

نحو توصيف لغوي لعمل الملخصات الآلية

إعداد

عبد الجبار الحروب

تقديم

- تحاول هذه الورقة طرح توصيف لغوي لعمل الملخصات الآلية يساعد على تحسين عملها، عبر تلخيص النصوص الطويلة من أخبار ومقالات ودراسات، خاصة في الحقل الإعلامي، إلى نصوص قصيرة في عدد من الجمل المركزية في النص، بالاعتماد على حذف الجمل وربط الجمل الأخرى بعضها مع بعض من خلال أدوات ربط مناسبة، من دون أن يخل ذلك بمضون النص ومعناه ودلالاته المقصودة. إضافة إلى معالجة الأخطاء التي تعترى عملية التلخيص المحوسب.

أسس عمل التوصيف اللغوي لعمل الملخص الآلي

- أولاً: تحديد بداية الجملة ونهايتها بالاستناد إلى علامات الترقيم والكلمات المفتاحية وأدوات الربط، فتكون النقطة (.) نهاية الجملة موحية باستيفائها للمعنى، وتكون (الدوال) الكلمات المفتاحية مثل: بناء على ما سبق، بسبب ذلك، وفي كل الأحوال، كلا، نعم، وكذلك، ولذلك استناداً إلى ذلك، إضافة إلى ذلك، يشار إلى أن، وأشار، ووضح، وبين، وغيرها بداية للجملة.
- كما أن هناك بعض الدوال تنبئ بنهاية الجملة وبداية جملة أخرى، مثل: قط، لا غير، بتاتا ... إلخ.

- ثانيًا: إيجاد الجمل المركزية في النص من خلال إيجاد الربط ما بينها وبين العنوانات ومفرداتها ودلالاتها، اعتماد على الكلمات المفتاحية، وعلامات الترقيم، مثل الفاصلة (؛) في علاقة السببية، والنقطتين الرأسيتين (:) في التقسيم، وعلامات التنصيص ("...") في الكلام المنقول أو المنسوب، إضافة إلى مفككات العنوان الرئيسي، وأدوات الربط عبر القيام بعمل إحصائي، وحساب التكرار. إضافة إلى تحديد أسماء الإعلام والأحداث المهمة والمفاهيم المتكررة. وحوسبة الكلمات المفتاحية على شكل شيفرات "كودات" تخزن في ذاكرة الحاسوب.

- ثالثاً: إيجاد العلاقات بين الجمل المختلفة أو المتتالية، مثل علاقة السببية، والتعليل والنتيجة، والعطف، والتوكيد، والوصفية، والتقسيم .. إلخ. ووضعها على هيئة شيفرات "كودات" تخزن في الحاسوب. وبناء على ذلك تقسم العلاقة إلى جملتين، جملة تبقى لمركزيتها بالعنوان، والأخرى تحذف لأنها تكون متتمة فقط.

- رابعًا: ترتيب الجمل حسب مركزيتها وارتباطها الوثيق بالعنوان الرئيسي والعنوانات الفرعية، وربط الجمل بعضها من خلال تغذية البرنامج بعشرات أدوات الربط المناسبة.

- استنادًا إلى هذا التوصيف، يقوم الحاسوبي بحوسبة الكلمات المفتاحية وعلامات الترقيم والعلاقات ومفككات العنوان على صورة "كودات"، تؤتي ثمارها عبر تطبيق البرنامج على محركات البحث، والمواقع الإلكترونية الإخبارية وذات الصلة بالنصوص. وتأمل هذه الورقة أن تشكل إضافة لعمل أكثر دقة للملخصات الآلية.

- وما يحسن من عمل الملخصات الحاجة إلى محلل سياقي دلالي يضاف إلى ما تتيحه المحللات الصرفية والنحوية ليسهل توصيف النص إلى الحاسوب.