



توظيف المعالجات الذكية لتسخير محتوى معجم "خزانة الأدب ولبّ لباب لسان العرب لتوسيع الذخيرة اللغوية: فضاء الألوان أنموذجاً

حسن مظفر الرزوي¹

¹المعهد العالمي لحوسبة القرآن الكريم والعلوم الإسلامية

¹halrizzo@gmail.com

الخلاصة:

أسهم الحضور اللافت لآليات المعالجات الذكية التي تستثمر القدرات الهائلة للحواسيب، في لفت انتباه العاملين في مجال اللسانيات المحوسبة الى توظيف هذه القدرات في عملية التنقيب بمحتوى المعاجم والقواميس، وإعادة تشكيلها وفق أنماط مستحدثة، لتوليد بيانات تفاعلية تتسم بانفتاحها على علوم اللغة العربية بجميع مجالاتها، نوحاً، وصرفاً، ودلالة، وسياًقاً من خلال استثمار المعالجات الذكية التي باتت قادرة على ممارسة جميع أدبيات علوم اللغة العربية دفعة واحدة على النص، لتعود الى ممارسة سلسلة من عمليات إعادة تشكيل مفردات هذه العلوم، وقواعدها، وشواهداها، لتودعها في بيئة تفاعلية ذكية يمكن استثمارها في توليد نصوص جديدة، أو الإجابة عن التساؤلات، أو التنقيب في المحتوى المعرفي للمادة العلمية الخصبية التي تزخر بها معاجم لغتنا العربية وقواميسها.

أهداف البحث: معالجة محتوى معجم خزانة الأدب ولبّ لباب لسان العرب الذي ألفه اللغوي عبد القادر بن عمر البغدادي (1030-1093 هـ)، والذي يعد موسوعة لغوية شاملة تميزت بـ ذخيرة لغوية ثرية، واحتوت على مفردات الكلم العربي، وأمثال العرب، مع بيان معانيها ومضاربيها وأصولها، بواسطة خوارزميات الذكاء الاصطناعي المحوسب لإعادة هيكلتها، وفق منظور معالجات اللغة الطبيعية باستخدام حزمة من الخوارزميات الذكية التي ستمارس عمليات التحاليل النحوي، والصرفي والدلالي المحوسب لمحتوى المعجم، وتحويلها الى فضاء معرفي. محوسب مفتوح يتيح للمستخدم التنقيب بالمادة المعرفية واللسانية المودعة في هذا العمل الجليل.

الأسئلة التي سيجيب عليها: ما هي طبيعة الفرص المتاحة أمام معالجات الذكاء التوليدي في إعادة هيكلة محتوى المعاجم العربية لإنتاج فضاء دلالي متعدد الأبعاد يمكن استثمار مادته لممارسة سلسلة متنوعة من عمليات التنقيب اللغوي واللساني التي ستشمل بناء مدونة لغوية محوسبة Corpus، يمكن استثمارها في دعم ذخيرة النماذج اللغوية الكبيرة Large Language Models، وإنشاء نماذج ذكية توليدية لتوليد النصوص، ومراجعة تعدد معاني مفردات الخزانة الأدب ولبّ لباب لسان العربي، وتتبع المرادفات اللغوية، ومسائل لغوية أخرى ذات صلة بعلوم اللغة العربية مما سيتمنح المستخدم العربي فضاءً معرفياً من نمط جديد.

منهج البحث: استخدمت خوارزميات محوسبة في تفكيك مادة نص المعجم الى وحدات أولية، ثم مورست عمليات التحليل الصرفي، والنحوي، والدلالي المحوسب على مادة النص مع وسم الكيانات اللغوية التي استوطنت النص (أعلاماً، وأماكن وبلدان) مع إجراء تحليل دلالي محوسب لمسألة حضور الألوان في نص المعجم، والفضاءات الدلالية التي حضرت من خلالها في مادته اللغوية لبيان الفرص المتاحة أمام إنشاء مستويات متعددة من الفضاءات الدلالية التي يمكن أن ينهل من لبابها المعرفي المستخدمون بمختلف توجهاتهم ومستوياتهم المعرفية.

الخلاصة والتوصيات: أظهرت المعالجات الذكية وجود أكثر من فرصة لإعادة تشكيل مادة المحتوى اللغوي واللساني الذي أودع في المعاجم وقواميس اللغة العربية لإنتاج فضاء معرفي محوسب، متعدد المستويات يمكن أن يستثمر كمورد خصب على صعيد الدراسات النحوية، والصرفية، والدلالية. مع إمكانية استثمار مادتها في إنشاء نماذج لغوية ضخمة LLM لدعم نماذج الذكاء المحوسب التوليدي التي ستصاغ لثوابت وقواعد لغتنا العربية، وستسهم في تطوير البحث اللساني وفتح آفاق جديدة أمام آليات تعليم اللغة العربية وعلومها.

الكلمات الجوهرية: معالجة اللغة الطبيعية، الذخيرة اللغوية، التحليل النحوي والصرفي المحوسب، الفضاء اللوني.

1. المقدمة

أوجب التحالف الجديد بين اللسانيات اللغوية والحوسبة الذكية مسألة إعادة تشكيل النص التقليدي بأدوات معالجات اللغة الطبيعية لتجاوز عقبة تمثيل المعرفة البشرية المودعة في النصوص وتضمينها ضمن القوالب التي باتت تفرضها مختلف أنماط المعالجات المحوسبة الذكية. لقد حصلت طفرة مفاهيمية بالتعامل مع مفردات النص نتيجة للخطاظة المعرفية الجديدة New Paradigm التي تشكلت من تكامل المعالجات الذكية، مع أدوات اللغة الطبيعية، والمبادئ التي أرستها اللسانيات المحوسبة خلال السنوات الخيرة بعد ولادة الذكاء التوليدي Generative A.I.

فبرزت من رحم هذه الخطاظة المعرفية مفاهيم جديدة، ارتكزت الى سلسلة من معالجات تفكيك مفردات النص التقليدي الى وحدات أولية أطلق عليها اصطلاح Token، والتي أضحت مادة خصبة تتعامل معها الخوارزميات المحوسبة، بالسبر والتحليل، للكشف عن الأنماط النحوية، والصرفية، والدلالية التي أودعت بمفردات النص، والكشف عن طبيعة العلاقات التي تربط فيما بينها ضمن السياقات التي وردت فيها هذه المفردات.

وقد شهدنا خلال السنوات الثلاث الأخيرة ولادة المدونات المحوسبة التي استمدت ذخيرتها اللغوية من نواتج معالجات اللغة الطبيعية للنص، التي باتت توفر لنا كما هائلاً غير مسبوق من مادة التحليل المحوسب التي باتت تعالج كل مفردة من مفردات النص بأنماط لسانية متعددة، تتباعد في التنقيب عن المعاني، والدلالات، وتشكيل شبكات المفاهيم، على التوازي مع التحليل النحوي، والصرفي، والدلالي لكل مفردة من مفردات النص وبزمن قياسي غير مسبوق.

ولقد سعينا في هذه الدراسة الى توظيف حزمة من المعالجات اللسانية الذكية لتحليل محتوى احد اهم معاجم اللغة العربية ومدوناتها اللغوية، ألا وهو معجم خزانة الأدب ولب لباب لسان العرب لعبد القادر بن عمر البغدادي ، فجعلنا مادته مورداً لسلسلة من المعالجات المحوسبة الذكية التي قامت بتفكيك النص الى وحدات أولية، تم تشكيلها في مدونة محوسبة لمادة

المعجم، ثم باشرنا سلسلة من التحليلات اللسانية الذكية على المحتوى لكي تتحول مادته المعجمية الى نسق معرفي جديد، وقد حاولنا أن نعرض بعض الفرص التي يمكن أن يوفرها للعاملين في مجال العلوم اللسانية للغتنا العربية، والتتقيب في أعماق أكبر داخل الخطاب اللغوي المودع في هذا المعجم الثري، واستنتج مفرداته في مجال حضور الألوان ودلالاتها اللغوية بطريقة محوسبة نأمل أن تسهم في فتح آفاق جديدة لكي ننهل من المعاني والدلالات المودعة في مفردات لغتنا العربية الخصبة، ونسخر مادتها مورداً لتعليم اللغة العربية.

2. معجم خزانة الأدب ولبّ لسان العرب: المصنّف والكتاب:

ولد مصنّف هذا المعجم اللغوي الكبير، عبد القادر بن عمر بن بايزيد بن الحاج أحمد البغدادي في مدينة بغداد عام 1030 هـ، وبعد إقامة دامت بضعة عقود في مدينة أدرنه التركية انتقل الى مصر عام 1091 هـ قبل أن توافيه المنية فيها عام 1093 هـ¹. وقد برع في علوم اللغة العربية، حتى ذهب البعض الى عدّه زعيم اللغويين في القرن العاشر الهجري². وكانت له أياد بيضاء في شرح شواهد اللغة العربية أودعها في مصنّفاته³ التي تجاوزت بضعة عشر معجم ورسالة أغنى بها المكتبة العربية اللغوية.

ويعدّ كتابه خزانة الأدب ولبّ لسان العرب أعلى موسوعة في علوم العربية وآدابها، وذلك لما أودع في هذا المعجم من نصوص نادرة حفظ لنا بها تراثاً لغوياً خصباً من كتب فقدت أو اندثرت⁴، مع حرص بالغ بالنقد والتحقيق لما أودعه في سفره الجليل الذي بلغ عدد أجزائه 13 جزءاً⁵. وقد بدأ بتأليف الكتاب عام 1073 هـ بمصر المحروسة وانتهى منه عام 1079 هـ كما ذكر في خزانته⁶.

وقد حصلنا على النسخة الرقمية لجميع أجزاء هذا السفر اللغوي المهم، من المستودع الرقمي للمكتبة الشاملة⁷، حيث عمدنا الى تحويله من صيغة نصوص الشاملة. Filename.bok الى صيغة نصية Text File لكي يكون مورداً للمعالجات المحوسبة التي تمارسها الخوارزميات البرمجية الذكية.

¹ . البغدادي، عبد القادر بن عمر البغدادي، خزانة الأدب ولبّ لسان العرب، تحقيق عبد السلام محمد هارون، الطبعة الرابعة، (القاهرة: مطبعة الخانجي، 1997)، الجزء الأول: 11.

² . محمد، أسماء أبو بكر، عبد القادر البغدادي زعيم اللغويين في القرن العاشر الهجري، (بيروت: دار الكتب العلمية، 1413 هـ)، ص: 69.

³ . الطباخ، محمد راغب، عبد القادر البغدادي مؤلف خزانة الأدب الكبرى، مجلة الزهراء، الجزء 4-5، القاهرة، شوال 1347 هـ، الصفحات 209-211.

⁴ . محمد، أسماء أبو بكر، عبد القادر البغدادي زعيم اللغويين في القرن العاشر الهجري، (بيروت: دار الكتب العلمية، 1413 هـ)، الصفحات: 70-72.

⁵ . سرد في مقدمة كتابه المراجع التي استقى منها مادة الخزانة والتي تجاوزت 945 مصتفاً، منها في علم النحو، ومنها ما هو في شرح الشواهد، ومنها ما هو في تفسير أبيات المعاني المشكّلة، ومنها ما يلتحق بديوان شعر العرب، ومنها ما يرجع الى كتب الأدب، ومنها ما يرجع الى كتب السير، وكتب الصحابة وأنساب العرب، وطبقات الشعراء والأدباء، ومنها ما يتعلق بطيف واسع من علوم اللغة العربية، وكتب الأمثال، وكتب الأماكن والبلدان.

⁶ . البغدادي، عبد القادر بن عمر البغدادي، خزانة الأدب ولبّ لسان العرب، تحقيق عبد السلام محمد هارون، الطبعة الرابعة، (القاهرة: مطبعة الخانجي، 1997)، الجزء الأول: 21.

⁷ . تتوفر نسخة الكتاب بجميع أجزائها على موقع المكتبة الشاملة: <https://shamela.ws/book/12732>.

3. معجم خزانة الأدب ولب لباب خزانة الأدب ولب لباب لسان العرب: إعادة تشكيل المحتوى:

ابتكرت معالجات اللغة الطبيعية المحوسبة (Natural Language Processing (NLP) لتحقيق آمال العاملين في مجال اللسانيات المحوسبة والذكاء الاصطناعي لتمنح الحواسيب فرصة سبر الخطاب البشري، وتفسير معانيه، وإنتاج سياق الخطاب بطريقة تقارب محتوى الخطاب التواصلية الذي يستخدمه الإنسان المعاصر، وذلك عن طريق تجسير الفجوة القائمة بين اللغة والخوارزميات المحوسبة التي تستخدم في ميادين الذكاء الاصطناعي Artificial Intelligence والذكاء التوليدي Generative A.I.

بصورة عامة، تتضمن عملية استثمار معالجات اللغة الطبيعية المحوسبة في إعادة تشكيل محتوى النص التقليدي وهيكله مفرداته بحيث تتوافق مع البيئة الذكية التي تعمل بدائرتها الخوارزميات المحوسبة Computational Algorithms من خلال سلسلة من المعالجات التي تتألف من مجموعة من المراحل. وتتضمن المرحلة الأولى عملية تفكيك محتوى النص الى وحدات أولية Tokenization يتم تحديد هويتها من خلال المسافة البادئة التي تفصل بين الكلمة والكلمة التي تليها. ويطلق على كل منها اسم وحدة أولية Token.

ووفق الخطاطة المعرفية لمعالجات اللغة الطبيعية فإن الخوارزميات المحوسبة في تعاملها مع هذه الوحدات الأولية لا تأخذ بعين الاعتبار البعد المعرفي لحضورها على المستوى النحوي، أو الصرفي، أو الدلالي، وتعتبرها كتلة لسانية صماء ينصب اهتمام الخوارزميات على إحصاء مستويات تكرارها بالنص، والموقع الذي تستوطنه في نسيج مادته.

ولتجاوز عقبة كثرة المفردات اللغوية المودعة بالنصوص، والتي تورث الخوارزميات البرمجية، والحواسيب، المزيد من الجهد الرياضي والمنطقي المحوسب يصار الى استبعاد الوحدات الأولية التي لا تمتلك معنى قائماً بذاتها (كالحروف، وحروف الجر، والضمائر، وغيرها) والتي يطلق عليها اصطلاح كلمات التوقف Stop-Words من قائمة تعدّ سلفاً لاستبقاء المفردات التي تحمل معانٍ، وتؤثر على التحليل المحوسب الدلالي.

تأتي بعد ذلك مرحلة المعالجات اللغوية المحوسبة، والتي تتضمن التحليل النحوي المحوسب للنص، ثم التحليل الصرفي المحوسب، وأخيراً التحليل الدلالي المحوسب. فتمارس الخوارزميات المحوسبة في الشق الأول تحليل وتحديد أقسام الكلام لمحتوى النص، أما في الشق الثاني فتمارس الخوارزميات مهمة تحليل المفردات عن طريق إزالة اللواحق (السوابق واللواحق) لبلوغ جذر الكلمة وجذعها، وإحصاء جميع تفاصيل البيانات الصرفية في النص. أما الشق الدلالي فتمارس الخوارزميات المحوسبة مهمة كشف اللبس الدلالي الذي يصاحب حضور الكلمة الواحدة في سياقات متعددة.

وتعالج مخرجات التحليلات المحوسبة الثلاث بسلسلة من المعالجات الرياضية لبيان مستويات جذور الكلمات وجذوعها، وما يرتبط بها من أنماط التي تحصل على الجذر باستخدام الحروف والأصوات لتغيير المعنى وتحديده والتي تمارس دوراً مهماً في إنتاج المعاني في مختلف مستويات السياقات التي استخدمها المؤلف في إنتاج مادة النص. بعدها بوشر بعملية وسم هوية الكيانات Named Entity Recognition والتي تمارس بواسطة نظم تعلم بواسطة الشبكات العصبونية Neural Networks لوسم أسماء الأعلام، والبلدان، والأماكن، وأية كيانات أخرى لغرض إلحاقها بالمدونة الرقمية واستثمارها في معالجات الذكاء المحوسب.

ثم تأتي بعد ذلك معالجات التمهيد لإنتاج بيانات الذكاء التوليدي حيث تبرز معالجات تضمين الكلمات Word Embeddings لتوليد فضاء دلالي متعدد الأبعاد يحتوي كل بعد من أبعاده الدلالة التي حملتها المفردة في السياقات المتكاثرة التي وردت فيها في الخطاب اللغوي المودع في خزانة الأدب ولبّ لباب لسان العرب، والتي تعد من المؤشرات المهمة على الذخيرة اللغوية التي امتلكها عبد القادر بن عمر البغدادي من جهة، وفرص استثمارها في توليد كلمات واصطلاحات مرادفة، ومتلازمات لفظية، مع توفر فرصة سانحة للكشف المعاني من خلال الارتباط الدلالي بين المفردة وبقية المفردات الموجودة في نص خزانة الأدب ولبّ لباب لسان العرب، واختصار مقاطع محددة من النص، أو سبر الفضاء الدلالي للمفردات والاصطلاحات المقيمة في نص مدونة هذا المعجم.

4. المستخرجات اللسانية من النص المحوسب:

استخدمت البيئة البرمجية للغة بايثون كحاضنة برمجية لمجموع الخوارزميات المحوسبة التي أعدت لممارسة حزمة المعالجات اللسانية تحت مظلة معالجات اللغة الطبيعية بمختلف مراحلها وبدعم من المكتبات البرمجية المتوفرة في هذه البيئة البرمجية الذكية مثل مكتبات: NLTK, STANZA, CAMEL, SpaCy, Gensim & Pytorch، وغيرها من المكتبات التي تحفل بها تطبيقات هذه اللغة الذكية⁸.

واعتمد مبدأ إنشاء مدونة محوسبة⁹ Corpus من مادة معجم خزانة الأدب ولبّ لباب لسان العرب لكي تكون مورداً للعاملين في مجال الذكاء المحوسب، بعد أن تعاد هيكلتها بواسطة مجموعة متنوعة من الخوارزميات الذكية التي ستمنحها عمقاً معرفياً من نمط جديد سيتيح لنا فرصة التوغّل في أعماق البنيات النحوية، والصرفية، والدلالية لمادة خزانة الأدب ولبّ لباب لسان العرب مع وسم هوية الكيانات المقيمة في فضائه اللغوي، وسبر الفضاء الدلالي لكل كلمة وردت فيه، لكي تكون مورداً معرفياً خصباً، لإنتاج المعاني، واستخلاص المفاهيم¹⁰.

4. 1. مستخرجات تفكيك محتوى خزانة الأدب ولبّ لباب لسان العرب:

اعتمدت خوارزمية محوسبة لتفكيك محتوى خزانة الأدب ولبّ لباب لسان العرب وتحويله الى وحدات لغوية أولية Tokens، وذلك باعتماد وجود الفراغ بين كلمة وأخرى دليلاً للتفرقة بين كل وحدة لغوية وبين الوحدة التي تليها. ثم بوسرّت عملية استبعاد كلمات التوقّف من الوحدات الأولية نظراً لعدم امتلاكها معنى قائماً بذاتها، كما أنها تشكّل عبئاً على موارد الخوارزميات المحوسبة.

⁸ . Bird, Steven, Klien, Ewan and Edward Loper, *Natural Language Processing with Python*, (Cambridge: O'Reilly, 2009), pp.179,221,261,291,327,361,407.

⁹ . تعرف المدونة بأنها عبارة عن نص أو مجموعة من النصوص المحوسبة التي تنتهي الى حقل من حقول المعرفة، انتخبت لكي تعبّر عن الذخيرة اللغوية والاصطلاحية التي يراد سبر عناصرها في هذه النصوص، ووظفت أدوات الحوسبة الذكية لمعالجة وإدارة محتواها، وتحليل عناصرها نحوياً، وصرفياً ودلالياً، ووسمت كياناتها لتمييز هويتها، وتحديد انتماءاتها.

¹⁰ . Dilts, Robert, *Modeling with NLP*, (California: Meta Publications, 1998), pp.29-32.

وقد احتسب عدد تكرارات كل فئة من فئات الوحدات الأولية، والكلمات المستبعدة، وكذلك الوحدات الأولية المتبقية بعد عملية إزالة كلمات التوقف. أنظر شكل 1.



شكل 1: مادة المدونة المحوسبة لمعجم خزانة الأدب ولبّ لباب لسان العرب.

بداية أعلن الشكل عن السمات التي تألف منها محتوى الكتاب الذي بلغ عدد كلماته 968,888 كلمة، توزعت على 13 مجلداً وبواقع 11,317 فقرة. أما بالنسبة لخصائص العامة للمدونة المحوسبة التي تم تشكيلها من محتوى خزانة الأدب، بتوظيف معالجات اللغة الطبيعية، فقد بلغ عدد الوحدات الأولية 102,423 وحدة ناهز مجموع تكراراتها 968,888، تم استبعاد 346,130 كلمة توقف Stop Words عن دائرة التحليل لعدم امتلاكها لمعان مؤثرة على الخطاب اللغوي.

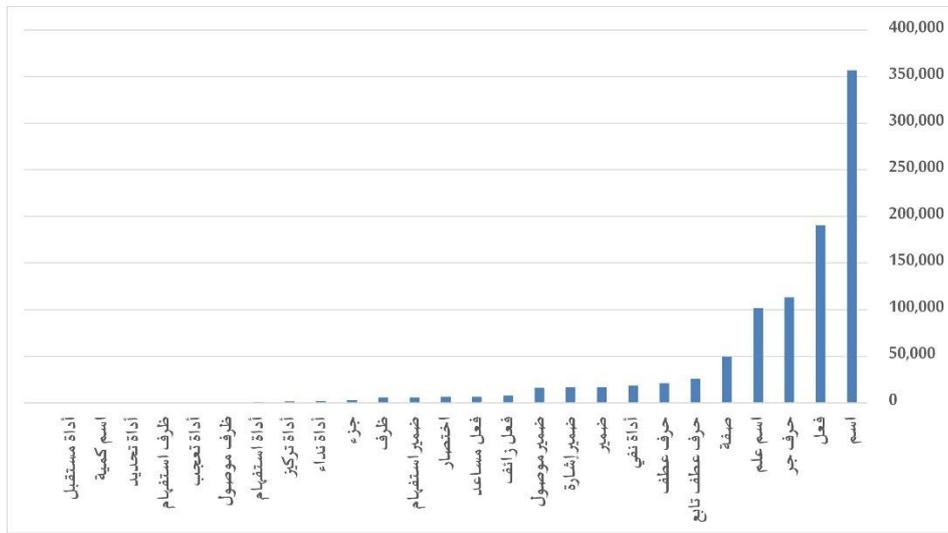
بالمقابل أثمرت معالجات التحليل النحوي المحوسب عن تشخيص 27 قسماً من أقسام الكلام، تباينت مستويات حضورها بحسب السياقات اللغوية التي تألف منها محتوى الخزانة. أما التحليل النحوي المحوسب فقد استطاع بعد إزالة اللواحق بكلماتها (السوابق واللواحق) عن الكشف عن 3,589 جذراً لغوياً، تنوّعت بنياتها الصرفية، وفي الوقت ذاته تم تشخيص 30,541 كياناً (أسماء أعلام، وأماكن وبلدان) أوردها البغدادي عند شرح الشواهد اللغوية ضمن خزانة الأدب.

4. 2. مستخرجات التحليل النحوي والصرفي والدلالي لمحتوى خزانة الأدب ولبّ لباب لسان العرب:

بوشرت عملية التحليل النحوي، والصرفي، والدلالي المحوسب لمحتوى معجم خزانة الأدب ولبّ لباب لسان العرب بواسطة مجموعة من الخوارزميات البرمجية الذكية. في البداية باشرنا بمعالجات التحليل الصرفي المحوسب بواسطة مجموعة أخرى من الخوارزميات البرمجية، التي حرصت على إزالة اللواحق (السوابق واللواحق) من المفردات للظفر بالجذور التي اشتقت منها المفردات، مع تحديد جذوع الكلمات، ثم أحصت عدد الجذور اللغوية التي وظّفها عبد القادر بن عمر البغدادي في خزانة معجمه لإنتاج السياق والمعاني، مع بيان أكثر الجذور اللغوية حضوراً في مادة خطابه اللغوي. أنظر شكل 2.

إن تنوع الجذور في معجم خزانة الأدب يمكن أن تكون له آثار واضحة على صعيد: تنوع المعاني، واستخدام الجذر في اشتقاقات متعددة مما يزيد ثراء اللغة، مع تعميق ظاهرة التنوع النحوي، مما يمنح مستخدم المعجم فرصة جيدة لانتخاب الجذر المناسب الذي يعبر عن المعنى الذي يريده في خطابه، فيعمق الإبداع اللغوي لدى المستخدم، ويمنحه مرونة إضافية في سبابة الكلمات بمختلف أنماط السياقات.

ثم مورست عملية التحليل النحوي المحوسب لتحديد أقسام الكلام في الخطاب اللغوي لعبد القادر بن عمر البغدادي، فتوزعت هذه المفردات على سبعة وعشرين فئة نحوية، تباينت أعداد حضورها بالنص بحسب السياقات التي ناقش فيها عبد القادر بن عمر البغدادي، الشواهد اللغوية في خزانته. أنظر شكل 4.



شكل 4: أقسام الكلام التي تم الحصول عليها من التحليل النحوي المحوسب لنص خزانة الأدب ولبّ لباب لسان العرب.

ويبدو واضحاً أن فئة الأسماء، هي الأكثر حضوراً، بينما تأتي بعدها فئة الأفعال، فالحروف، فأسماء الأعلام. وستوفر البيانات التي أنتجتها عملية التحليل النحوي المحوسب فضاءً بنوياً يمكن أن يمارس فيه المتخصصون في مجال اللسانيات وبالخصوص في مجال تحديد الأنماط النحوية السائدة بالنص وتوجهات استخداماتها في السياق اللغوي، الذي تضمن مجموعة متنوعة من الشواهد اللغوية التي امتدت على بعد زمني بلغ عدة قرون. وتشخيص الأخطاء اللغوية في النص، أو أنماط الشذوذ التي قد تظهر في هذا الشاهد وذلك. كما يمكن أن تستثمر هذه التحليلات بالكشف عن البنيات النحوية التي يمكن أن تدعم الذين يرومون تعلم اللغة العربية ممن لا ينطقون بها.

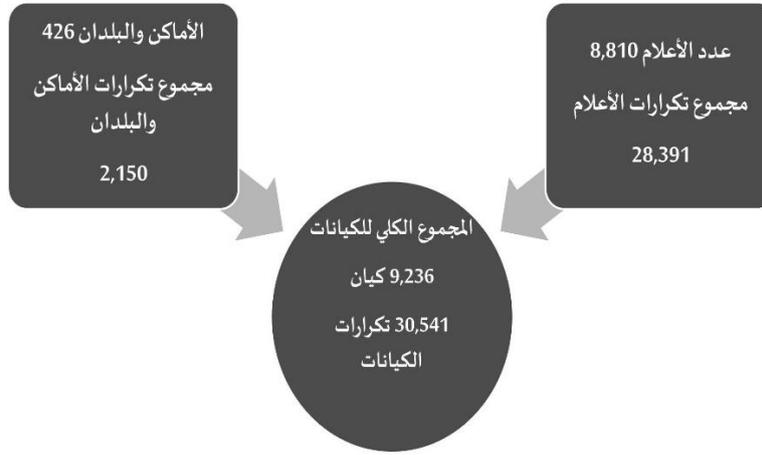
أما على صعيد التحليل الدلالي فقد برزت عقبة كشف اللبس الدلالي الذي صاحب حضور الكلمة في مختلف أنماط السياقات التي استخدمت في نص خزانة الأدب. وللتغلب على هذه العقبة فقد سخرنا مجموعة من الخوارزميات البرمجية التي وفّرت لنا قائمة تفصيلية بجميع اشتقاقات الجذور، والمعنى المحمول الذي يمكن أن تعبر عنه المفردة في هذا السياق، أو ذلك. أنظر الجدول 1.

ويظهر بجلاء من هذا الأنموذج المختصر، ثراء الذخيرة اللغوية التي امتلكها عبد القادر بن عمر البغدادي، وكيف حرص على استخدام مفردات متنوعة بحسب السياق الذي عالج فيه معاني الكلمات المودعة في خزانة الأدب ولبّ لباب لسان العرب، وسعيه للإفصاح عن معانيها ودلالاتها.

4. 3. مستخرجات وسم الكيانات المستودعة بخزانة الأدب ولبّ لباب لسان العرب:

ولا بد للمدونة المحوسبة أن تكتمل من خلال الكشف عن هوية الكيانات Named Entity Recognition سواء كانت هذه الكيانات تعبر عن أسماء أعلام، أو بلدان، أو أماكن، أو منظمات.

وقد وظفت خوارزميات التحليل العصبوني Neural Networks للتعقب عن حضور هذه الكيانات، وتمييز هويتها، وإحصاءها، ووسمها لكي تكون مادة خصبة يمكن أن تستثمر بواسطة معالجات الذكاء المحوسب. أنظر شكل 5:



شكل 5: الكيانات (أعلام وأماكن وبلدان) المقيمة في خزانة الأدب ولبّ لباب لسان العرب والتي وسمت بواسطة الخوارزميات المحوسبة.

5. استثمار المدونة المحوسبة في دراسة دلالة حقل الألوان في خزانة الأدب ولبّ لباب لسان العرب:

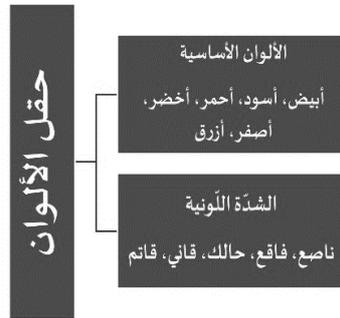
لعل من أبرز الفوائد المترتبة عن صناعة المدونة المحوسبة لمعجم خزانة الأدب ولبّ لباب لسان العرب وغيره من المعاجم، هو إعادة تشكيل المحتوى وفق أنساق رياضية ومنطقية توفّر للباحثين فرصة التنقيب في مادته والكشف عن المعاني والدلالات المودعة في مادة خطابه، وكشف عن شبكة المفاهيم التي أودعها المصنف في معجمه.

وتعد مسألة حضور ألفاظ حقل الألوان من المسائل المهمة التي تستأثر باهتمام الباحثين. لأن مفردات الألوان واصطلاحاتها، تتكاثر استخداماتها في وصف كل ما يحيط بنا، حيث نعبر بواسطتها عن معان ودلالات مختلفة. من أجل

هذا فقد اهتم بها أصحاب المعاجم والقواميس فخصصوا لها حضوراً في معاجمهم للكشف عن الكثير من خصائص ألفاظها، ومشتقاتها، مع بيان دلالاتها الحقيقية والرمزية¹².

بالمقابل فإن مسألة تعامل المعاجم مع ألفاظ الألوان، ودراسة الصيغ التي صبّت فيها مشتقاتها، ومستويات حضور معانيها، والتباين في دلالاتها، تعدّ من المسائل التي شغلت اهتمام العاملين في مجال اللسانيات والمعاجم¹³. ونظراً للخبرة اللغوية، والصرفية، والدلالية التي توفرت لنا من خلال المعالجة المحتوى المحوسب لمعجم خزانة الأدب ولبّ لباب لسان العرب، فقد وجهنا أنظارنا الى دراسة هذه المسألة مستثمرين الخوارزميات الذكية، وقدرتها على التنقيب في محتوى نص هذا المعجم، للتنقيب عن حضور المفردات اللونية فيه، وتحديد مواطنها، ومستويات حضورها، وارتباطاتها مع ألفاظ الألوان المختلفة، أو الألفاظ الأخرى التي تلازم حضورها مع مفردات الألوان.

ولبيان ما يمكن أن توفره لنا المدونة المحوسبة للمعجم، والخوارزميات البرمجية من مجال لساني لدراسة هذه المسألة، والتي ستشدد اهتمام الكثير من الباحثين في حقول معرفية متعددة، للاعتراف من الذخيرة اللغوية والدلالية التي استوطنت في هذا المورد، حاولنا مراجعة أهم المحاور التي وردت فيها ألفاظ الألوان الأساسية ومشتقاتها، مع توجيه الاهتمام نحو الألفاظ التي اهتمت ببيان درجات الألوان وشدتها، وفق معالجة تبنت مفردات النهج المحوسب في عرض نتائجها، تاركين للمتخصصين في مجال حقل الألوان، ومورفولوجيتها، ودلالات ألفاظها فرصة تفسير ما أنتجته المعالجات الذكية من بيانات.



شكل 6: مجال المعالجات المحوسبة في حقل الألوان بمعجم خزانة الأدب ولبّ لباب لسان العرب.

وقدّ وجهنا دفة الخوارزميات الذكية للتنقيب عن الجذور اللغوية التي استمدت منها ألفاظ الألوان الأساسية، وتلك التي استخدمت لبيان شدتها. فنجحت هذه الخوارزميات في عملية استقصاء حضور الجذور ومشتقاتها، لكل شق من شقّي حقل الألوان الأساسية.

¹² . عبيد، كلود، الألوان: دورها، تصنيفها، مصادرها، رمزيها، ودلالاتها، الطبعة الأولى، (بيروت: المؤسسة الجامعية للدراسات والنشر والتوزيع، 2013)، ص:10.

¹³ . زيادة، وفاء، حقل الألوان في اللغة العربية: بين الدلالة المورفولوجية والدلالة المعجمية، مجلة كلية الآداب، جامعة القاهرة، المجلد (81)، العدد (4)، أبريل 2021، الصفحات: 302.

ويظهر في الجدول 2 مستويات حضور ألفاظ حقل الألوان وأوصاف منتخبة من ألفاظ شدة الألوان، مع أعداد مشتقات ألفاظ هذه الألوان.

الجدول 2: . مستويات حضور ألفاظ حقل الألوان ومشتقاتها.

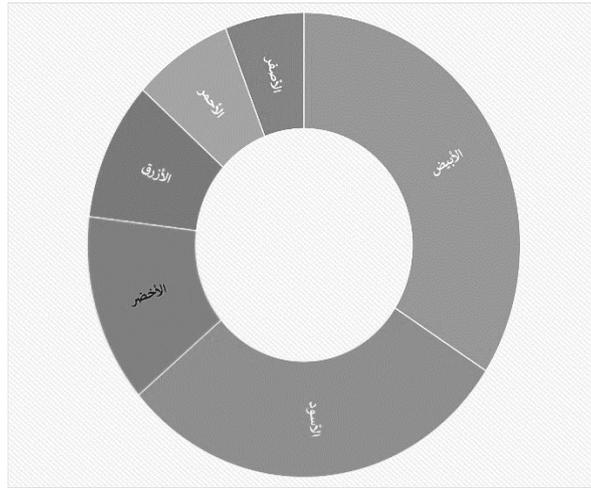
اللون	عدد المشتقات	المشتقات
لون	28	لونها / لونتته / واللون / للونها / ألواني / الألوان / ألوانهم / لون / بلونه / لونين / اللون / ألوانهن / وألوانا / لوانان / للألوان / لونه / ألوان / وألوان / لوني / ألوانه / ألوانا / ولون / ألوانها / وألوانكم / ولونك / كلون / لونا / للون.
صبغة	9	يصبغن / وصبغ / أصبغ / تصبغ / صبغ / ويصبغ / فصبغت / يصبغ / صبغت.
أسود	24	كأسود / سوداء / والسود / وأسودين / والأسود / سودا / للأسود / وسود / أسودان / أسودين / السود / بالأسود / أسود / وسودا / بأسود / بسود / وأسود / السوداء / وسوده / الأسود / سوده / كالأسود / بالسود / سود.
أبيض	12	وأبيض / الأبيض / البيضاء / وبيضاء / بأبيض / بأبيضهم / ببيضاء / لأبيض / والأبيض / وأبيضهم / أبيضهم / أبيض.
أحمر	28	حمرها / بأحمر / الحمر / بحمر / وأحمر / والحمر / كأحمر / الحمراء / وحمراء / وأحمرينا / وكأحمر / حمرا / والأحمر / الأحمر / وأحمر / أحمر / حمرة / بحمره / أحمرين / حمري / حمراء / أحمر / أحمره / وحمرة / أحمر / حمر / وحمرة / فأحمر.
أصفر	8	صفراء / بصفراء / وأصفر / أصفر / وأصفري / أصفرت / والصفراء / الأصفر.
أخضر	6	الأخضر / وأخضر / أخضر / خضراء / أخضرا / الخضراء.
أزرق	12	أزرق / والأزرق / أزرق / والزرق / وزرق / زرقه / زرقاء / الزرق / الزرقاء / زرق / الأزرق / وزرقاء.
الشدة	عدد المشتقات	المشتقات
ناصح	2	الناصعة / ناصع.
قاني	1	قان.
فاقع	1	والفواقع.
قاتم	5	للقاتم / قاتما / لقاتم / قاتم / وقاتم.
حالك	3	والحالك / الحالك / حالكا.

ويبدو واضحاً مما ورد في الجدول أعلاه، حضور ألفاظ الحقل اللوني بكثافة في المعجم، مع ثراء الذخيرة اللغوية لعبد القادر بن عمر البغدادي في معالجه لمعاني المفردات ذات الصلة بالحقل اللوني وحضورها في مدونة اللغة العربية.

ولاستكمال الشوط في توجيه دفة الخوارزميات المحوسبة في التنقيب عن مسائل أخرى تكشف النقاب عن مسألة الحقل اللوني وألفاظه في المعجم، سعينا الى الكشف عن ظاهرة المتلازمات والمتصاحبات اللفظية لحقل الألوان في المعجم، وكانت نتيجة المعالجات اللغوية التي مارستها الخوارزميات الذكية، البيانات التي أوردناها في كل من شكل 7، وشكل 8.



شكل 7: أعداد المتلازمات اللفظية للألوان في كتاب خزانة الأدب.



شكل 8: مستويات حضور المتلازمات اللفظية للألوان في معجم خزانة الأدب.

بصورة عامة، تعلن المتلازمات اللفظية عن ميل كلمة معينة الى الحضور مع كلمات أخرى في السياق، وبشكل متكرر. وقد ينشأ التلازم اللفظي عن إنشاء تعابير اصطلاحية، مع توسيع دائرة المعاني والدلالات من خلال إنشاء تعابير جديدة تعزز الذخيرة اللغوية.

لقد نجحت الخوارزميات الذكية في النقاط مجموعة من المتلازمات اللفظية من دائرة حقل ألفاظ الألوان، في المعجم، والتي تشكل ذخيرة لغوية جيدة، لتوظيف التعابير اللونية في سياقات متعددة، كما وستوفر للمستخدم فرصة توسيع مفرداته بالمزيد من التعابير الاصطلاحية التي شاع استخدامها في الخزانة الأدب ولبّ لباب لسان العربي، كما أنها ستشكل قاعدة للتعابير الاصطلاحية يمكن لنظم الذكاء الاصطناعي التوليدي استثمارها في إنتاج نصوص جديدة تتسم بمستوى بلاغي رصين في ميزان بلاغة الخطاب العربي.

ويمكن أن نستثمر هذه المتلازمات اللفظية واصطلاحاتها في توسيع دائرة الفضاء الدلالي¹⁴ لكل كلمة اشتقت من ألفاظ الألوان الستة، والتي يمكن أن تستخدم في نظم الذكاء الاصطناعي التوليدي، حيث يمكن أن تنتج نصوص جديدة من الذخيرة اللغوية لمعجم خزانة الأدب ولبّ لباب لسان العرب، أو تستخدم كمورد لغوي ودلالي يمكن أن توظف مفرداته والدلالات المتنوعة التي استوطنت فيها باختصار النصوص، وإنتاج المرادفات، وتوسيع دائرة النطاق الدلالي للتعبير اللونية المختلفة.

6. سبل استثمار معالجات الذكاء الاصطناعي في دعم دراسة اللسانيات وتعليم اللغة العربية:

بات واضحاً لدينا من المخرجات التي حصلنا عليها من استخدام الخوارزميات الذكية لمعالجات اللغة الطبيعية على نص معجم خزانة الأدب ولبّ لباب لسان العرب، أن هذه المعالجات المحوسبة قد أثرت عن إنتاج المدونة المحوسبة لهذا السفر اللغوي الجليل، مع قدرتنا على ممارسة التحليل النحوي، والصرفي، والدلالي على مفردات النص، مع الكشف عن هوية الكيانات التي أوردها المصنف (أعلام، وأماكن وبلدان)، يمكن أن تشكل ذخيرة لغوية ضخمة للعاملين في مجال اللسانيات وعلوم اللغة العربية، بالإضافة الى توفيرها لمورد ثري يمكن أن يسهم في دعم تعليم اللغة العربية، والارتقاء بالمهارات اللغوية للناطقين بها.

وقد حاولنا أن ندرج في الجدول 3 أهم الفوائد التي يمكن أن نجنيها من توظيف أدوات الذكاء الاصطناعي في معالجات اللغة الطبيعية، وإنشاء مدونة محوسبة لخزانة الأدب، حيث ستؤلف مادتها مورداً مهماً وذخيرة لغوية ومعرفية للعاملين في مجال لسانيات اللغة العربية، والذين يعكفون على إنشاء تطبيقات لتعليم اللغة العربية، وإثراء الذخيرة اللغوية لدى الناطقين بها، كذلك يمكن أن تستثمر في ضبط محتوى Fine Tuning النماذج اللغوية الكبيرة Large Language models وتطبيقات الذكاء التوليدي Generative A.I.

الجدول 3: . سبل استثمار مخرجات التحليل المحوسب لمحتوى معجم خزانة الأدب ولبّ لباب لسان العرب.

المعالجة المحوسبة	الوصف	المنافع المترتبة في مجال اللسانيات وتعليم اللغة العربية	فرص دعم النماذج اللغوية الكبيرة والذكاء التوليدي
التجزئة Tokenization	تفكيك محتوى خزانة الأدب الى وحدات لغوية أولية Tokens مثل الكلمات والاصطلاحات.	توفر فرصة لفهم بنية الجملة، واستخدام الكلمات والاصطلاحات، ومدخلاً الى تحليل السياق اللغوي.	توفير مجال لتحليل مستوى تكرار الكلمات في النص، وأنماط وتوجهات استخداماتها.
التحليل النحوي المحوسب	تحليل البنية النحوية لجمل المعجم، وتحديد أقسام الكلام السائدة فيه، والتي تمهّد وسبل إعرابها، وسبر القواعد لتحليلات لسانية معمقة.	تدعم المتعلمين على فهم البنات المعقدة للجمل العربية، وسبل إعرابها، وسبر القواعد النحوية التي تحكمها.	توفير أمثلة وشواهد للتطبيقات التي ستوجه المتعلمين اللغة العربية عن الحالات النحوية المختلفة.

14 . يستخدم في دائرة الذكاء الاصطناعي التوليدي اصطلاح الفضاء الدلالي للكلمة Word Space الذي يعبر عنه بمتجه Vector متعدد الأبعاد، يفصح كل بعد من أبعاده في فضاء المعنى، عن معنى تمتلكه الكلمة في السياق الذي ترد فيه. ويسهم تعدد متجهات دلالة الكلمة الواحدة عن تعدد المعاني التي يمكن أن تحملها في سياقات متعددة، حيث تستثمر هذه الدلالات في توليد نصوص جديدة.

المعالجة المحوسبة	الوصف	المنافع المترتبة في مجال اللسانيات وتعليم اللغة العربية	فرص دعم النماذج اللغوية الكبيرة والذكاء التوليدي
التحليل الصرفي المحوسب	دراسة جذوع الكلمات وجذورها الصرفية، وطبيعة السبائك الصرفية المستخدمة في الخطاب اللساني.	تعزيز الظفر بذخيرة لغوية تسترشد بالبنيات الصرفية التي استثمرت الجذور اللغوية لاشتقاق المزيد من المعاني والدلالات.	توليد تدريبات لغوية ولسانية توجه اهتماماتها نحو الأنماط والسبائك الصرفية وانعكاساتها على الفضاء الدلالي.
التحليل الدلالي المحوسب	سير المعاني والروابط الدلالية بين الكلمات وانعكاساتها على الفضاء الدلالي للسياق.	تساعد على فهم الفروق الدلالية وآثارها على محمولات المعاني والتي تنعكس إيجاباً على مهارات الكتابة.	توفر فضاءً دلاليًا يزيح آثار اللبس الدلالي عن المعاني المحمولة للكلمة في السياق وتعد مورداً مهماً لدعم الذكاء التوليدي في توليد نصوص متوافقة مع الخطاب الإنساني.
تحديد هوية الكيانات	تقوم بتشخيص هوية الكيانات (أعلام، أماكن وبلدان) الموجودة في النص وتصنيفها لتوفير فرصة لبحوث معمقة حول موارد النص، وتاريخية عناصره اللغوية.	فسح المجال أمام الدراسات اللسانية التي تعنى بتاريخية النص، وزمن ولادة المصطلح، وهوية المصنفين الذين لهم صلة بالموضوع.	توفر فضاءً محوسباً لوصم هوية الكيانات دون الحاجة الى تدخل بشري، وبزمن قصير لا يكاد يقارن بالجهد البشري المطلوب، كما توفر قواعد بيانات يمكن استثمارها في إجابات نظم الذكاء التوليدي.
المتلازمات اللفظية	تحديد هوية الكلمات التي تلازم حضورها مع بقية الكلمات، وأنماط التكرار بحسب السياق.	تعزيز قدرة المتعلم على إنتاج معانٍ متعددة وبمستويات بيانية مختلفة، كما توفر للمشتغلين في مجال اللسانيات فرصة للاطلاع على طبيعة أنواع التلازم اللفظي في هذا السفر اللغوي الجليل.	توفير فضاء واسع للمعاني أن يمكن أن توفرها نماذج اللغة الكبيرة لنظم الذكاء التوليدي، فتجعل السياق اللغوية المتولد أكثر قرباً من الخطاب الإنساني مع سيادة بلاغة مميزة في سياقاته.

إن ما عرضناه من بيانات في الجداول والأشكال الرسومية لا يكاد يشكل 1% من حجم البيانات الضخمة التي تولدت لدينا بواسطة الخوارزميات البرمجية الذكية، والتي أدرجناها ضمن قواعد بيانات يمكن أن تتكامل مادتها مع المعالجات التي تسري في النماذج اللغوية الكبيرة، والتي ستشكل مورداً وذخيرة لغوية ثرية يمكن أن تستثمر في تطبيقات، وآليات معالجة تدعم العاملين في مجال لسانيات اللغة العربية، من جهة، والقائمين على تدريس علوم اللغة العربية لغير الناطقين بها، وتعزيز الملكة اللغوية وترسيخ أركانها لدى الناطقين بها.

7. خلاصة:

بات واضحاً لدينا أن خوارزميات الذكاء المحوسب باتت تمتلك القدرة على التنقيب في محتوى المعاجم اللغوية العربية، والبحث المفردات اللغوية التي تستوطن فيها. كما أن هذه الخوارزميات قد نجحت في ممارسة سلسلة من التحليلات اللسانية المحوسبة التي شملت التحليل النحوي، والصرفي، والدلالي لمادة المحتوى مع وجود أكثر من فرصة لتحليل تدفق المعاني وتعدد الدلالات في حقول معرفية مختلفة، وبزمن قياسي لا يمكن أن يقارن بما يمكن أن يستغرقه المتخصص في العلوم اللسانية لكي يبلغ ما نجحت هذه الخوارزميات الذكية في تحقيقه.

كذلك فإن إعادة هيكلة محتوى معجم خزانة الأدب ولبّ لباب لسان العرب لعبد القادر بن عمر البغدادي ، ووسم هوية كياناته اللغوية قد أسهم في تحويله الى مادة خصبة يمكن لأدوات الذكاء الاصطناعي المحوسب أن تنتج عنه مجموعة متنوعة من الفضاءات الدلالية الخصبة، لتلبية احتياجاتنا في سير حقول معرفية محددة، وبمستويات تتباين في شمول المعالجة، وتوليد الاصطلاحات والتعبير التي ستحافظ على مفردات لغتنا العربية، وتتيح لنا أكثر من فرصة لإنتاج خطاب لغوي عربي رصين، حتى وإن غابت الفرصة عنا في قراءة ودراسة هذه المعاجم وما تحويه من ذخيرة لغوية خصبة.

مصادر الدراسة:

[1]

البغدادي، عبد القادر بن عمر البغدادي، خزانة الأدب ولبّ لباب لسان العرب، تحقيق عبد السلام محمد هارون، الطبعة الرابعة، (القاهرة: مطبعة الخانجي، 1997).

[2]

زيادة، وفاء، حقل الألوان في اللغة العربية: بين الدلالة المورفولوجية والدلالة المعجمية، مجلة كلية الآداب، جامعة القاهرة، المجلد (81)، العدد (4)، أبريل 2021، الصفحات: 301-365.

[3]

الطباخ، محمد راغب، عبد القادر البغدادي مؤلف خزانة الأدب الكبرى، مجلة الزهراء، الجزء 4-5، القاهرة، شوال 1347 هـ.

[4]

Bird, Steven, Klien , Ewan and Edward Loper , **Natural Language Processing with Python** , (Cambridge: O'Reilly ,2009).

[5]

Dilts, Robert , **Modeling with NLP** , (California: Meta Publications , 1998).

[6]

عبيد، كلود، الألوان: دورها، تصنيفها، مصادرها، رمزيتها، ودلالاتها، الطبعة الأولى، (بيروت: المؤسسة الجامعية للدراسات والنشر والتوزيع، 2013).

جدول الألفاظ

English	عربي
Abstract	الملخص
Intelligent Processes	المعالجات الذكية
Content	المحتوى
Dictionary	معجم
Linguistic Repository	الذخيرة اللغوية
Colors Semantic Space	فضاء الألوان
Key words	الكلمات الجوهرية
E-mail	البريد الإلكتروني
References	المراجع
Mining	التقيب
Computational Linguistics	اللسانيات المحوسبة
Computational Intelligence	الحوسبة الذكية
Generative A.I.	الذكاء التوليدي
Paradigm	الخطاطة المعرفية
Token	وحدة لغوية أولية
Natural Language Processing	معالجات اللغة الطبيعية
Tokenization	التجزئة الى الوحدات الأولية
Semantic Disambiguation	اللبس الدلالي
Named Entity Recognition	وسم هوية الكيانات
Neural Networks	الشبكات العصبونية
Intelligent Algorithms	الخوارزميات الذكية
Corpus	مدونة
Morphological Analysis	التحليل الصرفي
Syntactic Analysis	التحليل النحوي
Semantic Analysis	التحليل الدلالي
Roots	الجدور
Stems	الجنوع
Collocates	المتلازمات اللفظية
Fine Tuning	ضبط المحتوى
Large Language Model	الأنموذج اللغوي الكبير

السيرة الذاتية للباحث

عضو أقدم في المعهد العالمي لحوسبة القرآن الكريم والعلوم الإسلامية. تنصّب اهتماماته في مجالات: اقتصاد المعرفة، وتطبيقات الحوسبة الذكية في مجال العلوم الإسلامية، وتحليل النصوص. قام بإعداد 101 دراسة وتقريراً تقنياً لمنظمات دولية وعربية، وجهات حكومية و48 دراسة جدوى فنية واقتصادية لمشاريع استراتيجية. شارك بـ 110 بحثاً في مؤتمرات علمية دولية وعربية ووطنية، ولديه 147 بحث علمي ومعلوماتي وإسلامي نشر في مجلات علمية محكمة دولية وعربية ومحلية. وساهم بأربعة مشاريع دولية وعربية متخصصة في مجال اقتصاد المعرفة، والحفاظ على اللغة العربية، وسلاسل القيم الاقتصادية. لديه 28 كتاب مطبوع في مجالات اقتصاد المعرفة، وتطبيقات الحوسبة الذكية في مجال المعرفة الإسلامية، والأمن السيبراني، والجاهزية الرقمية. وقد رشحت إحدى كتبه ضمن أفضل 10 كتب على صعيد بناء الدول العربية من قبل جائزة الشيخ زايد الدولية للكتاب.



Employing Intelligent Processing to Harness the Content of the Lexicon "Khizanat al-Adab wa-Lubb Lubab Lisān al-'Arab" for Expanding Linguistic Repositories: A Case Study of the Semantic Space of Colors

Hasan M. Al-Rizzo

International Computing Institute of Quran & Islamic Sciences, IRAQ
halrizzo@gmail.com

Abstract:

This study focuses on processing the content of *Khizanat al-Adab wa-Lubb Lubab Lisān al-'Arab*, authored by the linguist Abdul Qadir bin Umar al-Baghdadi (1030–1093 AH). This comprehensive linguistic encyclopedia, distinguished by its rich linguistic repository, encompasses Arabic vocabulary, Arab proverbs, and their meanings, contexts, and origins. The study restructures the lexicon's content through natural language processing methodologies. A suite of intelligent algorithms is utilized to conduct computational syntactic, morphological, and semantic analyses, transforming the lexicon into an open, computational knowledge space. This transformation enables users to mine the vast linguistic and knowledge content within this monumental work.

Key Questions:

1. What opportunities do generative AI techniques offer in restructuring the content of Arabic lexicons to produce multidimensional semantic spaces?
2. How can these semantic spaces be utilized for diverse linguistic and lexical mining processes, including:
 - Building computational linguistic corpora to support large language models (LLMs).
 - Creating generative intelligent models for text generation.
 - Reviewing the polysemy of terms in *Khizanat al-Adab wa-Lubb Lubab Lisān al-'Arab*.
 - Tracking linguistic synonyms and other linguistic phenomena related to Arabic linguistic sciences.

This new knowledge space provides Arabic users with unprecedented, interactive, and dynamic linguistic tools.

Computational processing revealed numerous opportunities for reshaping the linguistic and lexical content stored in Arabic dictionaries and lexicons into a multi-level computational knowledge space. This space serves as a fertile resource for studies in syntax, morphology, and semantics. Moreover, the processed content can be utilized to develop large language models (LLMs) that support generative computational AI systems adhering to the principles and rules of the Arabic language. This advancement promises to enhance linguistic research and open new horizons for Arabic language education and its sciences.

Keywords: NLP, Linguistic Repositories, Syntactic Analyses, Morphological Analysis, Semantic Space of Colors.