



The Engineering Company  
for the Development of Digital Systems.

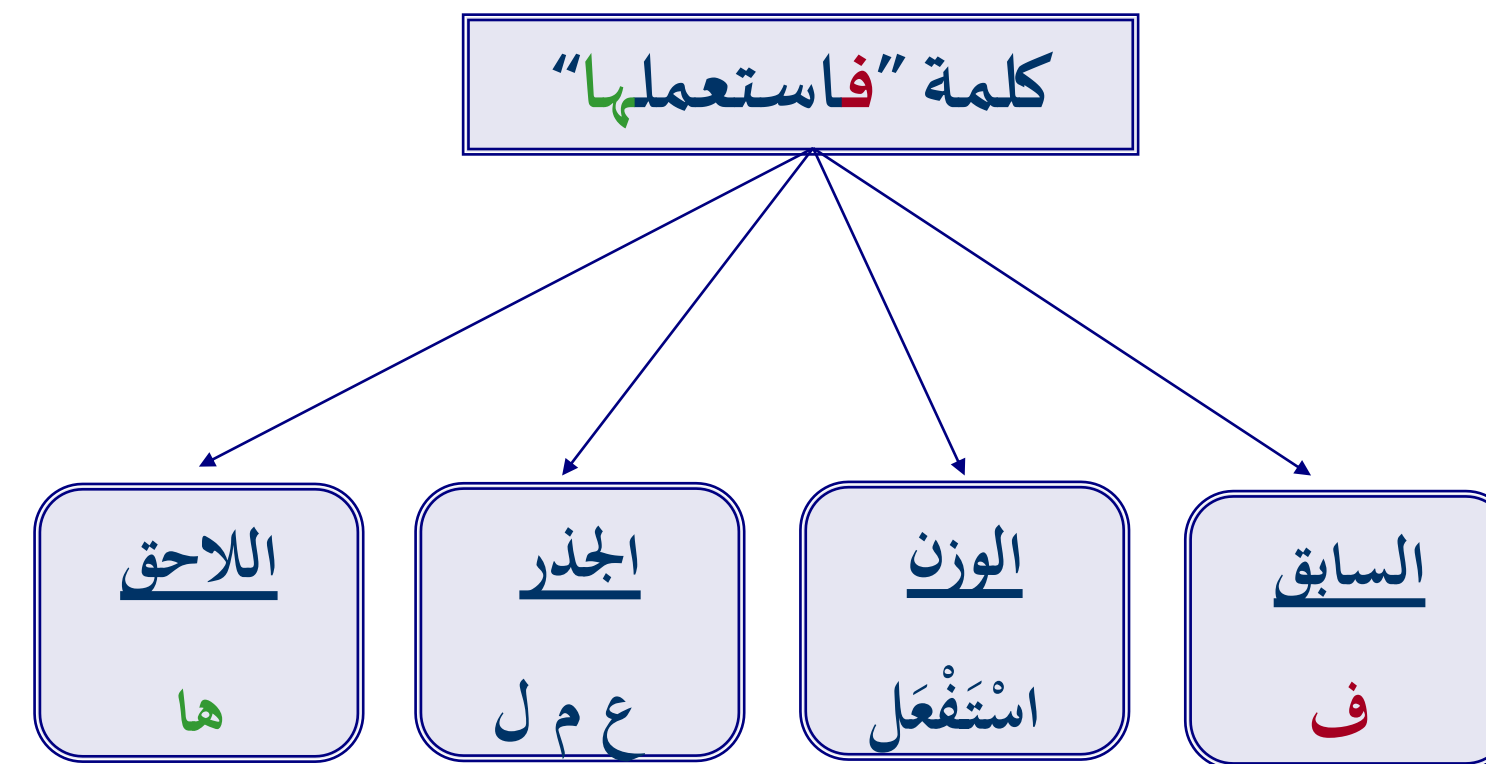
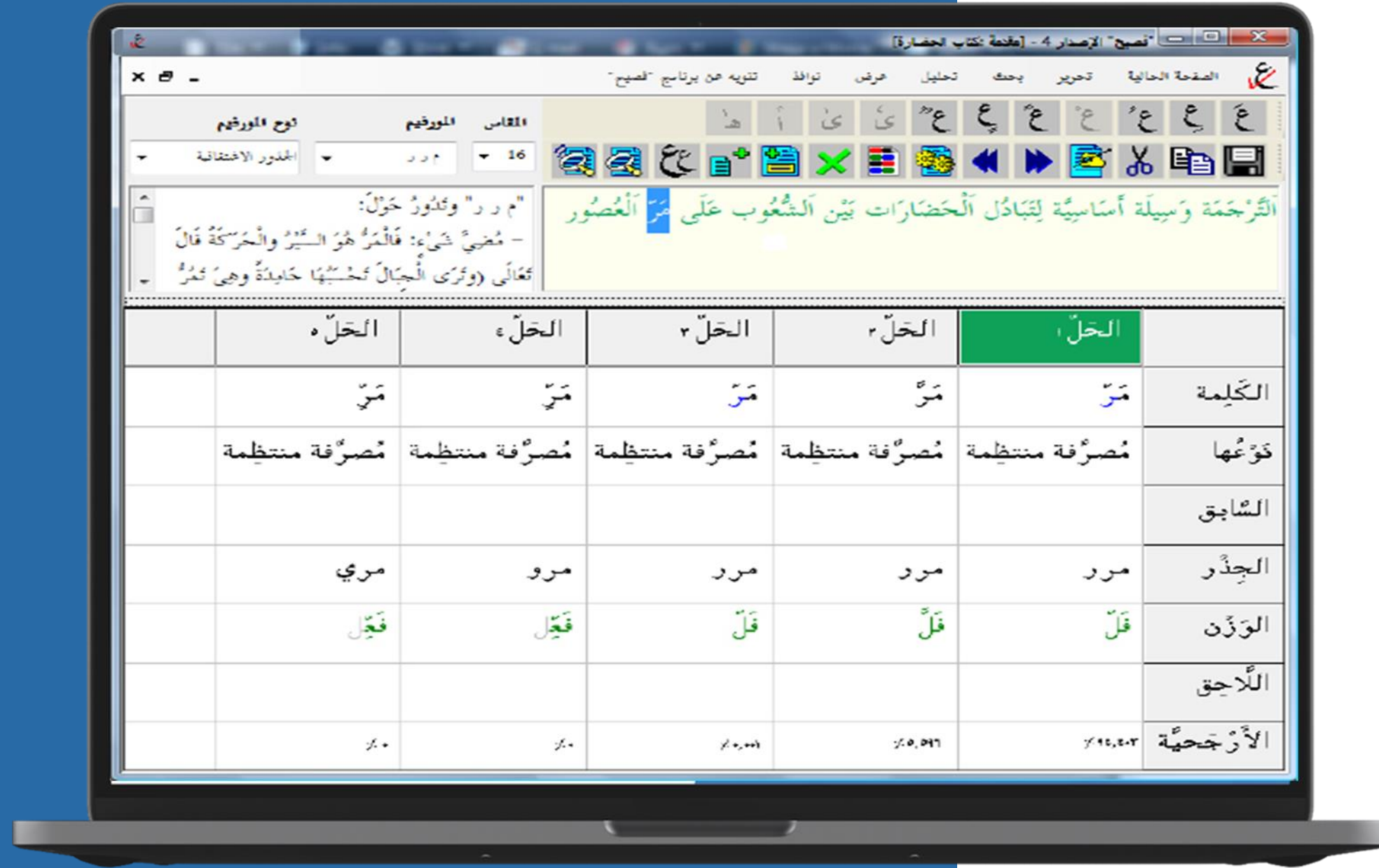
تقنيات معالجة اللغة العربية

**Arabic Natural Language Processing**

# الميزان

الميزان هو آلية للتحليل الصرفي العربي، يقوم بتغطية شاملة لمفردات اللغة العربية بمستوياتها التاريخية والجغرافية المختلفة. تتميز آلية التحليل الصرفي العربي «الميزان» بإمكانية زيادة تدريب النظام (حسب الطلب) على نوع معين من النصوص لزيادة دقة التحليل. والتعامل مع الأخطاء الهجائية الشائعة وتصويبها. والتعامل بمرونة مع الكلمات المدخلة؛ سواء أكانت مضبوطة أم لا. وكذلك توليد التصنيفات الدلالية المعجمية الممكنة لكل كلمة تم تحليلها صرفيا.

تقوم تقنية الميزان بتحليل كلماتها العربية إلى لبناتها الصرفية وتحديد نوع كل كلمة وجذرها الاشتقاقي ووزنها الصرفي وسوابقها ولواحقها.



# بعض الأبحاث ذات الصلة

- Attia, M., Rashwan, M., Al-Badrashiny, M., *Fassieh; a Semi-Automatic Visual Interactive Tool for the Morphological, PoS-Tags, Phonetic, and Semantic Annotation of the Arabic Text*, IEEE Transactions on Audio, Speech, and Language Processing (TASLP) Vol. 17 - Issue 5; pp. 916 to pp. 925, July 2009, [http://ieeexplore.ieee.org/xpl/freeabs\\_all.jsp?isnumber=5067414&arnumber=5075778&count=21&index=6](http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?isnumber=5067414&arnumber=5075778&count=21&index=6)
- Attia, M., *A Large-Scale Computational Processor of the Arabic Morphology, and Applications*, M.Sc. thesis, Dept. of Computer Engineering, Faculty of Engineering, Cairo University, 2000

# السراج

## آلية العنونة الصرفية للنص العربي

السراج هذه التقنية أساس لا غنى عنه كبداية أساسية لعمل أي محلل نحوي عربي **Arabic syntax analyzer** حيث يتم إدخال نص عربي خام مثل نصوص الجرائد، ومن ثم تقوم هذه الآلية بتوليد مجموعة من العناوين (tags) لكل كلمة تدل على صفاتها النحوية.

تتميز آلية العنونة الصرفية للنص العربي بأنها تعطينا إمكانية تمييز / عنونة أجزاء من الكلام، ومن هنا فهي عملية ضرورية لتوليد الأجزاء التي تحمل الخصائص الصرفية المتحررة من السياق من كلمات النص البسيطة. وإلى جانب مهام المعالجة اللغوية التي تعد العنونة اللغوية ذات نفع بالغ لها؛ فهناك منافع أخرى حيث تعد العناوين الناتجة مدخلا لجميع المعالجات الحاسوبية للتقطيع النحوي الذي يعد بدوره درجة في سلم فهم اللغات والترجمة الآلية.

أهمّ الخصائص والإمكانات :

- تحديد الترتيب الأصلي الافتراضي للكلمات داخل الجمل.
- إجراء التحويلات التركيبية على بنية الجملة، كالتقديم والتأخير، والحذف والإضمار .. وما إليها.
- إضافة التعديلات النحوية التي تحدد وظيفة الكلمات داخل الجمل، كعمليات الضبط الإعرابي، ومطابقة الفعل مع فاعله، وكذلك تلك التعديلات الناجمة عن التعديلات التركيبية.

وكانت بداية هذه التقنية من رسالة الماجستير للدكتور محمد عطية تحت إشراف د. محسن رشوان.

# العين

## قاعدة بيانات دلالية ومعجمية للغة العربية

العين تقوم هذه التقنية على ربط معظم مفردات اللغة العربية بعضها البعض، اعتمادا على نظرية الحقول الدلالية التي تقوم على أن مفردات اللغة تنقسم إلى مجموعات دلالية، تضم كل مجموعة منها جملة من المركبات الصرفية، ويمكن رصد علاقة هذه المفردات من خلال إدراج علاقات منطقية بين هذه المجموعات، وأبرز هذه العلاقات هي: الترادف، والتضاد، والكلية، والجزئية، والاشتمالية، والسببية، والمسببية، والحالية، والمحلية ... إلخ، وقد وصل حجم قاعدة البيانات إلى مائة ألف كلمة تقريبا.

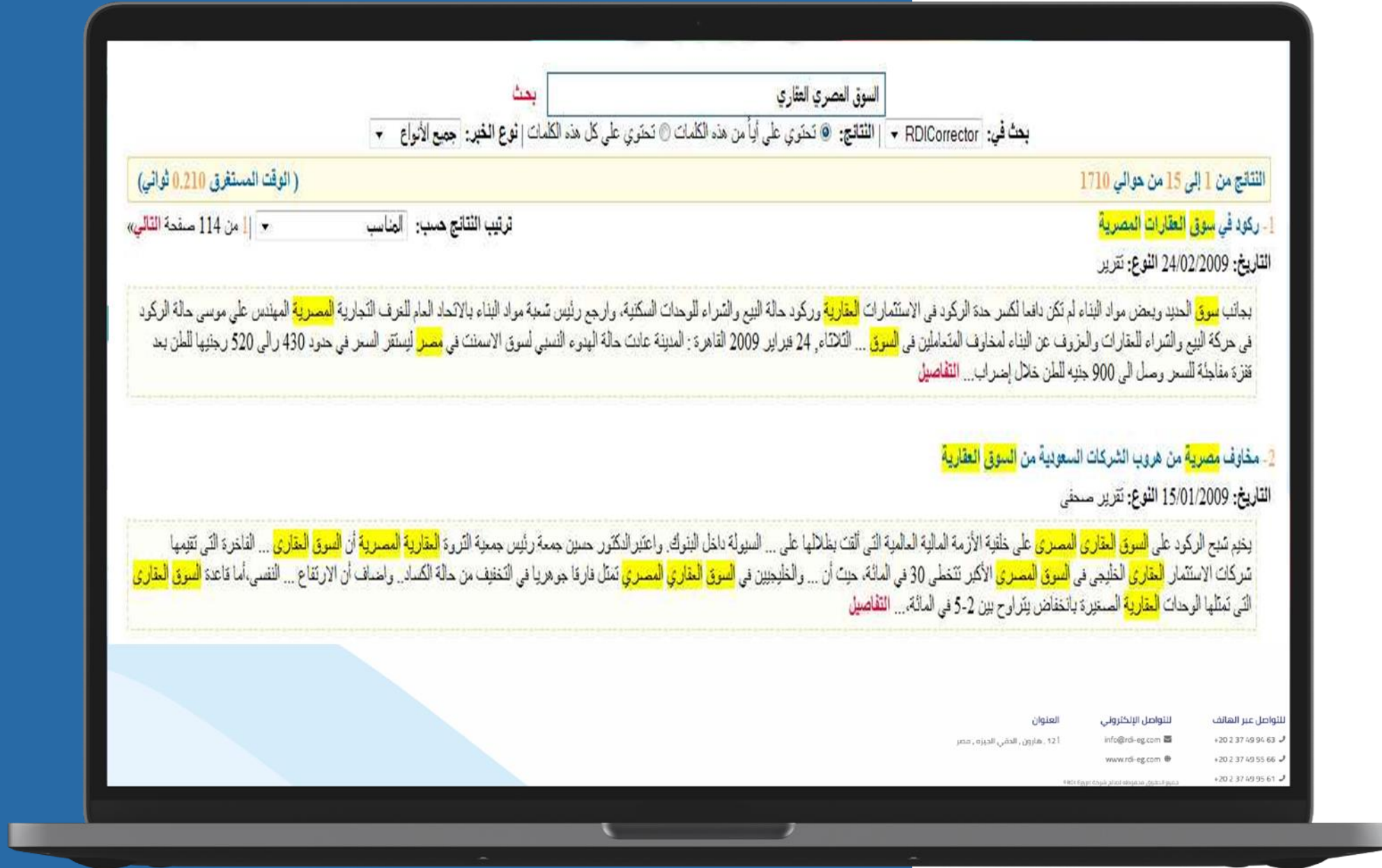
وقد اعتمدت قاعدة البيانات الدلالية والمعجمية على مجموعة من العلاقات التي تربط ألفاظ اللغة العربية تمثلت في عشرين علاقة هي:

علاقة الترادف	علاقة النوعية الاشتمالية	علاقة الأصلية الاشتمالية	علاقة المسببية
علاقة التضاد	علاقة الاشتمالية العضوية	علاقة الاشتمالية الأصلية	علاقة الحالية المكانية
علاقة الكلية	علاقة العضوية الاشتمالية	علاقة الشرطية	علاقة المحلية المكانية
علاقة الجزئية	علاقة الاشتمالية التكاملية	علاقة المشروطية	علاقة الحالية الزمانية
علاقة الاشتمالية النوعية	علاقة التكاملية الاشتمالية	علاقة السببية	علاقة المحلية الزمانية

Possible Semantic fields			Lexical Compound				
الحقل الدلالي 3	الحقل الدلالي 2	الحقل الدلالي 1	اللاحق	الجزر	السابق	النوع	الكلمة
المراسلة	التأليف	الاستثمار	لا يوجد	ك ت ب	لا يوجد	فعل	كتب
المراسلة	الإلزام	التأليف	لا يوجد	ك ت ب	لا يوجد	اسم	كتاب
		الاستثمار	لا يوجد	ك ت ب	لا يوجد	فعل	اكتب
		...	...	...	...	...	...

Reference:

<https://aclanthology.org/L08-1611/>



كاشف هو محرك بحث عربي ذكي لاسترجاع الملفات، والبحث عن المعلومات خلال البيانات الإلكترونية والمستندات وقواعد البيانات ومواقع الإنترنت. محرك البحث كاشف هو نتاج الأبحاث المتواصلة لشركة RDI في مجال المعالجة الطبيعية للغة العربية (NLP)، مما يجعله بالإضافة إلى تفرده في البحث باللغة العربية محركاً مرناً وواسع الاستخدام.

يستخدم محرك البحث الاشتقاقي والدلالي العربي «كاشف» تقنية المعالجة الطبيعية للغة في تصحيح الأخطاء الإملائية، كما يتيح لمستخدميه القيام بالبحث باللغة العربية باستخدام أصل الكلمة أو جذرها أو أحد مشتقاتها خلال كم هائل من البيانات؛ ليحصلوا على نتائج بحث غاية في الدقة.

يمكن للمؤسسات والشركات من خلال «كاشف» رسم صورة كاملة وذكية وآمنة ودقيقة لكل المعلومات لديهم في نظم الملفات أو مواقع الويب أو قواعد البيانات أو القواعد المعرفية.

## ورقة البحث:

- Aya Abed, Fatma Tarek, Marwa Farid, Mariam Emad, Mohsen A. Rashwan, Abdullah M. Rashwan; **SEARCH ENGINE FOR SCANNED ARABIC DOCUMENTS**; ALTIC -2011, Alexandria, Egypt  
( <https://cu.edu.eg/userfiles/ALTIC.pdf> )

# المميزات

فهرسة كم هائل من البيانات بسرعة فائقة.



البحث عن موضوعات مماثلة وذات صلة  
"Relevance Feedback Search".



يدعم «كاشف» القيام بجميع وظائفه من خلال الويب؛  
تمكن واجهة الاستخدام عن طريق الويب للمؤسسات بناء  
تطبيقاتها البحثية وتوزيعها عن بعد.



البحث بالمتراادفات اللغوية " Lexical Semantics  
"Search".



يمكن تخصيص محرك البحث الاشتقاقي والدلالي العربي  
«كاشف» وفقا لمتطلبات العميل



التعامل مع المصادر المختلفة للبيانات مثل (نظم الملفات -  
الأسطوانات المدمجة - الكتب التفاعلية - الشبكات LAN - قواعد  
البيانات - المواقع الإلكترونية)





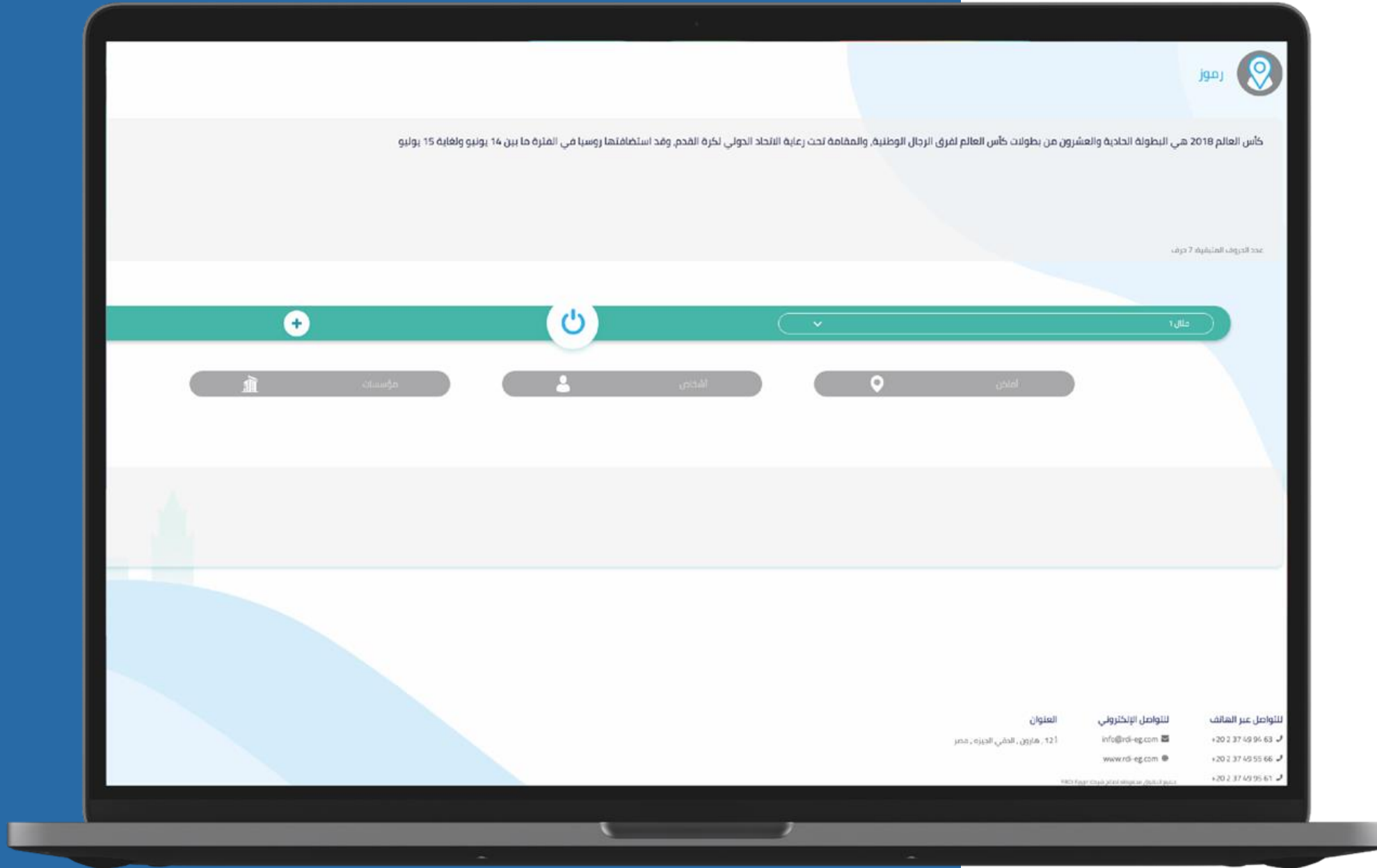
# رموز

رموز Named-Entity Recognition هي تقنية لتحليل النص والتعرف على الكيانات المسماة وتصنيفها إلى فئات محددة مسبقاً، ويستخلص "رموز" ثلاثة أنواع مختلفة من ملفات الكيانات المسماة:

- الأشخاص
- الأماكن
- المنظمات

يمكن توفير رموز بتنسيقات مختلفة طبقاً لمتطلبات العميل:

- الموقع الإلكتروني
- تطبيق IOS/Android
- واجهة برمجة تطبيقات API



ورقة البحث:

Amr M. Sayed, Sherif Abdou, **Mohsen Rashwan**, Hassanin Al-Barhamtoshy; *RANER: RDI Framework for Arabic Named Entity Recognition*; International Journal of Engineering & Technology, Vol. 8, pp. 161-164, 2019.

<https://www.sciencepubco.com/index.php/ijet/article/view/28187>

# المميزات

"رموز" يقبل النص المكتوب باللغة العربية الفصحى الحديثة.



يعتمد "رموز" على أحدث أساليب التعلم العميق، لضمان نتائج ذات دقة عالية.



ممكن استلام تنسيقات الإخراج للكيانات المسماة المعروفة بتنسيق JSON.



يستهدف رموز مجال الأخبار المصرية بشكل أساسي



## قابل للتطوير

## مميزات مستقبلية

سهل التكيف على المجالات وبيئات العمل المختلفة طبقاً لاحتياجات العميل.



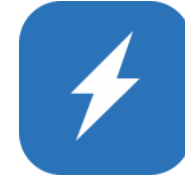
يحترم "رموز" خصوصية عملائه، ولا يشارك الملفات الخاصة بالمستخدمين مع أي طرف ثالث.



يمكن تطبيقه من جهة العميل مع خدمة التخصيص والدعم الكاملة.



سيكون الإصدار المستقبلي أكثر سرعة من الإصدار الحالي.

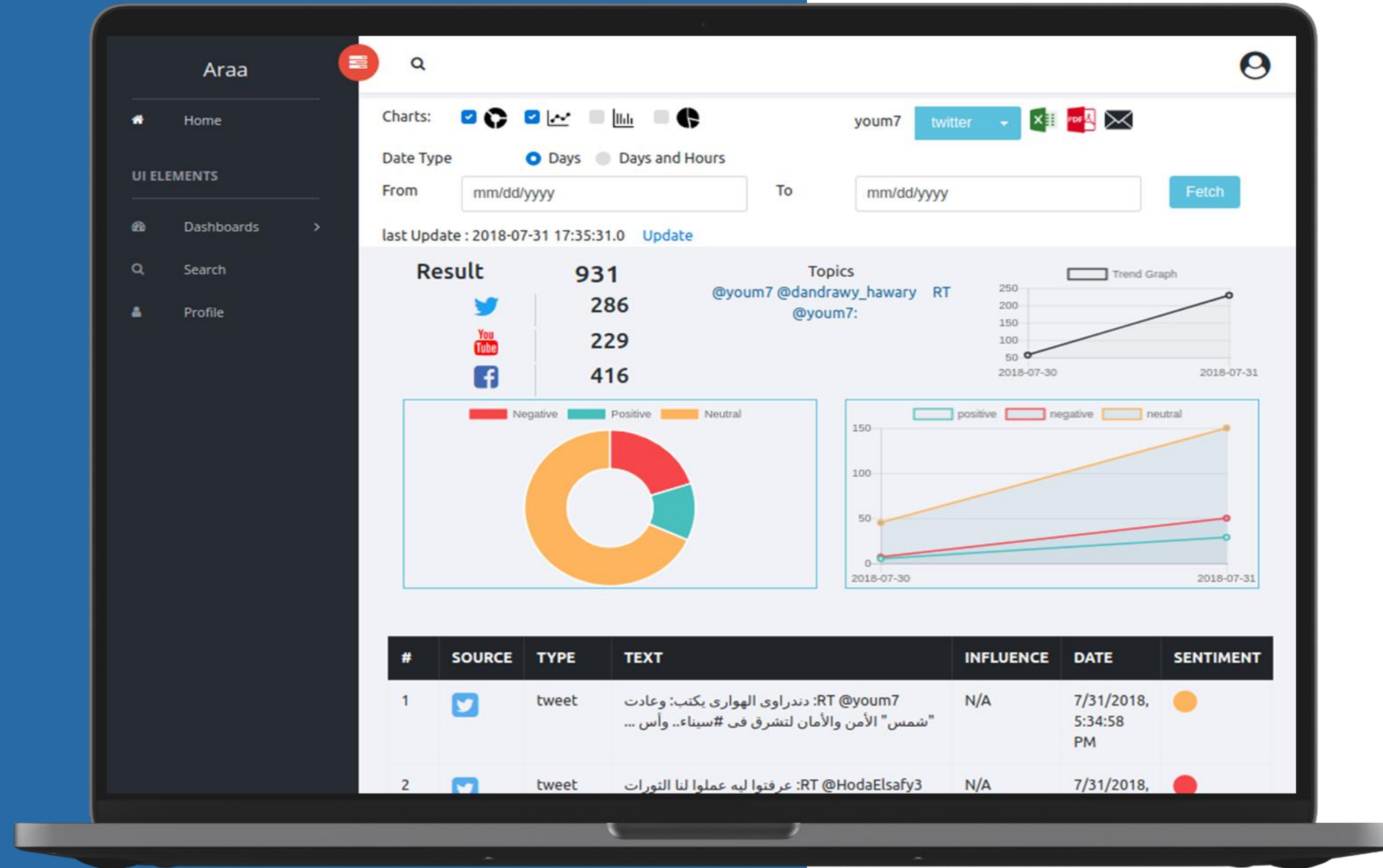


دعم التعرف على المزيد من أنواع الكيانات المسماة





## آراء



آراء هو حل متكامل متعدد الموارد، يدمج تقنيات تحويل الكلام إلى نص تلقائيًا، وتقنيات تحليل الآراء، وتقنيات معالجة اللغة العربية الطبيعية NLP، لتحليل الآراء من المحتوى العربي من نصوص وسائل التواصل الاجتماعي العربية، وتسجيلات الصوت والفيديو. يضمن آراء نظرة ثاقبة للمحتوى من خلال تصنيف المواضيع، وتحليل الآراء والتوجهات بالإضافة إلى التحليل الآلي للمشاعر.

يمكن توفير آراء بتنسيقات مختلفة طبقاً لمتطلبات العميل

- الموقع الإلكتروني
- SDK
- واجهة برمجة تطبيقات API

# المميزات

تشمل مصادر البيانات المدعومة - على سبيل المثال لا الحصر - Tweeter و Facebook و You Tube وقنوات التلفزيون والمكالمات الهاتفية المسجلة وقاعدة البيانات المملوكة للعملاء.



يمكن تكييف آراء لتناسب احتياجات العملاء، بحيث يضمن دقة تحليل أعلى مقارنة بالمحركات العامة.



التعرف على ما إذا كان الرأي المعبر عنه داخل المحتوى إيجابياً أم سلبياً أم محايداً.



يمكن استخدام آراء كمحرك أساسي لبناء حلول مخصصة كاملة مثل منصات تحليل المشاعر أو الآراء.



تساعدك آراء على اكتشاف رؤى حيوية ضمن مليارات المحادثات العربية التي تحدث كل يوم. يمكن استخدامه لتحسين تواصل العميل مع عملائه والكشف عن المعلومات البعيدة عن الأنظار.



# بعض الأبحاث ذات الصلة

- H. S. Ibrahim, S. M. Abdou, and M. Gheith, ***Sentiment Analysis for Modern Standard Arabic and Colloquial***, International Journal on Natural Language Computing (IJNLC), vol. 4, pp. 95-109, 2015.
- H. S. Ibrahim, S. M. Abdou, and M. Gheith, ***Idioms-Proverbs Lexicon for Modern Standard Arabic and Colloquial Sentiment Analysis***, International Journal of Computer Applications (IJCA), vol. 118, 2015.
- H. S. Ibrahim, S. M. Abdou, and M. Gheith, ***MIKA: A Tagged corpus for Modern Standard Arabic and Colloquial Sentiment Analysis***, In proceeding of the 2nd IEEE International Conference on Recent Trends in Information Systems (ReTIS-15), Jadavpur University, Kolkata, India, 2015.
- H. S. Ibrahim, S. M. Abdou, and M. Gheith, ***Automatic expandable large-scale sentiment lexicon of Modern Standard Arabic and Colloquial***, in 16th International Conference on Intelligent Text Processing and Computational Linguistics (CICLING), Cairo - Egypt, 2015.
- Abdelrahman, S., Mobarz, H., Farag, I., Rashwan, M., ***Arabic Phrase-Level Contextual Polarity Recognition to Enhance Sentiment Arabic Lexical Semantic Database Generation***, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 5, No. 10, PP. 32-36, 2014.



## التلخيص الآلي للنصوص

دعت الحاجة العملية إلى هذه التقنية وتزايدت بعد الانفجار المعرفي الحالي والمنتظر زيادته مستقبلا. فإذا أمكن التعرف على أهم ما ورد في ورقة بحثية أو كتاب في موضوع ما بدلا من استقطاع وقت كبير لقراءته بالكامل، فربما يفي الجزء المركز والمختصر بالغرض أو يقتنع المستخدم لهذه التقنية بقراءة البحث أو الكتاب كاملا لبلوغ الهدف.

تعمل تقنية التلخيص الآلي للنصوص على اختيار السياقات الأبرز في النص بأي نسبة يحتاجها المستخدم.