

المعالجة الآلية للغة

رضا زبدي

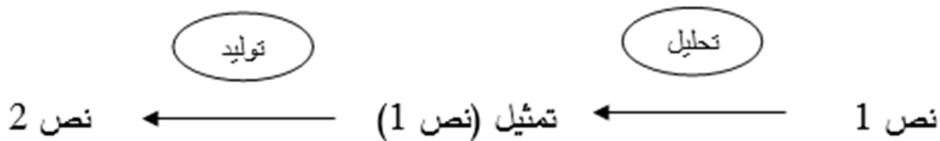
أستاذ بقسم الإعلام الآلي، المدرسة العليا للأساتذة، القبة

reda.zebdi@g.ens-kouba.dz

ما كنا نشاهده في أفلام الخيال خلال العقود السابقة بدأ يظهر في الواقع: روبوتات تحاور البشر، ترجمة فورية، بحث عن المعلومات بواسطة الصوت والصورة، التعرف على الحروف والكلام، إلخ. تمثل كل تلك الظواهر ثمرة أبحاث ودراسات دامت لسنوات ومرت بمراحل من البحث والتطوير في مجال من مجالات البحث المعلوماتي، يطلق عليه: المعالجة الآلية للغة.

1. ما المقصود بالمعالجة الآلية للغة؟

تعدّ المعالجة الآلية للغة Natural Language Processing من مجالات البحث متعددة التخصصات في ميدان المعلوماتية، والتي تُصنّف في إطار الذكاء الاصطناعي. فهي محاكاة لنشاط الإنسان اللغوي (قراءة، كتابة، حوار، فهم، ترجمة، إلخ)، حيث تدرس كيفية استخدام أجهزة الحاسوب لفهم وتحليل وتوليد النصوص المكتوبة أو المنطوقة باللغة البشرية، وذلك بهدف إنتاج برامج حاسوبية مفيدة. يُقصد بالمعالجة في هذا المجال، تحويل معطيات لغوية (نصوص) من صيغة لأخرى (تصحيح، تلخيص، ترجمة، إلخ)، وذلك بتحليلها ثم تمثيلها بشكل يسهل على الحاسوب استعماله، ثم توليد النص المطلوب.



تعتمد المعالجة الآلية للغة على عدد من التخصصات، من بينها: علم الحاسوب، اللسانيات، الرياضيات، الإلكترونيك، الذكاء الاصطناعي، وحتى علم النفس وعلم الاجتماع في بعض الأحيان.

2. أهداف المعالجة الآلية للغة

- هناك هدفان للمعالجة الآلية للغة، الأول نظري والثاني تطبيقي:
- الهدف النظري هو التحقق من النظريات اللسانية ودراسة كيفية التواصل بين البشر حيث يُستعمل الحاسوب لمحاكاة قدرات الإنسان لفهم وتوليد نصوص اللغة. تُقارن بعد ذلك النتائج بالأداء البشري من أجل التحقق من النظريات المدروسة.
 - الهدف التطبيقي هو إنتاج تطبيقات تجارية، مثل: الترجمة الآلية، تلخيص النصوص، استرجاع المعلومات، التعرف على الكلام، إلخ.

3. نبذة تاريخية

كانت التطبيقات الأولى في مجال المعالجة الآلية للغة متعلقة بالحوار الآلي، وذلك عندما أسس آلان تورينغ Alan Turing في الخمسينيات من القرن الماضي ما يسمى باختبار تورينغ الذي يقيس درجة ذكاء الآلة، إذ يقوم شخص بالتحاور كتابة، تارة مع حاسوب وتارة مع شخص آخر دون أن يدري في كل مرة أيهما سيجيبه، وإذا لم يستطع التمييز بين أجوبة الحاسوب وأجوبة الإنسان يكون الحاسوب قد نجح في الاختبار.

في عام 1952، أُقيم أول مؤتمر حول الترجمة الآلية أيام الحرب الباردة بين أمريكا والاتحاد السوفياتي في معهد ماساتشوستس للتكنولوجيا (MIT). وفي سنة 1954، أنجز أول مترجم آلي روسي إنجليزي في عدد من المجالات، منها السياسة والحقوق والرياضيات. بعد ذلك، وفي عام 1964، واعتمادا على أعمال تورينغ، تم إنشاء المحاور الآلي "إليزا" ELIZA بهدف العلاج النفسي، وهو تطبيق يحاكي عمل المستشار النفسي.

في فترة الستينيات كانت هناك خيبة أمل لسببين: (1) التعقيد الشديد للغة البشرية؛ (2) محدودية التجهيزات آنذاك من حيث سعة الذاكرات وسرعة المعالجات. فلم تكن التجهيزات كافية لإنتاج مثل هذه البرمجيات، حيث استثمرت مبالغ باهظة والنتائج لم تكن فعالة.

وقد ظهر خلال السبعينيات ما يسمى بعلم التوصيف (الأنطولوجيا) الذي يُعنى بطريقة هيكلة وتنظيم المفاهيم في مجال معين بهدف تمثيل المعارف وتخزينها في الحاسوب. ذلك ما سهل تمثيل المعنى وفك اللبس عند معالجة اللغة. كما ظهرت إبان تلك الفترة الأنظمة الخبيرة التي هي محاكاة لعمل الخبراء في مختلف المجالات، وذلك بتمثيل المعارف وتحديد قواعد لإنتاج معارف جديدة، حيث ساهمت كثيرا في تطوير عدة برمجيات في المجالات التي لها علاقة باللغة. وفي القرن الحالي، ومع تطور الأجهزة وشبكات الاتصال وزيادة سعة وسرعة الحواسيب وانتشار الحواسيب الشخصية، حدث تطور هائل في معالجة وفهم النصوص. ومع الكمّ الهائل للمعلومات المنتشرة على الشبكة العالمية، أصبح من الضروري تطوير تطبيقات قادرة على تحليل واستغلال كل هذا الكمّ من المعطيات (Data mining)، كما انتشرت تطبيقات المعالجة الآلية للغة على شبكة الانترنت وفي الهواتف الذكية انتشارا منقطع النظير.

4. مستويات المعالجة الآلية للغة

اتفق علماء اللغة على أن اللسانيات هي الدراسة العلمية للغة بكل مستوياتها. والمعالجة الآلية للغة هي دراسة وتحليل اللغة بواسطة الحاسوب، فهي كما قلنا محاكاة لدراسة الإنسان للغة، وهي تهدف إلى تحليل اللغة في كل مستوى من مستوياتها الستة من أجل تحقق الفهم الشامل للغة:

- المستوى الصوتي Phonetics الذي يتعامل مع النطق؛
 - المستوى الصرفي Morphology الذي يتعامل مع أصغر أجزاء الكلمات التي تحمل معنى، واللواحق والبادئات؛
 - المستوى المعجمي Lexicon الذي يتعامل مع الكلمات؛
 - المستوى النحوي Syntax الذي يتعامل مع قواعد اللغة وتركيب الجمل؛
 - المستوى الدلالي Semantics الذي يتعامل مع معنى الكلمات والجمل؛
 - المستوى التداولي Pragmatics الذي يتعامل مع المعرفة التي تأتي من السياق أو من العالم الخارجي.
- يمكن لنظام المعالجة الآلية للغة إشراك جميع أو بعض هذه المستويات من التحليل، ويمكن أن يحدث تداخل بين بعض هذه المستويات، مثل المستوى المعجمي والصرفي.

5. المقاربات النظرية

هناك تصوران في حلّ مشكلات المعالجة الآلية للغة، تفاوت نجاحهما مع مرور الوقت؛ المقاربة التجريبية والمقاربة العقلانية.

- **المقاربة التجريبية** Empirical approach وهي الأقدم: تعتمد على النصوص الموجودة (المدونات Corpus) وتُستعمل الأساليب الإحصائية لمعرفة العبارات المحتملة، وهي تُستعمل الآن لتحديد معنى الكلمة أو الجملة وفكّ اللبس.
 - **المقاربة العقلانية** Rational approach: نجدها في أعمال نعوم تشومسكي Noam Chomsky، وهي تعتمد على الفرضية التي تقول إن معارف اللغة يمكن أن تمثل على شكل معاجم وقواعد نحو تسمح بتوليد عبارات اللغة.
- في الآونة الأخيرة، أصبح الاعتماد على المقاربتين معا في حل مشكلات المعالجة الآلية للغة. استعمل توحيد المقاربتين في عدد من التطبيقات، وتسمى بالأنظمة الهجينة Hybrid systems، حيث تُدرس الظواهر اللسانية العامة باستعمال قواعد لغوية بحتة. أما الظواهر الواقعية فتُستعمل فيها المقاربة الإحصائية.

6. من أهم تطبيقات المعالجة الآلية للغة

- **الترجمة الآلية** Automatic translation: هي من التطبيقات الأولى التي ظهرت في الميدان، والتي نجدها الآن في جميع الأجهزة والهواتف الذكية.
- **استخلاص المعلومات** Information Extraction: هو استخراج المعلومات المفيدة وذات الصلة انطلاقا من النص؛ كأن نستخلص من نص يعالج مقابلة رياضية: اسم الفريقين، تاريخ ومكان المقابلة، عدد الأهداف، إلخ.
- **التلخيص الآلي** Automatic summary: يهدف إلى تقليص حجم النص المعطى، حيث يستخلص البرنامج المعلومات ذات الأهمية من النص ثم يقوم بتوليد النص الجديد الملخص.
- **استرجاع المعلومات** Information retrieval أو البحث عن المعلومات (نصوص، صور، إلخ) داخل مدونة انطلاقا من سؤال أو مجموعة كلمات. وهو التطبيق الأكثر استخداما في محركات البحث على الانترنت.
- **الحوار الآلي** Man-Machine dialog: هو نظام يسمح باستعمال واجهات من أجل التحوار مع الآلة. عادة ما يكون هذا الحوار على شكل سؤال-جواب، حيث يكتب المستعمل استعلامه على شكل سؤال، ثم يقوم النظام باستخراج الكلمات المفتاحية، وبعد ذلك يستعمل قاعدة معارف للإجابة على السؤال.
- **معالجة الكلام** Speech processing: هي التطبيقات التي لها علاقة بالكلام، وهي صنفان:
 - * التعرف على الكلام (الإملاء الآلي)، أي تحويل النص المنطوق إلى نص مكتوب على الآلة؛
 - * تركيب الكلام (النطق الآلي)، أي تحويل النص المكتوب على الآلة إلى نص منطوق.
- **التنقيب في البيانات** Data mining: يهدف إلى استخراج أكبر عدد من المعارف انطلاقا من كمّ هائل من المعطيات المنتشرة عبر شبكة الانترنت.

7. مشكلة اللبس اللغوي Ambiguity

يُقصد باللبس اللغوي أن يكون للكلمة أو الجملة الواحدة معنيان أو أكثر. فكلمة "قصر" في اللغة العربية تعني البناء الشامخ في مجال العمران، وهي رخصة في الصلاة في مجال الفقه، وتعتبر فنا من فنون البلاغة، ولها في النحو

معنى مختلف أيضا. كما أن لكلمة Flat باللغة الانجليزية عددا من المعاني. يمكن أن يكون اللبس أيضا في المستوى التركيبي، فيكون للجملة الواحدة عدة معان. في اللغة الفرنسية مثلا، يمكن أن تظهر ثلاثة تفاسير للعبارة "La belle ferme le voile".

مشكلة اللبس هذه تعتبر من أهم التحديات في المعالجة الآلية للغة، وهي تمثل حجر عثرة في هذا المجال، وذلك لأنها تؤثر على الفهم الحقيقي لمعنى النصوص المطلوب معالجتها. وبالتالي تؤثر على تمثيل المعنى الذي يعتبر أساس المعالجة الآلية للنصوص.

لحل مشكلة اللبس التي تعيق إنجاز تطبيقات المعالجة الآلية للغة، تُستخدم آليات وأدوات رياضية وإحصائية معقدة على النصوص المدروسة بهدف تحديد المعنى المقصود من الكلمة أو الجملة، كما تُستخدم خوارزميات الاستمثال (optimization) من أجل إيجاد الحل الأمثل عندما يكون هناك عدد هائل من الحلول.

8. المعالجة الآلية عبر اللغات Cross language NLP

نظرا لتعدد اللغات المستعملة في الانترنت، ومن أجل تعزيز التنوع اللغوي في استعمال تطبيقات المعالجة الآلية للغة، أصبح التوجه لإنجاز تطبيقات تعالج أكثر من لغة في نفس الوقت، ويمكن أن تقوم هذه التطبيقات بالتعرف على اللغة المصدر، ثم تقوم بالمعالجة المطلوبة. نلاحظ ذلك جيدا في تطبيقات الترجمة الشهيرة مثل google translate، وغيرها.

على الرغم من الوعود بإنتاج أدوات لمعالجة وفهم مجموعة واسعة من اللغات، فإن معظم الأعمال لا تزال تركز على عدد صغير من اللغات، وخاصة اللغة الانجليزية، وذلك نظرا للتحديات التي تتمثل في الاختلاف في النحو والطرق المعقدة في تركيب الجمل، خاصة في بعض اللغات.

9. تحديات المعالجة الآلية للغة

من أهم التحديات التي تواجه الباحثين في هذا المجال:

- صعوبة فهم اللغة؛
- الاستثناءات في القواعد؛
- تعدد المعارف التي يجب تخزينها؛
- كثرة أساليب تركيب الجمل؛
- اللبس في نصوص اللغة.

هذه التحديات تتعلق بالمعالجة الآلية للغة بصفة عامة. أما إذا تعلق الأمر باللغة العربية فهناك تحديات أخرى مرتبطة بخصوصية اللغة العربية والتي من بينها إسقاط الحركات عند الكتابة. وهذا يُحدث الكثير من اللبس أو الغموض في فهم النص المقروء. كما يمكن أن ينتج هذا اللبس أيضا عن المرونة الكبيرة في تقديم وتأخير مكونات الجملة، ويحدث أيضا من فنون البلاغة كالكناية والمجاز وغيرها من الفنون، والتي تسبب عدم فهم المعنى الحقيقي للعبارة اللغوية من قبل الآلة.

في الوقت الراهن، مازالت الأبحاث والدراسات في مجال المعالجة الآلية للغة مستمرة من أجل تطوير الأدوات والخوارزميات التي تُستخدم فيها، وذلك بهدف إنتاج تطبيقات أكثر دقة وتحسين فاعليتها وجعلها أكثر تعبيراً عن الواقع.

مراجع

1. عرار، مهدي أسعد: ظاهرة اللبس في اللغة، داروائل للنشر، عمان، 2003.
2. خطاب، شريف: المعالجة الآلية للغة العربية، دروس للطلبة، قسم علوم الحاسب، كلية الحاسبات والمعلومات، جامعة القاهرة.
3. Bouillon, P.: Traitement automatique des langues naturelles, Edition Duculot, Belgique, 1998.
4. Chowdhury, G. G.: Natural Language Processing, Dept. of Computer and Information Sciences, University of Strathclyde, Glasgow, UK, 2005.