هندسة اللغة العربية في خدمة البحث الطبي

الأستاذ محمد كريم بوزوبع

حاصل على ماجستر عام 1994 ودكتوراه عام 1998 من قسم تكنولوجيا المعلومات في جامعة لافال بكندا.

خبير لدى شركة eVision الكندية والوكالة الأمريكية للتعاون الدولي ولدى المنظمة العربية للتربية والثقافة والعلوم.

رئيس فريق البحث حول هندسة اللغة العربية بالمدرسة المحمدية للمهندسين، جامعة محمد الخامس بالرباط.

محكم للأعمال في مجال هندسة اللغة العربية ب15 ندوة ومجلة دولية.

المعلومات والمعرفة والشبكة العنكبوتية

كانت المعلومات قديما تتداول بطريقة شفهية ولا تدون، وإنما كان يقتصر فقط على ذاكرة الإنسان في تخزينها والحفاظ عليها. ومع ظهور الكتابة والطباعة، بقيت المعلومات تتداول بطريقة شفهية مع الإعتماد على الورق والطباعة في عملية التخزين والتداول أيضا.

ومنذ أواخر القرن العشرين، عرف تداول المعلومات واستخدامها تطورا كبيرا، حيث أصبحت المعلومة رقمية أكثر منها شفهية وورقية؛ بل أصبح الإنسان لا يتداول المعلومات إلا باستعمال أجهزة ناقلة للمعلومة الرقمية كالحواسيب والهواتف النقالة وخاصة منها الذكية. أما عملية التخزين فتعتمد استعمال الشبكة العنكبوتية أو الإنترنيت كخزان ضخم يحتوي على رصيد هائل من المعطيات والمعلومات الرقمية (تقدر بالمليارات) التي يتداولها الناس فيما بينهم.

أدى هذا الإنفجار المعلوماتي إلى ظهور مفهوم "المعطيات الضخمة" التي يصعب معالجتها واستغلالها بدون استعمال برامج معدة ومخصصة لذلك. هذا الانفجار المعرفي الرقمي والتدفق السريع للمعلومات في مختلف الوسائط وعلى رأسها شبكة الانترنيت أدى إلى تزايد الطلب

على الاستعانة بعالم التكنلوجيا والحاجة إلى استعمال تطبيقات تكنولوجيا المعلومات في كافة مجالات الحياة وخاصة منها المتعلقة بهندسة اللغات الطبيعية.

الهندسة اللغوية : التعريف والاستغلال

نقصد بهندسة اللغات الطبيعية، ويصطلح عليها أيضا بالمعالجة الآلية والحاسوبية للغة، المجال الذي يعنى باستعمال الحاسوب والخوارزميات لتحليل وفهم النصوص الرقمية مكتوبة بلغة معينة. هذا المجال هو الذي يسهل إذن معالجة واستغلال المعرفة الرقمية المتداولة في زمننا. إن مجالات تطبيق هندسة اللغة كثيرة ومتعددة منها ما نستعمله يوميا في عملنا، ونذكر منها محركات البحث التي تساعد على الولوج إلى المعلومات الرقمية المتواجدة على الشبكة انطلاقا من كلمات مفاتيح.

الترجمة الآلية هي أيضا من بين التطبيقات المتواجدة والتي تمكن المستعمل من الحصول على ترجمة نص معين بين لغتين. أضف إلى ذلك، تطبيقات التلخيص الآلي واستخلاص المعلومات والكشف عن رسائل البريد الإلكتروني المزعجة والأنظمة التي تجيب آليا على الأسئلة، ... إلخ.

جعل هذا الزخم الهائل من التطبيقات والأدوات الممكنة، مجموعة كبيرة من الجامعات ومراكز البحث والشركات العالمية تولي اهتماما كبيرا لهذا المجال، نظرا للأرباح الطائلة التي تجنيها من خلال استغلالها في مجال الإستشهار. كما أصبحت هندسة اللغات الطبيعية من إحدى الصناعات التي تعنى باهتمام كبير لدى جل الدول المتقدمة حيث تستغلها ليس فقط في التنقيب والبحث في الفضاء المعرفي الرقمي وإنما تسخرها أيضا لحماية لغاتها وتستعملها لإسهامها الفعال في الاقتصاد بالإضافة إلى دورها الفاعل في حماية أمن الدولة.

المحتوى الرقمي العربي

أما فيما يخص مفهوم المحتوى العربي على الشبكة، ويعرف أيضا بالمحتوى الرقمي العربي، فهو مجموع المواقع والصفحات التي كتبت باللغة العربية أو الفيديوهات أو الموسيقى أو غيرها على الشبكة العالمية. وهو يمثل تقريبا 3% من مجموع المحتوى لكل اللغات على الشبكة.

تعتبر نسبة حضور اللغة العربية على الشبكة ضعيفة جدا، إلا أن نموها على الشبكة العنكبوتية خلال السنوات الأخيرة وصل الى حد 2500% حيث حظيت مشكلات المعالجة الحاسوبية للعربية باهتمام غير مسبوق من طرف العديد من الجامعات والمعاهد والمؤسسات البحثية بالإضافة إلى الشركات في العالم كله.

إن ارتفاع نسبة حضور اللغة العربية على الشبكة العنكبوتية أوجب على الباحث العربي التمكن من هذه التقنيات واستعمالها على أوسع نطاق في جميع الميادين والأنشطة. من هنا فإن هندسة اللغة العربية تكتسي أهمية بالغة باعتبار هذه الأخيرة لغة أكثر من 400 مليون نسمة.

لقد شهدت هندسة اللغة العربية بدورها تطورا ملحوظا مما مكن من تطوير مجموعة من التطبيقات مماثلة لما هو معتمد في اللغات الأخرى خاصة منها اللاتينية، حيث تمكن محركات البحث من استرجاع الملفات المكتوبة باللغة العربية والأنظمة المترجمة والتصحيح الآلي في برامج تحرير الكتابة وتطبيقات التشكيل الآلي،... إلخ. ورغم الجهود المبذولة من لدن المختصين في هذا المجال ورغم تزايد عدد التطبيقات ورغم نمو حجم المحتوى الرقمي العربي على الشبكة العنكبوتية خلال السنوات القليلة الماضية، لم ترق اللغة العربية بعد إلى المستوى المطلوب ولم تصل هندسة حوسبتها الى نضج نظيراتها اللاتينية.

تطبيقات في المجال الطبي

يعد المجال الطبي من المجالات التي تحتاج إلى اعتماد التطبيقات التكنلوجية واستغلال الهندسة اللغوية العربية نظرا لغنى هذا المجال بالمفردات المتخصصة التي تكاد تفوق جميع المجالات الأخرى في عدد المصطلحات التي يتوجب على كل عضو من الهيئة الطبية معرفتها وفهمها. تعد إذن بلورة وتحميل معاجم طبية رقمية، عامة و متخصصة، على الأجهزة الالكترونية من حواسيب وهواتف نقالة ذات أهمية بالغة. تتم عملية تطوير هذه المعاجم الرقمية بطريقة آلية وذلك بتنقيب واستخلاص المعلومات الطبية من الموارد المكتوبة (المعاجم المتوفرة باللغات الأخرى أوالنصوص الطبية المتخصصة).

بالإضافة إلى ما سبق ذكره، يجب التفكير أيضا في تطوير الأنطولوجيات (المعاجم الدلالية) الطبية المتخصصة التي تتوفر على مفاهيم طبية وعلاقات دلالية ترتبط فيما بينها. نشير إلى أن هذه المعاجم متوفرة باللغة الإنجليزية في العديد من التخصصات، في حين أن اللغة العربية لا تتوفر على أية انطولوجيا طبية عربية.

أما بخصوص تطبيقات محركات البحث، فرغم أن هذه الأخيرة متوفرة وقادرة على استيعاب اللغة العربية إلا أنها لا تستجيب لخصوصيات اللغة العربية ولا لخصائص الميدان الطبي. يجب ألا يقتصر تطوير محرك بحث متخصص على إيجاد النصوص التي تحمل الكلمات المفاتيح المدخلة من طرف المستعمل ولكن يجب أن يتوفر أيضا على الكلمات القريبة من الناحيتين الصرفية والدلالية الطبية. فحين يبحث المستعمل مثلا على نصوص تتضمن كلمة "قلوب"، عليه أن يسترجع أيضا النصوص التي تتضمن كلمة "قلب" بما أن الكلمة الأولى تمثل جمع تكسير للكلمة الثانية، تعتبر هذه الخاصية من خصائص اللغة العربية التي لا تؤخذ بعين الاعتبار في محركات البحث الكلاسيكية. وفي نفس السياق، يجب أن يؤدي البحث عن كلمة "شرايين" للبحث عن كلمة "قلب" نظرا للعلاقة الدلالية الطبية الجزئية (جزء من) التي تربط الكلمة الأولى بالثانية.

نذكر أيضا، من بين الأدوات الأخرى المتعددة التي يمكن أن تخدم المجال الطبي، أنظمة الإجابة الآلية على الأسئلة التي تمكن المستعمل من الحصول المباشر على الإجابة على سؤاله عوض البحث في النصوص الطبية الوافرة. في نفس السياق، تتميز تطبيقات التلخيص الآلي بأهمية كبيرة لأنها تمكن من تقليص النصوص الطبية الطويلة بالإضافة إلى الاستفادة من عامل الوقت بالنسبة للطلبة.

تجدر الإشارة أن كل هذه التطبيقات ومثيلاتها المعدة باستعمال مجال هندسة اللغة، يمكن أن تستغل ليس فقط من المستعمل العادي ولكن أيضا من طرف كل أعضاء الهيئة الطبية من أطباء وصيادلة وطلبة وباحثين وأساتذة جامعيين حيث ستمكنهم من الحصول على المعلومة الطبية بلغتهم الأم وبأسرع وقت ممكن. وستسمح هذه العملية، أيضا، من توفير الوقت والجهد وتجاوز ما يعرف بظاهرة عبء المعلومات التي لا تقتصر على لغة الضاد ولكن تتعداها إلى باقي اللغات الأخرى.