلمات	النص الأول (TF)	النص الثاني (TF)	IDF	النص الأول TF-IDF	النص الثاني TF-IDF
1	أثر	0.077	0.000	0.693	0.301	0.000
2	اثراء	0.000	0.111	0.693	0.000	0.377
3	استخدام	0.077	0.000	0.693	0.301	0.000
4	البيانات	0.077	0.000	0.693	0.301	0.000
5	الجامعات	0.000	0.111	0.693	0.000	0.377
6	الخبيرة	0.077	0.111	0.000	0.214	0.268
7	العراقية	0.077	0.111	0.000	0.214	0.268
8	المكتبات	0.077	0.111	0.000	0.214	0.268
9	النظم	0.077	0.111	0.000	0.214	0.268
10	تجريبية	0.077	0.000	0.693	0.301	0.000
11	تصنيف	0.077	0.000	0.693	0.301	0.000
12	دراسة	0.077	0.000	0.693	0.301	0.000
13	في	0.154	0.222	0.000	0.428	0.536
14	المعلومات	0.077	0.000	0.693	0.301	0.000
15	دورها	0.000	0.111	0.693	0.000	0.377

نلاحظ من الجدول آنفًا أَنَّ هناك كلمات حازت على قيمة (0) في إِحدى النصين فيما يتعلق بحساب (TF) وهذا معناه ان الكلمة لا تنتمي إِلى هذا النص. مثلاً نجد أَنَّ كلمة "أثر" حازت في العمود الأول على قيمة (0.077) في عمود النص الأَوَّل، بينما كانت قيمتها (0) في عمود النص الثاني، وهذا يعني ان هذه الكلمة تنتمي إِلى النص الأول فقط ولا تظهر في النص الثاني، اي انه لا يوجد ارتباط بينها وبين النص الثاني. في الجانب الآخر نرى كلمات أُخرى مثلًا كلمة "النظم"، نراها حازت على قيمة (0.077) فيما يتعلق بقيمة (TF) في النص الأول، وفي النص الثاني على قيمة (0.111)، أي ان هذه الكلمة تنتمي إِلى النصين، وهذا صحيح حسب المثال آنفًا.

إضافة لذلك وكما نشاهد في الجدول آنفًا أَنَّ هناك بعض الكلمات جاءت عند القيمة (0) فيما يتعلق باحتساب قيمة (IDF)، وهي الكلمات (الخبيرة، العراقية، النظم، المكتبات، في)، وهذا يدل على ان هذه الكلمات ليست فريدة في النصوص آنفًا أي تتواجد في النصين وربما تقل أهميتها عند تكرارها في مجمل النصوص؛ إذ من الناحية العملية، نبحث عن الكلمات التي تميز نص معين دون الاخر لكي يتم تصنيفها بشكل جيد. فمثلًا توجد الضمائر وحروف الجر وأسماء الإشارة وغيرها قد يكون في أَغلبية النصوص متوفراً، لو فرضنا اننا نتعامل مع تصنيف الرسائل على وفق عناوينها، بالإمكان ان تتواجد الكلمات الشائعة في جميع الاختصاصات مما يحول دون تمييز تخصص عن آخر ومن ثُمَّ عدم دقة عملية التصنيف؛ لذلك وبالنظر إِلى بيانات الجدول آنفًا، نلاحظ انه في نهاية المطاف سيتم توليد مصفوفة تحتوي على جميع الكلمات الفريدة الموجودة في جميع النصوص الداخلة في عملية المعالجة، فضلًا عن توليد أَوزان لها تمثل صلة هذه الكلمات بباقي النصوص الموجود، ومن الجدير بالذكر أَنَّ القيم في الجدول آنفًا تبدو متقاربة ومتكررة إِلى حد ما وذلك بسبب اننا نحلل نصين فقط ممَّا يدعو إِلى تكرار بعض القيم خصوصاً ان عدد الكلمات محدود. ولكن عند استخدام كم كبير من البيانات، عندها سيكون هنالك تنوع في قيم تكرار الكلمات بشكل ملحوظ. وكما في الجدول رقم (6) الَّذي يبين امثلة عن القيم التي تم احتسابها بإِجراء معالجة لجميع العناوين التي تم استخدامها في بحثنا هذا.

جدول (6): نماذج من قيم (TF-IDF) لعنوانين من قاعدة البيانات المستخدمة

تسلسل النص	الكلمة	قيمة (TFIDF)
1	التركيب	0.423
1	الاقتصادي	0.482
1	للسكان	0.299
1	قضاء	0.386
1	الزبير	0.331
5500	دراسة	0.346
5500	البكتريا	0.211
5500	الملوثة	0.287
5500	والمكونة	0.275
5500	للأغشية	0.402
5500	الرقيقة	0.489

5500	الحيوية	0.354
5500	حاويات	0.414
5500	الحليب	0.283

وبانتهاء مرحلة استخلاص الخصائص، تكون قاعدة البيانات جاهزة لإدخالها ضمن خوارزمية التصنيف وتقييم نتائج التصنيف.

3-4 تصنيف العناوين باستخدام خوارزمية (Naïve Bayes)

ذكرنا مسبقاً أَنَّ بحثنا الحالي يتضمن تصنيف الرسائل الجامعية اعتماداً على عناوينها باستخدام خوارزمية ((Naïve Bayes (NB). تعمل خوارزمية (NB) على مبدأ أن وجود ميزة أَو صفة معينة في فئة محددة لا علاقة لها بوجود أي ميزة أُخرى، مثلا لكل اختصاص هنالك كلمات تكون احتمالية وجودها عالية، مثلا في القانون قد نلاحظ كلمات (مادة، دستور، عقوبة، مدني، جريمة، الخ) وفي مادة الطب قد تكون هنالك كلمات (علاج، صحة، دواء، مرض، الخ). وبتعبير آخر، يمكن اعتبار الفاكهة على أنها تفاحة إذا كانت تملك المواصفات (اللون الأحمر، الشكل الدائري والقطر بحدود 3 بوصات)، حتى إذا كانت هذه الميزات تعتمد على بعضها البعض أَو على وجود الميزات الأُخرى، أي من الممكن ان تكون هنالك فواكه أُخرى تحمل نفس الصفات فإِنَّ كل هذه الخصائص تساهم بشكل مستقل في احتمال أن تكون هذه الفاكهة تفاحة وهذا هو السبب في أَنَّها تُعرف باسم "السذاجة" أي تأخذ الأُمور من جوانب بسيطة جدًا خوارزمية (NB) سهلة البناء ومفيدة بشكل خاص لمجموعات البيانات الكبيرة جدًا إِلى جانب بساطتها، من المعروف أن هذه الخوارزمية تتفوق في الأَداء حتى على أَساليب التصنيف المعقدة للغاية. إِذ من السهل بناء نموذج هذه الخوارزمية بدون تقدير قيم تكرارية معقدة مما يجعله مفيدًا بشكل خاص لمجموعات البيانات الكبيرة جدًا على الرغم من بساطته، غالبًا ما تعمل (NB) جيدًا بشكل ملفت للنظر ويستخدم على نطاق واسع؛ لأَنَّه غالبًا ما يتفوق على طرائق التصنيف الأَكثر تعقيدًا. وتعمل هذه الخوارزمية على وفق الصيغة الرياضية الآتية:

$$P(c|x) = \frac{P(x|c).\ P(c)}{P(x)}$$

إذ إنَّ:

$P(c|x)$: هو الاحتمال اللاحق للفئة (الهدف) اعتماداً على القيمة المعطاة (السمة).

$P(x|c)$: هو الاحتمال السابق للفئة.

$P(c)$: هو احتمال توقع فئة معينة

$P(x)$: هو الاحتمال السابق للتنبؤ.

ويمكن توضيح الأُمور بشكل أَدق باستخدام المثال التالي، أَدناه مجموعة بيانات للطقس ومتغير الهدف (يشير إِلى احتمالات اللعب). نحتاج إِلى تصنيف ما إذا كان اللاعبون سيلعبون أم لا بناءً على حالة الطقس. ويمكن استخدام خوارزمية (NB) لإِجراء عملية التنبؤ. ويتم ذلك بإِجراء تصنيف للقيم المعطاة ومن ثم بناء نموذج التنبؤ بحالة اللعب. وكما في الجدول رقم (7) الذي يمثل البيانات المعطاة عن حالة الطقس.

جدول (7): البيانات التدريبية

ت	حالة الطقس	قرار اللعب
.1	مشمس	كلا
.2	غائم	نعم
.3	ممطر	نعم
.4	مشمس	نعم
.5	مشمس	نعم
.6	غائم	نعم
.7	ممطر	كلا
.8	ممطر	كلا
.9	مشمس	نعم
.10	ممطر	نعم
.11	مشمس	كلا
.12	غائم	نعم
.13	غائم	نعم
.14	ممطر	كلا

ويتم في بداية الأمر تلخيص معطيات النموذج آنفًا عن طريق احتساب التكرارات المختلفة وكما في الجدول الآتي:

جدول (8): تكرارات البيانات التدريبية

حالة الطقس	كلا	نعم	احتمالية اللعب	النسبة
غائم	–	4	4/14	0.29
ممطر	3	2	5/14	0.36
مشمس	2	3	5/14	0.36
المجموع	5	9		
الاحتمالية	5/14=0.36	9/14=0.64		

وعند تطبيق الصيغة الرياضية لخوارزمية (NB)، وعلى فرض حساب توقع (الطقس مشمس، وحالة اللعب نعم) وكما يأتي:

احتمالية (نعم | مشمس) = احتمالية (مشمس | نعم) × احتمالية (نعم) ÷ احتمالية (مشمس)

وعلى وفق البيانات المعطاة في جدول البيانات التدريبية يمكننا أن نجد ما يأتي:

احتمالية (مشمس | نعم) = 9/3 = 0.33

احتمالية (مشمس) = 14/4 = 0.36

احتمالية (نعم) = 14/9= 0.64

إذًا احتمالية (نعم | مشمس) = 0.33 × 0.64 ÷ 0.36 = 0.60

وهو ما معناه أنَّ هناك ما نسبته (60%) من احتمال اللعب في حالة كون الطقس مشمساً. ويتم أيضًا بنفس المبدأ احتساب جميع الاحتمالات الأُخرى والتنبؤ بأي معلمة تخص هذا النوع من الأمثلة. ومن الجدير بالذكر أنَّ هذه الخوارزمية تعمل بكفاءة عالية مع البيانات النصية وكذلك مع عدد كبير من البيانات، وعلى وفق ذلك تم اختيار هذه الخوارزمية لغرض تصنيف بيانات البحث الحالي.

تم تطبيق خوارزمية (NB) على بيانات البحث باستخدام لغة برمجة البايثون.

تم لغرض إجراء عملية التصنيف تقسيم بيانات البحث إلى قسمين:

الأول: بيانات التدريب (Training Data): وهي البيانات التي يتم تدريب الخوارزمية عليها وإجراء عملية التصنيف، وتم وضع نسبة (80%) من البيانات الاصلية للتدريب وهو ما يشكل عدد (4880) عنواناً من عناوين الرسائل.

الثاني: بيانات الاختبار (Testing Data) وهي بيانات يتم إدخالها للخوارزمية بعد تدريب الخوارزمية على بيانات التدريب وتمثل هذه البيانات قيماً لم تتعرف عليها الخوارزمية من قبل ولم تطلع عليها. وتشكل نسبة بيانات الاختبار (20%) من مجموع البيانات الكلية للبحث، ما يعادل (1220) قيدًا.

إنَّ فصل البيانات إلى تدريب واختبار هو أُسلوب متبع في جميع عمليات التصنيف التي يتم تنفيذها باستخدام الأنظمة الخبيرة، كونها تقيس مدى قدرة النظام الخبير على تمييز البيانات عند إدخالها له اول مرة؛ إذ يتم استخدام بيانات التدريب بغية بناء نموذج التصنيف الذي يعتمد في الأساس على ما تم تدريبه من بيانات. عليه فأن جودة البيانات المدخلة تعتبر الجزء المحوري والمؤثر في دقة التصنيف.

بعد تنفيذ عملية التصنيف كانت نتائج التدريب كما في الجدول الآتي:

جدول (9): دقة التصنيف لبيانات التدريب (Training)

ت	الاختصاص	دقة التصنيف
1	الجغرافية	0.99
2	الدين	0.92
3	الإدارة	0.98
4	الاحياء المجهرية	0.89
5	الطب	0.97
6	التاريخ	0.95
7	القانون	0.88
معدل الأداء		0.94

ونلاحظ من الجدول آنفًا أَنَّ المعدل العام لدقة التصنيف لبيانات التدريب جاء عند (94%) وهو يمثل النسبة العامة للتدريب أي أَنَّ الخوارزمية تدربت على ما

نسبته (94%) من البيانات واستطاعت تكوين معدلات للتصنيف أَنَّ النسبة آنفًا تعد نسبة جيدة ولاسيَّما أَنَّ البيانات التي يتم التعامل معها تحوي على عناوين قد تكون قصيرة بعض الشيء الأَمر الذي يرجح أَن تكون هنالك نسبة خطأ في تصنيف البيانات.

الجزء الآخر هو عملية اختبار دقة النظام الخبير، وهو عملية فحص واختبار الخوارزمية من خلال استخدام بيانات التدريب. وبشكل عام تشكل معدلات الاختبار في الأنظمة الخبيرة المعيار الأساسي لجودة العمل؛ إِذ إِنَّ الأَنظمة التي تنتج معدلات عالية في الاختبارات تعدُّ أكثر معنوية فيما يخص الأَداء، والجدول رقم (10) يمثل نتائج الاختبار للنظام الخبير المقترح.

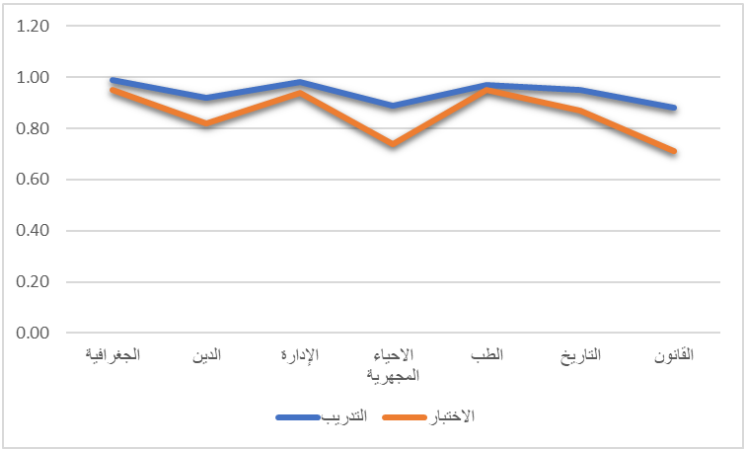
جدول (10): دقة التصنيف لبيانات الاختبار (Testing)

ت	الاختصاص	دقة التصنيف
1	الجغرافية	0.95
2	الدين	0.82
3	الإدارة	0.94
4	الاحياء المجهرية	0.74
5	الطب	0.95
6	التاريخ	0.87
7	القانون	0.71
	معدل الأَداء	0.86

وكما هو مبين في الجدول آنفًا، أَنَّ المعدل العام للتصنيف الذي يقدمه النظام الخبير المقترح يأتي عند معدل (86%) وهي نسبة تعد مقنعة إِلى حد بعيد، ولاسيَّما وان البيانات التي نتعامل معها تعد قصيرة من حيث عدد الكلمات الأَمر الذي ينتج عنه خصائص أَقل بالمقارنة مع استخدام التصنيف مع كم أكبر من النصوص العربية.

ويمكن القول إِنَّ النظام الخبير يستطيع أَن يصنف الرسائل الجامعية اعتماداً على العناوين المكتوبة باللغة العربية نسبة جيدة جداً. ولاسيَّما إذا ما أَخذنا بنظر الاعتبار التعقيدات التي تتسم بها اللغة العربية من حيث المعالجة المبدئية. فضلًا عن ذلك وعلى الرغم من انخفاض نسبة الاختبار مقارنة بنسبة التدريب، إِلَّا أَنَّنا يمكن ان

نلاحظ ان النظام الخبير يتسم بنوع من الاستقرار فيما يخص الأَداء العام وإذا ما اعتمدنا الأَداء لكل اختصاص (فئة) من الاختصاصات التي تمثل قاعدة بيانات البحث، وكما هو موضح في الشكل التالي.



شكل (3): الأَداء العام للنظام الخبير (التدريب والاختبار)

3-5 تقييم مدى تحقق جودة النظام الخبير المقترح:

نقدم في هذا الجزء من البحث تقييماً للنظام الخبير المقترح، إِذ تم بناء استبانة لغرض توزيعها على عينة من المختصين في علم المعلومات والمكتبات بغية تقييم النظام، وتم توضيح عمل النظام وعرض نتائج التصنيف للعينة وعلى وفق ذلك، تم تحليل نتائج الاستبانة الَّذي تم إِدراجه في ملحق رقم (1). وفيما يلي من فقرات نوضح عرضاً تفصيلياً لتحليل آراء عينة من المختصين فيما يخص عمل النظام الخبير؛ إِذ تم اعتماد متغيرين مستقلين في الاستبانة وهما (دقة النظام الخبير واهميته)، أَمَّا المتغير المعتمد فقد تم اعتماد الفائدة المتوخاة من النظام كمتغير معتمد.

3-5-1 وصف العينة:

تُعدُّ عملية اختيار عينة الدراسة من أَبرز الخطوات التي يجب أَخذها بعين الاعتبار في البحوث العلمية لما لها من أثر على نتائج الدراسة وتحديد المشاكل

وتحقيق الاهداف، وجرى اختيار العينة وفق مقتضيات البحث، وإجراء الاستبانة على عينة عشوائية من المتخصصين بمجال المعلومات والمكتبات فضلًا عن عدد من المختصين في مجال الحاسوب وتكنولوجيا المعلومات؛ إذ اشتملت العينة على 65 فردًا، جدول (11) يبين وصفاً تفصيلًا لخصائص العينة الَّذي يوضح ما يأتي:

1. إنَّ نسبة الذكور المشاركين في الاستبانة تفوق نسبة الإناث وبنسبة (54%) من مجموع المشاركين في الدراسة.
2. على صعيد التحصيل الدراسي، كان المشاركون الأكثر هم من حملة شهادة الماجستير وبنسبة (57%) من مجموع المشاركين في عينة الدراسة.
3. بالنسبة للفئة العمرية، كانت الفئة (36 فأكثر) تشكل ما نسبته (37%) وهو الفئة العمرية التي شكلت النسبة الأعلى من عينة الدراسة.

جدول (11): البيانات الوصفية لعينة

المعلومات التعريفية	الفئة	العدد	النسبة المئوية (%)
الجنس	الذكور	35	54
	الإناث	30	46
	المجموع	65	100
التحصيل الدراسي	دبلوم عال	7	11
	ماجستير	37	57
	دكتوراه	21	32
	المجموع	65	100
العمر	25-20	4	6
	30-26	17	26
	35-31	20	31
	36 فأكثر	24	37
	المجموع	65	100

3-5-2 عرض البيانات وتحليلها وتفسيرها وفقاً لإجابات العينة:

في هذا القسم من البحث نقوم بعرض إِمكانيَّة التحقق من صحة الفرضيات الموضوعة للدراسة وذلك بدراسة القوة التأثيرية ومعاملات الارتباط لكل من المتغيرات المستقلة والمتغير المعتمد، والجدول (2) يعرض إِجمالي إِجابات العينة على الأَسئلة المقدمة في الاستبانة الإِحصائيَّة.

جدول (12): إِجابات عينة الدراسة وفقا لمحاور الاستبانة

الأَسئلة حسب المحاور		توزيع إِجابات افراد عينة الدراسة (درجة مقياس)					المتوسط المرجح	الانحراف المعياري	النسبة المئوية
		اتفق تماما	اتفق	اتفق لحد ما	لا اتفق	لا اتفق تماما			
الفائدة المتوخاة من النظام الخبير	Y1	21.00	23.00	16.00	0.00	0.00	4.08	0.79	81.66
	Y2	7.00	40.00	10.00	3.00	0.00	3.85	0.69	77.00
	Y3	13.00	30.00	16.00	1.00	0.00	3.92	0.74	78.34
	Y4	12.00	33.00	15.00	0.00	0.00	3.95	0.68	79.00
	Y5	18.00	29.00	12.00	1.00	0.00	4.07	0.76	81.34
	المجموع	71.00	155.00	69.00	5.00	0.00	3.97	0.73	79.47
الأَهميَّة	X11	24.00	24.00	11.00	1.00	0.00	4.18	0.79	83.66
	X12	8.00	37.00	14.00	1.00	0.00	3.87	0.65	77.34
	X13	6.00	36.00	17.00	1.00	0.00	3.78	0.64	75.66
	X14	10.00	26.00	18.00	5.00	1.00	3.65	0.92	73.00
	X15	10.00	27.00	17.00	2.00	4.00	3.62	1.03	72.34
	المجموع	58.00	150.00	77.00	10.00	5.00	3.91	0.82	78.18
الدقة	X21	20.00	28.00	10.00	1.00	1.00	4.08	0.85	81.66
	X22	10.00	37.00	11.00	2.00	0.00	3.92	0.70	78.34
	X23	4.00	36.00	16.00	4.00	0.00	3.67	0.71	73.34
	X24	6.00	36.00	16.00	2.00	0.00	3.77	0.67	75.34
	X25	13.00	29.00	15.00	2.00	1.00	3.85	0.86	77.00
	المجموع	53.00	166.00	68.00	11.00	2.00	3.90	0.81	77.97

وبحسب الجدول آنفًا فيما يخص محور القائدة المتوخاة من النظام الخبير، نرى أَنَّ شدة الإِجابة جاءت عند مستوى (79.47) وبانحراف معياري قدره (0.73)، اما

محور الأَهميَّة فكانت شدة الإِجابة عند مستوى (78.18) وبانحراف قدره (0.82). اما فيما يخص محور الدقة فكانت شدة الإِجابة عند مستوى (77.97) وبانحراف قدره (0.81). ومن خلال ملاحظة الجدول آنفًا يمكن القول إِنَّ محاور الاستبانة والإِجابة عليه قد عكس طابعاً جيداً لدى عينة الدراسة وكونه يُسهم إِلى حد كبير في توظيف تقنيات النظم الخبيرة في خدمة مختلف التخصصات العلمية.

1.3 القوة التأثيرية للمتغيرات المستقلة على المتغير المعتمد:

لغرض اختبار فرضيات النظام تم احتساب مدى معنوية تأثير المتغيرات المستقلة على المتغير المعتمد. جدول (13) يوضح تفاصيل هذه التأثيرات الَّذي يمكن شرحها كما يأتي:

1. بالنسبة للمتغيرات المستقلة (الأَهميَّة، والدقة) فإِنَّ لها جميعا تأثيراً معنوياً على المتغير المعتمد (الفائدة المتوخاة من النظام الخبير). وهذه المتغيرات تعكس أَهميَّة النظام ودقته في تقديم المعلومة الجيدة للمتلقي من وجهة نظر عينة الدراسة. الأَمر الذي يبرز أَهميَّة عمل النظام الخبير في مجال تصنيف المعلومات.

جدول (13): تفاصيل القوة التأثيرية للمتغيرات المستقلة على المتغير المعتمد

المتغيرات المستقلة	المتغير المعتمد	معامل التفسير R^2	القيمة المحتسبةF	القيمة الجدوليةF	المعلمات	القيمة المحتسبةt	القيمة الجدوليةt	النتيجة
أهميَّة النظام الخبير	الفائدة المتوخاة	19.60%	14.1	4.09	0.38	3.755	2.002	معنوية
دقة النظام الخبير	الفائدة المتوخاة	11.90%	7.84	4.09	0.36	2.80	2.002	معنوية

4.3 تحليل نماذج الانحدار

نلاحظ من الجدول (13) أَنَّ قيمة (F) المحسوبة (14.1) أكبر من الجدولية (4.09) بمستوى (0.95) ودرجة حرية (1) فيما يتعلق بتأثير المتغير أَهميَّة النظام الخبير. وهذا يعني قبول الفرضية الاولى الَّتي تعني وجود تأثير معنوي للمتغير المستقل

(أهميَّة النظام الخبير) على المتغير المعتمد (الفائدة المتوخاة من النظام الخبير). كما أنَّ معامل التفسير (19.60%) يدل على ان هذه النسبة فقط هي من تأثير المتغير المستقل والباقي اي (81.40%) يعود إلى متغيرات أُخرى وهذا ما يثبت قوة المتغير وتأثيره الإيجابي على النتائج.

كما ونلاحظ من الجدول آنفًا ان قيمة (F) المحسوبة (7.84) أكبر من الجدولية (4.09) بمستوى (0.95) ودرجة حرية (1) فيما يتعلق بتأثير المتغير دقة النظام الخبير. وهذا يعني قبول الفرضية الثانية الَّتي تعني وجود تأثير معنوي للمتغير المستقل (دقة النظام الخبير) على المتغير المعتمد (الفائدة المتوخاة من النظام الخبير). كما أنَّ معامل التفسير (11.90%) يدل على أنَّ هذه النسبة فقط هي من تأثير المتغير المستقل والباقي اي (88.01%) يعود إلى متغيرات أُخرى، وهذا ما يثبت قوة المتغير وتأثيره الإيجابي على النتائج.

6-3 الاستنتاجات والتوصيات:

بعد تنفيذ مقترح النظام الخبير وعرض نتائجه، يقدم البحث في ضوء ذلك عددًا الاستنتاجات، وكما يأتي:

- يمكن بناء نظام خبير لتصنيف الرسائل الجامعية اعتمادًا على العناوين العربية لها.
- تمتاز اللغة العربية بعدد من التحديات التي تواجه عملية تحليل النصوص العربية نظراً لطبيعة اللغة الشكلية فضلًا عن امتلاكها للحركات الإعرابيَّة الَّتي تعد ميزة فريدة في اللغة العربية.
- تقدم خوارزمية (NB) نتائجاً جيدة فيما يخص الأَداء العام للتصنيف.
- يمتاز النظام الخبير المقترح بقدرته على تصنيف العناوين العربية وبنسبة جيدة.
- يمتاز النموذج المقترح بعدد من الخصائص الجيدة من وجهة نظر عدد من المختصين.

وفي ضوء نتائج البحث، نقدم في القسم التالي عددًا من التوصيات والاتجاهات المستقبلية للبحث وكما يأتي:

- **استخدام خوارزميات أُخرى للتصنيف ومقارنة النتائج.**
- **توسيع قاعدة البيانات لتشمل اختصاصات أُخرى وتقييم الأَداء.**
- **اختبار دقة الخوارزمية على ملخصات البحوث بدلا من العناوين ومقارنة النتائج.**
- **استخدام طرائق أُخرى لاستخلاص خصائص النصوص ومقارنتها مع النتائج الحالية.**

المصادر العربية:

1. عبد الجليل، بوداح، استخدام الأنظمة الخبيرة في مجال اتخاذ قرار منح القروض البنكية، (أطروحة دكتوراه). جامعة منتوري- الجزائر، (2017).
2. عبد الحسين، وجدان رشيد، نظام خبير هجين لتشخيص امراض القمح باستخدام المنطق المضبب، الشبكات العصبية وطريقة بايسن. مجلة علوم ذي قار، 5(2)، 80-87، (2015).
3. الشهري، احمد بن حسين، نظم المعلومات وتكاملها مع النظم الخبيرة. الفكر الشرطي، 78(376)، 1-81، (2012).
4. الصعيدي، محمد الشناوي والشرقاوي، جمال مصطفى وإبراهيم، رشا احمد، أثر تصميم نظام خبير تعليمي على تنمية مهارات إنتاج المقررات الإلكترونيَّة لدى طلاب الدراسات العليا بكلية التربية. مجلة دراسات عربية في التربية وعلم النفس (64)، 205-239، (2015).
5. الفقي، عبد الاله إبراهيم، الذكاء الاصطناعي والنظم الخبيرة. القاهرة: دار الثقافة والنشر والتوزيع. (2012).
6. القرة غولي، عفاف سامي، النظم الخبيرة والذكاء الاصطناعي وإمكانيَّة استخدامها في المكتبات ومراكز المعلومات. عمان: دار صفاء للنشر والتوزيع. (2005)
7. بسيوني، عبد الحميد، الذكاء الاصطناعي والوكيل الذكي. القاهرة: دار الكتب العلمية للنشر والتوزيع، (2005).
8. حسن، محمد المتولي علي وعبد الحميد، عبد العزيز طلبة و عفيفي، طارق محمد احمد، بناء نظام خبير للمساعدة في إنتاج المقررات الإلكترونيَّة. مجلة بحوث التربية النوعية، (36)، 943-970، (2014).

9. رشيد، زينب عبد اللطيف وقاسم، عمر صابر، بناء نظام طبي خبير باستخدام نظام مهجن بين المنطق المضبب والشبكات العصبية الاصطناعية مع التطبيق. المجلة العراقية للعلوم الإحصائيَّة، 13(25)، 474-487، (2013).
10. عفانة، سونا عبد الله إبراهيم، دور الأنظمة الخبيرة في تنمية وتطوير أداء الموظفين في مؤسسات القطاع الخاص الفلسطيني/رام الله. (ماجستير). جامعة القدس، (2012).
11. عفيفي، جهاد، الذكاء الاصطناعي والنظم الخبيرة. عمان – الاردن: دار أمجد للنشر والتوزيع، (2014).
12. عيسى، رهام مصطفى النظم الخبيرة "النظام الخبيرة – المعرفة – النظم الخبيرة والاستدلال". مصر: دار التعليم الجامعي، (2016).
13. مهدي، عامر، تصميم وتنفيذ نظام خبير مضبب لتشخيص امراض الغدد الصماء باستخدام خوارزمية الطيـــور. مجلة جامعة ذي قار العلمية، 9(3)، 1-14، (2014).

References in English

14. **The** ,Bodah ,Jalil- Abdel **Use of Expert Systems in the Decision-Making Field of Granting Ban L**Mentouri .(ishest PhD) , **oans** . (Algeria, (2017 -ty rsiUnive
15. **A ,**(2015)Wijdan Rashid ,Hussein- Abdel **Hybrid Expert System for Diagnosing Wheat Diseases Using fuzzy Logic** neural networks and , ., 87-Payson's method. Dhi Qar Sciences Journal, 5(2), 80
16. **and their stemsformation syIn ,**(2012)bin Hussein medAh ,Shehri-Al **with integration Expert Systems** 81-1 ,(376)78 ,**tional ThoughtCondi .** .,
17. **gence and Expert Artificial Intelli** .(2012Ilah Ibrahim-Abd El ,Feki-Al .e of Culture, Publishing and Distributionairo: HousC .**Systems**
18. ,Sami Afaf , Qara Gul-Al(2005**) Expert Systems and A rtificial Iand the ntelligence Pof ossibility Using them in Libraries and . Information Centers** ication and Amman: Dar Safaa for publ distribution.
19. **Artificial** .(2005)Abdel Hamid , iouniBass **Iand the ecnegilletn A gent** Book House for Publishing and Cairo: Scientific . **The smart one** , Distribution

20. Aziz -Abd al ,Hamid-Abd al Mutwali Ali and-Muhammad al ,Hassan **n Expert Building a** ,2014) ,ariq Mohamed AhmedT ,Afifi And Tolba Journal of . **Electronic Courses uction ofSystem to Assist in the Prod** .(970-Specific Education Research, (36), 943
21. an expert Latif and Qasim, Omar Saber, Building-Zainab Abdel ,Rashid gic and artificial on fuzzy lbrid system betweemedical system using a hy tion. Iraqi Journal of Statistical Sciences, neural networks with applica .(2013) , 487-474 ,(25)13
22. **The Role of Expert ,**(2012)and Wana Abdullah Ibrahim .Afana, S **es in theDeveloping the Performance of Employe Systems in** Quds -lRamallah. (Master's). A / **an Private Sector InstitutionslestiniPa** ,University
23. Amman **.Artificial Intelligence and Expert Systems** ,(2014)Hadd .Afifi, J , for Publishing and Distribution Amjad Jordan: Dar -
24. **-Systems xperts, "ESystem Expert** .(2016)Ham Mostafa .Issa, R University :ptEgy ." **and Inference Expert Systems -Knowledge** ,Education House
25. **Design** ,. (2014)Amer ,Mahdi **A ndimplementation of an Expert Fussy System for Diagnosing Endocrine Diseases Using the Avia Algorithm** , 14-urnal, 9(3), 1entific Joiniversity ScDhi Qar U **.**

المصادر الاجنبية

26. Al-Saleem, S. (2010). Associative classification to categorize Arabic data sets. International Journal of ACM Jordan, 1(3), 118-127 .
27. Alsayat, A., & Elmitwally, N. (2020). A comprehensive study for Arabic sentiment analysis (Challenges and Applications). Egyptian Informatics Journal, 21(1), 7-12.
28. Elnagar, A., Al-Debsi, R., & Einea, O. (2020). Arabic text classification using deep learning models. Information Processing & Management, 57(1), 102121 .
29. Ghanem, O. A., & Ashour, W. M. (2012). Stemming effectiveness in clustering of Arabic documents. International Journal of Computer Applications, 49 .(5)
30. Ibrahim, M. F., Alhakeem, M. A., & Fadhil, N. A. (2021). Evaluation of Naïve Bayes Classification in Arabic Short Text Classification. Al-Mustansiriyah Journal of Science, 32(4), 42-50.

31. Mostafa, A. M. (2017). An evaluation of sentiment analysis and classification algorithms for Arabic textual data. Int. J. Comput. Appl, 158(3), 29-36.
32. Oussous, A., Benjelloun, F.-Z., Lahcen, A. A., & Belfkih, S. (2020). ASA: A framework for Arabic sentiment analysis. Journal of Information Science, 46(4), 544-559.
33. Sharef, B. T., Omar, N., & Sharef, Z. T. (2014). An automated Arabic text categorization based on the frequency ratio accumulation. Int. Arab J. Inf. Technol., 11(2), 213-221 .
34. Zahidi, Y., El Younoussi, Y., & Al-Amrani, Y. (2021). Different valuable tools for Arabic sentiment analysis: a comparative evaluation. International Journal of Electrical & Computer Engineering (2088-8708), 11(1).

1. Abdul Jalil, Budah, the use of expert systems in the field of decision-making for granting bank loans, (PhD thesis). University of Montour-Algeria, (2017).
2. Abdul-Hussein, wajdan Rashid, a hybrid expert system for diagnosing wheat diseases using fuzzy logic, neural networks and the paysen method. Journal of Dhi Qar Sciences5 ،(2) ،87-80 ،(2015).
3. Ahmed bin Hussein, information systems and their integration with expert systems. Conditional thought78 ،(376) ،81-1 ،(2012).
4. Al-Saidi, Mohammed Al-Shennawi and Al-Sharqawi, Jamal Mustafa and Ibrahim, Rasha Ahmed, the impact of designing an educational expert system on the development of electronic course production skills among graduate students at the Faculty of Education. Journal of Arab Studies in education and psychology (64) ،239-205 ،(2015).
5. Elfeki, Abdellah Ibrahim, artificial intelligence and expert systems. Cairo: House of culture, publishing and distribution. (2012).
6. Qura gholi, Afaf Sami, expert systems, artificial intelligence and the possibility of their use in libraries and information centers. Amman: Safa publishing and distribution house. (2005)
7. Bassiouni, Abdul Hamid, artificial intelligence and intelligent agent. Cairo: House of scientific books for publishing and distribution, (2005).
8. Hassan, Mohammed Al-Metwally Ali and Abdul Hamid, Abdul Aziz Talaba and Afifi, Tarek Mohammed Ahmed, building an expert system to help in the production of electronic courses. Journal of qualitative education research ، (36) ،970-943 ،(2014).
9. Rashid, Zainab Abdullatif and Qasim, Omar saber, building an expert medical system using a hybrid system between fuzzy logic and artificial neural networks with application. Iraqi Journal of statistical sciences13 ،(25) ،487-474 ، (2013).
10. Afana, Sona Abdullah Ibrahim, the role of expert systems in the development and development of employee performance in the Palestinian private sector institutions/Ramallah. (Master). Al-Quds University, (2012).

11. Afifi, jihad, artificial intelligence and expert systems. Amman-Jordan: Amjad publishing house, (2014).
12. Isa, Reham Mustafa expert systems "expert system-knowledge-expert systems and inference". Egypt: House of university education, (2016).
13. Mehdi, Amer, design and implementation of a precise expert system for the diagnosis of endocrine diseases using the bird algorithm. Dhi Qar University scientific journal9 ،(3) ،14-1 ،(2014).

ملحق رقم (1)

م/ استبانة

السادة الباحثون المحترمون

تحية طيبة...

نضع بين ايديكم استمارة الاستبانة وانطلاقاً من ثقتنا بتعاونكم معنا في انجاز هذا البحث العلمي (بناء نظام خبير لتصنيف الرسائل الجامعية باستخدام خوارزمية (Naïve Bayes): دراسة تجريبية) من خلال إسهامكم بملء الاستمارة وفق الفقرات المثبتة فيها، اذ ان نجاح البحث مرهون بدقة اجابتكم، لذا يرجى اعتماد الدقة والموضوعية في تأشير الإجابة المعبرة فعلاً عن الواقع. على امل أن يخرج هذا البحث بنتائج تخدم البحث العلمي في هذا القطاع الحيوي.

شاكرين لكم حسن تعاونكم معنا

المعلومات الشخصية:

التحصيل الدراسي ☐ **دبلوم** ☐ **بكالوريوس** ☐ **دبلوم عالي** ☐ **ماجستير** ☐ **دكتوراه**

الجنس: ☐ **ذكر** ☐ **انثى**

العمر: ☐ **25-20** ☐ **30-26** ☐ **35-31** ☐ **36 فأكثر**

اولاً: الأَهميَّة:

ت	الفقرات	اتفق تماما	اتفق	محايد	لا اتفق	لا اتفق تماما
1	يقدم النظام أَهميَّة نسبية في تصنيف البيانات النصية					
2	تساهم الأنظمة الخبيرة في تذليل العقبات وتسهيل الإجراءات					
3	استخدام ادوات الأنظمة الخبيرة من شانها رفع كفاءة الأداء للمنظمات والافراد					
4	بنية النظام الخبير واداوته سهلة الاستخدام					
5	سرعة تصنيف البيانات الضخمة تساهم في تنفيذ من المهام التصنيفية المعقدة					

ثانيا: الدقة:

ت	الفقرات	اتفق تماما	اتفق	محايد	لا اتفق	لا اتفق تماما
1	يمكن من خلال النظام الحصول على دقة بيانات جيدة					
2	يمكن تفعيل النظام للاستفادة من عدد اكبر من البيانات					
3	يوفر النظام الخبير ما يحتاجه المستفيد من عملية التصنيف					
4	العلاقة بين الأدوات المستخدمة والطرائق البرمجية تسمح بزيادة دقة البيانات					
5	سهولة استخدام النموذج تساهم في زيادة دقة البيانات المنتجة					

ثالثا– الفائدة المتوخاة من النظام الخبير

ت	الفقرات	اتفق تماما	اتفق	محايد	لا اتفق	لا اتفق تماما
1	هل تعتقد ان النظام الخبير حقق الغاية المرجوة من استخدامه					
2	هل تؤمن بان النظام الخبير ممكن ان يضيف أهميَّة علمية					
3	هل يحقق النظام الهدف الذي صمم من أجله					
4	يمكن ان يكون النظام الخبير مساهماً بشكل فاعل في تصنيف البيانات					
5	تصنيف البيانات باستخدام النظام الخبير يمكن ان يرفد المؤسسات بالعديد من المميزات					

Building an expert system for classifying university theses using the Naïve Bayes algorithm: an empirical study

Inas Jassim Hadi *

Abstract

Data classification has become a widespread field of knowledge, especially in light of the tremendous development of information technology and expert systems. Among the most important fields of knowledge that are concerned with data classification is the classification of text data. In our research, we present a model for an expert system whose task is to classify university theses and theses based on their Arabic titles and using the Naïve Bayes algorithm. In the research, a database was used, which was compiled using the (Scrapping) technique, which is to collect data from repositories spread on the Internet. A database of (6100) titles for a master's thesis and a doctoral thesis was built. These titles were divided into (seven) categories representing the general specialization of the thesis - thesis. Among the most important challenges faced by the research is the nature of the Arabic language and its degree of complexity. The results of the research proved that the model was able to classify theses and theses at a rate of (86%), which is a very promising percentage. The proposed model was also evaluated by distributing a questionnaire to a number of specialists in computer science and information systems. The questionnaire included measuring the impact of two independent variables, namely, the accuracy and importance of the expert system. The results proved that there is a statistically significant effect of the independent variables on the dependent variable (the desired benefit of the expert system). The results of the research open new horizons in the process of classifying textual data, especially the Arabic language, and open the way towards implementing other algorithms and comparing the results.

Key words: Arabic text classification, expert systems, Naïve Bayes algorithm**.**

* **Lect/ College of Arts / Al-Mustansiriya University.**