

# **التلخيص الآلي... المفهوم والإجراءات والتطبيقات والتحديات**

**تحت إشراف:**

**الأستاذ الدكتور مُحسِنُ عَبْدِ الرَّازِقِ رَشْوَانَ الْخَبِيرِ بِالْمَجْمَعِ**

**بالاشتراك مع:**

**لجنة اللغة العربية والذكاء الاصطناعي بالمجمع**







## التلخيص الآلي .. المفهوم والإجراءات والتطبيقات والتحديات

### ملخص البحث<sup>(\*)</sup>

الأسلوب اللغوي، وأسلوب شبكات  
الكلمات... إلخ.

ورغم التقدم العلمي في مجال التلخيص  
الآلي فإن هناك تحدياتٍ كبرى ما تزال على  
الطريق تتعلق بطبيعة النصوص نفسها  
وبإيجاد معيارية واحدة أو واضحة لتقييم  
مخرجات التلخيص الآلي.

### الكلمات المفتاحية

التلخيص الآلي - الذكاء الاصطناعي -  
معالجة اللغات الطبيعية

مع كثرة النصوص التي ينتجها البشر  
باستمرار تزداد الحاجة إلى تلخيص لأهم  
المعلومات التي يحتاجها الإنسان، وهنا تظهر  
الحاجة إلى الذكاء الاصطناعي الذي يوظف  
تقنيات التلخيص الآلي التي تعتمد على  
التلخيص الاستخراجي أو التوضيحي أو  
الهجين عبر خوارزميات تعتمد على المفردات  
والجمل داخل النص، وهذه الخوارزميات تتخذ  
عدة أساليب، منها الأسلوب الإحصائي، أو

---

(\*) شارك في عمل هذا البحث الأستاذ الدكتور نعيم  
مُحَمَّد عَبْدُ الْغَنِيِّ.

## المقدمة

"الإيجاز في كلام العرب للدكتور مختار عطية... إلخ.

وممن ولع بالاختصار ابن منظور صاحب لسان العرب فقد ذكر في ترجمته أن الكتب التي علق عليها بخطه من مختصراته خمسمئة مجلد، فلا يُعرف كتاب مطول في الأدب وغيره إلا وقد اختصره (محمد علي كرد، ٤٩).

وأسرف طلاب العلم كثيرًا في الاختصار بحيث اختزلت العبارات بشكل استغلق على الألفهام فأعادوا شرح هذه المختصرات فتجد عناوين مثل: (شرح مختصر التحرير في أصول الفقه، شرح مختصر خليل في الفقه، شرح مختصر حديث مسلم... إلخ).

وامتدادًا لهذا الفريق ظهرت المطولات التي اتخذت أسماء مثل المبسوط، المفصل... إلخ، مثل كتاب (المبسوط في الفقه للإمام السرخسي، المفصل في تاريخ العرب للدكتور جواد علي وغيرها الكثير والكثير).

وما بين الاختصار والتطويل ظهر فريق من طلاب العلم عوان بين ذلك، يتبعون منهج "الاختصار غير المخل، والتطويل غير الممل".

التلخيص رغبة إنسانية يحتاجها المتلقي الذي يبحث عن خبر أو معلومة أو رقم معين أو خلاصة أحداث أو حوار أو غير ذلك، والمتحدث إذا لم يراع المقام الذي يتحدث فيه ورغبة المتلقي في الاختصار والتجوُّز فإنه يفقد انتباه السامع، من هنا نشأت في كل لغات العالم تقريبًا مصطلحات تعطي معنى التلخيص فمثلًا في لغتنا العربية عندنا: الإيجاز والاختصار، وفي المقابل لهما كانت مصطلحات مثل الإطالة والإطناب.

لقد كان الاختصار والتلخيص من أنواع التأليف المنهجي الذي ولع به الناس في كل زمان، وعندما نطالع التراث العربي مثلًا نجد عناوين مثل: (المختصر في النحو - المختصر المفيد في أحكام التجويد... إلخ) أو (تلخيص المفتاح) ونقصد بالمفتاح كتاب السكاكي مفتاح البلاغة الذي هو أيضًا مختصر وجامع لكل أنواع البلاغة في كتاب واحد، ونجد أيضًا (المقتضب) في النحو للمبرد، وعناوين كثيرة قديمًا وحديثًا تحمل اسم (الإيجاز) مثل: (الإيجاز في شرح سنن أبي داود للإمام يحيى بن شرف النووي، وكتاب

هذه ومضة خاطفة جداً عن حركة التأليف الدائرة في التراث العربي حول التلخيص والتطويل، وهي لا شك توحى بأن جهداً عقلياً كبيراً مبذولاً في معالجة النص لتلخيصه واختصاره، وفي عصر المعلومات باتت الحاجة إلى اختصار كم هائل من المعلومات والبيانات، وهذا الكم يحتاج إلى جهد بشري كبير لا يمكن توفيره أو تحقيقه، ومن ثم كانت الحاجة الماسة إلى الذكاء الاصطناعي لحل هذه المشكلة، فهل يمكن للآلة أن تقوم بهذه المهمة، هذا ما سنحاول أن نتعرفه من خلال رصد الأبحاث العلمية في هذا الشأن وكتابة خلاصة ما توصلت إليه، وما دمت إليه.

## المبحث الأول: التلخيص .. مفهومه وأنواعه وإجراءاته

### (١.١) مفهوم التلخيص:

التلخيص أو الاختصار: تقليل الشيء، فقد يكون اختصار الكتاب بتقليل مسائله، وقد يكون بتقليل ألفاظه مع تأدية المعنى، ومن ذلك قول النبي - صلى الله عليه وسلم - في صحيح البخاري: (أوتيتُ جوامعَ الكلم، واختصر لي الكلام اختصارًا)، وقال الخطيب الشربيني في مقدمة مغني المحتاج ص ١٣: "هو ما قل لفظه وكثرت معانيه"، وبالمعنى نفسه قال أيضًا: "الاختصار إيجاز اللفظ مع استيفاء المعنى".

والتلخيص معناه الموجز المنتج من نص أو أكثر، ويحتوي على نسبة كبيرة من المعلومات الموجودة في النص الأصلي، ولا يتعدى حجمه نصف النص الأصلي (علي فهمي، ٢٠١٩، ١٣٩).

### (١.٢) أنواع التلخيص

التلخيص البشري نوعان: إما أن يفهم المتلقي النص الأصلي ويعبر عن أفكاره بأسلوبه، وإما أن يقوم بحذف الكلمات التي لا تؤثر في جوهر الفكرة، ويترك النص كما هو،

والأول إعادة إنتاج للنص بأسلوب المتلقي، والثاني إعادة إنتاج للنص بأسلوب المؤلف، والنوع الأول أكثر دقة، وأوفر حظاً في التعبير عن الفكرة الرئيسية، لكن يشوبها أن المتلقي يلخص النص من وجهة نظره هو، وهذا قد يؤدي إلى توجيه المعنى بهدف مغاير لمنشئ النص نفسه.

وفي التلخيص الآلي نرى الباحثين لهم تصورات متباينة تجاه أنواع التلخيص، فهناك تصور يقسم أنواع التلخيص حسب الغرض، أو طبيعة الملخص، أو رغبة المتلقي، أو مصدر النصوص، أو لنوع التلخيص.

فإذًا لدينا خمسة منطلقات لأنواع التلخيص، فالمنطلق الأول حسب غرض النص نفسه عندنا نوعان: تلخيص معبر عن نوع النص، وتلخيص مؤشر عن نوع النص، والنوع الأول هو الأكثر دقة، والنوع الآخر يستخدم للتصنيف الآلي.

وأما المنطلق الثاني الذي يقسم نوعية التلخيص حسب طبيعة الملخص فيرى التلخيص في نوعين: تلخيص يقوم على استقطاع أجزاء من النص أو إعادة صياغة للنص الأصلي بما لا يخل بمعناه.

التلخيص يكون على أساس رقمي بحت، (كسكاس 2010) أو تلخيص يقوم على الرمزية التي تقوم على اللغة والعلاقات البلاغية بين أجزاء النص (مان وتومسان 1988)، أو تلخيص آلي مزدوج يجمع بين اللغة والعلاقات الرياضية (كسكاس وآخرون: 2012).

وهناك تصور ثالث لا يبعد كثيرًا عما سبق، ولكنه يختصر أنواع التلخيص في صنفين: التلخيص الاستخراجي (Extractive Summarization)، والتلخيص التوضيحي ((Abstractive Summarization))

في التلخيص الاستخراجي نختار أهم الجمل والعبارات في النص، دون إعادة لكتابة النص، وفي التلخيص التوضيحي نعيد بناء النص بعبارات جديدة، بناء على عدة معايير تكون على مستوى الكلمة المفردة أو الجملة أو على مستوى الفقرة، (N.Moratanch\*، S.Chitrakala، 2017)، وتفصيل ذلك على النحو التالي:

1- على مستوى الكلمة: حيث نستعين في التلخيص النهائي بالكلمات التي تعطي أكبر كم من المعلومات من غير تكرار، وهذه غالبًا تتحقق خصائصها فيما يأتي:

وأما المنطلق الثالث فينظر إلى رغبة المتلقي من حيث تعبير الملخص عن وجهة نظر منشئ النص أم المتلقي.

وأما المنطلق الرابع فينظر إلى مصدر النص ولغته، هل هو مصدر واحد أو الملخص من مصادر متعددة، وينظر إلى لغة النص الملخص، هل هي لغة النص الأصلي أو لغة أخرى، وهل هذه اللغة تلخيص لنصوص بأكثر من لغة تحمل أكثر من فكرة، فمثلاً عندما نتحدث عن انتخابات الرئيس الأمريكي ونريد أن ننشئ ملخصًا سريعًا في نشرة أخبار، قد نعتمد على مصدر واحد مثل صحيفة الواشنطن بوست وقد نعتمد أكثر من صحيفة أمريكية وغير أمريكية بلغة إنجليزية أو لغة أخرى، ثم نأتي بملخص للحدث أمام القارئ.

وأما المنطلق الخامس فينظر إلى نوع التلخيص: هل هو بأسلوب استخراج عدد محدود من الجمل في نص واحد أو عدة نصوص، أو إعادة إنتاج لنص آخر بجمل جديدة ومختصرة تعبر عن الفكرة الرئيسية للنص الأصلي. (علي فهمي: ١٤٠).

وهناك تصور ثانٍ يقارب التصور السابق ولكن ينظر إلى ناحية الإجراء، فيرى أن

- الكلمات المفتاحية التي قد تكون أسماء أو أفعالاً أو صفات أو ظروفًا
- موقع الفقرة في النص.
- التماسك بين الجمل في النص.

### (١.٣) أساليب التلخيص الآلي

يرى علي فهمي أن أساليب التلخيص الآلي تتنوع إلى أربعة أساليب هي: الإحصاء، أو المعالجة اللغوية في مستوياتها الصرفية والنحوية، أو المعالجة الدلالية وأساليب شبكات الكلمات، وأخيرًا أساليب الحسابات المرنة، مثل الشبكات العصبية والخوارزميات الجينية والمنطق الفازي وذكاء الأسراب. (علي فهمي: 2019).

وترى وفاء سامي أن مداخل التلخيص الآلي للنص تتمثل في:

#### 1- التلخيص الاستخراجي

(Extractive

Summarization)

#### ٢- التلخيص التوضيحي

(Abstractive

Summarization)

#### ٣- التلخيص الهجين (Hybrid).

بينما يقصرها وليد الصانع على النوعين الأولين؛ التلخيص الاستخراجي، والتوضيحي (الصانع ٢٠١٨)

➤ كلمات العنوان المأخوذ من الجمل الرئيسية في الوثيقة، ومن ثم يكون العنوان مليئًا بالمعلومات.

➤ الكلمات السببية: مثل (بسبب، لأن، من أجل ذلك... إلخ) وفيها إجابة عن سؤال لماذا حدث؟

➤ الكلمات المتحيزة التي تعطي رأيًا في الوثيقة سواء أكان سلبياً أم إيجابياً، ونعرفها بتكوين قائمة معدة سابقاً، وهذه الكلمات غالباً ما تشير إلى أهم المعلومات التي نحتاجها في التلخيص.

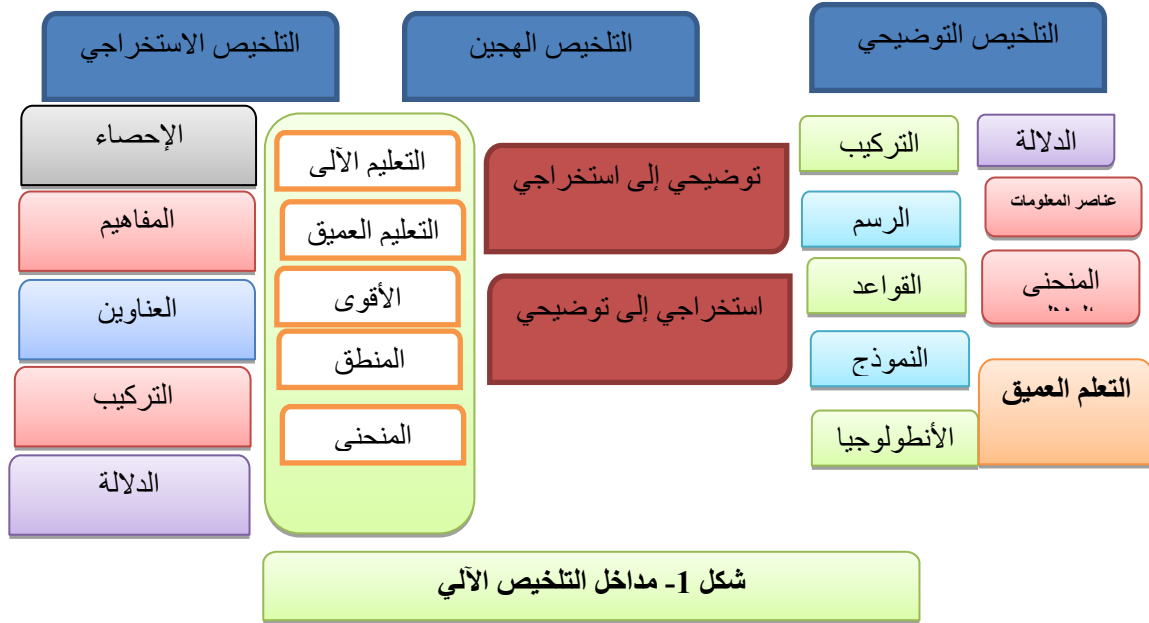
➤ حجم الخط وشكل الكلمة: حيث إن تمييز بعض الكلمات في النص بخط معين يوحي بأهمية هذه الكلمات التي قد تكون اسماً أو رقمًا أو مفتاحًا للنص.

#### 2- على مستوى الجملة:

على مستوى الجملة ننظر إلى عدة خصائص في النص هي:

- موقع الجملة.
- طول الجملة.

## مداخل التلخيص الآلي



- 4- الجمل التي تقع في عناوين الوثائق.
- 5- موقع الجملة داخل النص .
- 6- طول الجملة من حيث عدد كلماتها.

وعادة ما تعطى الأولوية للجمل التي تقع في عناوين الوثائق والجمل التي تحتوي على كلمات ذات معدل تكراري عال.

### ثانياً: الأسلوب اللغوي لعملية التلخيص:

في هذا الأسلوب يحدد أولاً موضوع النص (Topic Identification) بمجموعة من الطرق، منها طريقة بنية الخطاب

وهذه التتويجات تلتقي فيما بينها عند الخطوات العملية لعملية التلخيص الآلي، فأيًا ما كان الاسم فهناك خطوات غالبًا تؤدي إلى النتيجة، وسنحاول إجمالها على النحو التالي مستفيدين من التصورات الثلاثة السابقة:

### أولاً: الأسلوب الإحصائي:

في هذا الأسلوب نهتم بما يأتي:

- 1- الجمل التي تحتوي على كلمات ذات عدد تكراري عال في النص.
- 2- العبارات المميزة .
- 3- الجمل التي تقع على رأس الفقرات. والأجزاء داخل النص

2- تحديد العلاقات البلاغية بين مختلف هذه الوحدات.

3- تكوين البنية الخطابية المجزأة .

4- استخراج التلخيص الآلي.

ويتجاوز الباحثون المرحلتين الأوليين ويبدؤون مباشرة في معالجة المرحلتين الثالثة والرابعة؛ حيث يكونون البنية الخطابية المجزأة بطريقة تكرارية تعاقبية بالخطوات التالية:

1- تحديد كل وحدة صغرى في البنية الأخيرة للنص.

2- حذف الوحدات الخاطئة اعتمادًا على العلاقات البلاغية.

3- تحديث البنية الخطابية اعتمادًا على جملة من القواعد.

ولتوضيح هذه الخطوات بشكل عملي نعتمد على المثال الذي عرضه الباحثون آلياً، لكننا سنحاول عرضه بشكل أكثر سهولة، وسنتجاوز عن الخوارزميات التي عرضها الباحثون.

### تقسيم النص

تقلد خالد بندقية/ وخرج إلى الصيد/  
وصل إلى أرض/ ذات تلال وطيبة  
تتخللها أودية نبت فيها عشب وشجيرات  
الطرفاء والرتم/ تأمل خالد المكان/ ثم  
أطلق سلوقيه.

(Discourse Structure) وطريقة التسلسل المعجمي (Lexical Chains).

### (1.1) بنية الخطاب

في بنية الخطاب نكتشف مجموعة الجمل التي تغطي سياق النص، وتجب عن أهم الأسئلة التي يبحث عنها القارئ فإذا كان هناك حادث ما، فإنه يبحث عن الحدث نفسه، ثم يبحث عن إجابات لأسئلة من نوعية ماذا وكيف وأين ومتى حدث، وهنا نستخدم الكائنات الاسمية -مثل أسماء الجهات والأشخاص- بشكل كبير لأنها تجيب عن كثير من الأسئلة التي نبحث عنها. (علي فهمي: 2019)

ونستعرض هنا ملخصاً لبحث إسكندر كسكاس ملاك الليهوي وفرج بن عمارة ولمياء هدرش بلغيث (2012) وفي هذا البحث استخدام لبنية الخطاب المجزأة (SDRS) في التلخيص الآلي؛ وإبرادنا لهذه المحاولة أنها تعد المحاولة الأولى في اللغة العربية التي تعتمد بنية الخطاب المجزأة، وتضم هذه الطريقة أربع مراحل تتمثل في:

1- تقسيم النص إلى وحدات صغرى، ونقصد بالوحدة أن تكون عبارة مفهومة جملة قصيرة أو جزءاً من جملة.

تجاوز الباحثون هذه المرحلة من التقسيم وانتقلوا إلى تحليل بنية الخطاب مباشرة اعتماداً على مئة مقالة جاهزة، وحددوا بناء

عليها العلاقات البلاغية في النص السابقة على النحو التالي:

العلاقة	الجملة
جزئية	تقلد خالد بندقية
استئناف	وخرج إلى الصيد
جزء - تأطير مكاني	إلى أرض
جزء - تخصيص	ذات تلال وطيبة تتخللها أودية نبت فيها عشب وشجيرات الطرفاء والرتم

المرحلة الثالثة تحديد طرفي العلاقات البلاغية باستخدام المحلل الصرفي، فإذا كان اسم العلاقة البلاغية تخصيصاً، فيجب أن تحتوي الوحدة القابلة للإدراج على اسم حتى نحقق هذه القاعدة، وإذا وجدت عدة وحدات تحقق نفس القاعدة فإننا نلجأ إلى الترتيب من خلال إعطاء وزن لكل منها، ثم تحسب الأوزان بعد ذلك.

وإثر هذه المرحلة مباشرة نقوم بتحديث البنية الخطابية من خلال دمج الوحدة الحالية في المخطط (SDRS)، وهذه هي المرحلة التي نستطيع أن نحدد من خلالها شكل بنية النص (بيسكي وآخرون: ٢٠٠١)

وفي المرحلة التالية نطبق القواعد البلاغية لحذف الوحدات الخاطئة من خلال تقسيم العلاقات البلاغية إلى رأسية وأفقية هكذا:

العلاقات العمودية	العلاقات الرأسية
تأطير زمني	استئناف
تأطير مكاني	ترتيب
غرض	تعدد
تفصيل	تلازم
حال	ربط
تخصيص	عطف

## (1.2) التسلسل المعجمي

أما أسلوب التسلسل المعجمي فنستخلص فيه سلاسل الكلمات التي لها صلة بالنص؛ لنعرف الأفكار المحورية، وهذه الطريقة طورها موريس وهاريس في التسعينيات بناء على أبحاث سابقة في منتصف السبعينيات حول وظائف اللغة للعالم اللغوي مايكل هاليداي.

وتعتمد هذه الطريقة على التماسك النصي الذي يربط الجمل بعضها ببعض من خلال أدوات لغوية مثل: أسماء الإشارة والضمائر والاستبدال والحذف والاقتران وغيرها، أي أن السلاسل المعجمية تتبع كائناً اسمياً متغيراً، فقد يتتبع الضمائر أو أسماء الإشارة مثلاً.

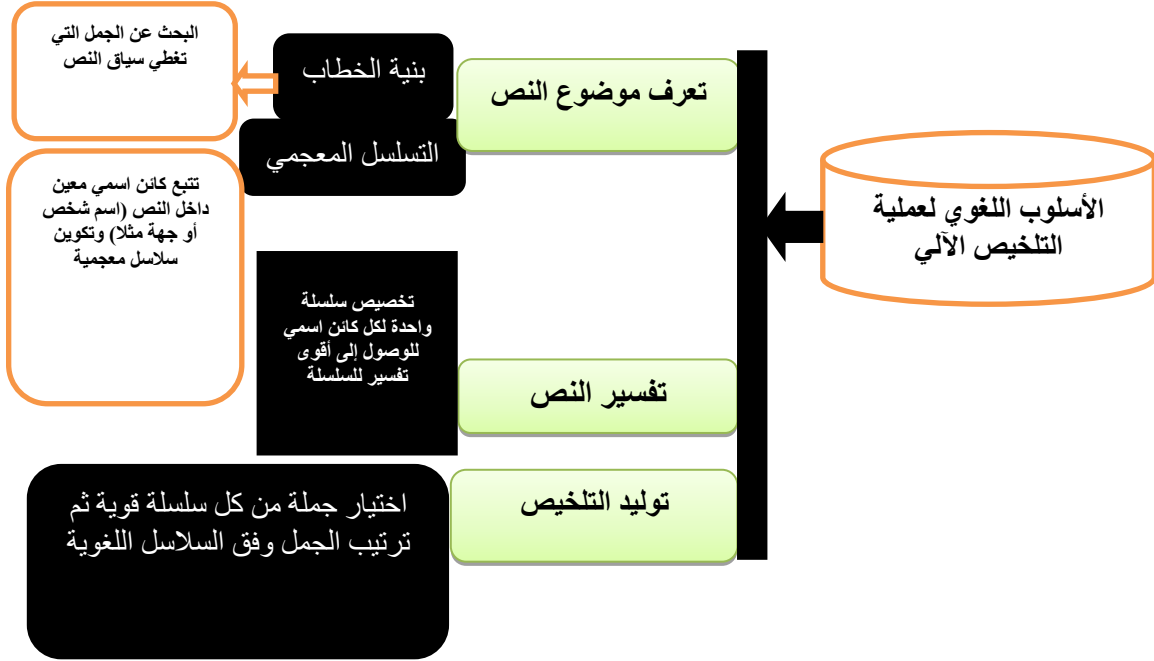
وفي هذه الطريقة تظهر مشكلة الالتباس الدلالي في إحالة الضمائر وأسماء الإشارة، ولا يمكن فك هذا اللبس الدلالي دون الرجوع إلى قواعد اللغة.

وبعد التعرف على موضوع النص تأتي مرحلة التفسير (Interpretation) التي تخصص سلسلة واحدة فقط لكائن اسمي في النص، وهذا يجعلنا أمام كم هائل من الاحتمالات واللبس الدلالي الذي لا يؤدي

إلى تفسير صحيح، وهي معضلة كبرى حاول الباحثون التغلب عليها من خلال ربط السلسلة المعجمية بمفهوم معين، وربط هذا المفهوم بمعنى مأخوذ من نظام وورد نت ( word Net) وهي قاعدة بيانات معجمية للغة الإنجليزية تجمع فيها الكلمات الإنجليزية في مجموعة مترادفات تسمى (Synsets)، وتوفر تعريفات قصيرة عامة وتسجل العلاقات الدلالية المختلفة بين مجموعة المترادفات، وهذه الطريقة اقترحها رجينا برزيبلا ومايكل الحداد وسيلبر ومكوكي، واعتمدوا على استخدام معربات سطحية بسيطة وأدوات لتعيين أقسام الكلم (معنونات) للكلمات للتعرف على الأسماء، حيث تأخذ دلالة هذه الأسماء في النص للوصول إلى أحسن التفسيرات للسلاسل المعجمية، بناء على: عدد مرات تكرار الكائن الاسمي وعلاقاته السابقة مع باقي كلمات السلسلة. ( Regina Barzilay & Michael ElHadad. 1997)

وأخيراً تأتي مرحلة توليد الملخص (Generation) حيث نأخذ فيها أفضل السلاسل المعجمية، ثم نختار جملة واحدة من كل سلسلة بناء على عدة معايير هي: اختيار أول جملة تشير إليها السلسلة المعجمية

وتضمينها بترتيبها في الملخص، أو اختيار أول جملة تشمل الممثل الاسمي الذي يعبر عنه وترتبط به السلسلة وتوضع أيضاً بالترتيب



شكل: 2 - الأسلوب اللغوي للتلخيص الآلي

### ثالثاً: التلخيص الهجين

3- ترتب المجموعات بحسب أهمية كل

مجموعة.

4- نختار جملة واحدة من المجموعات

الأولى.

5- تحذف الكلمات الزائدة والتكرارات من

الجملة المختارة.

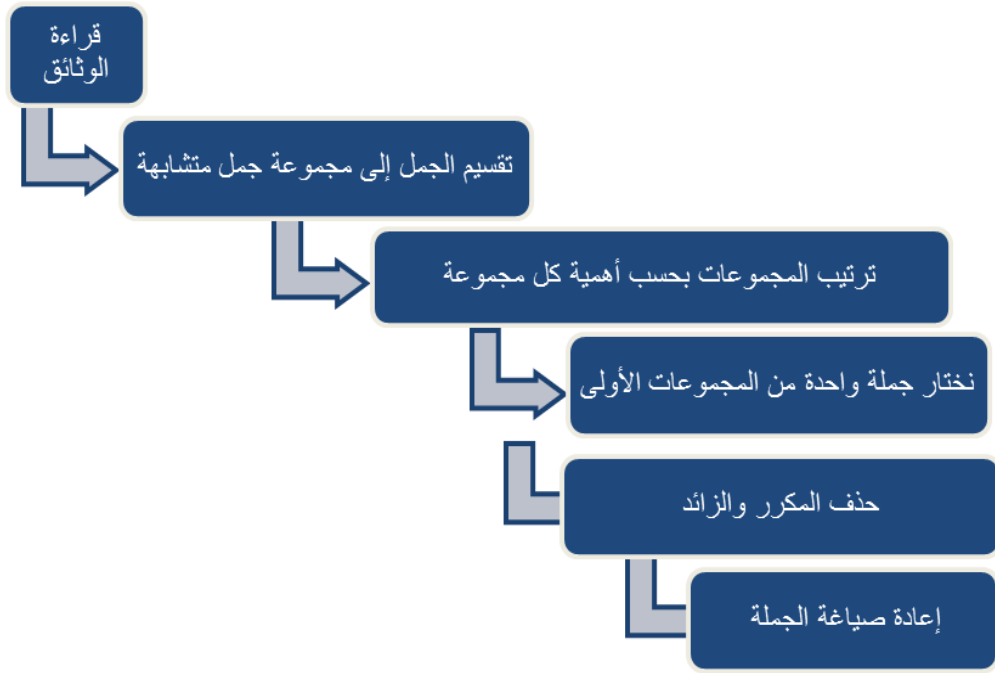
6- إعادة صياغة الجملة.

نقصد بالتلخيص الهجين الجمع بين تعلم الآلة والمعالجة اللغوية، وفي هذا الأسلوب نأخذ ثلاث خطوات هي: قراءة الوثيقة، ثم فهمها من خلال التمثيل الدلالي للمحتوى، ثم أخيراً توليد النص الملخص من التمثيل الدلالي.

1- قراءة الوثائق.

2- تقسيم الجمل في كل وثيقة إلى

مجموعات من الجمل المتشابهة.



شكل: (3) يبين طريقة الأسلوب الهجين بين اللغة والآلة

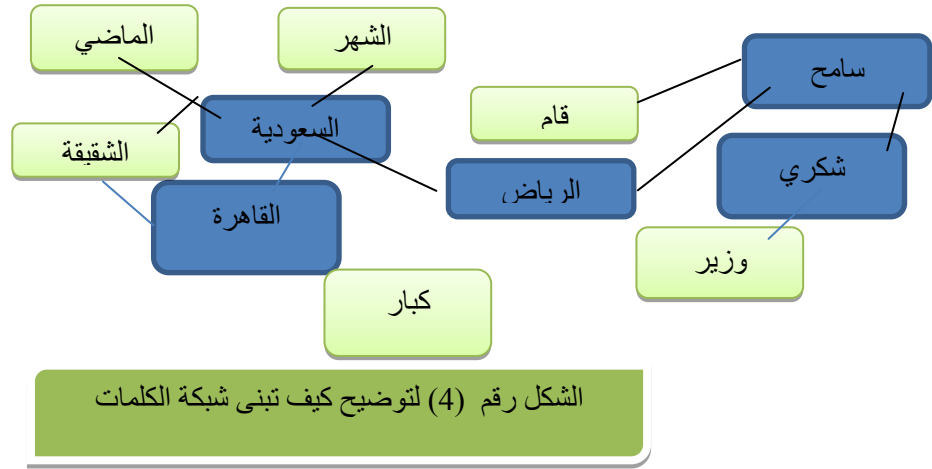
#### رابعًا: أسلوب شبكات الكلمات

ابتكر هذا الأسلوب "كاتجا فيلبوفا" بمؤسسة جوجل الشهيرة؛ حيث تدمج مجموعة من الجمل المتشابهة في موضوع واحد في جملة واحدة، ويعتمد على بناء شبكة بين كلمات مجموعة من الجمل المراد اختصارها؛ حيث تمثل كل كلمة بعقدة داخل الشبكة والخطوط التي تصل الكلمات تمثل تتابع الكلمات داخل كل جملة، مع ملاحظة أن الكلمات المكررة من الممكن دمجها في عقدة واحدة، وهذا النوع من التلخيص يعتمد على المعالجة اللغوية والتعرف على مواطن

الالتباس في الجملة وحروف الربط والإشارة وربطها بمدلولها.

ولتوضيح ما سبق دعنا نتأمل هذه الجمل:

- 1- زار وزير الخارجية سامح شكري المملكة العربية السعودية الشقيقة الشهر الماضي في إطار التعاون بين القاهرة والرياض.
- 2- قام سامح شكري بزيارة الرياض الشهر الماضي.
- 3- الشهر الماضي زار سامح شكري الرياض والتقى كبار المسؤولين السعوديين.



يعد موجودًا، وأننا استطعنا الوصول بأهم جملة في النص.

والخصائص التي نبحث عنها هي: موقع الجملة في النص، ومدى تشابه هذه الجملة مع العنوان، وما تمثله من محورية، وما تحويه من كلمات إيجابية وسلبية وكيانات اسمية وبيانات عددية، كما نعتمد خاصية طول بناء الجملة في النص.

وهذه الخواص يمكن رصدها بشكل آلي، ودقتها تعتمد على تكوين سلاسل معجمية يمكن تنميطها بشكل رياضي لبناء خوارزمية تستطيع استخلاص أهم ما في النص.

#### خامسًا: تلخيص النصوص باستخدام أساليب الحسابات المرنة:

هذه الأساليب تعد الأكثر محاكاة للعقل البشري، وتتخذ أدوات متنوعة مثل أشجار القرار التي تعد أداة للتنبؤ بالأحداث وما يصاحبها من آثار، كما توظف الشبكات العصبية التي يوزع فيها الباحث النص على ثلاث طبقات، ففي الطبقة الأولى مجموعة من العقد تمثل خصائص الجمل المختارة، ثم تأتي الطبقة الثانية بعقد أقل، ويكون التجريب فيما بينها للوصول إلى عقد واحدة في الطبقة الثالثة، وتكون هذه نتيجة التلخيص؛ لأن صغر حجم العقد يعني أن اللبس الدلالي لم

## المبحث الثاني: أنظمة التلخيص الآلي

### .. تطبيقات وتحديات

من المؤسف أن نقول: إن أنظمة التلخيص الآلي الموجودة حاليًا تخدم في غالبيتها اللغة الإنجليزية، أما العربية، فإن أنظمة التلخيص التي تخدمها قليلة، وليست بمستوى الدقة المطلوبة، وقبل استعراض هذه الأنظمة فإننا نعرض أولاً لأسس الأنظمة للتلخيص الأكثر استعمالاً، ثم نعرض موجزاً عن كل نظام في أكثر من لغة، ثم التحديات التي يواجهها التلخيص الآلي.

### (1.1) أسس أنظمة التلخيص

#### الأكثر استعمالاً:

- الموجز الدلالي: وفيه تصنيف سريع للنصوص، ويضبط مرجعية ملائمة لها؛ للقيام بدراسة معمقة فيما بعد.
- الموجز الإخباري: وفيه معالجة لمحتوى النص، ويشمل كل المعلومات المهمة بدرجة قابلة للضبط.
- الموجز المقتطف: يعتمد على استخراج وحدات من النص.

- الموجز المتماusk: وفيه إعادة صياغة لأهم ما ورد في النص.
- الموجز العام: وفيه اهتمام بتلخيص نظرة الكاتب.
- الموجز الموجه للطلب: وفيه يعكس حالة اهتمامات القراء.

(إسكندر كسكاس وآخرون 2012)

### (1.2) أنظمة التلخيص الآلي

#### ١. نظام (Swe Sum)

يشير علي فهمي (2019) إلى أن هذا النظام يعد أول نظام تلخيص للغة السويدية، وهو مختص بتلخيص الأنباء المكتوبة على شبكة الإنترنت، ونتيجة التلخيص تتراوح بين ١٠ إلى ٥٠ من الكلمات الحاكمة، وتتراوح دقة التلخيص ما بين ٤٠ إلى ٨٠٪ لنص طوله في المتوسط ١٨٠ كلمة، وهذا النظام متاح في اللغات الدنمركية والفارسية والنرويجية والإنجليزية والإسبانية والفرنسية والإيطالية واليونانية والألمانية.

ويستند هذا التلخيص على أسلوب الإحصاء اللغوي وأساليب الذكاء الاصطناعي، حيث يحسب تكرار الكلمات الرئيسية ويأخذ في عين الاعتبار حجم الخط وتقسيم الفقرات والجمل.

## ٢. نظام (SUMMARIST)

يصنف هذا النظام بأنه يتمتع بالقوة في تلخيص النصوص؛ لأنه يتعامل مع الأخطاء اللغوية بخوارزميات متقدمة، ويقوم على تقنية التلخيص الاستخراجي حيث يستخدم أنظمة التحليل والتفسير بشكل عميق، ويمكنه صهر المدخلات بشكل صحيح واستخراج موضوع متماسك من النص.

## ٣. نظام (Lex Rank Summarizer)

هو نظام طوره جامعة ميتشجين بالولايات المتحدة، ويعتمد على الأساليب الإحصائية والرسوم الشبكية؛ حيث تحول الجمل إلى متجهات عددية ثم تحسب قيمة الزاوية بينها، ثم تحسب مصفوفة الجوار

لجميع الجمل الموجودة في النص، ويكون ناتج التلخيص في الجمل الأعلى قيمة داخل المصفوفة من خلال تحكيم معياري الجوار والتشابه. (N.Moratanch: 2017).

## ٤. نظام (Intellexer)

### (Summarizer) ٢

في هذا النظام استخدام كثيف لمعالجة اللغات الطبيعية حيث يعتمد على التحليل النحوي والصرفي لاستخراج ملخص النص.

وبتجربة النسخة التجريبية المتاحة لهذا التطبيق، فإن السرعة في تلخيص النص لم تتجاوز ثواني معدودة، كما أنه استطاع أن يستخلص أهم معلومة في الخبر الذي كان عن استطلاع شهر شوال.

الخبر الأصلي 178 كلمة، والملخص

14 كلمة

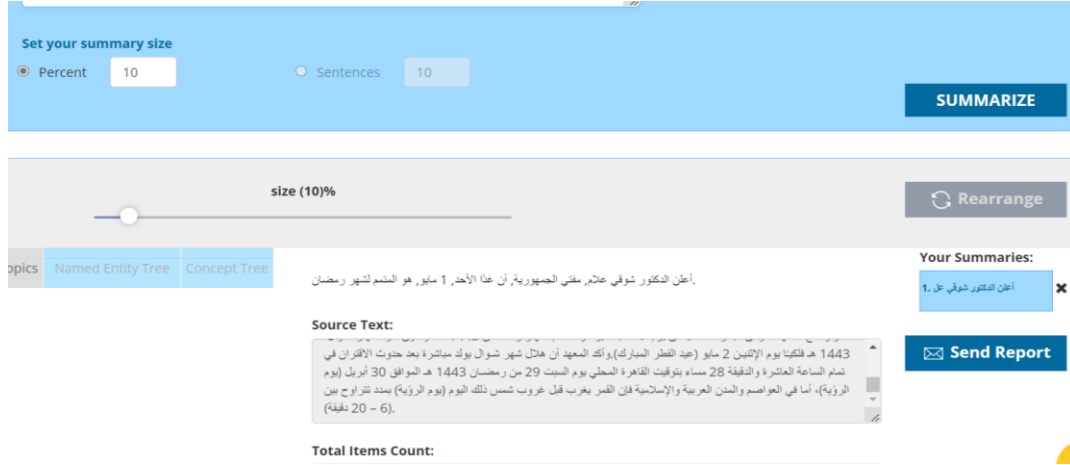
أعلن الدكتور شوقي علام، مفتي الجمهورية، أن غدًا الأحد، 1 مايو، هو المتمم لشهر رمضان. وقال المفتي، في كلمة عبر الصفحة الرسمية لدار الإفتاء، السبت، إنه لم تثبت رؤية هلال شهر شوال، وبذلك يكون الاثنين، الموافق 2 مايو، هو غرة شهر شوال أول أيام عيد الفطر المبارك. واستطلعت دار الإفتاء المصرية، من خلال لجانها الشرعية المنتشرة في 7 مراكز على مستوى الجمهورية، مساء السبت، هلال شهر شوال لعام 1443 هجريًا. وكانت الحسابات الفلكية قد أشارت إلى أن غرة شهر شوال وأول أيام عيد الفطر المبارك لعام 1443 هجريًا يوافق يوم الاثنين 2 مايو المقبل. وأوضح المعهد القومي للبحوث الفلكية أن يوم الأحد 1 مايو هو المتمم لشهر رمضان 1443 هـ، وتكون غرة شهر شوال 1443 هـ فلكيًا يوم الاثنين 2 مايو (عيد الفطر المبارك). وأكد المعهد أن هلال شهر شوال يولد مباشرة بعد حدوث الاقتران في تمام الساعة العاشرة والدقيقة 28 مساءً بتوقيت القاهرة المحلي يوم السبت 29 من رمضان 1443 هـ الموافق 30 أبريل (يوم الرؤية)، أما في العواصم والمدن العربية والإسلامية فإن القمر يغرب قبل غروب شمس ذلك اليوم (يوم الرؤية) بمدد تتراوح بين (6 - 20 دقيقة).

<sup>1</sup> - [LexRank method for Text Summarization \(opengen.org\)](http://LexRank%20method%20for%20Text%20Summarization%20(opengen.org))

<sup>2</sup> - [Summarizer \(intelleger.com\)](http://Summarizer(intelleger.com))

## التلخيص الآلي .. المفهوم والإجراءات والتطبيقات والتحديات

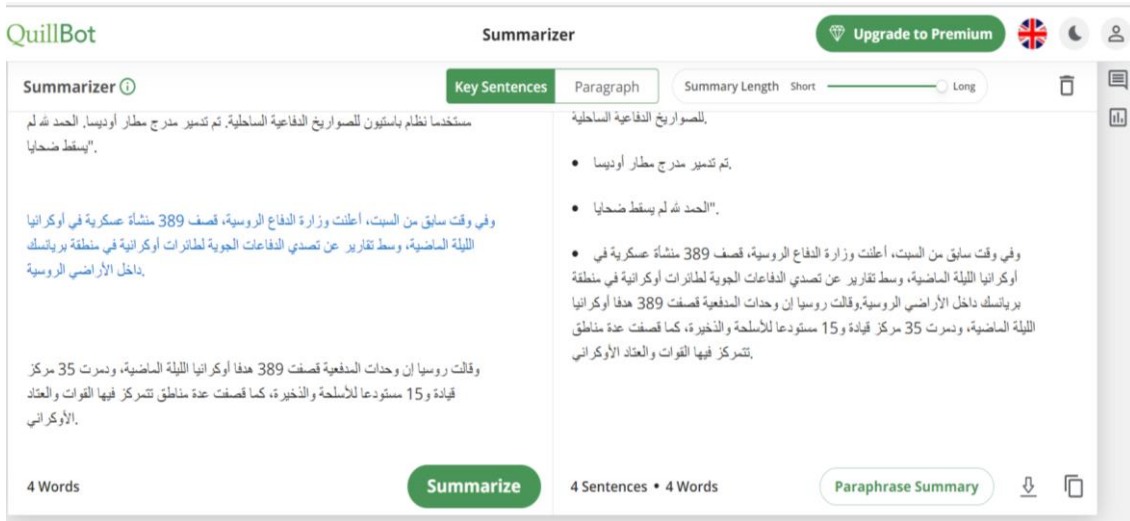
**المُلخَص:** أعلن الدكتور شوقي علام، مفتي الجمهورية، أن غداً الأحد، 1 مايو، هو المتمم لشهر رمضان.



### ٥. نظام ( Paraphrasing Tool | QuillBot AI

في هذا النظام الذي يدعم اللغة العربية تلخيص لأهم الجمل في الخبر اعتماداً على الإحصاء اللغوي أيضاً، وقد قمنا بتجريب هذا التطبيق على خبر فُلخص مضمونه في عنوانين فقط، هما تم تدمير مطار أوديسا - الحمد لله لم يسقط ضحايا

في هذا الملخص اعتماد على الكيانات الاسمية وتحليل لأهم المعلومات في النص، وتلخيص المعلومة المفيدة فقط، وهي أن الأحد هو المتمم لشهر رمضان، ولسنا في حاجة إلى أن نقول إن الاثنين هو أول أيام شوال، فالتلخيص هنا وافٍ وكافٍ.



واللواحق، ثم حساب المعدلات التكرارية للكلمات المستخلصة والوزن النسبي لكل جملة اعتماداً على معدلات التكرار للكلمات ومواقعها داخل الجملة، وصولاً لاستخلاص الجمل ذات الوزن النسبي العالي التي يتكون منها الملخص المطلوب.

## 6. نظام (الخص - Lakhas) (\*)

الذي طورته جامعة مونتريال الكندية يعتمد الإحصاء في عملية التلخيص عبر عدة مراحل تبدأ بتجزئة النص إلى جمل، ثم تقسيم الجمل إلى كلمات مع وضع حروف الكلمات في صور موحدة، وإزالة كلمات الوقف مثل: الذي والتي ومن وفي وغيرها، وحذف السوابق



النص الملخص بالنسبة للنص الأصلي، هل يريده 10% من النص الأصلي أو 20% أو أكثر أو أقل.

وبتجريب الموقع تبين لنا ما يأتي: 1- الموقع يتيح للمستخدم تحديد مقدار



2- يعرض للمستخدم النصين الملخص والأصلي معاً.



(\*) تلخيص النصوص العربية أونلاين | لخصلي

(lakhasly.com)

بالجودة المناسبة، وسوف نجمل هذه التحديات في النقاط التالية:

### (1.1) التحكم في المخرجات (Controlling the Outputs)

نقصد بالتحكم في المخرجات الحكم على جودة النص الملخص، ورأينا بعض التطبيقات التيح للمستخدم أن يتحكم في الحكم الذي يريد أن يختصره من النص، لكن صياغة الجمل، ووضوح المعنى، والاختيار الأمثل للمفردات والتراكيب يظل هو التحدي الأكبر الذي يواجهه هذه التقنية.

### (1.2) التلخيص متعدد الوثائق (Multi-document Summarization)

أحياناً يتطلب الأمر أن ننشئ ملخصاً لفكرة معينة اعتماداً على عدة وثائق، وهذه الوثائق قد تكون أحادية اللغة أو متعددة اللغات، وهذا يحتاجه العاملون في مجال الإعلام كثيراً حين يعدون مثلاً تقريراً إخبارياً معيناً، فيحتاجون إلى جمع مصادر الخبر واستخلاص أهم المعلومات الموجودة به، وفي كلتا الحالتين الأمر صعب جداً، فإذا كانت المدخلات إلى نظام التلخيص الآلي أحادية اللغة فإن الأمر يحتاج أولاً لتدريب النظام

### ٧. نظام (PEGASUS)\*

هو نظام تلخيص يعتمد على أسلوب التلخيص التوضيحي ويتعامل مع النصوص الطويلة نسبياً، حيث تستخدم فيه خوارزميات تتعامل مع الجمل الطويلة التي يتم تقسيمها وحساب معدلات التكرار فيها، وصولاً لمرحلة توليد النص الملخص.

### ٨. نظام (ACBTSS)

وهو اختصار لـ (Arabic Concept-Based Text Summarization System) (ACBTSS) الذي طورته جامعة (Essex) البريطانية، وهو مختص بتلخيص النص العربي، ويعتمد على مدونة لغوية مرمزة ومزودة بالحواشي لتدريب النظام على استخراج خصائص الجمل التي يتم الاحتفاظ بها في نتائج التلخيص.

### ثانياً: التحديات التي يواجهها التلخيص الآلي Challenges for Automatic Text Summarization

رغم التقدم النسبي في التلخيص الآلي فإن هناك بعض التحديات التي ينبغي على الباحثين تجاوزها حتى يخرج النص الملخص

(\*) [google/pegasus-xsum · Hugging Face](https://github.com/google/pegasus-xsum)

البحث لا يعدو أن يكون وسائل تستعين بالعنصر البشري في تقييم نتائج التلخيص، والعنصر البشري ليست دقته عالية رغم تفوقه على الآلة، فهناك فوارق نسبية بين كل قارئ، فإذا أعطيت نصًا لخمس قراء وطلبت منهم أن يلخصوه لك فستخرج حتمًا بخمس نتائج متنوعة، فكل سيختار حسب ما يراه أنه مهم، لكن بطبيعة الحال هناك قواسم مشتركة بين الخمسة، وهذه القواسم هي الأساس الذي يمكن أن تدرب الآلة عليه، وهذا ما يحدث بالفعل، ويرى علي فهمي أن قياس جودة التلخيص لها معياران: الأول نسبة ضغط النص، والثاني نسبة الاحتفاظ بالمعلومات المهمة، ونسبة ضغط النص يمكن حسابها بعدد الكلمات أو الجمل أو الفقرات، وقياس نسبة النص الأصلي إلى النص الملخص، أما نسبة الاحتفاظ بالمعلومة فهناك ست وسائل هي:

**الأولى:** التقييم اليدوي (Manual Evaluation) التحكيم البشري بين النص الأصلي والنص الملخص، وهذا التحكيم سيكون مختلفًا من شخص إلى آخر.

**الثانية:** استخدام طريقة حساب عدد المقاطع المشتركة بطول معين بين النص

على الكيانات الاسمية المهمة، سواء أكانت أشخاصًا أم هيئات أم أرقامًا، ثم بعدها يحتاج النظام - كما بينا من قبل - إلى تفسير، ثم توليد النص، وهنا سنواجه التحدي السابق في اختيار أهم الجمل وطريقة بناء النص الجديد، وفي حالة ما إذا كان المدخل متعدد اللغات ففي هذه الحالة لا بد من ترجمة آلية وتوحيد اللغة أولًا، ثم إجراء عمليات التلخيص، وهي عملية معقدة جدًّا، وبها تحديات تركيبية ودلالية لم تتمكن أي خوارزمية من حلها حتى الآن.

### (1.3) غموض النص الأصلي

من التحديات التي تواجه التلخيص الآلي وجود جمل غامضة في النص الأصلي، وهذا الغموض على درجات وأنواع، فمثلًا من أنواع الغموض الخطأ الطباعي أو عدم وضوح الخطوط، أو الخطأ التركيبي في الجملة، أو الصياغة الدلالية الملبسة، أو طول بناء الجملة، فالنص الأصلي إذا لم يكن واضحًا ومحددًا به الكيانات الاسمية فإن جودة التلخيص ستتأثر سلبيًا.

### (1.4) تقييم نظم التلخيص

حتى الآن لا توجد معيارية واحدة لتقييم نظم التلخيص الآلي، وما أشرنا إليه في هذا

ROUGE-2 يشير إلى أننا نريد مقارنة عدد التعابير المكونة من كلمتين بين ملخص النظام والملخص المرجعي.

**الثالثة:** اعتماد نظرية "كلود شانون" المعروفة باسم "نظرية المعلومات"، وتستخدم لضغط البيانات؛ بهدف سرعة نقلها، وهي طريقة لها القدرة على استرجاع البيانات الأصلية، وبسبب قدرتها على هاتين الخاصيتين يمكن أن نقارن ونعرف نسبة الاحتفاظ بالمعلومات المهمة.

**الرابعة:** طريقة الاستعلام، حيث تقرأ مجموعة من الأشخاص النص الأصلي للوثيقة ثم يضعون مجموعة أسئلة تعكس أهم عناصره، ثم تجيب مجموعة أخرى بالإجابة عن الأسئلة دون الاطلاع على أي شيء، لا النص الأصلي ولا الملخص، وهذه المجموعة تسمى خط الأساس، ثم تأتي مجموعة تجيب عن الأسئلة بعد الاطلاع على الملخص ومجموعة أخرى تجيب عن الأسئلة بعد الاطلاع على النص الأصلي، وتحتسب جودة التلخيص بنسبة الاحتفاظ بالمعلومة عن طريق الإجابات الصحيحة التي أجابت عنها المجموعة الثانية التي اطلعت على الملخص الآلي مقارنة بالإجابات التي كانت قبل النص الأصلي أو بعد تلخيصه.

**الخامسة:** جمع خمسة آلاف ( 5000 ) مقالة إخبارية من خمسة مجالات مختلفة، ثم

الأصلي والملخص ( BLEU (Bilingual Evaluation Understudy)، وفقاً للنحو العددي (N-gram) وكما زاد عدد النقاطات بين النص الأصلي والنص الملخص كان ذلك مؤشراً على جودة التلخيص.

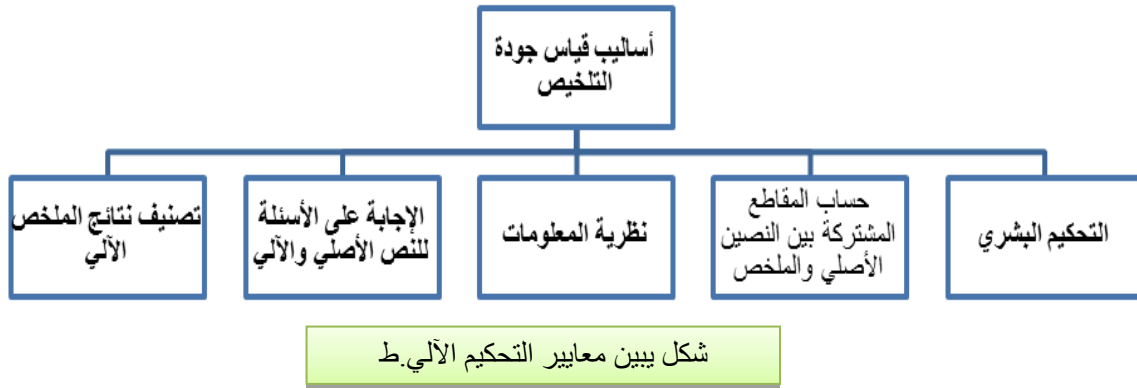
ويتصل بالنحو العددي تقنية (ROUGE(Recall-Oriented Understudy for Gusting Evaluation، هو مقياس لتقييم أداء التلخيص الآلي (Automatic Summarization)

وبرامج الترجمة الآلية (Machine Translation MT) في معالجة اللغات الطبيعية بناءً على عدد الكلمات المشتركة بين المخرج الآلي والمرجع الذي تتم المقارنة معه. تم تطوير أكثر من نسخة من هذا المقياس بناءً على طريقة المقارنة، هل هي بكلمة واحدة؟ أو بتعابير مكونة من أكثر من كلمة؟ والذي يعرف في أدبيات معالجة اللغة الطبيعية بـ n-gram. لذا فإن هذا المقياس يمكن الإشارة إليه بـ

ROUGE-N حيث يشير الرمز N إلى عدد الكلمات في التعبير المراد استخدامه في المقارنة ROUGE-1 . مثلاً يشير إلى أننا نريد مقارنة عدد الكلمات المشتركة بين ملخص النظام والملخص المرجعي،

إجراء التلخيص الآلي عليها، ثم حساب عدد الملخصات التي تم تصنيفها بشكل دقيق، وحساب عدد الملخصات المصنفة بشكل خاطئ، وكلما كان التصنيف صحيحاً كان ذلك مؤشراً على جودة التلخيص والقدرة على الاحتفاظ بالمعلومة بشكل مميز.

وفي أغلب الطرق السابقة نستخدم مقياس (ROUGE) ومقياس (Measure-F) للتعبير عن جودة التلخيص (علي فهمي: 2019، 141-142)، وليد الصانع: 2018 (\*)



(\*) التلخيص الآلي للنصوص - نماذجيات

(nmthgiat.com)، يشير وليد الصانع إلى 6 طرق لمعرفة جودة التلخيص تتفق مع علي فهمي، غير أنه يشفعها بالمعادلات الرياضية التي يمكن أن يحسب بها نتائجه بدقة.

## الخاتمة

3- أنظمة التلخيص الآلي الحالية تقوم

على طريقتين أساسيتين؛ التلخيص التوضيحي، والاستخراجي، وأحياناً الجمع بينهما فيما يسمى بالتلخيص الهجين.

4- لا توجد طريقة محددة لتقييم نتائج

التلخيص الآلي، وما يتم الآن هو الاستعانة بالعنصر البشري الذي تكون تقييماته نسبية.

خلصنا في العرض السابق إلى جملة من النتائج التالية:

1- لا تزال الحاجة ماسة في جميع

المجالات إلى تقنية التلخيص الآلي.

2- هناك تقدم في عدد من اللغات أبرزها

الإنجليزية، لكن اللغة العربية لا تزال

تصادف مشكلات بسبب طبيعة اللغة

العربية التي بها كثير من اللبس

التركيبى والدلالي.

9- التلخيص التوضيحي (Abstractive)

(Summarization): إعادة بناء

النص بعد تلخيصه بعبارات جديدة

10- التلخيص الهجين (Hybrid):

الدمج بين اختيار أهم الجمل

والعبارات في النص والتدخل المحدود

في إعادة بناء النص.

11- التسلسل المعجمي (Lexical

Chains): استخلاص سلاسل

الكلمات التي لها صلة بالنص؛

لنعرف من خلالها الأفكار المحورية،

وهذه الطريقة طورها موريس وهاريس

في التسعينيات بناء على أبحاث

سابقة في منتصف السبعينيات حول

وظائف اللغة للعالم اللغوي مايكل

هاليداي.

مصطلحات البحث

5- بنية الخطاب ( Discourse

Structure): اكتشاف مجموعة

الجمل التي تغطي سياق النص

وتجيب عن أهم الأسئلة التي يبحث

عنها القارئ.

6- التلخيص الآلي ( Automatic

Text Summarization): عملية

يحدث فيها ضغط للنص وحذف

المعلومات الهامشية والإبقاء على أهم

المعلومات التي يحتاجها القارئ.

7- التلخيص الاستخراجي (Extractive

Summarization):

8- استخراج أهم الجمل والعبارات في

النص، دون إعادة لكتابة النص

3- Using Lexical Chains for Text Summarization by Regina Barzilay & Michael El Hadad, 1997.

4- Mihalcea, R., & Tarau, P. (2004, July). TextRank: Bringing order into text. In *Proceedings of the 2004 conference on empirical methods in natural language processing* (pp. 404-411)

5- Zaheer, M., Guruganesh, G., Dubey, K. A., Ainslie, J., Alberti, C., Ontanon, S., ... & Ahmed, A. (2020, July). Big Bird: Transformers for Longer Sequences. In *NeurIPS*

6- N.Moratanch, 2017, A Survey on Extractive Text Summarization, IEEE International Conference on Computer, Communication, and Signal Processing.

7- Wafaa Samy Abdul-Hamed El-Kassas, 2021, Automatic Text Summarization Using Natural Language Processing And Artificial Intelligence Techniques,

### المراجع العربية

1- محمد علي كرد، خطط الشام، د.ط،

د. ت

2- علي فهمي وآخرون، 2019، المعالجة الآلية للنصوص العربية.

3- شمس الدين، محمد بن أحمد الخطيب الشربيني، 1994، مغني المحتاج، دار الكتب العلمية، بيروت.

4- إسكندر كسكاس وآخرون، التلخيص الآلي للنصوص العربية بين نظرية التلقين الرقمي ونظرية البنية البلاغية، الدورة السابعة للمؤتمر الدولي لعلوم وهندسة الحاسوب، الرياض 2012

### المراجع الأجنبية

1- "Lexical cohesion computed by thesaural relations as an indicator of the structure of text by Morris, J & G Hirst 1991

2- Luhn, H. P. (1958). The automatic creation of literature abstracts. *IBM Journal of research and development*, 2, 159-165