تصميم نظام لتحليل المزاج العام للغة العربية الفصحى باستخدام معجم حاسوبي

د٠ صلاح راشد الناجم (*)

مستخلص ..

يقدم هذا البحث نظاما لتحليل المزاج العام للغة العربية الفصحى قمنا بتصميمه باستخدام نظام Repustate Sentiment Analysis Engine المستخدَم في تصميم أنظمة تحليل المزاج العام لعدة لغات في العالم. لتصميم هذا النظام، قمنا ببناء قاعدة بيانات معجمية للمزاج العام (Sentiment Analysis Lexical Database) للصيغ الصرفية الأساسية (Lemmas) للأسماء والأفعال في اللغة العربية الفصحى عن طريق تطوير وتعديل قاعدة بيانات معجمية صرفية مفتوحة المصدر (Open Source) مخصصة للتحليل الصرفي للأسماء والأفعال العربية صُمِّمَت سابقا باستخدام تقنية الآلات منتهية الحالات (Finite State Automata) وبُنِيَت من مدونة نصية حاسوبية (Corpus) ضخمة مكونة من أكثر من مليار كلمة. بعد ذلك استخدمنا نظام Xerox Finite-State Tools لتوليد الكلمات المصرفة (Inflected Words) المصاغة باستخدام الصيغ الصرفية الأساسية للأسماء والأفعال من قاعدة البيانات المعجمية المذكورة مع بيانات المزاج العام (Sentiment Polarity) والسمات الصرف-نحوية (Morpho-syntactic Features) لهذه الكلمات المصرفة. ثم استخدمنا لغة Python للبرمجة للقيام ببرمجة نظام Repustate Sentiment Analysis Engine وذلك ببرمجة قانون حاسوبي للمزاج العام (Sentiment Rule) لكل كلمة مصرفة مع تحديد المزاج العام لهذه الكلمة بناء على معلومات المزاج العام التي حددناها في قاعدة البيانات

(*) أستاذ بكلية الآداب- قسم اللغة العربية- جامعة الكويت.

المعجمية المذكورة. باستخدام هذه التعليمات البرمجية، استطعنا تكوين معجم حاسوبي للمزاج العام (Sentiment Analysis Lexicon) للغة العربية الفصحى خاص بنظام Repustate Sentiment Analysis Engine بناء على بيانات قاعدتنا المعجمية، وذلك لاستخدامه في التعرف الآلي على المزاج العام للنصوص كما هو متبع في أنظمة المزاج العام المبنية على القوانين (Rule-based) والتي تُعرف أيضا بالأنظمة المبنية على المعجم (Lexicon-based)، وهي أنظمة تعتمد في عملها على معاجم حاسوبية.

**

تمهيـد

يُعرف تحليل النصوص (Text Analytics) بأنه اكتشاف وانتزاع معرفة مهمة من نصوص حرة، أي نصوص لا تسير وفق بنية منتظمة (Unstructured Text). لتحقيق ذلك، تُستَخدَم أنواع عديدة من التمثيل المعرفي (Knowledge Representation) للمعلومات اللغوية. نحصل على هذه المعلومات اللغوية عن طريق استخدام المعجم الحاسوبي (Lexicon) الذي يحوي كلمات لغة معينة وسماتها الصرف – نحوية (Morpho-syntactic Features) وقيمها الافتراضية من حيث المزاج العام (Default Sentiment Polarity)، إضافة إلى استخدام القوانين النحوية والمعلومات الدلالية كالتصنيف الدلالي (Ontology/Taxonomy) للكيانات (Entities) والأحداث (Events) فضلا عن استخدام مكانز (Thesaurus) المترادفات والاختصارات.

يُعد تحليل النصوص (ويطلق عليه أيضا مصطلح التنقيب في النصوص Text) (Mining) تطبيقا من تطبيقات المعالجة الحاسوبية للغة الطبيعية (Natural Language Processing). المعالجة الحاسوبية للغة الطبيعية هي فرع مشترك بين علم الحاسوب (Computer Science) واللسانيات الحاسوبية (Computational Linguistics). في هذا السياق، تعد المعالجة الحاسوبية للغة الطبيعية التطبيق العملي للجوانب النظرية لعلم الحاسوب واللسانيات الحاسوبية.

يمثل تحليل النصوص مجالا بينيا (Interdisciplinary) حديثا يدمج أكثر من مجال أكاديمي أهمها علم الحاسوب (Computer Science)، اللسانيات الحاسوبية، استرجاع المعلومات (Information Retrieval)، تحليل البيانات/التنقيب في البيانات (Data Analysis/Data Mining)، تعلُّم الآلة (Machine Learning)، والإحصاء (Statistics).

تُستخدم تقنيات تحليل النصوص في المجال التجاري والحكومي والأكاديمي؛ وذلك لأن أغلب المعلومات الرقمية المستخدمة عالميا مخزنة على شكل نصوص لا تسير وفق بنية منظمة (Unstructured) مقارنة مع البيانات التي تسير وفق بنية منظمة (Structured Data) مثل البيانات الموجودة في قواعد البيانات. (١) يدخل في هذا المجال استرجاع المعلومات (Information Retrieval) من وثائق أو مواقع على الشبكة العنكبوتية (Web) أو من وثائق مخزنة في أنظمة إدارة النصوص (Document Management). كما يدخل في هذا المجال أيضا تصنيف النصوص (Text Classification) وتحليل المزاج العام (Sentiment Analysis) وهو موضوع بحثنا هذا.

تتضمن عملية تحليل النصوص استخدام عدد من العمليات أهمها انتزاع المعلومات (Information Extraction) من مجموعة من النصوص وتطبيق مناهج إحصائية متقدمة عليها واستخدام المعالجة الحاسوبية للغة الطبيعية من خلال التحليل الصرفي الآلي (Morphological Processing) والوسم الآلي للفئات النحوية للكلمات (Part-of-Speech Tagging) والتحليل النحوي (Syntactic Parsing) و استخدام معلومات معجمية وتقنيات إحصائية لمعرفة الكيانات (Entity Recognition) في النصوص مثل أسماء الأشخاص والأماكن والشركات وغيرها. كذلك تشتمل هذه العمليات على ما يعرف بإزالة الغموض (Disambiguation) عن طريق استخدام معلومات سياقية لتحديد المعنى المقصود من الكلمة في حال وجود أكثر من معنى لها(٢). كذلك تتضمن عملية تحليل النصوص تحليل المزاج العام (Sentiment Analysis) للنص الذي يهدف إلى

(١) **Text Mining: A Guidebook for the Social Sciences.** Ignatow, Gabe, and Rada Mihalcea. SAGE Publications, ٢٠١٦.

(٢) Mihalcea, Rada. **Using Wikipedia for Automatic Word Sense Disambiguation**. In *HLT-NAACL*. ٢٠٠٧. pp. ١٩٦-٢٠٣.

تحديد معلومات متعلقة بالموقف والرأي، كتحديد أن كلمة أو عبارة معينة في النص تحمل مدلولا إيجابيا أو سلبيا أو محايدا و تحديد العاطفة (Emotion) المرتبطة بالكلمة أو العبارة [١].

١.١ مشكلة البحث وحدوده

يمثل التعرف الآلي على المزاج العام تحدياً يواجه الأنظمة الحاسوبية التي تحاول التعامل مع اللغة العربية الفصحى وذلك لندرة الأنظمة التي تستطيع التعرف الآلي على المزاج العام للغة العربية الفصحى بشكل سليم. كما تمثل ندرة مصادر البيانات المعجمية الحاسوبية المستخدمة في تحليل المزاج العام والتي تغطي نسبة كبيرة من كلمات اللغة العربية تحدياً آخر يواجه تلك الأنظمة. يقدم هذا البحث نظاما لتحليل المزاج العام للغة العربية الفصحى قمنا ببنائه باستخدام نظام Repustate Sentiment Analysis Engine المستخدم في تصميم أنظمة تحليل المزاج العام لعدة لغات في العالم. حيث ينتمي هذا النظام إلى فئة أنظمة تحليل المزاج العام المبنية على القوانين (Rule-based)، وهي أنظمة تعتمد في عملها على معاجم حاسوبية بُنِيَت بشكل يدوي أو بشكل آلي عن طريق استخراج الكلمات من مدونات نصية حاسوبية (Corpora) وإضافتها إلى المعجم الحاسوبي آليا. تجدر الإشارة إلى أن هذا البحث لا يتطرق إلى استخدام تقنية تَعلُّم الآلة (Machine Learning) في تحليل المزاج العام، وهي التقنية المستخدمة في فئة أخرى من فئات أنظمة تحليل المزاج العام، وهي تلك الأنظمة التي تقوم بتحليل المزاج العام للنصوص من خلال المعرفة التي تَدَرَّبَ الحاسوب عليها عن طريق المدونات النصية الموسومة بمعلومات المزاج العام باستخدام الذكاء الصناعي.

---

(١) **Learning to identify emotions in text**. Strapparava, Carlo, and Rada Mihalcea. In *Proceedings of the ٢٠٠٨ ACM symposium on Applied computing*. ACM. ٢٠٠٨. pp. ١٥٥٦-١٥٦٠.

٢.١ فرضية البحث

يمكن تصميم نظام لتحليل المزاج العام باللغة العربية الفصحى عن طريق تزويد هذا النظام بمعجم حاسوبي (Lexicon) يحوي مجموعة كبيرة من الكلمات باللغة العربية الفصحى مع تحديد المزاج العام الافتراضي (Default Sentiment Polarity) المقابل لهذه الكلمات (على سبيل المثال، كلمة "رائع" تحمل المزاج "إيجابي" بينما كلمة "سيئ" تحمل المزاج "سلبي"). يَستخدِم الحاسوب هذا المعجم بعد ذلك لتحديد المزاج العام للنصوص، مثل نصوص وسائل التواصل الاجتماعي ومواقع الشبكة العنكبوتية والنصوص المستخدمة في الأنظمة التي تحوي آراء الآخرين مثل الاستبانات الإلكترونية (Surveys). لزيادة قدرة هذا المعجم الحاسوبي على تغطية أكبر قدر من كلمات اللغة العربية الفصحى، يمكن إنشاء هذا المعجم الحاسوبي للمزاج العام باستخدام بيانات معجمية نحصل عليها من قاعدة بيانات معجمية تُبنى من مدونات نصية حاسوبية (Corpora) ضخمة مكونة من كميات كبيرة من النصوص المكتوبة باللغة العربية الفصحى.

٣.١ الدراسات السابقة

نظرًا لتطبيقاته الواسعة، أجريت العديد من الدراسات في مجال التحليل الحاسوبي للمزاج العام خاصة على نصوص اللغة الإنجليزية، بينما حظيت لغات أخرى مثل العربية باهتمام أقل في هذا السياق[١]. فقد تناولَت دراسة تطبيق التحليل الحاسوبي للمزاج العام على نصوص ١,٠٠٠ تغريدة باللغة العربية جُمعت من منصة التواصل الاجتماعي تويتر واستخدَمَت تقنية تَعلُّم الآلة (Machine

(١) **A comprehensive survey of arabic sentiment analysis**. Al-Ayyoub, Mahmoud, et al. In *Information processing & management.* ٢٠١٩. ٥٦(٢). p.p. ٣٢٠-٣٤٢.

Learning) لتحديد المزاج العام لتلك النصوص[١]. كذلك استخدَمَت دراسة أخرى تقنية تَعلُّم الآلة (Machine Learning) لتطبيق التحليل الحاسوبي للمزاج العام على نصوص تغريدات باللغة العربية جُمعت من منصة التواصل الاجتماعي تويتر وتعليقات جُمعت من منصة التواصل الاجتماعي فيسبوك. كما تناولت دراسة أخرى تطبيق التحليل الحاسوبي للمزاج العام باستخدام نظام مبني على القوانين (Rule-based) عن طريق معجم حاسوبي مصغر للمزاج العام (Sentiment Lexicon) مكون من ٣٠٠ كلمة، حيث استُخدم هذا النظام لتحديد المزاج العام لنصوص ٢,٠٠٠ تغريدة باللغة العربية جُمعت أيضا من منصة التواصل الاجتماعي تويتر[٢]. ,تناولت دراسة أخرى تطبيق التحليل الحاسوبي للمزاج العام على نصوص اللغة العربية الفصحى وعدد من اللهجات العربية باستخدام نظام مبني على المعجم (Lexicon-based) عن طريق معجم حاسوبي مصغر للمزاج العام بُنِيَ من مدونة نصية (Corpus) مكونة من نصوص ١,٠٨٠ تقييم (Reviews) جُمعَت من وسائل التواصل الاجتماعي ومواقع الشبكة العنكبوتية[٣]. من جهة أخرى تناولت دراسة أخرى وسائل تحسين أداء أنظمة تحليل المزاج العام لنصوص اللغة العربية والتي تنتمي إلى نوع أنظمة التحليل المبني على القوانين[٤].

---

(١) **Sentence-level Arabic sentiment analysis.** Shoukry, Amira, and Ahmed Rafea. In *International Conference on Collaboration Technologies and Systems (CTS)*. IEEE. ٢٠١٢. pp. ٥٤٦-٥٥٠.

(٢) **Arabic sentiment analysis: Lexicon-based and corpus-based**. Abdulla, Nawaf A., et al. In *IEEE Jordan conference on applied electrical engineering and computing technologies (AEECT)*. IEEE, ٢٠١٣. pp. ١-٦.

(٣) **An opinion analysis tool for colloquial and standard Arabic**. Al-Kabi, Mohammed, et al. In *The Fourth International Conference on Information and Communication Systems (ICICS ٢٠١٣)*. ٢٠١٣. pp. ٢٣-٢٥.

(٤) **Towards improving the lexicon-based approach for arabic sentiment analysis.** Abdulla, Nawaf A., et al. In *International Journal of Information Technology and Web Engineering (IJITWE)*. ٢٠١٤. ٩(٣). p.p. ٥٥-٧١.

٤.١ أنظمة تحليل المزاج العام (Sentiment Analysis Systems)

تعد أنظمة تحليل المزاج العام (Sentiment Analysis) – وتُعرَف أيضا بأنظمة تحليل الرأي (Opinion Mining) – من أهم تطبيقات تحليل النصوص. لقيت هذه الأنظمة اهتماما واسعا في السنوات الأخيرة مع تطور وزيادة استخدام أنظمة تحليل وسائل التواصل الاجتماعي (Social Media Analytics) والتي تشتمل على تحليل المزاج العام كأحد مكوناتها[١]. في هذه الأنظمة يقوم الحاسوب بتصنيف نص معين من حيث مزاج المحتوى إلى إيجابي أو سلبي أو محايد. في هذا السياق، توفر أنظمة تحليل المزاج العام للجهات الحكومية وسيلة لقياس تفاعل الجمهور وردود أفعالهم تجاه الأحداث والقضايا السياسية والاجتماعية والاقتصادية وغيرها. كذلك يُمكِن استخدام تحليل البيانات التاريخية لوسائل التواصل الاجتماعي لمعرفة التغير التاريخي في اتجاه المزاج العام تجاه قضية معينة (Trend Analysis and Benchmarking). من جهة أخرى أدركت الحكومات أن الحوار الذي يدور على وسائل التواصل الاجتماعي يمثل وسيلة حية لاستطلاع رأي الجمهور ومزاجه العام ولمعرفة اتجاه الرأي العام. كذلك يعد الحوار الذي يدور على وسائل التواصل الاجتماعي من المؤشرات الأساسية لقياس الأداء (Key Performance Indicators) والتي يستخدمها متخذو القرار والجهات الحكومية للتأكد من تحقيق الأهداف الاستراتيجية لاستراتيجياتهم السياسية والاقتصادية والإعلامية. كذلك تَستخدِم الشركات أنظمة تحليل المزاج العام لمعرفة

---

(١) **A system for real-time twitter sentiment analysis of ٢٠١٢ us presidential election cycle**. Wang, Hao, Dogan Can, Abe Kazemzadeh, François Bar, and Shrikanth Narayanan. In *Proceedings of the ACL ٢٠١٢ System Demonstrations*. Association for Computational Linguistics. ٢٠١٢. pp. ١١٥-١٢٠.

آراء الزبائن في منتجاتها وخدماتها من خلال تحليل مزاج الحوار الذي يدور على وسائل التواصل الاجتماعي بشأن تلك المنتجات والخدمات.

٥.١ أنواع أنظمة تحليل المزاج العام

يمكن تقسيم أنظمة تحليل المزاج العام إلى فئتين، الفئة الأولى هي الأنظمة المبنية على القوانين (Rule-based) وتُعرف أيضا بالأنظمة المبنية على المعجم (Lexicon-based) وهي أنظمة تعتمد في عملها على معاجم حاسوبية للمزاج العام (Sentiment Lexicons) بُنِيَت بشكل يدوي أو بشكل آلي عن طريق استخراج الكلمات من مدونات نصية حاسوبية (Corpora) وإضافتها إلى المعجم الحاسوبي آليا مع تحديد المزاج العام لكل كلمة (إيجابي، سلبي، محايد). حيث يستخدم الحاسوب هذه المعاجم للتعرف على جميع الكلمات التي تحمل مزاجاً إيجابيا أو سلبيا من الجملة، ثم يقوم بجمع القيمة الرقمية للمزاج العام لهذه الكلمات لتحديد ما إذا كانت الجملة تحمل مزاجا إيجابيا أو سلبيا بشكل عام. يُعرف هذا النوع من أنظمة تحليل المزاج العام أيضاً باسم الأنظمة ذات التوجه الدلالي (Semantic Orientation-Based). النوع الثاني من هذه الأنظمة يتمثل في الأنظمة التي تعتمد على تقنية تَعلُّم الآلة (Machine Learning) وهي أنظمة تقوم بتحليل المزاج العام للنصوص من خلال المعرفة التي تَدَرَّبَ الحاسوب عليها عن طريق مدونات النصوص الحاسوبية الموسومة بمعلومات المزاج العام باستخدام الذكاء الاصطناعي. من أمثلة أنظمة النوع الأول التي تعتمد على القوانين نظام OpinionFinder[١]. حيث طورت مجموعة من الباحثين من جامعات بيتسبيرج وكورنيل ويوتاه الأمريكية نظام OpinionFinder والذي يقوم بمعالجة

(١) **Annotating expressions of opinions and emotions in language**. Wiebe, Janyce, Theresa Wilson, and Claire Cardie. In *Language resources and evaluation*. ٢٠٠٥. ٣٩(٢). pp. ١٦٥-٢١٠.

الوثائق النصية والتعرف الآلي على جوانب الرأي الشخصي في جمل هذه الوثائق وتحديد المزاج العام لها. بُنِيَ المعجم الحاسوبي المرافق لنظام OpinionFinder باستخدام كلمات جُمِعَت وحُدِّدَ مزاجها العام يدويا، إضافة إلى استخدام كلمات استُخرِجَت من مدونات نصية حاسوبية وُسِمَ المزاج العام لكلماتها مسبقا. يحتوي المعجم على ٦,٨٥٦ مادة معجمية منها ٩٩٠ مادة مكونة من تعبيرات مركبة (Multiword Expressions). إضافة إلى تحديد المزاج العام (إيجابي، سلبي، محايد) لكل مادة معجمية في المعجم، حُدِّدَت الفئة النحوية (Part of Speech) لهذه المواد المعجمية. تجدر الإشارة إلى أنه في أنظمة تحليل المزاج العام التي تعتمد على القوانين، يقوم النظام بالتعرف الآلي على المزاج العام لكلمات وعبارات النصوص التي يحللها وفقا لوجود أو عدم وجود تلك الكلمات والعبارات في معجمه الحاسوبي. من أمثلة النوع الثاني لأنظمة تحليل المزاج العام والتي تعتمد في تحليلها للمزاج العام على تقنية تَعلُّم الآلة نظام تحليل المزاج العام للوثائق الذي طوره كل من بانج و لي[١]. لعل ما يميز الأنظمة التي تعتمد على التعلم الآلي عن طريق مدونات نصية حاسوبية موسومة مسبقا بمعلومات المزاج العام أن هذه الأنظمة يمكن استخدامها لتحليل المزاج العام لأي لغة متى ما توافرت نصوص المدونات النصية الحاسوبية الموسومة بمعلومات المزاج العام التي يتعلم منها الحاسوب.

(١) **A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts**. Pang, Bo, and Lillian Lee. In *Proceedings of the ٤٢nd annual meeting on Association for Computational Linguistics. Association for Computational Linguistics*. ٢٠٠٤.

٢ نظام لتحليلا لمزاج العام للغة العربية الفصحى

باستخدام نظام Repustate Sentiment Analysis Engine المستخدم في تصميم أنظمة تحليل المزاج العام لعدة لغات في العالم، قمت بتصميم نظام لتحليل المزاج العام للغة العربية الفصحى، وهو نظام حاسوبي يمكن من خلاله تحليل المزاج العام (إيجابي، سلبي، محايد) عن طريق التحليل الحاسوبي الذكي للغة المستخدمة في النصوص، مثل نصوص وسائل التواصل الاجتماعي ومواقع الشبكة العنكبوتية والنصوص المستخدمة في الأنظمة التي تحوي آراء الآخرين مثل الاستبانات الإلكترونية (Surveys). ينتمي هذا النظام إلى الفئة الأولى من فئات أنظمة تحليل المزاج العام المشار إليها سابقا، وهي الأنظمة المبنية على القوانين (Rule-based).

لتمكين نظام Repustate Sentiment Analysis Engine من التعرف على المزاج العام للتصريفات المذكورة، قمت بتوليد قانون معجمي حاسوبي للمزاج العام (Sentiment Rule) خاص بكل صيغة مصرفة بشكل آلي من قاعدة بيانات معجمية، ثم أُدخِلَت هذه القوانين إلى قاعدة بيانات قوانين المزاج العام لهذا النظام، وبالتالي صارت هذه القوانين تمثل معجماً حاسوبيا للمزاج العام يستخدمه النظام لكي يتعرف على المزاج العام للنصوص. يُعد هذا النظام من أكبر أنظمة تحليل المزاج العام التي تتعامل مع اللغة العربية الفصحى، سواء من حيث حجم معجمه الحاسوبي (Lexicon) ومجال تغطيته. يحتوي المعجم الحاسوبي لنظامنا على الصيغ الصرفية والاشتقاقية المحتملة لكلمات اللغة العربية الفصحى (على سبيل المثال: كتب، ويكتبان، كتبوا، سيكتبن، كاتب، المكتوب). يتوافر نظامنا للمبرمجين كواجهة لبرمجة التطبيقات (Application Programming Interface (API)). يمكن استخدام هذه الواجهة لتحليل المزاج العام عن طريق إتاحتها على شكل خدمة ويب سحابية (Cloud Web Service) يمكن دمجها مع أي نظام لتحليل النصوص العربية في وسائل التواصل الاجتماعي وغيرها من التطبيقات.

١.٢ *بناء قاعدة بيانات معجمية للمزاج العام*

من أجل بناء معجم حاسوبي لنظام تحليل المزاج العام باللغة العربية الفصحى، قمنا باستخدام قاعدة بيانات معجمية للمزاج العام (Sentiment Analysis Lexical Database) تحتوي على الصيغ الصرفية الأساسية (Lemmas) للأسماء والأفعال في اللغة العربية. في هذا السياق، ينبغي هنا أن نوضح المقصود بالصيغ الصرفية الأساسية للأسماء والأفعال. في مجال علم بناء الكلمة الحديث (Morphology)، تُستَخدَم عملية صرفية تُعرف بعملية استخلاص جذع الكلمة (Stemming) وهي عملية يُستَخلَصُ فيها الجزء الأساسي من الكلمة المشتقة (Derived Word) أو المصرفة (Inflected Word) بعد حذف السوابق (Prefixes) واللواحق (Suffixes) من تلك الكلمة. على سبيل المثال، الكلمات: كاتِبان، كاتِبات، كاتِبا، الكاتِبان، للكاتِبَيْن لها جذع (Stem) أساسي واحد وهو كاتِب. ينبغي أن نشير هنا إلى أن عملية استخلاص جذع الكلمة لا يُقصد بها إرجاع الكلمة إلى الجذر أو تجريد الكلمة من حروف الزيادة كما هو متعارف عليه في الدراسات الصرفية التقليدية. تجب الإشارة هنا أيضا إلى أن عملية استخلاص جذع الكلمة من خلال إزالة السوابق واللواحق قد لا تكون كافية لإرجاع بعض الكلمات إلى جذعها الأساسي، حيث إن إزالة السوابق واللواحق من كلمة مشتقة أو مصرفة قد تُنتِج جذعا غير مستخدم لغويا (أي ليس موجودا في معجم اللغة) أو قد تُرجِع أكثر من كلمة مرتبطة دلاليا وصرفيا إلى أكثر من جذع مع أنها في الأصل تشترك في جذع أساسي واحد. على سبيل المثال، في اللغة الإنجليزية، عندما تُطَبَّق عملية استخلاص جذع الكلمة على كلمة مثل having فإنها ستنتج لنا جذعا غير مستخدم لغويا وهو hav وذلك من خلال إزالة اللاحقة {ing-}. كذلك في اللغة العربية، قد نجد أنفسنا – على سبيل المثال – أمام كلمات جُمِعت جمع تكسير وهي كلمات لا نستطيع فقط الاعتماد على إزالة السوابق واللواحق

منهـا لاسـتخلاص جـذعها. علـى سـبيل المثـال، الكلمـات: الطالـب، طالبـان، الطالبات، الطلاب كلها تعود إلى جذع واحد وهو طالب. نلاحظ هنا أننا في كلمة الطلاب (جمع تكسير) نستطيع أن نرجعها إلى جذعها المفرد المستخدم لغويا وهو طالب وهو نفس جذع الكلمات الأخرى المرتبطة بها دلاليا وصرفيا (الطالب، طالبان، الطالبات) دون الاعتماد فقط على إزالة السواق واللواحق. حيث إننا لو اعتمدنا فقط على إزالة السوابق واللواحق، فسنحصل على جذعين مختلفين لهذه الكلمات المترابطة دلاليا وصرفيا؛ حيث سنحصل على الجذع طالب للكلمات الطالب، طالبان، الطالبات وسنحصل أيضا على الجذع طلاب لكلمة الطلاب (بعد حذف السابقة {ال –}. لمواجهة ذلك في مجال علم بناء الكلمة الحديث، تُستخدَم عملية صرفية أخرى وهي عملية استخلاص الصيغة الصرفية الأساسية للكلمة (Lemmatization). الصيغة الصرفية الأساسية للكلمة (Lemma) هي أصغر صيغة للكلمة مُستخدَمة لغويا؛ أي صيغة الكلمة دون وجود لسوابق أو لواحق تصريفية أو اشتقاقية أو ضمائر متصلة بشرط أن تكون هذه الصيغة الصرفية مُستخدَمة لغويا (أي موجودة في معجم اللغة). تقابل هذه الصيغة في اللغة العربية صيغة الماضي المفرد المذكر الغائب للأفعال (مثل كَتَب) وصيغة المفرد المذكر النكرة للأسماء (كاتِب). على سبيل المثال: كَتَب، استكتَب، كاتِب، مكتوب، كِتاب، مَكتَب تمثل صيغا صرفية أساسية (Lemmas) لأفعال وأسماء مشتقة من جذر واحد (ك ت ب).

للحصول على قاعدة بيانات معجمية للمزاج العام، قمنا بتطوير وتعديل قاعدة بيانات معجمية صرفية مفتوحة المصدر (Open Source) مخصصة للتحليل الصرفي للأسماء والأفعال العربية صُمِّمَت سابقا باستخدام تقنية الآلات منتهية الحالات (Finite State Automata) وهي قاعدة بيانات (Arabic AraComLex[1]

(١) لمعرفة المزيد عن قاعدة بيانات AraComLex المعجمية، يمكنكم الرجوع إلى:=

(Computer Lexicon والتي قمنا بتعديلها وتزويد بياناتها المعجمية ببيانات المزاج العام للصيغ الصرفية الأساسية للأسماء والأفعال. بعد ذلك قمنا بتوليد البيانات المعجمية المشتملة على بيانات المزاج العام للكلمات المصرفة المصاغة من الصيغ الصرفية الأساسية للأسماء والأفعال العربية من هذه القاعدة المعجمية باستخدام نظام Xerox Finite-State Tools [١].

بُنِيَت قاعدة بيانات AraComLex المعجمية من مدونة نصية حاسوبية (Corpus) ضخمة وهي مدونة مكونة من ١,٠٨٩,١١١,٢٠٤ كلمة، حيث استُخدِمت تقنيات تعلم الآلة وأدوات الوسم الآلي للكلمات (Annotation) وأدوات أخرى لاستخلاص وتنقية المعرفة المعجمية المتعلقة بالسمات الصرف-نحوية (Morpho-syntactic Features) وأشكال التصريفات (Inflection Paradigms) للصيغ الصرفية الأساسية في هذه القاعدة[٢]. تحوي هذه القاعدة المعجمية، إضافة

=**A lexical database for modern standard Arabic interoperable with a finite state morphological transducer**. Attia, Mohammed, et al. In *International Workshop on Systems and Frameworks for Computational Morphology.* Springer, Berlin, Heidelberg. ٢٠١١. pp. ٩٨-١١٨.

A corpus-based finite-state morphological toolkit for contemporary Arabic. Attia, Mohammed, et al. In *Journal of Logic and Computation*. ٢٠١٤. ٢٤(٢), pp. ٤٥٥-٤٧٢.

(١) لمعرفة المزيد عن استخدام نظام Xerox Finite-State Tools في المعالجة الحاسوبية للغة الطبيعية، يمكنكم الرجوع إلى:

Finite-State Morphology. Beesley, K. and Karttunen, L. CSLI, Stanford. ٢٠٠٣.

(٢) تتكون المدونة الحاسوبية النصية من ٩٢٥,٤٦١,٧٠٧ كلمة جُمِعَت من مدونة Arabic Gigword Corpus (Fifth Edition) و ١٦٣,٦٤٩,٤٩٧ كلمة جُمِعَت من موقع الجزيرة الإخباري (www.aljazeera.net). لمعرفة المزيد عن مدونة Arabic Gigword Corpus، يمكن الرجوع إلى:

Arabic Gigaword Fifth Edition. Linguistic Data Consortium. University of Pennsylvania. ٢٠١١. https://catalog.ldc.upenn.edu/LDC٢٠١١T١١. Accessed ٢٩ October ٢٠٢١.

إلى الكلمات الشائعة في اللغة العربية الفصحى، الكلمات المستخدمة في العصر الحديث من قبل المتحدثين والكتاب الذي يستخدمون اللغة العربية الفصحى في الكتب والصحف ووسائل الإعلام الحديثة الأخرى كالمواقع الإخبارية. كما تحوي كلمات متداولة في مجالات السياسة والعلوم إضافة إلى المصطلحات المسكوكة الحديثة (Coined Terms).

فيما يأتي نعرض أمثلة للقوانين المعجمية الخاصة بالأفعال والتي قمنا بتعديلها في قاعدة البيانات المعجمية AraComLex، حيث يظهر في القانون أننا قمنا بتحديد المزاج العام للصيغة الصرفية الأساسية للفعل بإضافة السمة POS+ (للدلالة على المزاج العام الإيجابي) و NEG+ (للدلالة على المزاج العام السلبي) إلى جانب السمات الصرف–نحوية للصيغ الصرفية الأساسية المحددة سابقا في قاعدة البيانات المعجمية AraComLex:

se^@D.V.P@@D.M.I@^احترق^NEG:^ss+ ؛Intransitive

^se^@D.V.P@@D.M.I@انهزم+NEG:^ss^ Intransitive;

^se^@D.V.P@@D.M.I@تدهور+NEG:^ss^ Intransitive;

^se^@D.V.P@@D.M.I@تسامح+POS:^ss^ Intransitive;

^se^@D.V.P@فرح+POS:^ss^ Intransitive;

^se^@D.V.P@@D.M.I@نجح+POS:^ss^ Intransitive;

^se^شكر+POS:^ss^ Transitive;

^se^ضرب+NEG:^ss^ Transitive;

^se^هدم+NEG:^ss^ Transitive;

في هذه القوانين المعجمية يُستخدم الرمزان Intransitive (فعل لازم) و Transitive (فعل متعد) لإرشاد نظام التحليل الحاسوبي الصرفي في قاعدة بيانات AraComLex إلى توليد التصريفات المناسبة لكل صيغة صرفية أساسية للأفعال وفقاً لتعدي أو لزوم الفعل. كما يُستخدم الرمز ^ss^ لتحديد بداية ونهاية

الـصيغة الـصرفية الأساسـية. كـذلك يُـستخدم الرمـز D.V.P@ للدلالـة علـى أن الصيغة الصرفية الأساسية للفعل لا تسمح بتوليد صيغة المبني للمجهول منها.

كما هو الحال في القوانين المعجمية الخاصة بالأفعال، فيما يأتي نعرض أمثلة للقوانين المعجميـة الخاصـة بالأسـماء والتـي قمنـا بتعـديلها فـي قاعـدة البيانـات المعجمية AraComLex، حيث يظهر أيضا في القانون أننا قمنا بتحديد المزاج العام للصيغة الصرفية الأساسية للاسم بإضافة السمة POS+ و NEG+ إلى جانب السمات الصرف–نحوية للصيغ الصرفية الأساسية المحددة سابقا في قاعدة البيانات المعجمية AraComLex:

FemduFempl; ^se^+femمؤامرة+NEG:^ss^

FemduFempl; ^se^+femابتسامة+POS:^ss^

NoNumber; ^se^+mascرخاء+POS:^ss^

NoNumber; ^se^+mascضجيج+NEG:^ss^

NoNumber; ^se^+mascضياع+NEG:^ss^

NoNumber; ^se^+mascطغيان+NEG:^ss^

NoNumber; ^se^+mascفساد+NEG:^ss^

MascduFempl; ^se^+mascانتصار+POS:^ss^

^se^+mascفخور+POS:^ss^

FemMascduFemduMascplFempl;

^se^+mascصادق+POS:^ss^

FemMascduFemduMascplFempl;

الجدول التالي يشرح المقصود بالرموز المذكورة في القوانين المعجمية الخاصة بالأسماء:

الرمز	الوظيفة
+masc	للدلالة على أن الصيغة الصرفية الأساسية للاسم هي صيغة مذكر
+fem	للدلالة على أن الصيغة الصرفية الأساسية للاسم هي صيغة مؤنث
FemduFempl	للدلالة على أن الصيغة الصرفية الأساسية للاسم تتصرف إلى المثنى المؤنث والجمع المؤنث السالم
MascduFempl	للدلالة على أن الصيغة الأساسية للاسم تتصرف إلى المثنى المذكر والجمع المؤنث السالم
FemMascduFemduMascplFempl	للدلالة على أن الصيغة الأساسية للاسم تتصرف إلى المفرد المؤنث والمثنى المذكر والمثنى المؤنث والجمع المذكر السالم والجمع المؤنث السالم
NoNumber	للدلالة على أن الصيغة الصرفية الأساسية للاسم لا تتصرف وفقا للعدد (مثنى، جمع)

جدول رقم (١): جدول يشرح المقصود بالرموز المذكورة في القوانين المعجمية الخاصة بالأسماء

تُستخدم الرموز FemduFempl ، MascduFempl ، FemMascduFemduMascplFempl ، و NoNumber لإرشاد نظام التحليل الحاسوبي الصرفي في قاعدة بيانات AraComLex إلى توليد التصريفات المناسبة لكل صيغة صرفية أساسية عن طريق إلصاق الوحدات الصرفية (Morphemes) المناسبة مثل اللواحق {– ات} و {–ان} و {–ون}. تجدر الإشارة هنا إلى أن قاعدة البيانات المعجمية AraComLex تستخدم قوانين معجمية خاصة للتعامل مع الأسماء التي تُجمَع جمع تكسير.

٢.٢ توليد الكلمات المصرفة من الصيغ الصرفية في قاعدة البيانات المعجمية

بعد تعديل القوانين المعجمية في قاعدة البيانات المعجمية عن طريق تحديد المزاج العام للصيغ الصرفية الأساسية للأسماء والأفعال، استخدمتُ نظام Xerox Finite–State Tools لتوليد الكلمات المصرفة المصاغة باستخدام الصيغ الصرفية الأساسية للأسماء والأفعال الموجودة في قاعدة البيانات المعجمية المذكورة مع توليد بيانات المزاج العام و السمات الصرف–نحوية لهذه الكلمات المصرفة. فيما يلي أمثلة على البيانات المعجمية التي قام نظام Xerox Finite–State Tools

بتوليدها للكلمات المصرفة والتي تحتوي على بيانات المزاج العام للكلمات و السمات الصرف–نحوية الخاصة بها:

الكلمة المصرفة	المزاج العام والسمات الصرف نحوية	التفسير
احترق	+verb+pres+active+1pers+sg+NEG	فعل مضارع مبني للمعلوم مسند إلى المتكلم (الشخص الأول) المفرد المذكر وهو فعل يدل على مزاج سلبي[1]
أنعمن	+verb+past+active+POS+3pers+pl+fem	فعل ماض مبني للمعلوم مسند إلى الغائب (الشخص الثالث) الجمع المؤنث وهو فعل يدل على مزاج إيجابي
أكرمتم	+verb+past+active+POS+2pers+pl	فعل ماض مبني للمعلوم مسند إلى المخاطب (الشخص الثاني) الجمع المذكر وهو فعل يدل على مزاج إيجابي
رخاء	+noun+inhuman+POS+sg	اسم مجرد لا يشير إلى عاقل وهو يدل على المفرد المذكر وهو اسم يحمل مزاجا إيجابيا
كريم	+adj+POS+sg	صفة (اسم) تدل على المفرد المذكر وهي صفة تحمل مزاجا إيجابيا
ضعيف	+adj+NEG+sg	صفة (اسم) تدل على المفرد المذكر وهي صفة تحمل مزاجا سلبيا
صادقة	+adj+POS+fem+sg	صفة (اسم) تدل على المفرد المؤنث وهي صفة تحمل مزاجا إيجابيا

جدول رقم (٢): جدول يشرح البيانات المعجمية التي قام نظام Xerox Finite–State Tools بتوليدها

٣.٢ توليد معجم حاسوبي للمزاج العام للغة العربية الفصحى بناء على بيانات قاعدتنا المعجمية

بعد استخدام نظام Xerox Finite–State Tools لتوليد الكلمات المصرفة المصاغة باستخدام الصيغ الصرفية الأساسية للأسماء والأفعال الموجودة في قاعدة البيانات المعجمية مع تحديد بيانات المزاج العام و السمات الصرف–نحوية لتلك الكلمات، استخدمتُ لغة Python للبرمجة للقيام ببرمجة نظام Repustate Sentiment Analysis Engine عن طريق واجهة برمجة التطبيقات (API)

الخاصة بالنظام وذلك من خلال برمجة قانون حاسوبي للمزاج العام (Sentiment Rule) لكل كلمة مصرفة مع تحديد المزاج العام لهذه الكلمة بناء على معلومات المزاج العام التي حُدِّدَت في قاعدة البيانات المعجمية. تُمَكِّن هذه التعليمات البرمجية نظام Repustate Sentiment Analysis Engine من توليد معجم حاسوبي للمزاج العام للغة العربية الفصحى بناء على بيانات قاعدتنا المعجمية وذلك لاستخدامه في التعرف الآلي على المزاج العام للنصوص العربية كما هو متبع في أنظمة المزاج العام المبنية على القوانين (Rule-based) والتي تعتمد في عملها على معاجم حاسوبية للمزاج العام. فيما يلي أمثلة على التعليمات البرمجية المستخدمة لصياغة قوانين المزاج العام الحاسوبية التي أدخلناها إلى نظام Repustate Sentiment Analysis Engine والتي تمثل عددا من الكلمات المصرفة التي قمنا بتوليدها عن طريق قاعدتنا المعجمية:

الكلمة	التعليمات البرمجية المستخدمة لصياغة قانون المزاج العام الحاسوبي للكلمة
احترق	client.add_sentiment_rule("احترق" , "neg" , lang="ar")
أنعمن	client.add_sentiment_rule("أنعمن" , "pos" , lang="ar")
أكرمتم	client.add_sentiment_rule("أكرمتم" , "pos" , lang="ar")
سأنهزم	client.add_sentiment_rule("سأنهزم" , "neg" , lang="ar")
كريم	client.add_sentiment_rule("كريم" , "pos" , lang="ar")
ضعيف	client.add_sentiment_rule("ضعيف" , "neg" , lang="ar")
صادقة	client.add_sentiment_rule("صادقة" , "pos" , lang="ar")
سخيفة	client.add_sentiment_rule("سخيفة" , "neg" , lang="ar")
رخاء	client.add_sentiment_rule("رخاء" , "pos" , lang="ar")
مؤامرة	client.add_sentiment_rule("مؤامرة" , "neg" , lang="ar")

جدول رقم (٣): جدول يبين التعليمات البرمجية المستخدمة لصياغة قوانين المزاج العام الحاسوبية التي أدخلناها إلى نظام Repustate Sentiment Engine والتي تمثل عددا من الكلمات المصرفة

يعرض الشكل التالي مسار العمليات التي قمنا بها والتي تشتمل على بناء قاعدة البيانات المعجمية للمزاج العام وتوليد الكلمات المصرفة من هذه القاعدة المعجمية ثم القيام ببرمجة نظام Repustate Sentiment Analysis Engine عن طريق برمجة قانون حاسوبي للمزاج العام (Sentiment Rule) لكل كلمة مصرفة من الصيغ الصرفية الأساسية للأسماء والأفعال الموجودة في قاعدة البيانات المعجمية مع تحديد المزاج العام لهذه الكلمة:

شكل رقم (١): مسار تكوين المعجم الحاسوبي للمزاج العام لاستخدامه في نظام Repustate Sentiment Analysis Engine

٤.٢ اختبار نظام تحليل المزاج العام بعد تزويده بقوانين المزاج العام الحاسوبية

تبين الشاشات الآتية تطبيقا فعليا (Sample Log) لاستخدام قوانين المزاج العام الحاسوبية التي قمنا ببرمجة نظام Repustate Sentiment Analysis Engine على أساسها، والتي تمثل كل كلمة مصرفة من الصيغ الصرفية الأساسية للأسماء والأفعال الموجودة في قاعدة البيانات المعجمية. حيث قمنا بإدخال جمل باللغة

العربية الفصحى إلى النظام باستخدام لغة Python للبرمجة عن طريق واجهة برمجة التطبيقات API الخاصة بنظام Repustate Sentiment Analysis Engine. كما هو مبين في الشاشات، إذا نَتَج عن إدخال الجملة إلى النظام قيمة إيجابية (مثل ٠.٩٧٥) فإن هذا يعني أن النظام قد تعرف على هذه الجملة على أنها جملة إيجابية من حيث المزاج العام بسبب وجود كلمة تحمل مزاجا إيجابيا فيها. أما إذا نَتَج عن إدخال الجملة إلى النظام قيمة سلبية (مثل ٠.٩٧٥-) فإن هذا يعني أن النظام قد تعرف على هذه الجملة على أنها جملة سلبية من حيث المزاج العام بسبب وجود كلمة تحمل مزاجا سلبيا فيها. أما إذا كانت الجملة تحمل مزاجا محايدا (ليس سلبيا أو إيجابيا)، فسينتج عن إدخال الجملة إلى النظام القيمة والتي تدل على أن النظام قد تعرف على هذه الجملة على أنها جملة محايدة من حيث المزاج العام بسبب عدم وجود كلمة تحمل مزاجا إيجابيا أو سلبيا فيها.

```
>>> client.sentiment(text='احترق البيت', lang='ar')
{'score': -0.975, 'status': 'OK'}
>>>
```

شكل رقم (٢): مثال (Sample Log) يبين تعرُّف النظام على المزاج العام لجملة "احترق البيت"

```
>>> client.sentiment(text='المعلمات أنعمن على الطالبات', lang='ar'
)
{'score': 0.975, 'status': 'OK'}
>>>
```

شكل رقم (٣): مثال يبين تعرُّف النظام على المزاج العام لجملة "المعلمات أنعمن على الطالبات"

```
>>> client.sentiment(text='علمت أنكم قد أكرمتم ضيوفكم', lang='ar')

{'score': 0.975, 'status': 'OK'}
>>>
```

شكل رقم (٤): مثال يبين تعرُّف النظام على المزاج العام لجملة "علمت أنكم قد أكرمتم ضيوفكم"

```
>>> client.sentiment(text='سأنهزم في المعركة', lang='ar')
{'score': -0.975, 'status': 'OK'}
>>>
```

شكل رقم (٥): مثال يبين تعرُّف النظام على المزاج العام لجملة "سأنهزم في المعركة"

```
>>> client.sentiment(text='محمد رجل كريم', lang='ar')
{'score': 0.975, 'status': 'OK'}
>>>
```

شكل رقم (٦): مثال يبين تعرُّف النظام على المزاج العام لجملة "محمد رجل كريم"

```
>>> >>> client.sentiment(text='هذا الولد ضعيف', lang='ar')
{'score': -0.975, 'status': 'OK'}
>>>
```

شكل رقم (٧): مثال يبين تعرُّف النظام على المزاج العام لجملة "هذا الولد ضعيف"

```
>>> client.sentiment(text='مريم طالبة صادقة', lang='ar')
{'score': 0.975, 'status': 'OK'}
>>>
```

شكل رقم (٨): مثال يبين تعرُّف النظام على المزاج العام لجملة "مريم طالبة صادقة"

```
>>> client.sentiment(text='سمعت قصة سخيفة', lang='ar')
{'score': -0.975, 'status': 'OK'}
>>>
```

شكل رقم (٩): مثال يبين تعرُّف النظام على المزاج العام لجملة "سمعت قصة سخيفة"

```
>>> client.sentiment(text='تسعى الحكومات إلى تحقيق الرخاء', lang='ar')
{'score': 0.975, 'status': 'OK'}
>>>
```

شكل رقم (١٠): مثال يبين تعرُّف النظام على المزاج العام لجملة "تسعى الحكومات إلى تحقيق الرخاء"

```
>>> client.sentiment(text='اكتشف المدير مؤامرة في الشركة', lang='ar')
{'score': -0.975, 'status': 'OK'}
>>>
```

شكل رقم (١١): مثال يبين تعرُّف النظام على المزاج العام لجملة "اكتشف المدير مؤامرة في الشركة"

```
>>> client.sentiment(text='كتب الطالب الدرس', lang='ar')
{'score': 0, 'status': 'OK'}
>>> client.sentiment(text='قرأ محمد الكتاب', lang='ar')
{'score': 0, 'status': 'OK'}
>>> client.sentiment(text='العصفور فوق الشجرة', lang='ar')
{'score': 0, 'status': 'OK'}
>>>
```

شكل رقم (١٢): مثال يبين تعرُّف النظام على المزاج العام لجمل محايدة لا تحمل مزاجا إيجابيا أو سلبيا

نتائج البحث

قدم هذا البحث نظاما لتحليل المزاج العام للغة العربية الفصحى قمنا بتصميمه باستخدام نظام Repustate Sentiment Analysis Engine المستخدم في تصميم أنظمة تحليل المزاج العام لعدة لغات في العالم. لتصميم هذا النظام، قمنا ببناء قاعدة بيانات معجمية للمزاج العام (Sentiment Analysis Lexical Database) تحتوي على الصيغ الصرفية الأساسية (Lemmas) للأسماء والأفعال في اللغة العربية الفصحى عن طريق تطوير وتعديل قاعدة بيانات معجمية صرفية مفتوحة المصدر (Open Source) مخصصة للتحليل الصرفي للأسماء والأفعال العربية صُمِّمَت سابقا باستخدام تقنية الآلات منتهية الحالات (Finite State Automata) وبُنِيَت من مدونة نصية حاسوبية (Corpus) ضخمة مكونة من أكثر من مليار كلمة. بعد ذلك استخدمنا نظام Xerox Finite-State Tools لتوليد الكلمات المصرفة (Inflected Words) المصاغة باستخدام الصيغ الصرفية الأساسية للأسماء والأفعال من قاعدة البيانات المعجمية المذكورة مع بيانات المزاج العام (Sentiment Polarity) و السمات الصرف-نحوية (Morpho-syntactic Features) لهذه الكلمات المصرفة. ثم استخدمنا لغة Python للبرمجة للقيام

ببرمجة نظام Repustate Sentiment Analysis Engine وذلك ببرمجة قانون حاسوبي للمزاج العام (Sentiment Rule) لكل كلمة مصرفة مع تحديد المزاج العام لهذه الكلمة بناء على معلومات المزاج العام التي حددناها في قاعدة البيانات المعجمية المذكورة. باستخدام هذه التعليمات البرمجية، استطعنا توليد معجم حاسوبي للمزاج العام (Sentiment Analysis Lexicon) للغة العربية الفصحى خاص بنظام Repustate Sentiment Analysis Engine بناء على بيانات قاعدتنا المعجمية وذلك لاستخدامه في التعرف الآلي على المزاج العام للنصوص كما هو متبع في أنظمة المزاج العام المبنية على القوانين (Rule-based) والتي تُعرف أيضا بالأنظمة المبنية على المعجم (Lexicon-based)، وهي أنظمة تعتمد في عملها على معاجم حاسوبية. بعد ذلك، قمنا باختبار هذا النظام، حيث أدخلنا جملا باللغة العربية الفصحى إليه واستطاع النظام أن يتعرف على المزاج العام لهذه الجمل بنجاح بناء على المعرفة المعجمية التي زودناه بها. تجدر الإشارة إلى أن النظام يستطيع التعرف على المزاج العام لتصريفات (Inflections) متعددة لأفعال مصاغة من قرابة ٦,٥٠٠ صيغة صرفية أساسية (Lemmas) للأفعال، كما يستطيع التعرف على المزاج العام لتصريفات متعددة لأسماء مصاغة من قرابة ٢٤,٩٠٠ صيغة صرفية أساسية (Lemmas) للأسماء.

**

المصادر والمراجع

A comprehensive survey of arabic sentiment analysis. Al-Ayyoub, Mahmoud, et al. In *Information processing & management.* ٢٠١٩. ٥٦(٢). p.p. ٣٢٠-٣٤٢.

A corpus-based finite-state morphological toolkit for contemporary Arabic. Attia, Mohammed, et al. In *Journal of Logic and Computation*. ٢٠١٤. ٢٤(٢), pp. ٤٥٥-٤٧٢.

A lexical database for modern standard Arabic interoperable with a finite state morphological transducer. Attia, Mohammed, et al. In *International Workshop on Systems and Frameworks for Computational Morphology.* Springer, Berlin, Heidelberg. ٢٠١١. pp. ٩٨-١١٨.

A sentimental education: Sentiment analysis using Pang, **subjectivity summarization based on minimum cuts**. Bo, and Lillian Lee. In *Proceedings of the ٤٢nd annual meeting on Association for Computational Linguistics. Association for Computational Linguistics*. ٢٠٠٤.

A system for real-time twitter sentiment analysis of ٢٠١٢ us presidential election cycle. Wang, Hao, Dogan Can, Abe Kazemzadeh, François Bar, and Shrikanth Narayanan. In *Proceedings of the ACL ٠١٢ System Demonstrations*. Association for Computational Linguistics. ٢٠١٢. pp. ١١٥-١٢٠.

An opinion analysis tool for colloquial and standard Arabic. Al-Kabi, Mohammed, et al. In *The Fourth International*

Conference on Information and Communication Systems (ICICS pp. ٢٣-٢٥..٢٠١٣*)*. ٢٠١٣

Annotating expressions of opinions and emotions in language. Wiebe, Janyce, Theresa Wilson, and Claire Cardie. In *Language resources and evaluation*. ٢٠٠٥. ٣٩(٢). pp. ١٦٥-٢١٠.

Arabic Gigaword Fifth Edition. **Linguistic Data Consortium**. **University of Pennsylvania**. ٢٠١١. **https://catalog.ldc.upenn.edu/LDC٢٠١١T١١. Accessed ٢٩ October ٢٠٢١**.

Arabic sentiment analysis: Lexicon-based and corpus-based. Abdulla, Nawaf A., et al. In *IEEE Jordan conference on applied electrical engineering and computing technologies* pp. ١-٦..*(AEECT)*. IEEE, ٢٠١٣

Finite-State Morphology. **Beesley, K. and Karttunen, L. CSLI, Stanford**. ٢٠٠٣.

Learning to identify emotions in text. Strapparava, Carlo, and Rada Mihalcea. In *Proceedings of the ٢٠٠٨ ACM symposium on Applied computing*. ACM. ٢٠٠٨. pp. ١٥٥٦-١٥٦٠.

Sentence-level Arabic sentiment analysis. Shoukry, Amira, and Ahmed Rafea. In *International Conference on Collaboration .Technologies and Systems (CTS)*. IEEE. ٢٠١٢. pp. ٥٤٦-٥٥٠

Text Mining: A Guidebook for the Social Sciences. Ignatow, Gabe, and Rada Mihalcea. SAGE Publications. ٢٠١٦.

Towards improving the lexicon-based approach for arabic sentiment analysis. Abdulla, Nawaf A., et al. In *International*

Journal of Information Technology and Web Engineering

.(IJITWE). ٢٠١٤. ٩(٣). p.p. ٥٥-٧١

Using Wikipedia for Automatic Word Sense

Mihalcea, Rada. In *HLT-NAACL*. ٢٠٠٧. **Disambiguation**.

pp.١٩٦-٢٠٣.

* * *