

ملخص:

ان النمو السريع لمصادر المعلومات المتصلة بالانترنت من قواعد بيانات وغيرها ادى الى حدوث ثورة في عالم البيانات. هذا أدى الى وجود كم هائل من البيانات المخزنة في وثائق ونصوص في لغات مختلفة . حيث اصبحت من اهم القضايا التي تؤرق العالم، الامر الذي ادى الى زيادة الطلب على التطبيقات البرمجة اللغوية الذكية من فهم واسترجاع وترجمة وتلخيص هذه الكميات الكبيرة من النصوص والتي تتطلب دقة في معالجتها، خاصة الجانب اللغوي. تعتبر المكنز اللغوية من الأمثلة على المصادر اللغوية المهمة ومن المكونات الأساسية للعديد من التطبيقات الذكية خاصة لدعم تعدد اللغات.

على الرغم من وجود العديد من المكنز الخاصة باللغة العربية، الا ان غالبيتها غير دقيقة خاصة وان علاقة المترادفات فيها تقريبية وهي اقرب لان تكون علاقة عامة بين الكلمات. مثلا قاموس غوغل للعربية، والمترادفات للكلمة "دولة" هي {بلد، قطر، وطن، الريف، ريف}. هنا الكلمات التي تحتها خط خاطئة، لأنها ليست مرادفات صحيحة.

تهدف هذه الأطروحة الى بناء مكنز للغة العربية بطريقة آلية، ومن ثم ربط مجموعات المترادفات في هذا المكنز بمقابلاتها في شبكة الدلالية للكلمات الانجليزية (English WordNet) بمعنى آخر، نأخذ الكلمات العربية المرادفة التي تم بناءها من المكنز و من ثم ربطها بالمترادفات (English WordNet) ، مثلا {ب.ع.....، ع₂، ع₁} := {ج.ن.....، ن₂، ن₁}. لعمل ذلك، قمنا بتطبيق خوارزمية SynsetGenerator لبناء المكنز بطريقة آلية ، حيث تحتاج الخوارزمية الى مجموعة من القواميس ثنائية اللغة عربي - إنجليزي كمدخل، وتم ربط النتائج مع (English WordNet) وتقييمها باستخدام مبدأ تشابه جيب التمام (Cosine Similarity).