مجلة الألكسو العلمية للفتيان

مجلة علمية نصف سنوية

العدد الثامن والعشرون - ربيع الأول 1439 هـ - ديسمبر 2017 م

المنظمة العربية للتربية والثقافة والعلوم

المحتويات

- كلمة العدد
- ملف العدد

المعالجة الآلية للغة العربية

بحوث ودراسات

- إنترنت الأشياء لزراعة ذكية
- تجربة تشاركية عربية ـ متوسطية واعدة

الميثاق المتوسطي للوظائف الخضراء

- هل من حلول لمواجهة أزمة المياه في الوطن العربي ؟
- مواقع التواصل الاجتماعي والوطن العربي: قراءة للمشهد الراهن
- شخصية العدد :

د. أحمد زويل : دروس من حياة هذا العالم العربي

- مؤسسة العدد :

كلية العلوم ـ جامعة محمد الأول ـ وجدة ـ المملكة المغربية

المشرف العام
د. سعود هلال الحربي

إشراف
أ. د. أبو القاسم حسن البدري

تنسيق
أ. خيرية السلامي
أ. صفوان الحكيم

تصميم وانجاز
أ. بلال العامري
أ. صفوان الحكيم
أ. طارق الدريدي

توجه المراسلات إلى البريد الإلكتروني الخاص بالمجلة
fetyan.scres@gmail.com

هاتف : 900 013 70 216+

إن كافة الآراء التي تنشر بأسماء كتّابها تعبّر عن وجهة نظر أصحابها
ولا تحمل بالضرورة وجهة نظر المنظمة العربية للتربية والثقافة والعلوم

بعض الصور الإضافية والأشكال تؤخذ من مصادر مفتوحة والمجلة تشكر الجميع لجهودهم

يسمح باستعمال ما ورد في هذه المجلة من مواد علمية،
أو ثقافية، أو تربوية، أو فنية، بشرط الإشارة إلى مصدرها

مجلة الألكسو العلمية للفتيان : مجلة علمية نصف سنوية /
تصدر عن المنظمة العربية للتربية والثقافة والعلوم. إدارة العلوم
والبحث العلمي . ـ العدد 28، ديسمبر 2017 . ـ تونس : المنظمة...
ردمد : 6489 – ISSN : 0330 = مجلة الألكسو العلمية للفتيان
ع / / /

جميع حقوق الطبع محفوظة للمنظمة

كلمة العدد

يُسعد المنظمة العربية للتربية والثقافة والعلوم (الألكسو) أن تضع بين أيدي قرائها الأعزاء، العدد الثامن والعشرين (ديسمبر/2017) من المجلة العربية العلمية للفتيان تحت اسمها الجديد: «مجلة الألكسو العلمية للفتيان». ويتضمّن العدد ملفا خاصا بالمعالجة الآلية للغة العربية، سعيا من المنظمة لتهيئة السليقة اللغوية لدى الشريحة العمرية التي تستهدفها المجلة (14 - 22 سنة)، بما تمكنهم من فهم ما يتعلق بالتطبيقات الآلية الحديثة للغة العربية، خاصة مع انتشار الإنترنت، واستخدام الحاسوب الآلي في كثير من المؤسسات العربية بما فيها المؤسسات المعرفية والثقافية. وقد تضمّن العدد مواضيع علميّة حول «الذخائر اللغوية...والمعالجة الآلية للغة العربية»، و«التعرف الضوئي على الحروف»، و«محرّكات البحث»، إضافة إلى مقالات علمية متنوّعة من بينها «إنترنت الأشياء لزراعة ذكيّة» و«هل من حلول لمواجهة أزمة المياه في الوطن العربي؟»، و«الميثاق المتوسطي للوظائف الخضراء»، ساهم في كتابتها خبراء علميّون من الجزائر والمغرب وتونس ومصر وسوريا؛

ويزدان هذا العدد بتقديم نبذة عن حياة العالم العربي الفذ المرحوم الدكتور أحمد زويل، وإنجازاته وأعماله، والجوائز التي تحصّل عليها، وهي قصة عالم مبدع، كان يحلم بتحقيق نهضة عربية تستند إلى العلم والبحث العلمي.

وفي إطار التعريف بالمؤسسات العلمية، يقدم هذا العدد كلية العلوم بجامعة محمد الأول ـ وجدة، المملكة المغربية، التي تعتبر ـ منذ إنشائها في العام 1979 ـ من أهم الكليّات بجامعة محمد الأول وأقدمها؛ وتتميّز بإعطائها للبحث العلمي مكانة بارزة في سياستها، حيث يتم التركيز فيها على مواضيع بحثية تخدم التنمية؛ وقد حققت هذه الأبحاث نتائج مهمّة في مختلف التخصصات، وتم نشرها في مجلات متخصصة من أجل التعريف بها، وتمكين الباحثين من الاستفادة من نتائجها.

ويعدّ هذا العدد (28) أول عدد يصدر منذ تولّي الدكتور سعود هلال الحربي مهامه مديرا عاما للمنظمة ؛ ويسرّ هيئة تحرير المجلة إصداره تحت إشرافه ورعايته، والإشادة برؤيته الداعية إلى تحقيق رسالة المنظمة وأهدافها من خلال ما تصدره من كتب ودوريات وتقارير، سعيا لتنشيط دورها في توحيد العمل العربي المشترك في مجالاتها التربوية والثقافية والعلمية.

وإذ تقدّم المنظمة هذا العدد إلى قرائها الأعزاء، تُزجي شكرها للكتّاب الأفاضل الذين أثروه بمقالاتهم، كما تعوّل على جهود الأكاديميين والخبراء للكتابة في هذه المجلة، وإيصالها للمستفيدين بشكل يلبّي تطلعات أجيالنا العربية.

هيئة التحرير

ملف العدد :

المعالجة الآلية للغة العربية

الذخائر اللغوية... والمعالجة الآلية للغة العربية 6

تقنية «ويكي Wiki» في ميادين المعالجة الآلية للغة العربية 14

تقنيات معالجة الصوت اللّغوي 20

د. المعتز بالله السعيد

واقع وآفاق الترجمة الآلية 27

التعرف الضوئي على الحروف 35

أ. د. عز الدين مزروعي

التعمية والأمن المعلوماتي 43

أ. د. عزيزي عبد المالك

المعالجة الآلية للكلام : التعرف الآلي على الكلام وتوليده 53

د. عبد الوافي مزيان

محركات البحث 59

أ. د. عبد الحق لخواجة

المعالجة الآلية للغة والكلام 65

د. عادل داود

جهود الألكسو في مجال النهوض باللغة العربية للتوجه نحو مجتمع المعرفة 71

أ. صفوان الحكيم

ملف العدد :

المعالجة الآلية للغة العربية :
آفاق التطوير للحاق بركب التقنيات المعاصرة

ملف العدد

الذخائر اللغوية... والمعالجة الآلية للغة العربية

د. المعتزّ باللّه السعيد

ما هي الذَّخيرة اللُّغَويَّة؟!

إذا أردتَ أن تبنيَ بيتًا، فهل تستطيعُ أن تفعلَ ذلكَ دونَ أن تضعَ أساسًا قويًّا ليحميَ بناءَك؟! وإذا أردتَ أن تقودَ سيَّارة، فهل بالإمكان قيادتُها دونَ أن تُزوِّدَها بمورد الطَّاقة اللَّازمة لحركتها من الوقود؟! بالطَّبع لا يُمكنُ هذا.

الذَّخيرةُ اللُّغَويَّة يا عزيزي هي الأساسُ الَّذي تقومُ عليه مُعالجة اللُّغات الطَّبيعيَّة، وهي المورد الَّذي يُساعدُنا على بناء أدوات مُعالجة اللُّغة والتَّحكُّم فيها وتطويرها.

نستطيعُ تعريفَ الذَّخيرة اللُّغويَّة بأنَّها مجموعةٌ كبيرةٌ من النُّصُوص المكتوبة أو المنطوقة للُّغة، يُمكنُ التَّعاملُ معها باستخدام الآلة، ويتمُّ التَّحكُّمُ في بياناتها بالإضافة أو الحذف أو التَّعديل، وذلك من خلال قواعد بياناتٍ صُمِّمَت لتكونَ قادرةً على التَّعامُل مع هذه النُّصوص؛ وتُمَثِّلُ هذه القواعدُ مخزنًا كبيرًا للُّغة، يُرجَع إليه وقتَ الحاجة ويَتحمَّل أيَّ قدرٍ من النُّصوص التَّي يُمكن أن تُضافَ إلى المادَّة الأساسيَّة للذَّخيرة اللُّغَويَّة مُستَقبلًا. وإذا كانت اللُّغةُ عبارة عن كلماتٍ وعباراتٍ ونُصُوص، فالذَّخيرةُ اللُّغويَّةُ (Linguistic Corpus) تضمُّ أيضًا كلماتٍ وعباراتٍ ونُصُوصًا؛ وهي بهذا نموذجٌ مُصَغَّرٌ للُّغة، نستخدمُهُ في تمكين الآلة من فهم طبيعة اللُّغة وعناصرها، بما يُساعدُ على مُعالجتها وتطوير الأدوات والبرامج المُساعدة على ذلك.

يُمكنُ القولُ إنَّ الذَّخائرَ اللُّغويَّة هي المورِدُ اللُّغويُّ الأساسيُّ المُستخدم في ميادين المُعالجة الآليَّة للُّغة، لأنَّها تُمثِّلُ شكلًا حقيقيًّا للُّغة الطَّبيعيَّة الَّتي يستخدمُها الإنسانُ ويتفاعلُ بها مع الآخرين. وإذا كانت المُعالجة الآليَّة للُّغة لا تتمُّ إلَّا من خلال تفسير الظَّواهر اللُّغويَّة لتمكين الآلة من مُحاكاة ذكاء الإنسان؛ فإنَّ هذا التَّفسيرَ لم يَعُد قاصرًا على النَّظريَّات اللُّغويَّة التقليديَّة؛ بل تخطَّى ذلك إلى التَّعامُل مع مناهجَ تجريبيَّةٍ تقومُ على نُصُوصٍ حيَّة. ونحنُ لا نستطيعُ إهمالَ منطق الآلة في فهم حقيقة الأشياء؛ لأنَّنا بهذا لن نصلَ إلى تحقيق التَّواصُل المأمول مع الآلة. وبالتَّالي، فإنَّها لن تستجيبَ إلى أوامرنا. ومن مُتطلَّبات إفهام الآلة

للُّغة الَّتي تُعَدُّ وسيلةَ التَّواصُل عندَ الإنسان أن نُزوِّدَها بأشكالٍ مُختلفة من عناصر اللُّغة، من الحُرُوف المكتوبة والأصوات المنطوقة والكلمات والجُمَل والتَّراكيب. وهذه الأشكالُ هي الَّتي تُعَبِّرُ عنها الذَّخيرةُ اللُّغَويَّة.

الذَّخائر اللُّغَويَّة العربيَّة

ظهرَت فكرةُ الذَّخائر اللُّغويَّة مع ظهور الحواسيب، في نهاية السِّتِّينيَّات من القرن العشرين؛ وانتشرَت في أمريكا وأوروبا، فأمكنَ استخدامُها في مُختلف أغراض المُعالجة الآليَّة للُّغات، ونخصُّ بالذِّكر ميادين: التَّرجمة الآليَّة، وبناء المعاجم والقواميس الآليَّة، والتَّعلُّم الآليّ. ومع هذا، فإنَّ اللُّغةَ العربيَّة لم تعرف الطَّريقَ إلى الذَّخائر اللُّغويَّة إلَّا قريبًا من القرن العشرين. فمع مطلع الألفيَّة الثَّالثة ازدادت الحاجة إلى تطوير أدوات المُعالجة الآليَّة للُّغة العربيَّة، نتيجةَ الطَّفرة المعلوماتيَّة الهائلة الَّتي فَرَضَت علينا أن نتعاملَ مع مجموعاتٍ ضخمة من النُّصُوص. وأمكنَ استثمارُ الذَّخائر اللُّغويَّة في بناء العديد من أدوات مُعالجة اللُّغة العربيَّة، كما أمكنَ استخدامُها في تطوير أبحاث اللُّغة العربيَّة

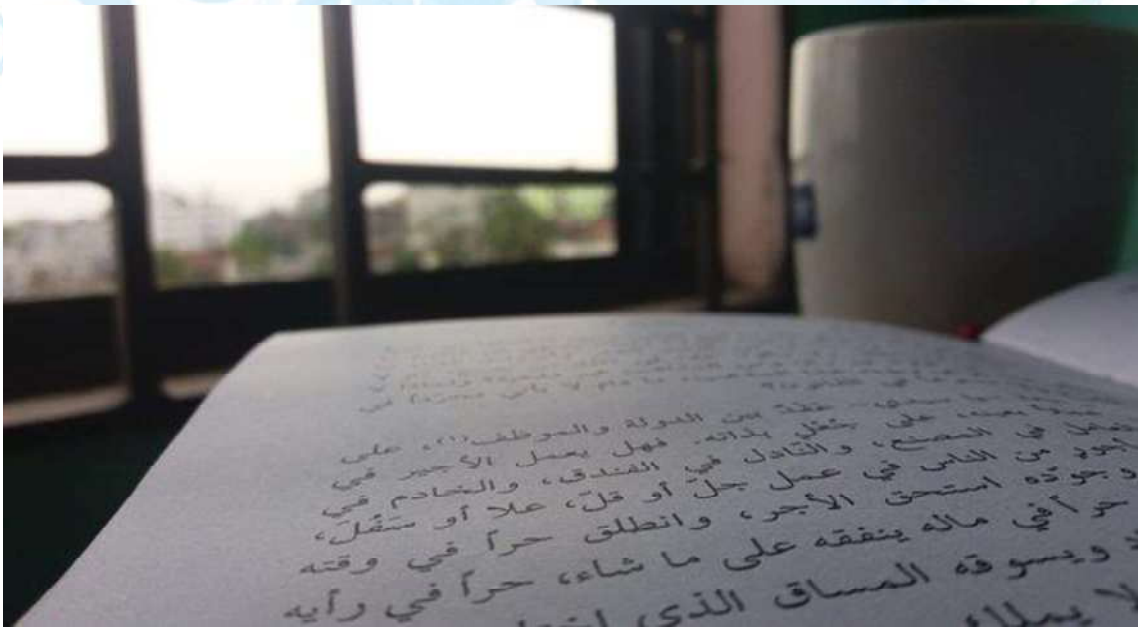
وبناء عدد من المعاجم اللُّغَويَّة المُعاصرة.

من اللَّافت للانتباه أنَّ أكثرَ المُؤسَّسات المُهتمَّة ببناء الذَّخائر اللُّغويَّة العربيَّة توجدُ خارجَ نطاق الوطن العربيّ. ولعلَّ هذا يعودُ إلى حداثة عهدنا بالذَّخائر اللُّغويَّة العربيَّة وعدم إدراك أهمِّيتها على النَّحو المنشود. وإنَّكَ لتجدُ مُؤسَّسات مثل: شركة مايكروسوفت، وشركة جوجل، ومُؤسَّسة البيانات اللُّغويَّة في جامعة بنسلفانيا الأمريكيَّة، وجامعة كولومبيا الأمريكيَّة، وغيرها، تهتمُّ جميعًا ببناء الذَّخائر اللُّغويَّة العربيَّة وتطويرها لأغراضٍ مُختلفة في البحث والصِّناعة والتَّدريس. وتتلاقى هذه الأغراضُ كلُّها في أنَّها تقومُ على المُعالجة الآليَّة للُّغة العربيَّة.

ومع هذا، فإنَّ هُناكَ عددًا من المُؤسَّسات العربيَّة تهتمُّ بالذَّخائر اللُّغويَّة العربيَّة؛ وقد أنجزَت مجموعةً من الذَّخائر العربيَة المُهمَّة؛ نذكرُ منها: المُنظَّمة العربيَّة للتَّربية والثَّقافة والعُلُوم «ألكسو ALECSO"، ومدينة الملك عبد العزيز للعُلُوم والتِّقنية بالرِّياض «KACST"، ومركز تقنيات اللُّغة العربيَّة في القاهرة «ALTEC"، وغيرها. وقد قامَ كذلكَ عددٌ من الباحثينَ العرب ببناء العديد من الذَّخائر اللُّغويَّة العربيَّة وتطويرها، لأغراضٍ بحثيَّةٍ مُتعدِّدة.

الطَّريقُ إلى حَوسبة الذَّخائر اللُّغويَّة

قبلَ مئات السِّنين، كانَ العربُ والهُنُودُ والصِّينيُّونَ يستخدمونَ النُّصُوصَ في تعليم اللُّغة واستنباط قواعد النَّحو وإعداد فهارس الكُتُب وصناعة المعاجم، وغير ذلك. واستطاعَ أجدادُنا وأسلافُنا أن يترُكوا تُراثًا خالدًا من إبداعاتهم في هذه الميادين. ومع أنَّهم بذلُوا جُهودًا كبيرةً في التَّعامُل مع النُّصُوص، إلَّا أنَّ عَملهم كانَ محدودًا دائمًا،



لعدم القُدرة على التَّعامُل مع مجموعاتٍ كبيرةٍ من النُّصُوص، لما يتطلَّبُهُ ذلكَ من الوقت والجُهد. الواقعُ أنَّهم كانوا يستخدمونَ الذَّخائرَ اللُّغويَّة، ولكن بصورةٍ مُبَسَّطة. دعنا نضربُ مثلًا على ذلك. هل تعرفُ أنَّ بعض فهارس القُرآن الكريم والكتاب المُقدَّس قد وُضِعَت منذُ مئات السِّنين؟! لقد قامَ العُلماءُ باستخدام مُحتوى القُرآن الكريم والكتاب المُقدَّس كصورةٍ من الذَّخيرة اللُّغويَّة؛ وصنعوا من هذا المُحتوى فهارسَ الموضوعات والآيات والأماكن والأعلام وغيرها. لكنَّهم تعاملوا مع هذه النُّصُوص يدويًّا، في حُدُود ما تُتيحُه إمكاناتُ عُصُورهم.

ولمَّا ظهرَ الحاسُوب، تطوَّرَ الأمرُ بصورةٍ كبيرة، لأنَّه ساعدَ الإنسانَ على تخزين النُّصُوص واستدعائها وقتما يشاء؛ وساعدَهُ أيضًا على التَّعامُل مع مجموعاتٍ هائلة منها، فاستطاعَ أن يجمعَها ويُدقِّقَها ويُرتِّبَها ويُجري عليها عمليَّاتٍ عديدة. وكُلَّما تطوَّرَ عملُ الحاسُوب، أمكنَ استخدامُ الذَّخائر اللُّغويَّة بصورةٍ أكبر، وأمكنَ توظيفُها في تطبيقاتٍ أكثر، حتَّى وصلنا إلى المُستوى الَّذي يسمحُ بالتَّعامُل مع بلايين الكلمات والجُمَل، وإجراء العمليَّات عليها في ثوانٍ معدودة.

ولا شكَّ أنَّ هذه الطَّفرة المعلوماتيَّة قد ساهَمَت بصورةٍ كبيرة في تطوير الصِّناعات والأبحاث الَّتي تعتمدُ على الذَّخائر اللُّغويَّة، وعلى رأسها المُعالجة الآليَّة للُّغات الطَّبيعيَّة.

ولأنَّ اللُّغة العربيَّة واحدة من أكثر لُغات العالم انتشارًا، حيثُ تأتي في المرتبة الخامسة من حيثُ عدد المُتحدِّثينَ الأصليِّينَ بعدَ الصِّينيَّة والإسبانيَّة والإنجليزيَّة والهنديَّة، فقد استفادَت من ظُهور الحواسيب الَّتي فتحَت المجالَ للتَّعامُل مع تُراثنا العربيّ قديمًا وحديثًا. ومع أنَّنا لم نصل بعدُ إلى مُستويات بعض اللُّغات الأخرى في استخدام الذَّخائر اللُّغويَّة، إلَّا أنَّنا استطعنا تحقيقَ نتائجَ إيجابيَّة. وأصبحَ بإمكاننا التَّعامُلُ مع ذخائرَ لُغويَّة عربيَّة لأغراض تقنية المعلومات وإنتاج أدوات مُعالجة اللُّغة العربيَّة. وإذا تتبَّعنا تاريخَ الذَّخائر اللُّغويَّة العربيَّة منذُ ظُهورها إلى الآن، فسنعرفُ أنَّها تتطوَّرُ وتتنامى يومًا بعدَ يوم؛ ولا نستبعدُ أن تُصبحَ الذَّخائرُ اللُّغويَّةُ العربيَّةُ أداةَ البحث الأولى في عُلوم العربيَّة في وقتٍ قريب.

من تطبيقات الذَّخائر اللُّغويَّة

التَّرجمة الآليَّة

تُعَدُّ التَّرجمة الآليَّة من أهمّ تطبيقات الذَّخائر اللُّغويَّة؛ إذ لا يُمكنُ تطويرُ أدواتٍ فعَّالة للتَّرجمة الآليَّة بينَ اللُّغات إلَّا باستخدام ذخائرَ مُتوازية

تضمُّ مجموعاتٍ كبيرة من النُّصُوص. فإذا أردنا مثلًا أن نصنعَ نظامًا للتَّرجمة الآليَّة بينَ العربيَّة والإنجليزيَّة، فإنَّ علينا أن نصنعَ ذخيرةً لُغويَّةً كبيرةً تحوي النُّصُوصَ العربيَّةَ وما يُقابلُها في اللُّغة الإنجليزيَّة، ثُمَّ نقومُ بتدريب الآلة على فهم قواعد النَّحو وأنماط الجُمَل ومعاني الكلمات والتَّراكيب بينَ اللُّغَتَين. وحتَّى تتمَّ الاستفادةُ الكاملة من الذَّخيرة اللُّغويَّة في التَّرجمة الآليَّة بينَ اللُّغات، فإنَّنا سنكونُ بحاجةٍ إلى إجراء مُعالجاتٍ إحصائيَّةٍ دقيقة على مادَّة الذَّخيرة، سعيًا إلى استخلاص إحصاءاتٍ تتَّفقُ مع واقع الأنظمة اللُّغويَّة لمجموعة اللُّغات الَّتي نُترجمُ منها وإليها.

بناء مُحرِّكات البحث

إذا قُمتَ باستخدام مُحرِّك البحث المشهور «جوجل Google" للبحث عن كلمة (الأمم) فستأتيكَ العديدُ من النَّتائج؛ لكنَّ النَّتائجَ الأسبقَ والأكثرَ وُرُودًا ستكونُ لعبارات مثل «الأمم الإفريقيَّة، الأمم المُتَّحدة»؛ وفي أغلب الأحيان يكونُ هذا بالفعل ما تبحثُ عنه. فهل فكَّرتَ في الوسيلة الَّتي يهتدي بها مُحرِّك البحث إلى ما في ذهنك؟!

الأمرُ ببساطة أنَّ القائمينَ على بناء مُحرِّكات البحث يقومونَ بإعداد فهرسٍ كبيرٍ لكلمات اللُّغة؛ ويأخذونَ مادَّةَ هذا الفهرس من ذخيرةٍ لُغويَّةٍ ضخمة، تضمُّ نُصُوصًا مُستمدَّة من واقع اللُّغة المُعاصرة، ثُمَّ يقومُونَ ببعض العمليَّات الإحصائيَّة لاستخلاص الكلمات الَّتي تلحقُ بكلمات البحث أو تسبقها. ويضعونَ هذه المخرجات في قاعدة بياناتٍ يتمُّ التَّحكُّم في مُحرِّك البحث من خلالها. وبالتَّالي فإنَّ عملَ مُحرِّك البحث يقومُ أساسًا على الذَّخائر اللُّغويَّة.

ولكَ أن تتخيَّلَ كيفَ سيكونُ الأمرُ لو لم تعتمد مُحرِّكاتُ البحث على ذخيرةٍ لُغويَّةٍ تضبطُ مُخرجاتها؟! حينَها ستكونُ النَّتائجُ عشوائيَّة إلى درجةٍ مُزعجة. ومن المُفيد أن تعلمَ أيضًا أنَّ الذَّخائر اللُّغويَّة تُساعدُ على مُضاعفة سُرعة مُحرِّكات البحث، لأنَّها تكونُ مادَّةً لبياناتٍ مُخزَّنة يتمُّ استدعاؤُها؛ وهذا أسرعُ كثيرًا ممَّا لو لم تكن هذه البياناتُ موجودةً أصلًا.

تعليم اللُّغات

هل سَمعتَ عن قوائم الكلمات الشَّائعة في اللُّغة العربيَّة أو غيرها من اللُّغات؟! إنَّها مجموعةٌ من

الكلمات الَّتي تتردَّدُ في النُّصُوص أكثرَ من غيرها، فيكثرُ بذلكَ استخدامُها. وتعيينُ مثل هذه القوائم يتمُّ باستخدام الذَّخائر اللُّغويَّة. دعنا نشرحُ ذلكَ بشيءٍ من التَّفصيل.

إذا كُنتَ مُهتمًّا بتعلُّم اللُّغات، فالذَّخيرة اللُّغويَّة هي المورِدُ الأساسيُّ الَّذي يُساعدُكَ على ذلك؛ سواءٌ أكُنتَ مُعلِّمًا أم مُتعلِّمًا. فهي تُوفِّرُ الكثيرَ من الوقت والجُهد في عَمَليَّتَي التَّعليم والتَّعلُّم، لأنَّها تُساعدُ على استخلاص أكثر الكلمات والتَّراكيب والقواعد استخدامًا في اللُّغة، فيتمكَّنُ المُتعلِّمُ بهذا من التَّعرُّف على القدر الأكبر من المعارف اللُّغويَّة بأقلّ جُهد مُمكن. ومن ناحيةٍ أخرى، فإنَّ الذَّخائرَ اللُّغَويَّة تُساعدُ بصورةٍ كبيرةٍ على تطوير المناهج التَّعليميَّة للُّغات في مراحل التَّعليم المختلفة؛ إذ من خلالها يُمكنُ توجيهُ الطُّلَّاب إلى واقع اللُّغة الملموس، بعيدًا عن التَّعقيدات الَّتي قد لا يستفيدون منها في تواصُلهم مع أبناء المُجتمَع فهمًا وإفهامًا. وبطبيعة الحال فإنَّ المُعلِّمين والمُتعلِّمين يستقبلونَ نتائجَ مُعالجة الذَّخائر اللُّغويَّة في صُورةٍ جاهزة. ويقعُ العبءُ في استخلاص المعارف على القائمينَ على إعداد المُقرَّرات التَّعليميَّة من الأساتذة المُختصِّينَ في عُلُوم التَّربية ومن الأساتذة القادرينَ على توجيه الحاسُوب إلى مُعالجة الذَّخائر اللُّغويَّة بما يُلبِّي حاجةَ المُتعلِّمين.

ولَعَلَّنا نستشعرُ أهمِّيَّةَ استخدام الذَّخائر اللُّغويَّة في تعليم اللُّغة العربيَّة بالنَّظَر إلى مُعاناة طُلَّاب المراحل الأساسيَّة في فهم مناهجها، خُصُوصًا قواعد النَّحو العربيّ الَّتي تُؤدِّي إلى نفور كثيرٍ من الطُّلَّاب من تعلُّم العربيَّة. ففي كثيرٍ من الأحيان يدرسُ الطَّالبُ القاعدةَ النَّحويَّة بصرف النَّظَر عن الاستخدام الفعليّ لها، لتبدو المادَّة العلميَّةُ بعيدةً عن اللُّغة الَّتي يُمارسُها قراءةً أو استماعًا. أضف

إلى ذلك ضَعف مُستوى بعض المعلِّمينَ نتيجةَ القُصُور في تأهيلِهم، ممَّا يُؤدِّي إلى عدم قُدرَتِهم على التَّواصُل مع طُلاَّبِهم.

الذَّخائرُ اللُّغويَّة ببساطة مقياسٌ لضبط العمليَّة التَّعليميَّة، ووسيلةٌ مثاليَّة لإحداث التَّجانُس بينَ اللُّغة الَّتي يدرسُها الطَّالبُ واللُّغة الَّتي يستخدمُها في ميادين الحياة. وحتَّى يتمَّ ذلك على أفضل وجه، فإنَّنا نعملُ على تمكين الآلة من التَّعامُل مع ذخائرَ لُغويَّةٍ تضمُّ قدرًا كبيرًا من النُّصُوص المُناسبة للمُستوى العُمريّ والمعرفيّ للطُّلاَّب، من خلال إجراءات لُغويَّة وإحصائيَّة وحاسوبيَّة دقيقة. ونستطيعُ استخدامَ نتائج هذه الإجراءات في إعداد المُقرَّرات التَّعليميَّة الَّتي تتناسبُ مع مُستوى الطَّالب وتُحقِّقُ أهدافَ العمليَّة التَّعليميَّة، ونستطيعُ استخدامَها كذلكَ في صناعة مُعجمٍ مدرسيٍّ يُساعدُ الطَّالبَ على فهم مُفرَدات اللُّغة واستخدامها على الوجه الأمثل.

بناء أدوات تحرير النُّصُوص

لا شكَّ أنَّنا جميعًا نستخدمُ مُحرِّرات النُّصُوص (مثل: مايكروسوفت أوفيس MS Office) في أغراضٍ مُتعدِّدة خلالَ تفاعُلنا مع الآلة، مثل: الاستذكار وكتابةِ الرَّسائل وإعداد الأبحاث وإرسال الرَّسائل عبرَ البريد الإلكترونيّ. فهل تعلمُ أنَّ أدوات التَّحرير المُلحقة بمُحرِّرات النُّصُوص (مثل: المُدقِّق الإملائيّ، والمُصحِّح اللُّغويّ، والمُشكِّل الآليّ) تعتمدُ أساسًا على الذَّخائر اللُّغويَّة؟!

إذا أردتَ مثلًا أن تصنعَ أداةً للتَّدقيق الإملائيّ، فأنتَ بحاجةٍ إلى ذخيرةٍ لُغويَّة خالية من أخطاء الإملاء، لتكونَ مادَّةً لقاعدة بيانات الأداة الَّتي تصنعُها. وإذا أردتَ أن تصنعَ أداةً للتَّصحيح اللُّغويّ، فأنتَ بحاجةٍ إلى ذخيرةٍ لُغويَّةٍ تحوي قدرًا كبيرًا من

التَّراكيب والجُمَل الصَّحيحة لُغويًّا، لتكونَ مصدرًا للقواعد اللُّغويَّة الَّتي يقومُ عليها نظامُ التَّصحيح. وكذلك إذا أردتَ أن تصنعَ أداةً للتَّشكيل الآليّ، فأنتَ بحاجة إلى ذخيرة لُغويَّة مشكولة، تُجري عليها العمليَّات الإحصائيَّة اللَّازمة لاستخلاص ضوابط التَّشكيل، وهكذا. فنحنُ بحاجة مُستمرَّة إلى مُعالجة الذَّخائر اللُّغويَّة لتطوير أدوات تحرير النُّصُوص والتَّأكُّد من سلامة مُخرَجاتِها.

ببليوغرافيا مرجعيَّة

- Barbara Dancygier. (2017). the Cambridge Handbook of Cognitive Linguistics. Cambridge University Press.
- Graeme Kennedy. (2014). an Introduction to Corpus Linguistics. Routledge.
- Keith Johnson. (2017). An Introduction to Foreign Language Learning and Teaching. Routledge.
- Shaimaa El Abasiry. (2017). Investigating Explicitation in Corpora of Translated Texts. Noor Publishing.
- Wendy Anderson, John Corbett. (2017). Exploring English with Online Corpora. Palgrave Macmillan.

ملف العدد

تقنية «ويكي Wiki» في ميادين المعالجة الآلية للغة العربية

د. المعتزّ باللّه السعيد

ما هي «الويكي»؟!

إذا كُنتَ تبحثُ عن معلومةٍ في موضوعٍ مُعَيَّن أو ترغب في كتابة بحثٍ ما، فإنَّ أوَّلَ ما يأتي في ذهنكَ أن تدخُلَ إلى موقع الموسوعة الحُرَّة «ويكيبيديا Wikipedia" لتجدَ فيه ما تُريد. فهل سألتَ نفسَكَ عن مصادر المعلومات الَّتي تجدُها وكيفَ تتنامى بسُرعة هائلة ؟!

تُشير «الويكي» إلى نوعٍ من المواقع الإلكترونيَّة التَّفاعُليَّة مُتعدِّدة اللُّغات الَّتي تسمحُ للزُّوَّار بصناعة مُحتواها، وتُتيحُ لهم التَّحكُّم في هذا المُحتوى بالإضافة أو الحذف أو التَّعديل؛ وتُشيرُ كذلكَ إلى الأدوات والبرامج الَّتي تُستخدَمُ لإدارة وتشغيل هذه المواقع. وتمتازُ هذه المواقعُ بأنَّها تحتوي على معلوماتٍ ومواردَ عديدة، يُمكنُ استخدامُها في العديد من أغراضِ البحث والصِّناعة والتَّدريس. وتمتازُ أيضًا بتنوُّع مادَّتها وشُموليَّتها ومُناسبتها لمُختلف الفئات الثَّقافيَّة والمُستويات العُمريَّة.

إنَّ كلمة «ويكي Wiki" تُطلَقُ في لُغة (هاواي) على الشَّيء السَّريع أو بالغ السُّرعة. وهي كلمة تُعبِّرُ فعلًا عن حال مواقع «الويكي» الَّتي تتطوَّرُ بسرعة كبيرة؛ من حيثُ المُحتوى، ومن حيثُ وسائل نشر هذا المُحتوى.

متى ظَهَرَت «الويكي»؟!

في العِقد الأخير من القرن العشرين، وتحديدًا في عام 1995، قامَ المُبرمج الأمريكيّ وورد كانينغهام Ward Cunningham بإطلاق أوَّل موقعٍ إلكتروني باسم «Wiki»؛ وهُو موقع «مُستودع بورتلاند للأنماط Portland Pattern Repository". ومع مطلع الألفيَّة الثَّالثة، بدأت أنظارُ العالم تتَّجه إلى هذا النَّوع من المواقع. وظهرَت العديد من مشروعات صناعة المُحتوى التَّفاعُليّ «Wiki"، الأمر الَّذي أدَّى إلى إحداث طفرة معلوماتيَّة كبيرة. وكانَ لهذه الطَّفرة الأثر الأكبر في بناء مُجتَمَع المعرفة الجديد.

المُحتوى الرَّقميّ للويكي

يعتقدُ الكثيرُونَ أنَّ المُحتوى الرَّقميّ للويكي يقتصرُ على النُّصُوص الموجودة في الموسوعات الحُرَّة، مثل: ويكيبيديا، وموسوعة المعرفة، ونحو ذلك. لكنَّ الحقيقة أنَّ مُحتوى الويكي أكبرُ من ذلكَ بكثير؛ حيثُ يتَّسع ليشملَ مُختلف أشكال المُحتوى الرَّقميّ من النُّصُوص المكتوبة والكُتُب المخطوطة والرُّسُوم المُتحرِّكة والصُّوَر والصَّوتيَّات والمرئيَّات، وغيرها. ومن الأمثلة الَّتي تدلُّ على تنوُّع المُحتوى الرَّقميّ للويكي، بخلاف الموسوعة الحُرَّة «ويكيبيديا»:

– **ويكيميديا كومُنز Wikimedia Commons**: تشتملُ على وسائطَ مُتعدِّدة حُرَّة بمُختلف لُغات العالم. وتأتي هذه الوسائطُ في صُورة ملفَّاتٍ صوتيَّة أو مرئيَّة أو أشكالٍ أو صُوَر توضيحيَّة. ويُعَدُّ هذا المُحتوى مخزنًا كبيرًا من الموادّ الرَّقميَّة؛ حيثُ يضمُّ أكثرَ من 37 مليون ملفّ، يُمكنُ استخدامُها في أغراضٍ مُتعدِّدة في إطارٍ تفاعُليّ وتشارُكيّ.

– **ويكي مصدر Wikisource**: تشتمل على مجموعةٍ كبيرة من النُّصُوص الَّتي لا تخضع للملكيَّة الفكريَّة، مثل النُّصُوص الدِّينيَّة، والقوانين الدَّوليَّة، والمُعاهَدات والاتِّفاقات. وأكثرُ النُّصُوص في "ويكي مصدر" لا تُنسَب إلى أفراد؛ وإنَّما إلى كياناتٍ كبيرة.

– **ويكي إجابات Wikianswers**: تشتمل على إجاباتٍ مُتعدِّدة عن الأسئلة الشَّائعة الَّتي يطرحُها مُستخدمو الحواسيب في مُختلف العُلُوم والمجالات؛ حيثُ يقومُ المُستخدمُ بتوجيه سُؤالٍ مُعَيَّن، ويقومُ آخرونَ بتوفير الإجابات المُناسبة بصورةٍ تفاعُليَّة. وتحتفظ الويكي بالأسئلة والأجوبة في صُورة قواعد بيانات تُشبهُ بُنوك الأسئلة.

– **ويكي الأطفال Wikijunior**: تشتمل على سلاسل موسوعيَّة مُبسَّطة ومُناسبة للأطفال. ويتنوَّعُ مُحتواها بينَ ما يُناسبُ مرحلة الطُّفولة المُبكِّرة، مثل: حُرُوف الهجاء والألوان والأرقام، وما يُناسبُ مرحلةَ التَّعليم الابتدائيّ، مثل: النِّظام الشَّمسيّ وتجارب العُلُوم وعناصر الطَّبيعة.

- **ويكي الأخبار Wikinews**: تشتمل على مُحتوى إخباري مُتنوِّع. ويُفتَرَضُ في هذا المُحتوى أن ينقلَ الخبرَ بصورة مُحايدة، بصرف النَّظر عن وجهات النَّظر المُختلفة بشأن قَبوله أو رفضه. وبطبيعة الحال، فإنَّ مُحتوى "ويكي الأخبار" يتنامى باستمرار، نتيجةَ التَّغيُّرات والتَّطوُّرات الَّتي تحدُثُ حولَ العالم.

- **ويكي الجامعة Wikiversity**: تشتمل على موادّ تعليميَّة تفاعُليَّة لهدف إتاحة الفُرصة للتَّعليم العالي، وتتنوَّعُ أشكالُها لتشملَ المُحاضرات المكتوبة والمسموعة والمرئيَّة.

الويكي .. واللُّغة العربيَّة

مع أنَّ تقنية الويكي مُنتشرة في العديد من لُغات العالم، إلَّا أنَّها محدودةٌ للغاية في اللُّغة العربيَّة. ولعلَّ هذا يعودُ إلى عدم إدراك القيمة الحقيقيَّة لهذه التِّقنية الَّتي تقومُ على أنَّ المعرفةَ حقٌّ للجميع، والَّتي تسمحُ لمُستخدميها بالوُصُول إلى المعارف والأفكار والبيانات الرَّقميَّة بسُرعة فائقة.

ومع أنَّ مواقعَ الويكي العربيَّة قليلة، مُقارنةً بغيرها في اللُّغات الأخرى، إلَّا أنَّها تحظى بنشاطٍ كبير، وتتنامى بصُورةٍ لافتة. ومن أهمّ هذه المواقع:

- **«ويكيبيديا Wikipedia" العربيَّة**: موقع الويكي الأشهر والأكبر بينَ مواقع الويكي العربيَّة. وتحظى ويكيبيديا العربيَّة بسُمعة طيِّبة بينَ الويكيبيديات الأخرى؛ حيثُ وصلَ عددُ مقالاتها في مطلع عام 2017 إلى أكثر من 450 ألف مقالة موسوعيَّة في ميادين ومجالات مُتعدِّدة. وجاءَت ويكيبيديا العربيَّة بذلك في المرتبة 19 بينَ الويكيبيديات الصَّادرة بمُختلف اللُّغات.

- **موسُوعة «المعرفة»**: موسوعة ويكي عربيَّة، قامَت في البداية كمشروع لصناعة المُحتوى الرَّقميّ العربيّ، تُشبه ويكيبيديا في مُحْتواها ونظام إدارتها، وتحتوي على أكثر من 100 ألف مقالة، بالإضافة إلى نحو 60 ألف صورة. وتحتوي الموسوعةُ أيضًا على آلاف الكُتُب والمخطوطات العربيَّة النَّادرة. ومع انفراد موسوعة المعرفة بهذا المُحتوى، فإنَّها تُمثِّلُ مصدرًا غنيًّا ومُهمًّا للباحثينَ العرب على اختلاف اهتماماتهم وتخصُّصاتهم.

- **«ويكي هاو "Wikihow العربيَّة**:

موسوعة ويكي تُساعدُكَ على أن تتعلَّمَ كيفَ تُصنَعُ الأشياء، وتُجيبُ عن العديد من الأسئلة الشَّائعة في أذهان رُوَّاد مُجتمع المعرفة. ستُجيبُكَ هذه الموسوعةُ مثلًا عن أسئلة على شاكلة: كيفَ تكونُ مُديرًا ناجحًا، وكيفَ تكونُ مُنظَّمًا، وكيفَ تنالُ محبَّةَ الآخَرين، وكيفَ تستخدم الحواسيب المحمولة، وغير ذلك من الأسئلة العلميَّة والحياتيَّة.

ومع أنَّ مُحتوى «ويكي هاو» العربيَّة مُمتعٌ ومُفيدٌ للغاية، إلَّا أنَّهُ لا يزالُ قليلًا، مُقارنةً بمُحتوى اللُّغات الأخرى.

– **"جوريسبيديا JurisPedia" العربيَّة:** إحدى الموسوعات الخاضعة لتقنية الويكي؛ وهي موسُوعة قانونيَّة مُختصَّة، يتمُّ تحريرُها بطريقة أكاديميَّة تتناسبُ مع الفئات المُستفيدة منها. وتأتي جوريسبيديا العربيَّة في المرتبة الأولى بينَ إصدارات اللُّغات الأخرى. ومع هذا فهي ليسَت مُنتشرة مثلَ ويكيبيديا، لأنَّ مادَّتها مُوجَّهة للمُشتغلينَ بالقانون، بخلاف ويكيبيديا الَّتي تُعَدُّ موسُوعةً عامَّة تُغطِّي جميعَ ميادين الحياة.

تطبيقات «ويكي Wiki" في ميادين المُعالجة الآليَّة للُّغة العربيَّة

لعلَّكَ حاولتَ أن تكتُبَ رسالةً أو خطابًا باستخدام حاسُوبك أو جوَّالك، فأخطأتَ في كتابة بعض الكلمات، لتجدَ الآلة تكتشفُ الخطأ وتقترحُ عليكَ حُلُولًا أخرى. فهل فكَّرتَ يومًا كيفَ اكتشَفَت الآلةُ هذا الخطأ؟! وكيفَ استطاعَت أن تُصَحِّحَه؟!

يُسَمَّى هذا التَّطبيق «المُدقِّق الإملائي»؛ ويُعَدُّ أحدَ التَّطبيقات المُهمَّة للمُعالجة الآليَّة للُّغة العربيَّة. وتقومُ فكرةُ هذا التَّطبيق ببساطة على بناء مُعجم من الكلمات الصَّحيحة ومُطابقة الكلمات الَّتي يكتُبُها المُستخدمُ بالكلمات الموجودة في هذا المُعجم؛ وإذا لم تتطابق الكلماتُ فهذا يعني وُجُودَ الخطإ فيما كتَبَهُ المُستخدم. وحينئذ تجدُ تنبيهًا واقتراحات للتَّصويب. وليسَ هذا المُعجمُ إلَّا موردًا من مواردَ عديدة تُستخدَمُ في المُعالجة الآليَّة للُّغة العربيَّة. فمن أينَ يُمكن الحُصول على هذه الموارد؟! هذا بالضَّبط ما تُقدِّمُهُ لنا «الويكي».

إنَّ الويكي مخزنٌ كبيرٌ للمُحتوى الرَّقميّ المُتعدِّد؛ ويحتوي هذاالمخزنُ على نُصُوص بلُغات مُتعدِّدة، ويحتوي أيضًا على صوتيَّاتٍ ومرئيَّاتٍ وأشكالٍ وصُوَرٍ وكُتُب، إلخ.

وإذا عرَفتَ أنَّ بناءَ تطبيقات مُعالجة اللُّغة يقومُ على أساس وجود مواردَ لُغويَّة تكونُ بمثابة قواعد البيانات الَّتي تجعلَ الآلةَ قادرةً على فهم منطق اللُّغة ومُحاكاة ذكاء الإنسان، فسيكونُ بمقدورك أن تُدركَ أهمِّيَّة الويكي في المُعالجة الآليَّة للُّغة.

إنَّنا نستطيعُ استخدامَ الويكي في بناء تطبيقات مُعالجة اللُّغة، ونستطيعُ استخدامَها أيضًا في مُعالجة أخطاء التَّطبيقات عندَ السَّعي إلى تطويرها.

سنعرضُ لكَ بعضَ الأمثلة للتَّطبيقات الَّتي يُمكنُ تطويرُها باستخدام الويكي.

– **تطبيقات مُعالجة النَّصّ**

ومن هذه التَّطبيقات: التَّرجمة الآليَّة، والتَّدقيق الإملائيّ، والتَّشكيل الآليّ، والإجابة الآليَّة عن الأسئلة، والبحث في النُّصُوص، والتَّنقيب عن الآراء.

عندَ تطوير أيِّ تطبيقٍ من هذه التَّطبيقات ستكونُ بحاجة إلى مجموعاتٍ كبيرة من النُّصُوص المُستمدَّة من واقع اللُّغة؛ إذ تقومُ بتدريب الآلة على فهمها أوَّلًا باستخدام أساليب التَّحليل الإحصائيّ، ثُمَّ تقومُ

ببناء البرمجيَّة المُناسبة اعتمادًا على الموارد المُتاحة، ثُمَّ تقومُ باختبار التَّطبيق بتجربة العمل على نُصُوصٍ أخرى.

تستطيعُ توفيرَ هذه النُّصُوص في صُورة مُنتظمة ودقيقة من خلال الويكي؛ حيثُ تُساعدُكَ نُصُوص الويكي على بناء المعاجم اللَّازمة والذَّخائر اللُّغويَّة للنُّصُوص العربيَّة وقواعد بيانات الحُرُوف والجُمَل والتَّراكيب وغيرها. ولأنَّ هذه النُّصُوص تُعبِّرُ عن واقع اللُّغة العربيَّة المُستخدَمة، فسيكونُ التَّطبيقُ الَّذي تصنعُهُ على درجةٍ كبيرة من الدِّقَّة والكفاءة.

– تطبيقات مُعالجة الصَّوت

ومن هذه التَّطبيقات: تحويل النَّصّ المكتوب إلى كلام منطوق، والتَّعرُّفً الآليّ على الكلام، والبحث في الصَّوتيَّات والمرئيَّات، والتَّعرُّف على اللُّغة، والاستجابة الصَّوتيَّة التَّفاعُليَّة.

ويستدعي تطوير هذه التَّطبيقات وُجُودَ مادَّة صوتيَّة نقيَّة وعالية الجودة، لتكونَ موردًا لقواعد البيانات الَّتي نصنعُها، ولتكونَ مادَّةً للمُعالَجة والتَّقييم. ويكفي أن تعرفَ أنَّ هناكَ مُؤسَّسات تُكرِّسُ جُهُودَها في توفير المُحتوى الصَّوتيّ الرَّقميّ، لهدف الاستفادة منه في مُختلف تطبيقات مُعالجة الصَّوت. وحينَ نتحدَّثُ عن الصَّوت العربيّ، فينبغي أن نُدرِكَ اختلافَه عن أصوات اللُّغات الأخرى في كثيرٍ من مخارج الحُرُوف وصفاتها؛ الأمر الَّذي يدفعُنا إلى توفير مادَّة صوتيَّة ذات طبيعة خاصَّة.

تُساعدُنا الويكي في الحُصُول على المادَّة الصَّوتيَّة الَّتي نحتاجُ إليها، لأنَّها غنيَّة بالمُحتوى الصَّوتيّ الحُرّ. وعندَ مُعالجة هذا المُحتوى، فسوفَ نحصل على نتائجَ أوَّليَّة، يُمكنُ من خلالها أن نُحدِّدَ أساليب تطوير التَّطبيقات. ونستطيعُ أيضًا في مرحلة لاحقة أن نتأكَّدَ من جودة التَّطبيقات الَّتي قُمنا بتطويرها باستخدام هذه المادَّة. وبالإضافة إلى إمكانيَّة الاستفادة من هذا المُحتوى في تطوير تطبيقات مُعالجة الصَّوت وتقييمها، فإنَّنا نستطيعُ الاستفادةَ منها في توليد الصَّوت لهدف القراءة الآليَّة، على النَّحو الَّذي نجدُهُ في أدوات القراءة الآليَّة للنُّصُوص المكتوبة.

– تطبيقات مُعالجة الصُّورة

ومن هذه التَّطبيقات: التَّعرُّف على الكتابة المطبوعة، والتَّعرُّف على الكتابة المخطوطة، والتَّعرُّف الذَّكيّ على الحُرُوف، وتمييز التَّوقيعات الصَّحيحة من المُزَيَّفة، وتعليم الخطّ باستخدام الحاسوب.

والصُّورةُ الَّتي تهتمُّ بها هذه التَّطبيقاتُ هي صُورةُ الكتاب المطبوع أو المخطوط؛ حيثُ تُعَدُّ موردًا أساسيًّا يُساعدُ الآلة على تمييز أنواع الخُطُوط وأشكالها وأحجامها، تمهيدًا لمُعالجتها، لهدف التَّحكُّم فيها. وإذا

استطعنا توجيهَ الآلة إلى التَّحكُّم في الصُّورة المكتوبة المطبوعة أو المخطوطة، فإنَّنا ننقلُ النَّصَّ من صُورته السَّاكنة إلى صُورةٍ مُتحرِّكة.

لقد ذكرنا أنَّ الويكي تضمُّ مجموعات كبيرة من الكُتُب. وتُعَدُّ هذه الكُتُبُ موردًا مُهمًّا للغاية في مُختلف تطبيقات مُعالجة الصُّورة، نظرًا لتنوُّعها واختلاف أشكال طباعتها. قد يبدو الأمرُ مُعقَّدًا بعضَ الشَّيء، لكنَّهُ يتمُّ بالتَّدريج؛ حيثُ يرتكزُ عملُ الآلة أوَّلًا على استكشاف الكلمات وتعيين المقاطع وتمييز الحُرُوف؛ ثُمَّ تتمُّ المُعالجاتُ اللَّازمة للوُصُول إلى التَّحكُّم الكامل في الصُّورة.

– **تطبيقات التَّعليم الإلكترونيّ**

لا شكَّ أنَّكَ تحتاجُ إلى تعلُّم العديد من الأشياء الَّتي لا تجدُها في المدرسة والجامعة. وقد تحتاجُ أيضًا إلى فهم أعمقَ لبعض ما تدرسُه، فتجد صُعُوبة في الحُصُول على المعلومات الَّتي تُريدُها. ولأجل هذا، فإنَّ التَّعليمَ الإلكترونيّ سيُساعدُكَ على تحقيق أهدافكَ وتنمية مهاراتك العلميَّة واللُّغويَّة والإبداعيَّة؛ وسيجعلُكَ قادرًا على التَّعلُّم الذَّاتيّ دونَ الحاجة المُستمرَّة إلى التَّلقين.

وحتَّى تحصُلَ على تعليم إلكترونيٍّ جيِّد، فأنتَ بحاجة إلى برامجَ تعليميَّةً مُناسبة، لا تُسبِّبُ لكَ الرَّتابةَ والملل، وتشعرُ من خلالها أنَّ العالمَ كتابٌ مفتوحٌ بينَ يديك، وأنَّ بإمكانكَ أن تحصُلَ على ما تشاء من المعرفة بسُهُولةٍ ويُسر.

هذا ما يُقدِّمُهُ لكَ مُجتمع الويكي اللَّامحدود؛ حيثُ يُمكنُ الاستفادة من مُحتوى الويكي من الموسوعات والمعارف والصَّوتيَّات والمرئيَّات والصُّوَر والرُّسُوم المُتحرِّكة وغيرها في الحُصُول على تعليم إلكترونيٍّ تفاعُليّ. وممَّا يتميَّزُ به المُحتوى التَّعليميّ للويكي أنَّهُ لا يُهملُ الجوانبَ التَّرفيهيَّة؛ كما أنَّهُ يسمحُ لكَ بالتَّفاعُل مع الآخرين وتبادُل المعرفة معهم.

ببليوغرافيا مرجعيَّة

– الموسوعة الحُرَّة «ويكيبيديا Wikipedia"، 2017.

– Anja Ebersbach, Markus Glaser, Richard Heigl, Alexander Warta. (2008). Wiki: Web Collaboration. Springer Science & Business Media.

– Daniel J. Barrett. (2008). MediaWiki: Wikipedia and Beyond. O›Reilly Media, Inc.

– Rod Collins. (2010). Leadership in a Wiki World: Leveraging Collective Knowledge to Make the Leap to Extraordinary Performance. Dog Ear Publishing.

ملف العدد

تقنيات معالجة الصوت اللّغوي

د. المعتزّ باللّه السعيد

الأصوات العربيَّة

حينَ كُنَّا صغارًا، تعلَّمنا في المدرسة أنَّ اللُّغة العربيَّة تتكوَّنُ من 28 حرفًا؛ وتعلَّمنا أيضًا كيفَ نكتُبُ الحُروفَ وكيفَ نستخدمُها في كلماتٍ مُختلفة. ولكن، هل سألتَ نفسَكَ عن عدد الأصوات الَّتي ننطقُها في اللُّغة العربيَّة؟! قد يبدو السُّؤالُ غريبًا، لأنَّ الكثيرينَ يظُنُّونَ أنَّ عددَ الأصوات العربيَّة يُساوي عددَ حُروفِها؛ لكنَّ الأمرَ ليسَ كذلك.

تتكوَّنُ اللُّغةُ العربيَّة من 34 صوتًا، تشملُ أصوات الحُرُوف الثَّمانية والعشرين (ء، ب، ت، ث، ج، ح، خ، د، ذ، ر، ز، س، ش، ص، ض، ط، ظ، ع، غ، ف، ق، ك، ل، م، ن، هـ و، ي) بالإضافة إلى ستَّة أصوات أخرى، هي أصوات الحركات (الفتحة القَصيرة والضَّمَّة القَصيرة والكسرة القَصيرة، والفتحة الطَّويلة والضَّمَّة الطَّويلة والكسرة الطَّويلة).

ونحنُ في علم الأصوات، نُسَمِّي الصَّوتَ اللُّغَويّ (الفُونيم)، ونُسمِّي أصوات الحُرُوف (الصَّوامت) ونُسَمِّي أصواتَ الحركات (الصَّوائت). وبهذا، يُمكنُ أن نقولَ إنَّ اللُّغة العربيَّة تتكوَّنُ من 34 فونيمًا، هي مجموع الصَّوامت (28 صامتًا) والصَّوائت (6 صوائت). وهذه الفونيمات هي المادَّة الأساسيَّة الَّتي نعتمدُ عليها في مُعالجة الصَّوت العربيّ.

لماذا نُعالجُ الصَّوتَ اللُّغَويّ؟!

لا شكَّ أنَّكَ تقضي ساعاتٍ طويلة مع الآلة (الحاسوب، الجوَّال، سمارت بوك، ...) دونَ أن تشعُرَ بمُرُور الوقت. لقد أصبَحَت هذه الأجهزة جُزءًا من حياتنا، نعيشُ معها أكثرَ ممَّا نعيشُ مع أهلنا وأصحابنا. ولأنَّنا في عصر السُّرعة، كانَ علينا أن نصنَعَ عالمًا افتراضيًّا تفاعُليًّا بينَ الإنسان والآلة. ومن مُتطلَّبات هذا العالم التَّفاعُليّ تمكين الآلة من فهم أصوات الإنسان وتمييزها عن أصوات الأشياء والكائنات الأخرى. واستطاعَ الإنسانُ من خلال التَّجربة أن يُحَقِّقَ الهدفَ ليجعلَ الآلةَ كيانًا مُتحرِّكًا يستمعُ لأوامره ويفهمها ويُنفِّذها. وشيئًا فشيئًا، قامَ العُلماءُ بتطوير أداء الآلة، لتقومَ بمُعالجة أصوات اللُّغة وتوليدها والتَّفاعُل مع الإنسان من خلالها، على النَّحو الَّذي يجعلُكَ تشعُرُ أنَّ الآلةَ إنسانٌ آخَر، يفهمُكَ ويستجيبُ لأوامركَ ويُساعدُكَ في عمليَّات البحث والتَّعليم والتَّعلُّم، ويُوفِّرُ عليكَ الوقتَ والجُهدَ في إنجاز العديد من الأهداف. وحتَّى تتَّضحَ لكَ أهمِّيَّةُ الأمر، فسوفَ نعرضُ بعضَ تقنيات مُعالجة الصَّوت اللُّغويّ.

التَّعرُّف على الكلام العَرَبيّ

إذا كُنتَ حريصًا على الاستماع إلى المُحاضرات في المدرسة أو الجامعة، فلا بُدَّ أنَّ بعضَ هذه المُحاضرات قد حازَت إعجابَكَ فأردتَ أن تكتُبَها، لتقرأها وتستفيدَ منها وتعودَ إليها عندَ الحاجة إليها. هذا ما تُقدِّمُهُ تقنيةُ التَّعرُّف على الكلام؛ حيثُ تقومُ بتحويل الكلام المنطوق إلى نصٍّ مكتوب، بعدَ تمكين الآلة أو الحاسُوب من تمييز الأصوات وترجمتها إلى حُرُوف مكتوبة. والواقعُ أنَّ هذه التِّقنية ليسَت بسيطةً كما يبدو؛ بل هي مُعَقَّدة إلى درجة كبيرة.

لقد عَرَفنا أنَّ اللُّغة العربيَّة تتكوَّن من أربعة وثلاثينَ صوتًا؛ وحتَّى تتمكَّنَ الآلةُ من تمييز هذه الأصوات، فلا يكفي أن تتعرَّفَ عليها بصورتها المُجرَّدة؛ وإنَّما ينبغي أن تتعرَّفَ عليها بتنويعاتها المُختلفة، وبأشكالها المُفرَدة والمُركَّبة. ولا يكفي مثلًا أن تتعرَّفَ الآلةُ على صوت «اللَّام». بل لا بُدَّ من تمكينها من التَّعرُّف على «اللَّام» المفتوحة

والمضمومة والمكسورة والسَّاكنة، ولا بُدَّ من تمكينها أيضًا من التَّعرُّف على نوع الحركة المُصاحبة للَّام، سواءٌ أكانَت حركةً قصيرة أم طويلة؛ وكذلك لا بُدَّ من تمكين الآلة من التَّعرُّف على صفات اللَّام بينَ التَّفخيم والتَّرقيق ... وهكذا. وقِس على ذلكَ جميعَ الأصوات الأخرى. وحينئذ، سنجدُ أنفُسَنا أمامَ آلاف الأشكال والتَّنويعات الصَّوتيَّة. وبطبيعة الحال، إذا أردنا توجيهَ الآلة إلى التَّعرُّف على هذه الأشكال والتَّنويعات، فلا بُدَّ من مُراعاة أمورٍ عديدة، أهمُّها: وُضُوحُ الصَّوت، وشدَّة الصَّوت، ونقاء الصَّوت من التَّشويش والضَّوضاء، وسلامة مخارج الأصوات من الأخطاء. وهُناكَ العديدُ من التَّطبيقات الَّتي تُستَخدَمُ فيها تقنيةُ التَّعرُّف على الكلام، نعرضُ لكَ بعضًا منها فيما يأتي.

– **أنظمة الاستجابة الصَّوتيَّة التَّفاعُليَّة:** وهي الأنظمة الَّتي تسمحُ بالتَّفاعُل بينَ الإنسان والآلة. ونجدُ هذه الأنظمة عندَ طلب الخدمات الإلكترونيَّة في تطبيق الحجز الآليّ لتذاكر السَّفر، وفي عمل الحكومات الإلكترونيَّة الَّتي تُساعدُكَ على إنهاء الإجراءات عبرَ الآلة، وفي المُكالمات المُسَجَّلة عبرَ الهاتف للرَّدّ على الاستفسارات، وغير ذلك.

– **التَّعرُّف على المُتكلِّم:** وهُو تطبيقٌ مُهمٌّ يُساعدُ على تمييز أصوات المُتكلِّمين والتَّعرُّف عليهم؛ ويُستخدَمُ للأغراض الاستخباريَّة والأمنيَّة، خُصُوصًا في ميادين البحث الجنائيّ الَّتي تسعى إلى منع حُدُوث الجرائم أو كشف غُموضها والتَّعرُّف على مُرتكبيها.

– **إعداد الوثائق المُنظَّمة:** ويُساعدُ هذا التَّطبيقُ في كتابة المُحاضرات وتحرير محاضر الاجتماعات وإعداد التَّقارير. ويُفيدُ بصورةٍ كبيرة في ميادين الرِّعاية الصِّحِّيَّة بإعداد سجلَّات المرضى؛ ويُفيدُ أيضًا في إدارة وحدات التَّحكُّم في حركة السُّفُن والطَّائرات.

– **التَّعليم الآليّ:** وهُوَ من أشهر التَّطبيقات المُستخدمة لتقنية التَّعرُّف على الكلام؛ حيثُ يُساعدُ هذا التَّطبيقُ في تنمية مهارات المُتعلِّمينَ في مراحلهم العُمريَّة المُختلفة، من خلال توفير التَّفاعُل المُباشر بينهم وبينَ الآلة الَّتي تُؤدِّي دَورَ المُعلِّم.

تحويل النَّصّ إلى كلام

هل سَمعتَ عن القارئ الآليّ؟! إنَّهُ تطبيقٌ حاسُوبيّ، يُساعدُكَ على قراءة الكُتُب والنُّصُوص باختلاف أشكالها، ويعتمدُ على تقنية تحويل النَّصّ إلى كلام.

تقومُ فكرةُ التِّقنية ببساطة على ترجمة الحُرُوف إلى إشاراتٍ رقميَّة مُناظرة للنَّصّ المكتوب، وترجمة الأصوات إلى إشاراتٍ رقميَّة مُناظرة للكلام المنطوق، ثُمَّ تحويل إشارات النَّصّ المكتوب إلى ما يُناظرُها من إشارات الكلام المنطوق. يبدو الأمرُ سهلًا؛ أليسَ كذلك؟!

الأمرُ سهلٌ بالفعل إذا كُنَّا نتعاملُ مع الصَّوت اللُّغويّ بمعزلٍ عن الانفعالات المُصاحبة له؛ ولكن إذا أردتَ التَّعامُلَ مع الأصوات كما ننطقُ بها، فالأمرُ بالغُ الصُّعوبة. فأنتَ تنطقُ الكلمةَ الواحدة، فينتقل معها شُعورٌ لدى مَن يسمعُكَ بأنَّكَ سعيدٌ أو حزينٌ أو غاضب .. وقد تنطقُ العبارةَ

الواحدة فيظهرُ للسَّامع أنَّها استفهامٌ أو إخبارٌ أو استنكار. والواقعُ أنَّ تمكينَ الآلة من استيعاب هذه الانفعالات يجعلُ الأمرَ صعبًا؛ لكنَّ التَّجاربَ لا تزالُ قائمة من أجل تطوير هذه التِّقنية.

ولكن، إذا كنتُ أعرفُ القراءة، فما الفائدة من مثل هذا التَّطبيق؟! الحقيقةُ أنَّه مُهمٌّ للغاية، ويُستخدمُ في العديد من الأغراض. من ذلكَ مثلًا: قراءة الكُتُب لذوي الاحتياجات الخاصَّة من فاقدي القُدرة على الإبصار، وتعليم اللُّغة العربيَّة للنَّاطقينَ بغيرها، وقراءة الرَّسائل النَّصِّيَّة عبرَ الهاتف للإخبار عن معلوماتٍ مُعَيَّنة، وقراءة الرَّسائل التَّحذيريَّة عندَ التَّنبُّؤ بالخطر، وغير ذلك.

تحويل الكلام إلى لُغة الإشارة

لهذه التِّقنية طبيعة خاصَّة؛ لأنَّها تُساهمُ في مُساعدة ذوي الاحتياجات الخاصَّة من فاقدي القُدرة على السَّمع والكلام على مُمارسة الحياة بطريقة طبيعيَّة، دونَ مُعاناة أو شُعور بالاختلاف عن غيرهم، الأمرُ الَّذي يُساعدُهم على التَّفاعُل والتَّعايُش مع المُحيط الَّذي يعيشُونَ فيه، ويفتحُ لهم البابَ أيضًا لخدمة المُجتمع والمُشارَكة في إعماره، من خلال توفير وسيلة للتَّواصُل وإدارة الحوارات وفهم الكلام المُوجَّه إليهم.

يُستفادُ من هذه التِّقنية في بناء تطبيق تحويل الكلام إلى لُغة الإشارة؛ وهي اللُّغة الَّتي نتحكَّمُ فيها باليَدِ لإفهام هذه الفئة من فاقدي القُدرة على السَّمع والكلام، ونراها في بعض المُسلسلات والبرامج التلفزيونيَّة.

تمرُّ هذه التِّقنية بمرحلَتَين؛ حيثُ نقومُ في المرحلة الأولى بتمكين الآلة من تحويل الكلام المنطوق إلى نصٍّ مكتوب، ثُمَّ نقومُ في المرحلة الثَّانية بتمكين الآلة من تحويل النَّصّ المكتوب إلى لُغة الإشارة. ونعتمدُ في هذه التِّقنية على قواعد بياناتٍ تتضمَّنُ مُختلَف أشكال الإشارات المُستخدَمة في التَّعبير عن الحُروف والكلمات، من خلال ما يُعرَفُ بـ «القامُوس الإشاريّ».

التَّرجمة الشَّفهيَّة الآليَّة

أحيانًا نقرأُ كتابًا بإحدى اللُّغات الأخرى، كالإنجليزيَّة أو الفرنسيَّة؛ وحينَ تستعصي علينا بعضُ العبارات، فإنَّنا نجتهدُ في ترجمتها إلى العربيَّة؛ ونستعينُ ببعض المواقع الإلكترونيَّة الَّتي تُوفِّرُ خدمةَ التَّرجمة الآليَّة.

ونظرًا لأهمِّيَّة التَّرجمة الآليَّة وحاجتنا إليها في عصر المعلومات، فقد تنوَّعَت أشكالُها. ومن هذه الأشكال «التَّرجمة الشَّفهيَّة الآليَّة»؛ وهي

تقنيةٌ مُتطوِّرة، تهدفُ إلى ترجمة الكلام المنطوق بينَ لُغَتين، وتُستخدَمُ فيها تقنيةُ التَّعرُّف على الكلام باعتبارها وسيلةً وسيطةً للتَّرجمة.

إذا أردنا ترجمةَ كلام منطوق من اللُّغة العربيَّة إلى اللُّغة الإنجليزيَّة مثلًا، فإنَّنا نُطلقُ على العربيَّة «اللُّغة المصدر» ونُطلقُ على الإنجليزيَّة «اللُّغة الهدف». وما نفعلُهُ ببساطة أنَّنا نقومُ بالتَّعرُّف على الكلام العربيّ أوَّلًا، ثُمَّ نَضَعُهُ في صُورة نُصُوص عربيَّة مكتوبة، ونقومُ بترجمة هذه النُّصُوص إلى ما يُقابلُها في نُصُوص اللُّغة الإنجليزيَّة؛ وأخيرًا نقومُ بتحويل الإشارات الرَّقميَّة للنُّصُوص الإنجليزيَّة المكتوبة إلى ما يُناظرُها من إشارات رقميَّة للكلام المنطوق بالإنجليزيَّة. ونفعلُ نفسَ الخُطوات إذا أردنا التَّرجمةَ من الإنجليزيَّة إلى العربيَّة، مع تغيير اللُّغة المصدر واللُّغة الهدف.

ولكن، ما الدَّاعي إلى التَّرجمة الشَّفهيَّة الآليَّة، ما دامَت الآلة قادرةً على ترجمة النُّصُوص المكتوبة؟!

ستعرفُ أهمِّيَّةَ هذه التِّقنية إذا حضرتَ محفلًا أو مُؤتمرًا دوليًّا يضمُّ عددًا كبيرًا من المُشاركينَ الَّذينَ يتحدَّثونَ بلُغات مُختلفة. انظر مثلًا إلى اجتماعات الجمعيَّة العامَّة للأمم المُتَّحدة الَّتي يُشاركُ فيها مُلوكٌ ورُؤساء ومُمثِّلونَ عن جميع الدُّوَل الأعضاء في الأمم المُتَّحدة. ستُلاحظُ أنَّ كُلَّ ملك أو رئيس يتحدَّثُ بلُغة بلاده، مع أنَّها قد لا تكونُ مفهومةً لأكثر الحُضُور. وستُلاحظُ أيضًا أنَّ المُستمعينَ يُنصتونَ ويضعونَ أجهزة سمعيَّة خاصَّة على آذانهم. وظيفةُ هذه الأجهزة أن تقومَ بالتَّرجمة الشَّفهيَّة من لُغة المُتحدِّث إلى لُغة المُستمع.

في مثل هذه الاجتماعات، يتواجدُ عددٌ من المُترجمينَ المُجيدين الَّذينَ تكونُ وظيفتُهم القيام بعمليَّة التَّرجمة الشَّفهيَّة الفوريَّة. ومع تمكين الآلة من القيام بدَور المُترجم، أصبحَ بمقدورنا الاعتمادُ بصورة كبيرة على تقنية التَّرجمة الآليَّة الشَّفهيَّة. والحقيقة أنَّ هذه التِّقنية لم تصل بعد إلى الشَّكل المأمول؛ لكنَّ التَّجاربَ القائمة حولَها تدعو إلى التَّفاؤل بشأن المُستقبَل.

البحث في الصَّوتيَّات والمرئيَّات

ماذا تفعلُ إذا أردتَ البحثَ عن موضوع مُعَيَّن على صفحات الويب؟! حتمًا ستستخدمُ أحدَ مُحرِّكات البحث (مثل: جوجل Google، أو ياهو Yahoo، أو بينغ Bing، ...) للوُصُول إلى ما تبحثُ عنه. أمَّا إذا كُنتَ تبحثُ عن ملفٍّ على حاسُوبكَ أو جوَّالك، فإنَّكَ ستقومُ بكتابة عُنوان الملفّ، وتستخدمُ أدوات البحث المُتاحة على الحاسوب أو الجوَّال. ستُلاحظُ أنَّ مُحرِّكات البحث على صفحات الويب

أو في الأجهزة الذَّكيَّة تبحثُ عن النُّصُوص. فهل فكَّرتَ أنَّ بالإمكان البحثَ عن موضوعٍ مُعَيَّنٍ في عددٍ كبيرٍ من الملفَّات الصَّوتيَّة أو المرئيَّة؟!

دعنا نشرحُ الأمرَ بصورةٍ أخرى. افتَرِض أنَّ حاسُوبَكَ يضُمُّ عددًا كبيرًا من المحاضرات الصَّوتيَّة المُتنوِّعة، وأنَّكَ تبحثُ عن موضوعٍ مُعَيَّنٍ حولَ كُرة القَدم، وتُريد استدعاء جميع الملفَّات الصَّوتيَّة الَّتي تناوَلَت هذا الموضوع. هذا ما تقومُ به تقنيةُ البحث في الصَّوتيَّات والمرئيَّات.

الواقعُ أنَّنا نستخدمُ هذه التِّقنية كثيرًا دونَ أن نشعُر. مثلًا، عندَ البحث عن مقطعٍ صوتيٍّ أو مرئيٍّ على موقع الويب (يوتيوب Youtube)، يكفي أن تكتُبَ الموضوعَ الَّذي تبحثُ فيه، ليستدعي الموقعُ العديدَ من المقاطع الَّتي تُساعدُكَ على الوُصُول إلى هدفك، حتَّى لو كانت الكلمة الَّتي أدخَلتَها غيرَ موجودة في بعض عناوين المقاطع الَّتي استدعاها الموقع.

الأمرُ ببساطة أنَّ مثلَ هذا الموقع يكونُ مُزوَّدًا بتقنيةٍ دقيقة للبحث في الصَّوتيَّات والمرئيَّات؛ حيثُ يتمُّ مُعالجة الأصوات اللُّغويَّة الموجودة في المقاطع، ثُمَّ تحديد مجموعة من الكلمات المفتاحيَّة المُهمَّة لكلّ مقطع. وعندَ البحث، تقومُ الآلة بمُطابقة كلمات البحث الَّتي قُمتَ أنتَ بإدخالها بالكلمات المفتاحيَّة المحفوظة في قواعد البيانات، ثُمَّ تستدعي جميعَ المقاطع الَّتي يحدُثُ فيها التَّطابُق.

ومن أهمّ تطبيقات هذه التِّقنية تطبيق صناعة الأرشيف الصَّوتيّ والمرئيّ في محطَّات الإذاعة والتلفزيون. وهو التَّطبيق الَّذي يُستخدَمُ في صناعة الأفلام الوثائقيَّة وإعداد التَّقارير حولَ الأشخاص والأحداث؛ حيثُ يقومُ هذا التَّطبيقُ باستدعاء المادَّة الصَّوتيَّة أو المرئيَّة عندَ الحاجة إليها، ويُساعدُ أيضًا على توثيق المعلومات الموجودة في هذه المادَّة.

تعلُّم النُّطق

تُعَدُّ هذه التِّقنيةُ إحدى تطبيقات التَّعلُّم الآليّ. ونقومُ فيها بمُعالجة الصَّوت اللُّغويّ لهدف تمكين الآلة من القيام بدور مُعلِّم النُّطق الصَّحيح للُّغة. ويُمكنُ الاستفادة من هذه التِّقنية في تطبيقاتٍ عديدة، منها: تعليم الأطفال، ومُعالجة أمراض الكلام لمن يجدُ صُعوبةً في نُطق الحُـــرُوف أو الكلمات، وكذلكَ في تعليم النُّطق الصَّحيح لمُتعلِّمي اللُّغة من النَّاطقينَ بغيرها؛ وسنُلاحظُ

أهمِّيَّةَ ذلكَ في تعليم اللُّغة العربيَّة؛ لأنَّها تضمُّ عددًا من الحُرُوف غير الموجودة في كثيرٍ من اللُّغات الأخرى، مثل: العَين والغَين والحاء والخاء. ونظرًا لارتباط هذه التِّقنية بأساليبَ تربويَّة، فإنَّها تعتمدُ على بعض النَّظريَّات اللُّغويَّة الَّتي تهتمُّ بوسائل اكتساب اللُّغة وتعليمها. وتستدعي هذه التِّقنيةُ توفيرَ بيئة تفاعُليَّة افتراضيَّة، لإتاحة الفُرصة للمُتعلِّم والآلة على حدٍّ سواءٍ من التَّواصُل الفعَّال.

ببليوغرافيا مرجعيَّة

- Gregor Donaj, Zdravko Kačič. (2017). Language Modeling for Automatic Speech Recognition of Inflective Languages: An Applications-Oriented Approach Using Lexical Data. Springer.
- John W. Schwieter, Aline Ferreira. (2017). the Handbook of Translation and Cognition. John Wiley & Sons.
- K. Sreenivasa Rao, Manjunath K E. (2017). Speech Recognition Using Articulatory and Excitation Source Features. Springer.
- Michael M. Richter, Paul Sheuli. (2017). Signal Processing and Machine Learning with Applications. Signal Processing and Machine Learning with Applications.
- Thomas Fang Zheng, Lantian Li. (2017). Robustness Related Issues in Speaker Recognition. Springer Singapore.
- TruthBeTold Ministry, Joern Andre Halseth. (2017). American Standard Bible 1901 - TTS: Text To Speech Friendly. TruthBeTold Ministry.