

## سيرة ذاتية

### المعلومات الشخصية

الاسم: عبدالمحسن بن عبيد بن عبدالله الثبيتي  
جهة العمل: معهد تقنيات الذكاء الاصطناعي والروبوت ، مدينة الملك عبدالعزيز للعلوم والتقنية.  
المرتبة العلمية: أستاذ بحث  
تاريخ الالتحاق بالمدينة: 1432 هـ  
هاتف: 00966114884148  
جوال: 00966505516387  
بريد الكتروني: aalthubaity@kacst.edu.sa

### التأهيل العلمي

1. بكالوريوس الهندسة الكهربائية والحاسبات- جامعة ام القرى - المملكة العربية السعودية - 1990 م .
2. ماجستير هندسة الحاسبات- جامعة الملك عبدالعزيز - المملكة العربية السعودية -العام الدراسي 1996م 1997م.
3. دكتوراه علوم الحاسب الآلي - جامعة ساري - بريطانيا - 20 سبتمبر 2004 م .

### الخبرات العملية

1. مهندس حاسب- الشركة السعودية الموحدة للكهرباء - من 6-9-1990 حتى 14-7-1992م.
2. ضابط مهندس- وزارة الدفاع-من 16-7-1992- حتى 14-3-2011م.
3. أستاذ بحث مساعد- مدينة الملك عبدالعزيز للعلوم والتقنية من 9-5-2011م حتى 31-8-2015م.
4. أستاذ بحث مشارك- مدينة الملك عبدالعزيز للعلوم والتقنية من 1-9-2015م - حتى 24-10-2020م.
5. أستاذ بحث - مدينة الملك عبدالعزيز للعلوم والتقنية من 25-10-2020م -حتى تاريخه.
6. مستشار غير متفرغ- الهيئة السعودية للبيانات والذكاء الاصطناعي- 2-5-2023- حتى تاريخه

### الاهتمامات العلمية

1. المصادر اللغوية العربية (المدونات اللغوية وتوسيمها، أدوات معالجة المدونات، تقييم المصادر اللغوية)
2. التنقيب في النصوص العربية (التصنيف والتجميع الآلي للنصوص، استخراج المعلومات)
3. اللسانيات الحاسوبية العربية (التحليل الصرفي، الوسم النحوي، التشكيل الآلي)
4. معالجة اللغة العربية (النماذج اللغوية، تحليل الآراء في شبكات التواصل الاجتماعي، التصحيح الآلي، التعرف على أسماء الأعلام)

## المشاريع العلمية

1. باحث - مشروع التصنيف الآلي للنصوص العربية -2007م-2008م. مدينة الملك عبدالعزيز للعلوم والتقنية - مشروع رقم 104-27-30.
2. باحث رئيس - مشروع المدونة العربية لمدينة الملك عبدالعزيز للعلوم والتقنية 2011م - 2012م - مدينة الملك عبدالعزيز للعلوم والتقنية - مشروع رقم 31-531.
3. باحث رئيس - مشروع إدارة وتطوير المدونة اللغوية العربية لمدينة الملك عبدالعزيز للعلوم والتقنية - 2013م - 2016م - مدينة الملك عبدالعزيز للعلوم والتقنية - مشروع رقم 33-824.
4. باحث رئيس - مشروع الاستخلاص الآلي للمصطلحات العلمية من النصوص العربية المتخصصة -2013م-2014م - مدينة الملك عبدالعزيز للعلوم والتقنية - مشروع رقم 33-821.
5. باحث رئيس - تحليل الآراء - 2017م- 2018م - مدينة الملك عبدالعزيز للعلوم والتقنية- مشروع بحثي داخلي.
6. باحث - مشروع المساعد العربي الافتراضي-2019م -2021م - مدينة الملك عبدالعزيز للعلوم والتقنية - ضمن مبادرة البيانات الضخمة.
7. باحث رئيس - مشروع مدونة كوفيد 19 - 2020م- حتى الآن - مدينة الملك عبدالعزيز للعلوم والتقنية- مشروع بحثي داخلي.
8. باحث رئيس - مشروع منصة لتوسيم بيانات تطبيقات الذكاء الاصطناعي - 2020م- حتى الآن- مدينة الملك عبدالعزيز للعلوم والتقنية- مشروع بحثي داخلي.
9. باحث رئيس- مشروع الكشف عن أسماء الأعلام- 2021م- حتى نهاية شهر مايو 2023م- مدينة الملك عبدالعزيز للعلوم والتقنية مع جهات أخرى- مشروع بحثي داخلي.
10. باحث رئيس- مشروع الكشف عن المصطلحات العلمية في النصوص المتخصصة باستخدام المحولات- 2023م- حتى تاريخه - مدينة الملك عبدالعزيز للعلوم والتقنية - مشروع بحثي داخلي.

## الابحاث العلمية المنشورة باللغة العربية

1. الشبتي، عبدالمحسن و الحربي، سامي. (2006). تقنيات إدارة المعرفة في المؤسسات الأمنية . مؤتمر الأمن والديموقراطية وحقوق الإنسان. جامعة مؤتة. عمان (الأردن).
2. الشبتي، عبدالمحسن. (2007). استخدام ذخائر النصوص لاستخلاص المصطلحات المتخصصة. الندوة الدولية الأولى عن الحاسب واللغة والعربية (ص ص 31-38). مدينة الملك عبدالعزيز للعلوم والتقنية بالتعاون مع جمعية الحاسبات السعودية. الرياض (المملكة العربية السعودية).
3. الشبتي، عبدالمحسن. الطوالة، ندى. المرشدي، بشاير و العتيبي، سعد. (2012). طريقة تعتمد على المدونات اللغوية لتجهيز بيانات تدريب واختبار أنظمة الوسوم النحوية. المؤتمر الدولي لعلوم وهندسة الحاسوب باللغة العربية في دورته الثامنة. القاهرة (جمهورية مصر العربية).
4. الشبتي، عبدالمحسن. هزاع، اعتصام. و الحوشان، منيرة. (2013). تأثير استخدام المتتابعات اللفظية كصفات ممثلة للنصوص على دقة التصنيف الآلي للنصوص العربية. الدورة التاسعة للمؤتمر الدولي لعلوم وهندسة الحاسوب باللغة العربية: ايكأ 2013. الحمامات (تونس).
5. الشبتي، عبدالمحسن. (2014). جهود مدينة الملك عبدالعزيز للعلوم والتقنية في خدمة اللغة العربية تقنياً. ملتقى التعليم الإلكتروني في خدمة اللغة العربية. جامعة الأميرة نورة بنت عبدالرحمن. الرياض (المملكة العربية السعودية).

6. الثبيتي، عبدالمحسن. (2014). نحو إطار عام لمدونة لغوية للمعجم التاريخي للغة العربية. فصل من كتاب نحو معجم تاريخي للغة العربية. (ص ص 291-313) المركز العربي للأبحاث ودراسة السياسات. الدوحة (قطر).
7. الثبيتي، عبدالمحسن. (2015). تصميم المدونات اللغوية وبنائها. في صالح فهد العصيمي (محرر.)، المدونات اللغوية العربية: بناؤها وطرق الإفادة منها. (ص ص 147-178) مركز الملك عبدالله بن عبدالعزيز الدولي لخدمة اللغة العربية. الرياض (المملكة العربية السعودية).
8. الشمري، عقيل. الثبيتي، عبدالمحسن. (2015). لسانيات المدونات: نماذج وتطبيقات في الصحافة العربية. في صالح فهد العصيمي (محرر.)، المدونات اللغوية العربية: بناؤها وطرق الإفادة منها. (ص ص 179-233) مركز الملك عبدالله بن عبدالعزيز الدولي لخدمة اللغة العربية. الرياض (المملكة العربية السعودية).
9. الثبيتي، عبدالمحسن. (2016). الكلمات المُمَيَّزَةُ للمدونات اللغوية: قضايا تقنية. في سلطان المجبول (محرر.)، لغويات المدونة الحاسوبية، تطبيقات تحليلية على العربية الطبيعية. (ص ص 92-35) مركز الملك عبدالله بن عبدالعزيز الدولي لخدمة اللغة العربية. الرياض (المملكة العربية السعودية).
10. العصيمي، صالح. الثبيتي، عبدالمحسن. (2019). نحو قائمة للألفاظ الأكاديمية في العربية المعاصرة: دراسة قائمة على المدونات. مجلة العلوم الشرعية واللغة العربية. ع 14، (4)، 403-431.
11. المجدوب، عزالدين. الثبيتي، عبدالمحسن. الاحم، إبراهيم. كرونة، سندس. (2019). الوسم النحوي الآلي للعربية في منهجية بنك المشجرات النحوية في جامعة بنسيلفانيا. مجلة اللسانيات العربية. ع 9، 6-78.
12. العصيمي صالح، الثبيتي عبدالمحسن. (2021). الكتابة العلمية المتخصصة في المقترحات البحثية لعلم اللغة التطبيقي: دراسة قائمة على المدونات. مجلة جامعة الشارقة للعلوم الانسانية والاجتماعية، 18، 195-222. (2A), <https://doi.org/10.36394/jhss/18/2A/8>
13. الثبيتي، عبدالمحسن. (2021). دليل لتوسيم البيانات اللغوية. منشورات مدينة الملك عبدالعزيز للعلوم والتقنية. (متوقع ظهوره مع اليوم العالمي للغة العربية 18 ديسمبر)
14. الثبيتي، عبدالمحسن، والقحطاني، وداد. (2022). طريقة مقترحة لاستخلاص المصطلحات من المدونات اللغوية العربية المتخصصة. مجلة اللسانيات العربية، 14، 35-56.
15. الثبيتي، عبدالمحسن، والقحطاني، وداد. (2023). الخواص الممكّنة لاستخلاص المصطلحات من المدونات العربية المتخصصة. مجلة اللسانيات العربية. 16، 179-209.

## الابحاث العلمية المنشورة باللغة الإنجليزية

1. Rushdi, A. M., & Al-Thubaity, A. O. (1993). Efficient computation of the sensitivity of k-out-of-n system reliability. *Microelectronics Reliability*, 33(13), 1963-1979.
2. Ahmad, K., Schierz, A., & Al-Thubaity, A. (2002). Discovery and terminology. In *TKE 2002: terminology and knowledge engineering (Nancy, 28-30 August 2002)* (pp. 1-6).
3. Al-Thubaity, A., & Ahmad, K. (2002, July). Tracking knowledge of emergent domains. In *Proceedings Sixth International Conference on Information Visualisation* (pp. 685-690). IEEE.
4. Al-Thubaity, A., & Ahmad, K. (2003, July). Knowledge maps as lexical signatures of journals papers and patent documents. In *Proceedings on Seventh International Conference on Information Visualization, 2003. IV 2003.* (pp. 582-588). IEEE.

5. Ahmad, K., & Al-Thubaity, A. (2003, July). Can text analysis tell us something about technology progress?. In *Proceedings of the ACL-2003 workshop on Patent corpus processing* (pp. 46-55).
6. Al-Thubaity, A., Almuhareb, S., Al-Harbi, A., Al-Rajeh, M. S., Khorshed, (2008). KACST Arabic Text Classification Project: Overview and Preliminary Results. *9th IBIMA Conference on Information Management in Modern Organizations* (pp. 1239-1244).
7. S. Al-Harbi, A. Almuhareb, A. Al-Thubaity, M. S. Khorshed, A. Al-Rajeh. (2008). Automatic Arabic Text Classification. *Proceedings of The 9th International Conference on the Statistical Analysis of Textual Data* (pp. 77-83).
8. Al-Thubaity, A., Alanazi, A., Hazzaa, I., & Al-Tuwaijri, H. (2012, November). Weirdness Coefficient as a Feature Selection Method for Arabic Special Domain Text Classification. In *Asian Language Processing (IALP), 2012 International Conference on* (pp. 69-72). IEEE.
9. Khorshed, M. S., & Al-Thubaity, A. O. (2013). Comparative evaluation of text classification techniques using a large diverse Arabic dataset. *Language Resources and Evaluation*, 47(2), 513-538.
10. Al-Thubaity, A., Abanumay, N., Al-Jerayyed, S., Alrukban, A., & Mannaa, Z. (2013, July). The Effect of Combining Different Feature Selection Methods on Arabic Text Classification. In *Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), 2013 14th ACIS International Conference on* (pp. 211-216). IEEE.
11. Al-Thubaity, A., Khan, M., Al-Mazrua, M., & Al-Mousa, M. (2013, August). New Language Resources for Arabic: Corpus Containing More Than Two Million Words and a Corpus Processing Tool. In *Asian Language Processing (IALP), 2013 International Conference on* (pp. 67-70). IEEE.
12. Al-Thubaity, A. M., Khan, M., Alotaibi, S., & Alonazi, B. (2014, October). Automatic arabic term extraction from special domain corpora. In *2014 International Conference on Asian Language Processing (IALP)* (pp. 1-5). IEEE.
13. Al-Thubaity, A., Alhoshan, M., & Hazzaa, I. (2015). Using word n-grams as features in Arabic text classification. In *Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing* (pp. 35-43). Springer, Cham.
14. Al-Thubaity, A. O. (2015). A 700M+ Arabic corpus: KACST Arabic corpus design and construction. *Language Resources and Evaluation*, 49(3), 721-751.
15. Al-Thubaity, A., Al-Khalifa, H., Alqifari, R., & Almazrua, M. (2014). Proposed framework for the evaluation of standalone corpora processing systems: an application to Arabic corpora. *The Scientific World Journal*, 2014, ., Article ID 602745, 10 pages, 2014. doi:10.1155/2014/602745
16. Al-Thubaity, A., & Al-Subaie, A. (2015, October). Effect of word segmentation on Arabic text classification. In *2015 International Conference on Asian Language Processing (IALP)* (pp. 127-131). IEEE.
17. Al-Thubaity, A., & Almujaivel, S. (2018). A quantitative inquiry into the keywords between primary and reference Arabic corpora. *Journal of Quantitative Linguistics*, 25(2), 122-141.
18. Al-Thubaity, A., & Baazeem, I. (2017). A collocation extraction tool and two language resources for MSA. *Procedia Computer Science*, 117, 23-29.
19. Al-Thubaity, A., Alqarni, A., & Alnafessah, A. (2018, April). Do words with certain part of speech tags improve the performance of arabic text classification?. In *Proceedings of the 2nd International Conference on Information System and Data Mining* (pp. 155-161).
20. Al-Thubaity, A., Alqahtani, Q., & Aljandal, A. (2018). Sentiment lexicon for sentiment analysis of Saudi dialect tweets. *Procedia computer science*, 142, 301-307.

21. **Al-Thubaity, A., Alharbi, M., Alqahtani, S., & Aljandal, A.** (2018, April). A Saudi dialect Twitter Corpus for sentiment and emotion analysis. In *2018 21st Saudi computer society national computer conference (NCC)* (pp. 1-6). IEEE.
22. **Al-Thubaity, A., & Alhoshan, M.** (2018, September). ARARSS: A System for Constructing and Updating Arabic Textual Resources. In *International Conference on Advanced Intelligent Systems and Informatics* (pp. 261-269). Springer, Cham.
23. **Almuhareb, A., Alsanie, W., & Al-Thubaity, A.** (2019). Arabic word segmentation with long short-term memory neural networks and word embedding. *IEEE Access*, 7, 12879-12887.
24. **Al-Thubaity, A., Alkhalifa, A., Almuhareb, A., & Alsanie, W.** (2020). Arabic diacritization using bidirectional long short-term memory neural networks with conditional random fields. *IEEE Access*, 8, 154984-154996.
25. **Al-Thubaity, A., Alkhereyf, S., & Bahanshal, A. O.** (2022, June). AraNPCC: The Arabic Newspaper COVID-19 Corpus. In *Proceedings of the 5th Workshop on Open-Source Arabic Corpora and Processing Tools with Shared Tasks on Qur'an QA and Fine-Grained Hate Speech Detection* (pp. 32-40).
26. **Mubarak, H., Al-Khalifa, H., & Al-Thubaity, A.** (2022, June). Overview of OSACT5 Shared Task on Arabic Offensive Language and Hate Speech Detection. In *Proceedings of the 5th Workshop on Open-Source Arabic Corpora and Processing Tools with Shared Tasks on Qur'an QA and Fine-Grained Hate Speech Detection* (pp. 162-166).
27. **Al-Khalifa, H., Elsayed, T., Mubarak, H., Al-Thubaity, A., Magdy, W., & Darwish, K.** (2022, June). Proceedings of the 5th Workshop on Open-Source Arabic Corpora and Processing Tools with Shared Tasks on Qur'an QA and Fine-Grained Hate Speech Detection. In *Proceedings of the 5th Workshop on Open-Source Arabic Corpora and Processing Tools with Shared Tasks on Qur'an QA and Fine-Grained Hate Speech Detection*.
28. **Alfraidi, T.; Abdeen, M.A.R.; Yatimi, A.; Alluhaibi, R.; Al-Thubaity, A.** The Saudi Novel Corpus: Design and Compilation. *Appl. Sci.* 2022, 12, 6648. <https://doi.org/10.3390/app12136648>.
29. **Al-Thubaity, A., Alkhereyf, S., Alzahrani, W., & Bahanshal, A.** (2022, December). CAraNER: The COVID-19 Arabic Named Entity Corpus. In *Proceedings of The Seventh Arabic Natural Language Processing Workshop (WANLP)* (pp. 1-10).

## تنظيم المؤتمرات وورش العمل

1. عضو اللجنة المنظمة لورشة عمل " Workshop on Free/Open-Source Arabic Corpora and Corpora " Ninth International Conference on Language Resources and Evaluation. في مؤتمر. Processing Tools
2. عضو اللجنة المنظمة لورشة عمل " 2nd Workshop on Arabic Corpora and Processing Tools " في 10th International Conference on Language Resources and Evaluation.
3. عضو اللجنة المنظمة للمؤتمر السعودي الدولي الرابع لتقنية المعلومات KACSTIT-16.
4. رئاسة جلسة علمية في المؤتمر الدولي الرابع لتقنية المعلومات عام 2016م.
5. عضو اللجنة المنظمة لورشة عمل " 5th Workshop on Arabic Corpora and Processing Tools " في مؤتمر 13th Edition of its Language Resources and Evaluation Conference.

## اللجان والاستشارات

1. عضوية اللجنة الاستشارية بمركز الملك عبدالله بن عبدالعزيز لخدمة اللغة العربية اعتباراً من 1439/3/18 هـ. (منتهى)
2. عضوية مجلس مركز الملك سلمان للغويات التطبيقية لعام 1438 هـ. (منتهى)
3. مستشار غير متفرغ لمشروع " تصميم وبناء مدونة لغوية-حاسوبية للروايات السعودية" في الجامعة الإسلامية بالمدينة المنورة اعتباراً من 1442/4/1 هـ. (منتهى).
4. مستشار غير متفرغ في الهيئة السعودية للبيانات والذكاء الاصطناعي. اعتباراً من 2023-5-2م. (مستمر)

## المحاضرات والندوات

1. محاضرة بعنوان "المدونة اللغوية العربية لمدينة الملك عبدالعزيز للعلوم والتقنية" في الملتقى التنسيقي للجامعات والمؤسسات المعنية باللغة العربية في دول مجلس التعاون لدول الخليج العربية - مركز الملك عبدالله الدولي لخدمة اللغة العربية بتاريخ 1434/6/29 هـ.
2. محاضرة بعنوان "المدونات اللغوية العربية: التحليل الصرفي والتوسيم النحوي" بمركز اللغويات التطبيقية - جامعة الإمام محمد بن سعود الإسلامية بتاريخ 1435/8/18 هـ.
3. محاضرة بعنوان "مبادرة الملك عبدالله للمحتوى العربي" في المنتدى الإقليمي الثاني للمحتوى الرقمي العربي في مدينة القاهرة بتاريخ 1436/1/20 هـ.
4. محاضرة بعنوان "المدونات اللغوية وأدواتها: تجارب ومحاولات" في جامعة الملك سعود لمجموعة إيوان البحثية بتاريخ 1437-2-18 هـ.
5. محاضرة بعنوان "جهود مدينة الملك عبدالعزيز للعلوم والتقنية في خدمة اللغة العربية" في مدينة الملك عبدالعزيز للعلوم والتقنية بتاريخ 1438/3/27 هـ.
6. محاضرة بعنوان "حوسبة اللغة العربية ومعالجتها آلياً: نماذج من أعمال مدينة الملك عبدالعزيز للعلوم والتقنية" في المؤتمر الدولي للغة العربية والنص الأدبي بجامعة الملك خالد بأبها بتاريخ 1438/5/18 هـ.
7. محاضرة بعنوان "نماذج من الأدوات الحاسوبية للتحليل اللغوي" في مؤسسة العربية للجميع بالرياض بتاريخ 1439/4/6 هـ.
8. محاضرة في جامعة القصيم عنوان "حوسبة اللغة العربية: الوسم النحوي أمودجا" بتاريخ 1440/6/15 هـ.
9. محاضرة بعنوان "اللسانيات الحاسوبية: مقدمة ونماذج من العربية" في جامعة الملك سعود بمعهد اللغويات العربية بتاريخ 1441/3/22 هـ.
10. محاضرة بجامعة الأميرة نورة بنت عبدالرحمن خلال ورشة عمل مع المركز بعنوان "نحو مصدر لغوي عربي مرجعي" بتاريخ 1441-3-28 هـ.
11. محاضرة بعنوان "هل نحن بحاجة لمصدر لغوي جديد لحوسبة اللغة العربية؟" في مكتبة الملك عبدالعزيز العامة عن بتاريخ 1441/4/21 هـ.
12. محاضرة بعنوان "اللغة العربية والذكاء الاصطناعي" في الجامعة السعودية الإلكترونية بتاريخ 1441/4/22 هـ.
13. محاضرة بعنوان "اللسانيات الحاسوبية ومعالجة اللغة مقدمة ونماذج من العربية" في المركز الوطني لتحليل البيانات والذكاء الاصطناعي بالمدينة لمنسوبي المركز ومتدربيه بتاريخ 1441/12/22 هـ.

14. المشاركة في حلقة نقاش بعنوان "بناء مصادر البيانات اللغوية والمعالجة الآلية للغة العربية" ضمن فعاليات القمة العالمية للذكاء الاصطناعي في نسختها الثانية (2022) بتنسيق من مجمع الملك سلمان العالمي للغة العربية.
15. حلقة تدريبية بعنوان "مراحل بناء نماذج معالجة اللغة باستخدام تعلم الآلة" ضمن فعاليات "برمجان العربية" الذي نظمه مجمع الملك سلمان العالمي للغة العربية بتاريخ 2022/5/23م.
16. محاضرة علمية بعنوان "البيانات ما تحيا به معالجة اللغة وحوسبتها" ضمن أنشطة مجمع الملك سلمان العالمي للغة العربية بتاريخ 2022/10/25م.
17. محاضرة بعنوان "جهود مدينة الملك عبدالعزيز للعلوم والتقنية في حوسبة اللغة العربية" في مؤتمر اللسانيات الحاسوبية واللغة العربية: التصورات، والتطبيقات، والتطلعات" بتاريخ 2022/10/25م ، أبوظبي، الإمارات العربية المتحدة.

### الدورات التدريبية التي قدمتها

1. تدريب خلال ورشة عمل بعنوان "نماذج لأشهر المدونات اللغوية ومجالات البحث فيها" في جامعة الأميرة نورة بنت عبدالرحمن- كرسي بحث صحيفة الجزيرة للدراسات اللغوية الحديثة بتاريخ 1433/12/20 هـ.
2. دورة تدريبية بعنوان "تصميم وتحليل المدونات اللغوية" في معهد تعليم اللغة العربية بجامعة الملك سعود لمدة أسبوع اعتباراً من 1434/7/1 هـ.
3. دورة تدريبية ليوم واحد بعنوان "استخدام برمجيات التصاحب اللفظي من المدونات المحوسبة" لمعهد اللغويات بجامعة الملك سعود.
4. دورة تدريبية بعنوان "التحليل الإحصائي للمدونات اللغوية باستخدام نظام معالجة المدونات غواص" في مركز اللغويات التطبيقية بجامعة الإمام محمد بن سعود الإسلامية لمدة أسبوع اعتباراً من 1438/1/22 هـ.

### البرامج والمنتجات التقنية

1. المدونة اللغوية العربية لمدينة الملك عبدالعزيز للعلوم والتقنية: تصميم المدونة والإشراف على تنفيذها. قاعدة بيانات تحوي أكثر من مليون ومائة ألف نص مجموع كلماتها أكثر من بليون ومائة وثمانين ألف كلمة من العصر الجاهلي وحتى الوقت الراهن مع توفير أدوات للتحليل اللغوي والإحصائي لاستخدام اللغة العربية. متوفرة على الموقع [corpus.kacst.edu.sa](http://corpus.kacst.edu.sa). بالتعاون مع الأستاذة منيرة الحوشان والأستاذة إيمان النخيلان.
2. غواص: نظام حاسوبي لمعالجة المدونات العربية على أجهزة الحاسب الشخصي. بالتعاون مع متدربات خلال التدريب الصيفي والأستاذة مروة خان.
3. مصاحب: نظام لاستخراج المتلازمات اللفظية من المدونات اللغوية. بالتعاون مع الأستاذة ابتهاج باعظيم.
4. المستخلص العربي: نظام لجمع النصوص من مواقع الإنترنت اعتماداً على تقنية RSS. بالتعاون مع الأستاذة منيرة الحوشان.
5. المشكل العربي: نظام آلي مبني على تقنيات التعلم العميق لتشكيل النصوص العربية. حقق وقت النشر أفضل المعدلات في نسبة الأخطاء على مستوى الأحرف مقارنة بالنماذج الأخرى المتاحة.

6. نظام "وسم": منصة على الإنترنت لتوسيم البيانات اللغوية العربية لتطبيقات الذكاء الاصطناعي. موجودة على الرابط [./\(https://wssem.pythonanywhere.com/login\)](https://wssem.pythonanywhere.com/login).
7. المصادر اللغوية: مجموعة من البيانات اللغوية التي يمكن أن تستخدم في بناء وتطوير أنظمة الذكاء الاصطناعي قمت بتصميمها والتخطيط لجمعها وتقييمها بعد الجمع، كما قمت بالمشاركة في جمع جزء كبير منها وهي كالتالي:
  - 1) مدونة تصنيف النصوص العربية: 17,658 نصا مجموعة ومصنفة يدويا من 7 أنواع من الكتابات يبلغ عدد كلماتها أكثر من 11.5 مليون كلمة.
  - 2) مدونة الصحف العربية 2012: تحوي 2910 نصا مصنفة يدويا حسب الموضوعات من الصحف العربية خلال 2012. تحتوي المدونة على أكثر من مليوني كلمة.
  - 3) مدونة عاصفة الحزم: 63 ألف نص مصنف موضوعيا، تحوي أكثر من 20 مليون كلمة من الصحف الصادرة في 11 بلدا عربيا خلال فترة عاصفة الحزم. (مصاحب)
  - 4) المتلازمات اللفظية في الصحافة العربية: 35,934 متلازمة لفظية من كلمتين يمكن استخدامها في تمذجة اللغة وتصحيح الأخطاء والدراسات اللغوية. (مصاحب)
  - 5) مدونة عاصفة الأمل: تغطي الأسبوع الأول من عملية عاصفة الحزم (9-5-2015 حتى 14-5-2015). جمعت نصوصها من 17 بلدا عربيا. تحتوي المدونة على أكثر من 14 ألف نص مصنفة موضوعيا وتحتوي على أكثر من 4.5 مليون كلمة.
  - 6) مدونة لحظية للصحافة العربية: تغطي الفترة من 4 فبراير وحتى 8 مايو عام 2018م. 117,819 نصا مصنف موضوعيا، تحتوي 28,334,332 كلمة من الصحف الصادرة في 13 بلدا عربيا. (المستخلص العربي)
  - 7) المدونة العلمية العربية: تغطي ستة مجالات هي التعليم، والمعلومات والمكتبات، وأصول الفقه، والفقه، وعلوم اللغة، والعلوم والهندسة. مكونة من 732 نصا تحوي أكثر من 15 مليون كلمة.
  - 8) مدونة مصنفة للأسئلة العامة: تحتوي على أكثر من 5000 سؤال مصنفة إلى 6 أقسام تجيب عن الاختصارات، الوصف، الموجودات، الإنسان، المكان والأعداد. تستخدم في أنظمة الإجابة على الاستفسارات.
  - 9) مدونة تغريدات الوسوم الرائجة في تويتر السعودية: تم جمعها خلال 2016م و 2017م. تحتوي أكثر من 19 مليون تغريدة تغطي 3,456 وسما رائجا في تويتر السعودية.
  - 10) مدونة لتحليل الآراء في تويتر: 15 ألف تغريدة مصنفة إلى إيجابي، سلبي، محايد، مزعج، خبر، غير محدد.
  - 11) مدونة لتحليل المشاعر في تويتر بحسب نظرية إيكمان: 5 آلاف تغريدة مصنفة إلى غضب، اشمئزاز، خوف، حزن، سعادة، تفاجئ.
  - 12) مدونة لتحليل المشاعر في تويتر بحسب نظرية بولتشك: 10 آلاف تغريدة مصنفة إلى قبول، غضب، توقع، اشمئزاز، خوف، فرح، حزن، تفاجئ.
  - 13) قاموس لتحليل الآراء في تويتر: يحوي 14428 مفردة وعبارة منها 4448 إيجابية و 9980 سلبية.
  - 14) مدونة كوفيد-19: مدونة من الصحف العربية من 12 بلدا عربيا- تغطي الفترة من 2019/1/1م وحتى 2021/12/31م، تحوي حتى الآن على 1.7 مليون نصا مصنفة موضوعيا وأكثر من 2 بليون كلمة.
  - 15) الخمسة آلاف كلمة الأكثر تكرارا في العربية مقطعة يدويا مع الوزن الصرفي والجذر الخاص بكل كلمة.

- 16) مدونة التعرف على أسماء الأعلام. مدونة مكونة من 1277 جملة منتقاة عشوائياً من الصحف السعودية من مواضيع متعددة. تحتوي على أكثر من 55 ألف كلمة موزعة على التصنيفات التالية: اسم شخص، منظمة، موقع، لقب، أخرى، ليست اسم علم).
- 17) مدونة التعرف على أسماء الأعلام. مدونة مكونة من 2589 جملة منتقاة عشوائياً من الصحف السعودية من مواضيع متعددة. تحتوي على أكثر من 100 ألف كلمة موزعة على التصنيفات التالية: اسم شخص، منظمة، موقع، لقب، حدث، جيوسياسي، ديني، صحي، أخرى، ليست اسم علم).
- 18) مدونة التعرف على المصطلحات اللغوية. مدونة موسمة بالمصطلحات العلمية في مجال علوم اللغة العربية تحتوي 4,148 جملة تحوي 155,502 كلمة وبها 12,895 مصطلحا. موسما.

## التدريس:

1. تدريس مادة "قواعد البيانات" لطلاب مرحلة البكالوريوس بكلية علوم الحاسب بجامعة الإمام محمد بن سعود الإسلامية للفصل الدراسي الثاني عام 1430/1429 هـ.
2. تدريس مادة "اللغة والحاسوب" لطلاب مرحلة الدكتوراه بمعهد تعليم اللغة العربية بجامعة الإمام محمد بن سعود الإسلامية للفصل الدراسي الثاني عام 1436/1435 هـ.
3. تدريس مادة "اللغة والحاسوب" لطلاب مرحلة الدكتوراه بمعهد تعليم اللغة العربية بجامعة الإمام محمد بن سعود الإسلامية للفصل الدراسي الثاني عام 1437/1436 هـ.

## التحكيم العلمي

1. تحكيم مشروع علمي رقم (ا ت-32-218) بعنوان "النظام الذكي متعدد الوسائط لتشخيص الأمراض باستخدام تقنية Cloud Based". الإدارة العامة لمنح البحوث، مدينة الملك عبدالعزيز للعلوم والتقنية.
2. تحكيم مشروع علمي رقم (ا ت-34-200) بعنوان "استخدام النص والصورة ووصف البيانات المتعدد وتعليقات المستخدمين في الترجمة الآلية لأغراض تصنيف نصوص عربية على وسائل التواصل الاجتماعي" الإدارة العامة لمنح البحوث، مدينة الملك عبدالعزيز للعلوم والتقنية.
3. تحكيم عدد (3) من الأعمال المقدمة لجائزة خادم الحرمين الشريفين لتكريم المخترعين والموهوبين في دورتها الأولى 2012م.
4. تحكيم كتاب بعنوان "تصميم واجهات المستخدم الرسومية وربطها بقواعد البيانات". الإدارة العامة للتوعية العلمية والنشر، مدينة الملك عبدالعزيز للعلوم والتقنية.
5. تحكيم ترجمة كتاب "مقدمة في معالجة اللغة العربية". مركز الترجمة، جامعة الملك سعود.
6. تحرير سجل أبحاث ورشة عمل OSACT في مؤتمر LREC2014
7. تحرير سجل أبحاث ورشة عمل OSACT في مؤتمر LREC2016
8. تقويم مقترح بحثي للمشروع رقم (م ص -36-32) بعنوان ""تصميم نظام لتحديد هوية الترددات اللاسلكية بدون رقاقة بالاعتماد على تقنية الإمضاء الترددي للتعريف الراديوي للحروف العربية ". الإدارة العامة لدعم البحث والابتكار.

9. تقويم مقترح بحثي للمشروع رقم (م ص 36-227) بعنوان " قاعدة بيانات الإلكترونية للمخطوطات العربية والإسلامية لخدمة بحوث العرف على الأنماط". الإدارة العامة لدعم البحث والابتكار.
10. تقويم تقرير بحثي سنوي للمشروع رقم (و ع 38-1300) بعنوان " التعرف الآلي على الكتابة العربية اليدوية الآنية ". الإدارة العامة لدعم البحث والابتكار.
11. تقويم تقرير فني نهائي للمشروع رقم (و ع 38-1365) بعنوان " التعرف الآلي على الأخطاء المطبعية وتصحيحها". الإدارة العامة لدعم البحث والابتكار.
12. تحكيم كتاب "الذخائر النصية ودورها في المعالجة الآلية للغة العربية". مركز الملك عبدالله بن عبدالعزيز الدولي لخدمة اللغة العربية. الرياض (المملكة العربية السعودية).
13. تحكيم كتاب "البنك الشجري النحوي: بناؤه وتوظيفه في إطار تقنيات الذكاء الاصطناعي". مركز الملك عبدالله بن عبدالعزيز الدولي لخدمة اللغة العربية. الرياض (المملكة العربية السعودية).
14. تحكيم عملين لجائزة خادم الحرمين الشريفين (فئة الموهوبين فوق 25 سنة) في دورتها الثالثة 2014م.
15. تحكيم ترجمة كتاب "لغويات المدونة الحاسوبية" لمركز الترجمة بجامعة الملك سعود.
16. تحكيم ثلاث أوراق علمية لمجلة IEEE ACCESS .
17. تحكيم ورقتين علميتين لمجلة King Saud University - Computer and Information Sciences.
18. تحكيم ورقتين علميتين لمجلة Language Resources and Evaluation.
19. تحكيم ثلاث أوراق علمية للمؤتمر الدولي الرابع لتقنية المعلومات KACSTIT-16.
20. تحكيم ورقة علمية للمؤتمر الدولي الرابع للسانيات الحاسوبية العربية ACLing 2018.
21. تحكيم ورقة علمية لمجلة اللسانيات العربية (2022).
22. تحرير سجل أبحاث ورشة عمل OSACT في مؤتمر LREC2022
23. تحكيم 3 أوراق علمية في ورشة عمل OSACT في مؤتمر LREC2022

### الإشراف على رسائل الدكتوراة

1. رسالة بعنوان " نحو بناء معجم للمتلازمات اللفظية في المعاجم العربية المعاصرة: دراسة تحليلية في ضوء مدونة لغوية " للطالبة "بدرية بنت براك العنزي". معهد تعليم اللغة العربية، جامعة الإمام محمد بن سعود الإسلامية. 1438 هـ
2. رسالة بعنوان " التوسيم النحوي للمدونات العربية: نماذج توسيمية مقترحة ". للطالبة "أفراح بنت عبدالعزيز التميمي". معهد تعليم اللغة العربية، جامعة الإمام محمد بن سعود الإسلامية. 1440 هـ
3. رسالة بعنوان "المتلازمات اللفظية في نماذج من القوانين الوطنية والاتفاقيات الدولية المتعلقة بالإرهاب في ضوء مدونة لغوية عربية محوسبة " للطالب "وافي بن عبدالله المطيري". كلية الآداب، جامعة القصيم. 1442 هـ
4. رسالة بعنوان "الأفعال العماد في العربية الفصيحة القديمة والعربية المعاصرة: دراسة تحليلية مقارنة باستخدام المدونات اللغوية " للطالب "فايز بن سعد آل لجم". كلية الآداب، جامعة القصيم. 1442 هـ