مجلة جامعة دمشق للعلوم الهندسية
المجلد 39- العدد 2 (2023) 124-140

Damascus university Journal for engineering sciences
Vol 39 No. 2 (2023): 124-140

التلخيص الآلي للنصوص العربية باستخدام الأنطولوجيا

م.سندس محمد سعيد الحلبي[1] د. محمد مازن محايري[2] أ.د. فيصل العباس[3]

1 مهندسة إلكترون حاصلة على ماجستير في الحاسبات والتحكم من جامعة دمشق.

2 أستاذ مساعد في قسم هندسة الحواسيب والأتمتة – كلية الهندسة الميكانيكية والكهربائية – جامعة دمشق.

3 أستاذ دكتور ورئيس الجامعة الدولية الخاصة للعلوم والتكنولوجيا– رئيس شعبة الكهرباء والحاسوب في الموسوعة العربية.

الملخص

تلخيص النصوص هو عملية استخلاص أو تجميع المعلومات الهامة من النصوص الأصلية وعرضها بشكل ملخص. يساعد التلخيص في الحصول على المعلومات المطلوبة بأقل وقت. يوجد الكثير من البحوث والدراسات والأنظمة الخاصة بالتلخيص الآلي للنصوص باللغات الأوروبية المتعددة وغيرها من اللغات العالمية، إلا أن البحوث والدراسات في مجال تلخيص النصوص العربية ضئيلة وتنمو ببطئ (El-Haj *et al.*, 2011). واتبعت في معظمها الأسلوب الاستخراجي واستخدام تقنيات إحصائية وابتعدت عن الأسلوب التجريدي واستخدام التقنيات الدلالية استناداً إلى مصادر معرفية مثل الأنطولوجيا. لذا اقتُرِحَ في هذا البحث نظام تلخيص آلي للنصوص العربية باعتماد الأنطولوجيا، وهو نظام تلخيص استخراجي عام أحادي اللغة وأحادي المستند يستخدم تقنيات دلالية بالاعتماد على الأنطولوجيا، ووُظِّفَت خصائص إحصائية ودلالية لتثقيل الجمل وتقييم أهميتها استناداً إلى توارد وتكرار مفاهيم المصطلحات الواردة فيها بدلاً من المصطلحات نفسها باستخدام علاقات دلالية بين المصطلحات تستنبط من الأنطولوجيا العربية وهي الترادف synonym والاشتمال holo_part والتضمين hyponym والعضو holo_member. وقد استخدمت أنطولوجيا الكلمات العربية Arabic Wordnet (AWN) المستخدمة على نحو واسع بعد توسيعها وإغناءها بمفاهيم وعلاقات مستنبطة من الموسوعة العربية باعتماد تقنية التقابل mapping. وقُيّم النظام باستخدام مجموعة الملخصات العربية من Essex (EASC) (El-Haj M. *et al.*, 2010) واعتماد مقاييس الدقة والاستذكار ومعامل F. فبينت النتائج أن متوسطات الدقة والاستذكار ومعامل F في التلخيص باعتماد الأنطولوجيا المقترح هي نسب جيدة مقارنة مع نسب أنظمة التلخيص الآلي للنصوص العربية المقارن معها.

الكلمات المفتاحية: التلخيص الآلي، التلخيص الآلي للنصوص العربية، التلخيص الاستخراجي، التلخيص التجريدي، التلخيص الدلالي، الأنطولوجيا، شبكة الكلمات العربية، التقابل الموسوعة العربية.

تاريخ الإيداع:2022/2/7
تاريخ القبول:2022/6/13

حقوق النشر: جامعة دمشق –سورية، يحتفظ المؤلفون بحقوق النشر بموجب الترخيص
CC BY-NC-SA 04

Automatic Summarization of Arabic Text Using Ontology

Eng. Sundos M. Said Al-Halabi[1]
Dr. Mohamad Mazen Almahayri[2]
Prof. Faysal Al-Abbas[3]

1 An electronic engineer with a master's degree in computers and control from Damascus University.
2Dr. at the Faculty of Computer Engineering at Damascus University
3Professor, and President of the International University for Science and Technology (IUST)- Head of the Electricity & Computer Department in the Arabic Encyclopedia

ABSTRACT

Text summarization is the process of extracting or compiling important information from the original texts and presenting them in a summary form. Summarizing helps in obtaining the required information in the least time.

There has been a lot of research, studies and systems for the automatic summarization of texts in various European and other international languages, but research and studies in the field of summarizing Arabic texts are few and growing slowly (El-Haj *et al.*, 2011) Most of them followed the extractive approaches and the use of statistical techniques and moved away from the abstract approaches and the use of semantic techniques based on knowledge sources such as ontology. Therefore, in this paper, we proposed an automatic summarization system for Arabic texts by adopting ontology, which is a general, monolingual and single-document extractive summarization system that uses semantic techniques based on ontology. Semantic relations between concepts derived from the Arabic ontology, which are synsets, holo_part, hyponym, and holo_member. The widely used Arabic Wordnet ontology has been used after expanding and enriching it with concepts and relationships derived from the Arabic encyclopedia using the mapping technology.The system was evaluated using Essex Arabic Abstracts Collection (EASC) (El-Haj *et al.*,2010), and measures of Precision, Recall and F measure. The results showed that the average precision, recall, and F measure in summarizing using the proposed ontology are good proportions compared to the proportions of automated summarization systems for Arabic texts.

Keywords: Automatic Sumarization, Arabic text summarization, Extractive Summarization, Abstractive Summarization, Semantic Summarization, Arabic Wordnet, Ontology, Arabic Encyclopedia, Mappin.

Received: 7/2/2022
Accepted: 13/6/2022

Copyright: Damascus University- Syria, The authors retain the copyright under a CC BY- NC-SA

1- المقدمة

مع السرعة الكبيرة لانتشار البيانات على الانترنت، أصبح هناك كم هائل من المصادر المتاحة للمستخدمين ومتصفحي الانترنت، لذا أصبحت عملية التلخيص الآلي للنصوص ضرورية للتغلب على مشكلة تزايد وتضخم المعلومات. تعرَف الملخص عموماً على أنها: "النص المولد من نص أو مجموعة نصوص، والذي يتضمن المعلومات الهامة من النص الأصلي، وهو بطول لايزيد عن نصف طول النص الاصلي وعادة يكون أقل من ذلك" (Radev *et al.*, 2002). ويمكن أن تتضمن الملخصات أي نوع آخر من الوسائط مثل الصوت والفيديو.

الهدف من هذا البحث هو توليد ملخص آلي لتلخيص النصوص العربية باعتماد الطريقة الدلالية باستخدام الأنطولوجيا. سوف يبدأ هذا البحث بخلفية نظرية عن أنظمة التلخيص الآلي ومراجعة الأدبيات والأعمال والدراسات السابقة لمعرفة الوضع الحالي في المواضيع المرتبطة ببناء ملخص آلي للنصوص العربية، وبعدها يقدم شرحاً عن نظام التلخيص الآلي للنصوص العربية باعتماد الأنطولوجيا المقترح في هذا البحث وطريقة تقييمه كما يتضمن بيان لآلية إيجاد أنطولوجيا عربية وذلك بإغناء وتوسيع أنطولوجيا شبكة الكلمات العربية باستخدام معارف الموسوعة العربية، وأخيراً الخاتمة والآفاق المستقبلية.

2- أنظمة التلخيص الآلي:

يمكن تصنيف أنظمة التلخيص حسب عدة محددات مثل: الدخل، والخرج، والعمومية، واللغة، والطريقة (Lloret *et al.*, 2008). يمكن أن يكون التلخيص أحادي المستند أومتعدد المستندات، ويمكن تصنيفه حسب العمومية إلى ملخص عام وملخص مقاد بالاستعلام. علاوة على ذلك يمكن التحدث عن تلخيص أحادي اللغة وتلخيص متعدد اللغات. كما تصنف طرائق التلخيص إلى استخراجية Extractive وتجريدية Abstructive. تختار الجمل الهامة في طرق التلخيص الاستخراجي من المستند المصدر اعتماداً على خصائص إحصائية ولغوية ودمجها على شكل ملخص (Gupta *et al.*, 2010). بينما تختلف طريقة التلخيص التجريدي بشكل أساسي عن الاستخراجي بتوفير ملخصات لها درجة من الاستدلال حول المعرفة الخلفية الموجودة في المستند الأصلي (Bawakid, 2011). تستخدم طرائق لغوية لفحص النص وإيجاد مفاهيم وتعابير جديدة لوصفه على نحو أفضل.

طُورَت عدة تقنيات لتلخيص النصوص صنفها الباحثون في (Nazari *et al.*, 2019) إلى أربع فئات هي: تقنيات إحصائية statistical، وتقنيات التعلم الآلي machine learning، وتقنيات دلالية semantic، وتقنيات الذكاء الجماعي Swarm intelligence. ومن طرائق التلخيص التي تعتمد التقنيات الدلالية، التلخيص باعتماد الأنطولوجيا كمصدر معرفي لاستنتاج تلخيص دلالي يأخذ بعين الاعتبار المفاهيم والعلاقات الدلالية.

تستخدم تقنيات تلخيص النصوص مجموعة من الخصائص الإحصائية والدلالية من أجل اختيار الجمل الهامة التي يمكن تضمينها في الملخص منها: العبارات المفتاحية، والموقع، والعبارات البارزة، والتشابه مع العنوان، وطول الجملة، والتشابه بين الجمل، والأسماء، ونوع الخط، والمعطيات الرقمية، ومركزية الجملة، والمعلومات غير الهامة.

تُقيّم الملخصات عادة إما يدوياً بالاستعانة بخبراء في المجال وهذا النوع من التقييم يكلف وقت وجهد كبيرين، أو آلياً باعتماد على معايير Recall-Oriented Understudy for Gisting Evaluation (ROUGE) لتقييم محتوى الملخصات وجودتها، أوباعتماد المقاييس المستندة إلى التوارد المشترك باستخدام مقاييس الدقة precision والاستذكار recall ومعاملF (Steinberger *et al.*, 2009) ويمكن

تقييم الملخصات خارجياً بتقييم أداء الملخص استناداً إلى تطبيق معين.

3- الدراسات والأبحاث في مجال التلخيص الآلي للنصوص العربية:

يوجد الكثير من البحوث والدراسات والأنظمة الخاصة بالتلخيص الآلي للنصوص باللغات الأوروبية المتعددة وغيرها من اللغات العالمية، إلا أن البحوث والدراسات في مجال تلخيص النصوص العربية ضئيلة وتنمو ببطئ. سوف نراجع فيما يلي الأبحاث والدراسات الحديثة في مجال تلخيص النصوص العربية مصنفة حسب التقنية المستخدمة في التلخيص.

التقنيات الإحصائية: يوجد العديد من مشاريع ودراسات تلخيص النصوص العربية التي تعتمد على التقنيات الإحصائية التي استخدمت مجموعة من الخصائص لتقييم أهمية الجمل إحصائياً منها: ((El-Ghannam *et al.*, 2013)، (El-Haj *et al.*, 2013)، (Abdelkrime *et al.*, 2015)، (Litvak *et al.*, 2016)). التقنيات الإحصائية سهلة التنفيذ ويمكن استخدامها لتحسين اختيار الجمل المهمة أو لإلغاء التكرار. ومع ذلك، فإنها تفشل في فهم النص لأنها تعتمد فقط على المقاييس الإحصائية.

تقنيات التعلم الآلي: الفكرة وراء التعلم الآلي هو استخدام مجموعة بيانات تدريب من أجل تدريب نظام التلخيص، حيث ينمذج التلخيص في هذه التقنيات كمشكلة تصنيف، تصنف فيها الجمل إلى مجموعتين هي: جمل ملخص وجمل ليست جمل ملخص اعتماداً على خصائص إحصائية بوجود مجموعة تدريب. وَظَّفَت بعض أنظمة تلخيص النصوص العربية التعلم الآلي باعتماد عدة خوارزميات تعلم الآلة مثل الخوارزميات الجينية والشبكات العصبونية ومن هذه الأنظمة: (Belguith) (Belkebir *et* ،(Yazan *et al.*, 2016) ،*et al.*, 2014) (Al Maleh *et* ،(Qaroush *et al.*, 2021) ،*al.*, 2015) (Tanfouri *et al.*, 2021) ،*al.*, 2020)) . يمكن أن ينتج عن هذه التقنيات نتائج تلخيص جيدة إلا أنها تعتمد على تعقيد قواعد بيانات التدريب في الحساب.

التقنيات الدلالية: تهتم هذه التقنيات بمعنى الكلمات والعلاقات بين الكلمات والعبارات والجمل لبناء المفاهيم المقصودة من النص. استخدمت عدة تقنيات دلالية في تلخيص النصوص العربية منها: التحليل الدلالي الكامن، واستخدام الأنطولوجيا والسلاسل المعجمية، والتجميع، والبيان، ونظرية الخطاب. استخدم التحليل الدلالي الكامن في كل من المشاريع التالية: ((Barzilay *et al.*, 2015)، (Froud *et al.*, 2013)). كما تم استخدام الأنطولوجيا كتلخيص دلالي يأخذ بعين الاعتبار المفاهيم والعلاقات الدلالية بين المصطلحات كما في نظام OSSAD (Imam *et al.*, 2013) وهو نظام مقاد بالاستعلام ويستخدم أنطولوجيا شبكة الكلمات العربية لتوسيع الاستعلام. كما استخدمت السلاسل المعجمية التي لها علاقات دلالية في (AL-Khawaldeh *et al.*, 2015). كما دمج بين التحليل الإحصائي واللغوي في (El-Shishtawy *et al.*, 2012). ساهم استخدام هذه التقنيات في التلخيص الآلي للنصوص في تحسين الجودة على نطاق واسع من خلال إنشاء ملخصات أكثر تماسكاً وأقل تكراراً وأكثر إفادة. ومع ذلك، فهي مهمة صعبة نظراً لأنها تواجه صعوبات في استخدام أدوات التحليل الدلالي عالية الجودة والموارد اللغوية لأنها تتطلب ذاكرة لحفظ المعلومات الدلالية مثل الأنطولوجيا وسعة المعالج بسبب المعالجة اللغوية المعقدة. تقنية التجميع من التقنيات الدلالية المستخدمة لتلخيص النصوص العربية التي تجمّع الجمل في فئات بالاعتماد على أوجه التشابه، استخدمت هذه التقنية لتلخيص النصوص العربية في المشاريع:

((Oufaida *et al.*, 2014)، (Fejer *et al.*, 2014)، (Haboush *et al.*, 2012)، (El-Haj *et al.*,2011)). على الرغم من استخدام تقنيات التجميع لتقليل تكرار البيانات، إلا أن الملخص الذي أنشئ قد يفتقر إلى المعنى بما يكفي لأن الجمل المختارة يتم تصنيفها بشكل أساسي اعتماداً على القرب من

النقطه الوسطى العنقودية دون الأخذ بالمعاني. كما استخدم البيان كتقنية دلالية يمثل فيها المستند ببيان غير موجه كل عقدة فيه تمثل جملة، تمثل الحواف بين العقد المتصلة علاقة التشابه بين الجمل. تعتبر الجملة مهمة إذا كانت مرتبطة ارتباطاً وثيقاً بالعديد من الجمل الأخرى. ومن المشاريع التي استخدمت البيان في التلخيص المشاريع التالية:((Al-Taani) et al., 2014) ، (Thomas et al., 2015) ، (Moawad et al., 2012)). إن استخدام التقنيات المعتمدة على البيان له مساهمة إيجابية في التلخيص متعدد المستندات لقدرته على اختيار موضوعات متميزة من البيانيات الفرعية غير المتصلة. ومع ذلك، فإن إنشاء البيانيات الفرعية اعتماداً على مقاييس التشابه الإحصائي دون إيلاء أي اهتمام لمعنى النص، يخاطر بإنتاج ملخص أقل إفادة (Lagrin et al., 2017). استخدمت أيضاً نظرية الخطاب كتقنية دلالية التي تمثل فيها النصوص بهيكل بطريقة ترتبط فيها وحدات الخطاب ببعضها البعض، بدلاً من التعامل مع النص على أنه تسلسل كلمات وجمل. استخدمت هذه النظرية في العديد من أنظمة تلخيص النصوص العربية مثل أنظمة: ((Elghazaly et al., 2012)، (Azmi et al., 2012)، (Keskes, 2015)، (Ibrahim et al., 2013)، (Maaloul,2012)). ومع ذلك، فقد ثبت أن بناء محللات آلية لمعلومات الخطاب مهمة صعبة ومكلفة من الناحية الحسابية.

تقنيات الذكاء الجماعي: اعتمدت خوارزمية حشد الجسيمات PSO في أنظمة التلخيص التالية: ((AlZahrani et al., 2015)، (Al-Abdallah et al., 2017)) لاستخراج ملخص لمستند واحد. يمكن أن تتميز هذه التقنيات بأداء جيد إلا أنها تتطلب تعقيد في القياس.

5- إيجاد أنطولوجيا عربية:

لا يوجد أنطولوجيا عربية تضم المفاهيم العربية ومجموعة ترادف المصطلحات العربية وعلاقاتها الدلالية إلا في بعض المشاريع البحثية غير المكتملة. وتعتبر شبكة الكلمات العربية أكثر الأنطولوجيات العربية استخداماً وهي متاحة ومتوفرة كمصدر معرفي مفتوح، إلا أنها أنشأت كترجمة لشبكة الكلمات الإنكليزية وللمفاهيم الإنكليزية. لذا حاولنا في هذا البحث توسيع هذه الأنطولوجيا آلياً باستخدام تقنية التقابل Mapping مع خصائص ومفاهيم الموسوعة العربية التي تعتبر مصدر معرفي غني بالمعلومات والمفاهيم العربية في كافة المجالات والاختصاصات. وقد استخدمت هذه الطريقة في مشاريع سابقة مثل: ((Szymanski et al., 2012)، (Fernando et al., 2009)، (John,2017))، حيث وسعت أنطولوجيا شبكة الكلمات الإنكليزية بإغنائها بمفاهيم وخصائص موسوعة الويكيبيديا باللغة الإنكليزية. لذا حاولنا في هذا البحث تطبيق تقنية التقابل بالاستفادة من هذه المشاريع وتطبيقها على الموسوعة العربية لتوسيع أنطولوجيا شبكة الكلمات العربية. وقد تمت عملية التقابل على مرحلتين، المرحلة الأولى هي مرحلة توليد خصائص الموسوعة العربية والتي تتضمن المعالجة الأولية لنصوص مقالات الموسوعة وفهرستها فهرسة آلية لاستخراج المصطلحات وتثقيلها اعتماداً على تواردها في نصوص الموسوعة، والمرحلة الثانية هي التقابل بين الخصائص المستخرجة وشبكة الكلمات العربية وفقاً لما يلي:

- مطابقة مفاهيم شبكة الكلمات مع عناوين مقالات الموسوعة، وإضافة عناوين المقالات غير الموجودة في الشبكة كمفاهيم جديدة.
- إضافة المصطلحات الهامة في المقالات إلى الشبكة بضمها إلى المفهوم المناسب في الشبكة.

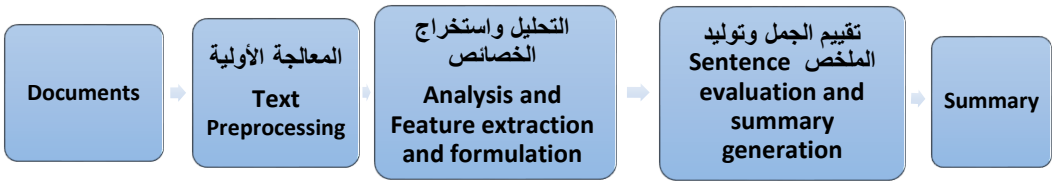
وبذلك حصلنا على أنطولوجيا عربية أفضل من شبكة الكلمات العربية يمكن الاعتماد عليها في عملية التلخيص الآلي المقترحة في هذا البحث.

6- نظام التلخيص الآلي للنصوص العربية:

اقترح نظامان لتلخيص النصوص العربية وهما نظام تلخيص باعتماد الأنطولوجيا ونظام تلخيص إحصائي لايعتمد الأنطولوجيا، واستخدم الأخير من أجل المقارنة وتقييم النظام المعتمد على الأنطولوجيا. وقد اعتمدت الأنطولوجيا العربية المولدة بالتقابل بين شبكة الكلمات العربية وخصائص الموسوعة العربية. وفيما يلي نقدم شرحاً عن نظامي التلخيص المقترحين.

يطبق النظامان المقترحان تلخيص استخراجي عام أحادي المستند وأحادي اللغة، ويتشاركان في مراحل التلخيص الأساسية مع اختلافات في تنفيذها توضح أثناء الشرح، وهذه المراحل هي: المعالجة الأولية Text Preprocessing لمستندات المصدر، والتحليل واستخراج الخصائص وصياغتها Analysis and Feature extraction and formulation، وتقييم الجمل وتوليد الملخص Sentence evaluation and summary generation.

يوضح الشكل (1) مراحل التلخيص الآلي المطبقة. يستقبل نظام التلخيص المستند المطلوب تلخيصه كدخل، ومتغيرات اختيارية يحددها المستخدم وهي نسبة التلخيص ونوع التلخيص، وينتج عن النظام ملخص مستند الدخل كخرج.



الشكل (1) خطوات التلخيص الآلي المقترح

يُعد المستند في مرحلة المعالجة الأولية ويُقدم بطريقة بنيوية موحدة لتسهيل العمل في المراحل القادمة. تحسب في المرحلة الثانية مجموعة من الخصائص الإحصائية والدلالية لكل جملة لتبين أهميتها، وتستخدم هذه الخصائص في مرحلة تقييم الجمل وتوليد الملخص.

6-1 المعالجة الأولية للنصوص:

هذه المرحلة هي المرحلة الأولية في جميع طرق التلخيص تقريباً. والغرض الرئيسي منها هو إعداد مستند الدخل النصي للمعالجة في المراحل الأخرى. يتضمن نظام تلخيص النصوص المقترح العمليات المتسلسلة التالية للمعالجة الأولية: التقطيع Tokenization، والتقييس Normalization، وإلغاء كلمات التوقف StopWord Removal، والتجريد Stemming (Abdelkrime *et al.*, 2015)، (Litvak *et al.*, 2016)، ((Thomas *et al.*, 2015)) يوضح الشكل (2) تسلسل خطوات المعالجة الأولية المقترحة.

الشكل (2) تسلسل خطوات المعالجة الأولية

التقطيع: تبدأ المعالجة الأولية للنص بعملية التقطيع التي تقسم مستندات الدخل إلى وحدات بمستويات مختلفة لتسهيل الوصول إلى جميع أجزاء مستند الدخل. هذه الوحدات عبارة عن جمل أو رموز أو أرقام أو أي وحدة مناسبة أخرى. التقطيع المقترح عبارة عن تحليل صرفي يعتمد على علامات الترقيم لتقسيم المستند إلى جمل بناءً على النقطة (.) وعلامة الاستفهام (؟) وعلامة التعجب (!) والفواصل (،) كمحددات. ثم يتم تقسيم هذه الجمل إلى مقاطع استناداً إلى محددات مثل المسافة البيضاء والفاصلة المنقوطة والاقتباسات. من خلال إجرائية تطبق خوارزمية الأداة AraNLP مع بعض التعديلات للتعامل مع تسلسل العمليات أعلاه (Althobaiti *et al.*, 2014).

التقييس: في اللغة العربية، قد تظهر بعض الحروف العربية بأشكال مختلفة، وتستخدم حروف بدلاً من حروف لأن أشكالها متشابهة. كما يستخدم الكتّاب علامات التشكيل في نصوصهم، التي يمكن أن تؤدي إلى اختلافات للمصطلح نفسه، وبالتالي تؤثر على حساب بعض الخصائص مثل التردد (TF). لذلك فإن عملية التقييس مطلوبة لتوحيد الأشكال المختلفة للحرف نفسه لتجنب مثل هذه الاختلافات. استخدمت إجرائية التقييس المقترحة الخوارزمية المطبقة في أداة AraNLP للقيام بالمهام التالية (Althobaiti *et al.*, 2014): إزالة الأحرف غير العربية مثل الرموز وعلامات الترقيم الخاصة، وإزالة علامات التشكيل، واستبدال "ا" بـ "أ إ آ"، و "ي" بـ "ى"، و "ه" بـ "ة"، و "و" بـ "ؤ"، و "ي" بـ "ئ" (Ayedh *et al.*,2016)، وإزالة التطويل.

إزالة كلمات التوقف: كلمات التوقف (مثل الضمائر وحروف الجر وحروف العطف وما إلى ذلك) هي كلمات غير مهمة تظهر كثيرًا في المستندات لتكوين الجمل (Kanan *et al.*, 2004). نظراً لأن هذه الكلمات ليست مفيدة (لا تضيف معلومات)، فيمكن حذفها من الجمل دون التأثير على المحتوى الأساسي للجملة. في الواقع، هناك العديد من طرق قائمة التوقف المستخدمة لإزالة كلمات الإيقاف من النص بما في ذلك، قائمة توقف عامة، قائمة توقف بمجال معين، وقائمة توقف مدمجة. تعتمد الطريقة المقترحة في هذا البحث على قائمة التوقف العامة المستخدمة في أداة AraNLP، وهي القائمة المعتمدة في معظم التطبيقات مع بعض الإضافات.

التجريد: التجريد بالعربية عملية شاقة ودقيقة، لأن اللغة العربية هي لغة صرفية واشتقاقية بحتة، مما يعني أن الكلمات العربية يمكن أن يكون لها العديد من الأشكال المختلفة ولكنها تشترك في المعنى المجرد نفسه. إن عملية التجريد هي عملية إزالة بعض أو كل اللواحق (مثل البادئات، واللواحق، واللواحق اللاحقة) من الكلمة وذلك بتحويل الأشكال والمشتقات المختلفة للكلمة إلى شكل موحد واحد مثل الجذر أو الجذع. في اللغة العربية، هناك طريقتان شائعتان للتجريد، التجريد المعتمد على الجذور والتجريد الخفيف (Mustafa *et al.*,2017). قورنت هذه التقنيات في العمل المقدم في (Alami *et al.*,2016) فيما يتعلق بتلخيص النصوص باستخدام اثنين من المجردات العربية المعروفة. وأظهرت تجاربهم أنه في تلخيص النص العربي يفضل تجريد الجذر على التجريد الخفيف. بناءً على هذه النتائج، قمنا بتطبيق إجرائية التي تعتمد على خوارزمية مجرد خوجة وقواعد بياناتها مع بعض الإضافات على تلك القواعد للتعامل مع عملية الاشتقاق كمهمة معالجة مسبقة للعمل المقترح.

6-2 التحليل واستخراج الخصائص وصياغتها:

يجري اختيار الجمل ذات الأهمية أو التي لها المواءمة العليا استناداً على توظيف مجموعة خصائص لتوليد ملخص متماسك يحوي الأفكار الرئيسة في مستند معين. لذا فإن تصميم واختيار هذه الخصائص يؤثر على جودة الملخص المتولد. اعتمد في عملية التصميم والاختيار على فرضيات وتجارب وبعض الدراسات السابقة ((Meena *et al.*, 2014)،

(Meena *et al.*, 2016)). ويمكن بيان الخصائص المختارة لعملية التلخيص بما يلي:

6-2-1 خاصية العبارات المفتاحية: العبارات المفتاحية هي قائمة مختصرة بالكلمات المفتاحية الهامة التي توفر ملخصاً موجزاً للموضوع الرئيسي في المستند (Turney *et al.*, 2000)، قد تكون كلمة واحدة أو مصطلحاً من عدة كلمات. إن إمكانية وجود فكرة أساسية في الجملة مشروطة باحتوائها على عبارة أو كلمات مفتاحية الذي من شأنه زيادة أهميتها بالنسبة للجمل الأخرى (El-Shishtawy *et al.*, 2012). تعتمد درجة خاصية العبارة المفتاحية على العديد من العوامل، بما في ذلك: تكرار العبارة، وعدد الكلمات فيها، وتكرار الكلمة ذات التوارد الأكبر فيها، وطول العبارة بالنسبة للجملة التي تحتويها (El-Shishtawy *et al.*, 2012). استخدمنا في هذا البحث اجرائيات معتمدة على الخوارزمية المطبقة في أداة Kp-Miner (El-Beltagy *et al.*, 2009) لاستخراج العبارات المفتاحية مع التعديل عليها وذلك بتثقيل العبارات في الجمل بطريقتين طريقة تخص التلخيص الإحصائي المقترح وطريقة تخص التلخيص باعتماد الأنطولوجيا المقترح.

تحسب درجة خاصية العبارة المفتاحية في نظام التلخيص الإحصائي المقترح باستخدام إجرائية بناءً على ثلاثة من أهم العوامل وهي: تردد المصطلح في الجملة، عدد الجمل في المستند، وعدد الجمل التي تحتوي المصطلح في المستند. استخدمت المعادلة رقم (1) وهي علاقة التثقيل المشهورة من أجل حساب وزن العبارات المفتاحية والتي تم تقييسها بضرب وجمع العدد '0.1' الذي تم اختياره تجريبياً.

$$S_{key} = 0.1 + 0.1\left(\frac{t_{fi}}{max_{t_{fi}}}\right)\log(\frac{N}{n_i}) \quad (1)$$

حيث:tfi تردد المصطلح في الجملة، و N عدد الجمل في المستند، و ni عدد الجمل التي تحتوي المصطلح i.

اعتُمد في التلخيص باعتماد الأنطولوجيا على حساب الثقل استناداً إلى المفاهيم بدلاً من المصطلحات، حيث يعاد المصطلح إلى مفهومه الدلالي باستخدام الأنطولوجيا العربية المقترحة، ويحسب عدد مرات تكرار المفهوم الذي يدخل في علاقة حساب الثقل. وبذلك يكون الحساب أدق حيث يكون للمصطلح ومرادفاته الثقل نفسه. لذا طبقت إجرائية التثقيل في التلخيص الآلي باعتماد الأنطولوجيا باستخدام الخطوتين التاليتين:

أولاً- إعادة المصطلح إلى مفهومه وفقاً للخوارزمية التالية:

1. يسند إلى المصطلح المفهوم المناسب له حسب علاقة الترادف synset في الأنطولوجيا.
2. يسند إلى مفهوم المصطلح أحد المصطلحات المرتبطة به وفق علاقات الاشتمال أو العضو أو التضمين إذا لم توجد له علاقة ترادف حسب البند 1.
3. يسند إلى مفهوم المصطلح أحد مفاهيم المصطلحات المشابهة له، والتي ترتبط معه بعلاقات تشابه إذا لم يرتبط المصطلح بعلاقات البند 1 و 2.

ثانياً- تثقيل المفاهيم: تستخدم علاقة التثقيل وفقاً للمعادلة (2) المشابهة للعلاقة المطبقة في التلخيص الإحصائي مع حساب تكرار المفهوم في الجملة والمستند بدلاً من المصطلحات.

$$S_{key} = 0.1 + 0.1\left(\frac{c_{fi}}{max_{c_{fi}}}\right)\log(\frac{N}{n_i}) \quad (2)$$

حيث:cfi تردد المفهوم في الجملة، و N عدد الجمل في المستند، و ni عدد الجمل التي تحتوي المفهوم i.

6-2-2 خاصية موقع الجملة: تم اقتراح هذه الخاصية أولاً في (Baxendale,1958) لتقييم الجملة حيث تعتمد أهمية الجملة على موقعها في الفقرة / المستند بغض النظر عن مجال / موضوع المستند. اقترحت العديد من العلاقات فيما يتعلق بموقع الجملة في مجال تلخيص النص الاستخراجي

Gupta *et al.*,) ،(Abuobieda *et al.*, 2012))
.((2011

اعتمد في طرق التلخيص المقترحة أن الجملتين الأوليتين في المستند لها أهمية أكثر من باقي المستند في النص لذلك أعطيت $S_{Location}$ درجة تساوي 2، وأعطيت الجملتين الثالثة والرابعة درجة تساوي 1، وباقي الجمل أخذت درجة تساوي الصفر. وقد اختيرت هذه القيم تجريبياً.

6-2-3 خاصية التشابه مع العنوان: اقترحت هذه الخاصية أولاً في (Edmundson ,1969)، وعُرفت على أنها التشابه أوالتداخل بين جملة معينة وعنوان المستند ((Fattah *et al.*, 2009)، (Abuobieda *et al.*, 2012)، (Nobata *et al.*, 2009)). تأتي أهمية هذه الخاصية من فكرة أنه إذا كانت الجملة تتكون من كلمات تظهر في العنوان، فقد تكون جملة مهمة. علاوة على ذلك، إذا كانت الجملة تشترك في العبارة المفتاحية مع العنوان، فسيؤدي ذلك إلى زيادة درجتها بشكل كبير. اعتبرت في هذا البحث الجملتين الأوليتين في المستند هما عنوان المستند لكون هناك مستندات بلاعنوان. تحسب درجة خاصية العنوان في التلخيص الإحصائي المقترح وفقاً للمعادلة (3) باعتماد ثقل وتكرار المصطلحات الواردة في الجملة والعنوان.

$$S_{TitleSim} = Similarity(Title, S_i) * \sqrt{1 + (KPT \cap KPS_i)} \quad (3)$$

حيث: Si هي الجملة الحالية، و KPT عدد العبارات المفتاحية الموجودة في العنوان، وKPSi عدد العبارات المفتاحية الموجودة في الجملة Si، و$(KPT \cap KPS_i)$ هو عدد العبارات المفتاحية المشتركة بين الجملة Si والعنوان. و Similarity(Title,Si) هي درجة تشابه العنوان مع جملة Si التي تحسب بمقياس تشابه التجيب باستخدام العلاقة المشهورة لحساب تشابه النصوص (Gomaa *et al.*,. 2013) وفقاً للمعادلة رقم (4)، تمثَّل كل جملة Si في هذا النموذج بشعاع بـ n بعد.

$S_i = \{w_{i1}, w_{i2}, \dots, w_{ik}, \dots, w_{in}\}$، حيث wik هو وزن المصطلح tk الموجود في الجملة Si والذي تم حسابه وفقاً للمعادلة (2)، و n هو عدد الكلمات الفريدة جميعها الموجودة في المستند الهدف. لذلك، بناءً على هذا التمثيل، يحسب تشابه التجيب على النحو التالي:

$$Sim_{Title,Si} = \frac{\sum_{t=1}^{n} w_{tSi} w_{tTitle}}{\sqrt{\sum_{t=1}^{n} w_{tSi}^2} \sqrt{\sum_{t=1}^{n} w_{tTitle}^2}} \quad (4)$$

حيث: SimTitle,Si التشابه بين الجملة Si والعنوان Title، و n عدد المصطلحات الفريدة في المستند D، WtSi وزن المصطلح t في الجملة Si، و WtTitle وزن المصطلح t في العنوان Title.

بالنسبة للتلخيص باعتماد الأنطولوجيا: تحسب علاقة التشابه بين العنوان والجملة وفقاً للمعادلة رقم (5) وذلك بحساب تجيب الزاوية بين شعاع مفاهيم الجملة (بدلاً من شعاع مصطلحات الجملة) وشعاع العنوان.

$$Sim_{Title,Si} = \frac{\sum_{c=1}^{n} w_{cSi} w_{cTitle}}{\sqrt{\sum_{c=1}^{n} w_{cSi}^2} \sqrt{\sum_{c=1}^{n} w_{cTitle}^2}} \quad (5)$$

حيث: $Sim_{Title,Si}$ التشابه بين الجملة Si والعنوان Title، و n عدد المفاهيم الفريدة في المستند D، و WcSi وزن المفهوم c في الجملة Si، وWcTitle وزن المفهوم c في العنوان Title.

6-2-4 خاصية طول الجملة: قد يؤثر طول الجملة على أهميتها. وبالتالي، قد تُستبعد الجمل الطويلة جداً أو القصيرة جداً من الملخص. فيمكن أن تزيد الجملة الطويلة جداً من محتواها المعلوماتي ولكن يمكن أن يتجاوز طولها الحد المقبول عندما تصبح الجملة مفرطة في التفاصيل. أيضاً، تميل الجمل القصيرة إلى تضمين معلومات أقل مقارنة بالجمل الأخرى، وبالتالي فهي أقل أهمية.

يحسب طول الجملة الطبيعي في التلخيص الإحصائي المقترح باستخدام المعادلة رقم (6)، وذلك بحساب حاصل قسمة عدد المصطلحات في الجملة على عدد المصطلحات في أطول جملة.

$$S_{length} = \frac{number\ of\ words\ in\ the\ sentence}{number\ of\ words\ in\ the\ longest\ sentence} \quad (6)$$

ويحسب طول الجملة الطبيعي في التلخيص باعتماد الأنطولوجيا باستخدام المعادلة (7)، وذلك

بحساب حاصل قسمة عدد المفاهيم في الجملة على عدد المفاهيم في أطول جملة.

$$S_{length} = Length\ Score = \frac{number\ of\ concepts\ in\ the\ sentence}{number\ of\ concepts\ in\ the\ longest\ sentence} \quad (7)$$

6-2-5 خاصية العبارات البارزة: يمكن أن يكون وجود بعض العبارات في الجملة (مثل: الأهم، بالتحديد، كملخص، كنتيجة، الأفضل) مؤشراً جيداً على أهمية المحتوى ((Lakshmi *et al.*, 2015)، (Barzilay *et al.*, 2015)). وبالتالي، تُعطى الجمل التي تحتوي على هذه العبارات درجة أعلى مقارنة بالجمل الأخرى. تحسب درجة هذه الخاصية باستخدام المعادلة رقم (8)، وذلك بحساب حاصل قسمة عدد العبارات البارزة في الجملة على عدد العبارات البارزة في المستند.

$$S_{Cue} = \frac{number\ of\ Cue\ phrases\ in\ the\ sentence}{number\ of\ Cue\ phrases\ in\ the\ document} \quad (8)$$

6-2-6 خاصية توارد المعلومات غير الهامة: بعض العبارات مثل (بالإضافة إلى ذلك، على سبيل المثال) هي علامات للكلام بمثابة كلمات شرح. مثل هذه العبارات تضعف الجملة لأنها توحي بأن الجملة القادمة هي معلومات إضافية تتعلق بالفكرة الأساسية (Gupta *et al.*, 2010). لذلك يجب تقليل درجة الجمل التي تحتوي على هذه العبارات باستخدام المعادلة رقم (9) وذلك بحساب مقلوب عدد الكلمات الضعيفة في الجملة.

$$S_{Weak} = \frac{1}{number\ of\ words\ in\ the\ sentence} \quad (9)$$

6-3 تقييم الجمل وتوليد الملخص:

في تلخيص نص استخراجي، تُحدد مقاطع نصية مهمة (مثل الجمل) من المستند الأصلي عادة بناء على مجموعة من الخصائص الهامة المستخرجة من مستويات مختلفة (على سبيل المثال الجملة، والمستند). تستخدم هذه الطرق على نطاق واسع في تلخيص المستندات الأحادية الاستخراجية وتتأثر نتائجها بشكل كبير بالخصائص المختارة مع تصميمها وصياغتها. في التلخيص الإحصائي، يتم استخراج الجمل الهامة من النص بناء على مجموع الدرجات التي يتم تحديدها لهم (Meena *et al.*, 2016).

تُستخرج الجمل الهامة من المستند في التلخيص الإحصائي المقترح بناءً على المجموع الخطي لدرجات خصائص كل جملة في المستند والتي حسبت اعتماداً على المصطلحات الواردة فيها، بينما تستنتج الجمل الهامة من المستند في التلخيص المعتمد على الأنطولوجيا المقترح استناداً إلى المجموع الخطي لخصائص كل جملة في المستند والتي حسبت اعتماداً على مفاهيم المصطلحات الواردة فيها بدلاً من المصطلحات نفسها.

تحسب درجة ثقل جملة معينة بحساب المجموع الخطي لدرجات الخصائص المحسوبة أعلاه وذلك بتطبيق المعادلة رقم (10)

$$Sentence\ Score = S_{weak} * (S_{key} + S_{Location} + S_{TitleSim} + S_{length} + S_{cue}) \quad (10)$$

حيث S_{key} درجة خاصية العبارات المفتاحية المحسوبة بالمعادلة رقم (1) في التلخيص الإحصائي أوالمعادلة رقم (2) في التلخيص باعتماد الأنطولوجيا، و$S_{Location}$ درجة خاصية موقع الجملة، و$S_{TitleSim}$ درجة خاصية التشابه مع العنوان المحسوبة بالمعادلة رقم (3) في التلخيص الإحصائي أو المعادلة رقم (5) في التلخيص باعتماد الأنطولوجيا، و S_{length} درجة خاصية طول الجملة المحسوبة بالمعادلة رقم (6) في التلخيص الإحصائي أو المعادلة رقم (7) في التلخيص باعتماد الأنطولوجيا ، وS_{cue} درجة خاصية العبارات البارزة المحسوبة بالمعادلة رقم (8)، وS_{weak} درجة خاصية الكلمات الضعيفة المحسوبة بالمعادلة رقم (9).

تُرتب الجمل بعد حساب النتيجة الإجمالية ترتيباً تنازلياً بناءً على المجموع الإجمالي لدرجات خصائصها. ثم تُختار الجمل ذات الترتيب الأعلى لتضمينها في ملخص الخرج بناءً على نسبة التلخيص المطلوبة.

أخيراً، يعاد ترتيب الجمل المستخرجة بناءً على موضعها الأصلي في المستند للحفاظ على تناسق النص في الملخص الذي أنشئ.

7- التقييم والنتائج:

إن اختبار وتقييم نظام تلخيص النصوص الآلي عملية صعبة نظراً لعدم وجود ملخص مثالي لمستند معين أو مجموعة من المستندات المرتبطة. علاوة على ذلك، فإن الافتقار إلى مجموعات البيانات القياسية العربية الحالية جعل عملية التقييم أكثر تعقيداً وربما شخصية في بعض الحالات لأن الباحثين يميلون إلى جمع مجموعات البيانات الخاصة بهم (Al-Saleh *et al.*, 2016). هناك طريقتان لعملية التقييم: يدوية وآلية. الطريقة اليدوية مكلفة وتستغرق وقتاً كثيراً لذلك اعتمدت الطريقة الآلية في التقييم.

طبق التقييم الآلي باعتماد التوارد المشترك واستخدام مقاييس الدقة precision والاستذكار recall ومعامل F وهي المقاييس الأكثر شهرة واستخداماً (Oufaida *et al.*, 2014)، تقيس الدقة مقدار صحة correction الملخص أو مقدار صحة المعلومات الموجودة في الملخص، وتحسب بتقاطع الجمل المستخرجة والجمل الموائمة مقسوماً على الجمل المستخرجة كلها وتحسب حسب المعادلة رقم (12)، ويستخدم الاستذكار لقياس الاستكمال completeness أومقدار كون المعلومات الموجودة في الملخص تغطي المعلومات الهامة في المرجع المرتبط بالملخص. ويحسب بتقاطع الجمل المستخرجة والجمل المثالية مقسوماً على الجمل المثالية الموائمة كلها. ويحسب باستخدام المعادلة رقم (13)، ويدمج المقياس F-Score بين الدقة Precision P والاستذكار Recall R (Salton,1989) وهو متوسط مقاييس الدقة والاستذكار ويعطى بالمعادلة رقم (14).

$$P = \frac{relevant\ Sentences \cap Retrieved\ Sentences}{retrieved\ Sentences} \quad (12)$$

$$R = \frac{Relevant\ Sentences \cap Retrieved\ Sentences}{relavant\ Sentences} \quad (13)$$

$$F = \frac{2PR}{P+R} \quad (14)$$

مجموعة البيانات:

استُخدمَت مجموعة الملخصات العربية من Essex (EASC) (El-Haj *et al.*,2010) لاختبار وتقييم نظم التلخيص المقترحة. مجموعة EASC عبارة عن ملخص استخراجي من صنع الإنسان تم نشره بواسطة مجموعة من الباحثين في جامعة إسيكس. يضم 153 مستنداً حول مواضيع مختلفة تم جمعها من الصحف العربية وويكيبيديا. لكل مستند في مجموعة EASC هناك خمسة ملخصات مرجعية مختلفة أنشئت كل واحدة من قبل إنسان مختلف. تتميز مجموعة البيانات هذه بأنها مجموعة البيانات العربية الوحيدة المُنشأة والتي تجعل التقييم أكثر واقعية مقارنة بالمناهج الأخرى مثل الاعتماد على مجموعة بيانات مترجمة أو اعتماداً على خرج الملخصات المطورة مسبقاً.

ولحل مشكلة الشخصنة في وجود خمس ملخصات مرجعية لكل مستند في مستندات EASC، أنشئ ملخص الأغلبية يسمى ملخص المعيار الذهبي الذي يضم الجمل الموجودة في معظم المراجع الخمسة (ثلاثة أو أكثر)، وهذا مايسمى الملخص المرجعي الذهبي (El-Haj, 2012). كما حددت عتبة نسبة التلخيص لإنشاء ملخص الخرج بـ 50%، وهي النسبة الوسطية لطول ملخص المعيار الذهبي المرجعي لكل مستند.

التجارب والنتائج:

حُسبت مقاييس الدقة والاستذكار والمعامل F وذلك باستخدام مجموعة الملخصات المرجعية EASC، من أجل تقييم نظام التلخيص الإحصائي المقترح، ونظام التلخيص باعتماد الأنطولوجيا المقترح، ونظام التلخيص المشهور لخصلي Lakhasly والمتوافر على الشابكة من خلال موقع ذي العنوان https://lakhasly.com، والذي يعتبر أحد أبرز المواقع التي تقوم بتلخيص الفقرات

والنصوص العربية على الانترنت Online وأكثرها استخداماً.

يبين الشكل (3) مثالاً لأحد النصوص الملخصة وهو "لودفيج فان بيتهوفن" عدد جمله 54 جملة وعدد جمل الملخص المرجعي فيه 20 جملة. كما تبين الأشكال (4 و 5 و 6) تلخيص هذا النص باستخدام نظام التلخيص باستخدام الأنطولوجيا والتلخيص الإحصائي وتلخيص نظام لخصلي على الترتيب، مع إظهار الجمل الموائمة بلون شفاف، ونسب الدقة والاستذكار ومعامل F في كل منها.

يبين الجدول رقم (1) جدولاً بجمل التلخيص باعتماد الأنطولوجيا مع قيم ثقلها ونسب خصائصها لهذا النص.

لودفيج فان بيتهوفن مؤلف موسيقي ألماني ولد عام 1770 م في مدينة بون. يعتبر من أبرز عباقرة الموسيقى في جميع العصور، وأبدع أعمالاً موسيقية خالدة. له الفضل الأعظم في تطوير الموسيقى الكلاسيكية. قدم أول عمل موسيقي وعمره 8 سنوات. تشمل مؤلفاته للأوركسترا تسعة سيمفونيات وخمس مقطوعات موسيقية على البيانو ومقطوعة على الكمان. كما ألّف العديد من المقطوعات الموسيقية كمقدمات للأوبرا. بدأ بيتهوفن يفقد سمعه في الثلاثينيات من عمره إلا أن ذلك لم يؤثر على إنتاجه الذي ازداد في تلك الفترة وتميز بالإبداع. من أجمل أعماله السمفونية الخامسة والسادسة والتاسعة. وقد توفي في فيينا عام 1827م. شهدت مدينة بون الألمانية ميلاد الفنان العبقري لودفج فان بيتهوفن في 16 ديسمبر عام 1770، وتم تعميده في 17 ديسمبر 1770. ظهر تميزه الموسيقي منذ صغره، فنشرت أولى أعماله وهو في الثانية عشرمن عمره عام 1783 م. اتسعت شهرته كعازف بيانو في سن مبكرة، ثم زاد إنتاجه وذاع صيته كمؤلف موسيقى. عانى بيتهوفن كثيراً في حياته، عائلياً وصحياً، فبالرغم من أن أباه هو معلمه الأول الذي وجه اهتمامه للموسيقى ولقنه العزف على البيانو والكمان، إلا أنه لم يكن الأب المثالي، فقد كان مدمناً للكحول، كما أن والدته توفيت وهو في السابعة عشر من عمره بعد صراع طويل مع المرض، تاركة له مسؤولية العائلة. مما منعه من إتمام خطته والسفر إلى فيينا، عاصمة الموسيقى في ذلك العصر. فهل كان التأليف الموسيقي هو نوع من أنواع العلاج والتغلب على المشاكل بالنسبة لبيتهوفن. في 1789 م تحقق حلمه أخيراً، فقد أرسله حاكم بون إلى فيينا، وهناك تتلمذ على يد هايدن. ولكن بيتهوفن، صاحب الألحان واجه بعض الخلافات مع معلمه، وعندما سافر هايدن إلى لندن، تحول بيتهوفن إلى معلمين آخرين مثل ساليري وشينك وألبريشتبيرجر. وقد أسهمت كل هذه الدروس والاحتكاكات في تكوين شخصية بيتهوفن الفنية. وحاول أن يشق لنفسه طريق كعازف في عاصمة الموسيقى، وسرعان ما لاقى مكانة كبرى خاصة في الأوساط الأرستقراطية. فقد حاز بإعجاب الأسرة الملكية وعومل كصديق أكثر منه مؤلفاً. بالرغم من ذلك فقد عاش ومات فقيراً، غناه هو أعماله الفنية المتميزة. فقد جاء إنتاجه الفني غزيراً حتى بعد إصابته بالصمم. بدأت إصابة بيتهوفن بصمم بسيط عام 1802، فبدأ في الانسحاب من الأوساط الفنية تدريجياً، وأمضى حياته بلا زواج يرتبط بعلاقات عدة مع سيدات صغيرات. إلا أنه لم يتوقف عن الإنتاج الفني، ولكن أعماله اتخذت اتجاه جديد. ومع ازدياد حالة الصمم التي أصابته، امتنع عن العزف في الحفلات العامة، وابتعد عن الحياة الاجتماعية واتجه للوحدة، وقلت مؤلفاته، وأصبحت أكثر تعقيداً. حتى أنه رد على انتقادات نقاده بأنه يعزف للأجيال القادمة. وبالفعل مازالت أعماله حتى اليوم من أهم ما أنتجته الموسيقى الكلاسيكية العالمية. واكتسبت اثنان من السيمفونيات التي كتبها في صممه أكبر شعبية، وهما السيمفونية الخامسة والتاسعة. كما أنه أحدث الكثير من التغييرات في الموسيقى، وأدخل الغناء والكلمات في سيمفونيته التاسعة.فجاءت رسالته إلى العالم كل البشر سيصبحون إخوة.

الشكل (3) النص الأصلي (54 جملة، 20 جملة موائمة ملونة بلون شفاف)

لودفيج فان بيتهوفن مؤلف موسيقي ألماني ولد عام 1770 م في مدينة بون، يعتبر من أبرز عباقرة الموسيقى في جميع العصور، وأبدع أعمالاً موسيقية خالدة، له الفضل الأعظم في تطوير الموسيقى الكلاسيكية، تشمل مؤلفاته للأوركسترا تسعة سيمفونيات وخمس مقطوعات موسيقية على البيانو ومقطوعة على الكمان، بدأ بيتهوفن يفقد سمعه في الثلاثينيات من عمره إلا أن ذلك لم يؤثر على إنتاجه الذي ازداد في تلك الفترة وتميز بالإبداع، وقد توفي في فيينا عام 1827م، شهدت مدينة بون الألمانية ميلاد الفنان العبقري لودفج فان بيتهوفن في 16 ديسمبر عام 1770، ظهر تميزه الموسيقي منذ صغره، فنشرت أولى أعماله وهو في الثانية عشرمن عمره عام 1783 م، ثم زاد إنتاجه وذاع صيته كمؤلف موسيقى، فبالرغم من أن أباه هو معلمه الأول الذي وجه اهتمامه للموسيقى ولقنه العزف على البيانو والكمان، عاصمة الموسيقى في ذلك العصر، فهل كان التأليف الموسيقي هو نوع من أنواع العلاج والتغلب على المشاكل بالنسبة لبيتهوفن، فقد أرسله حاكم بون إلى فيينا، وهناك تتلمذ على يد هايدن ولكن بيتهوفن، تحول بيتهوفن إلى معلمين آخرين مثل ساليري وشينك وألبريشتبيرجر، وحاول أن يشق لنفسه طريق كعازف في عاصمة الموسيقى، غناه هو أعماله الفنية المتميزة، بدأت إصابة بيتهوفن بصمم بسيط عام 1802، وأمضى حياته بلا زواج يرتبط بعلاقات عدة مع سيدات صغيرات، ومع ازدياد حالة الصمم التي أصابته، امتنع عن العزف في الحفلات العامة، وابتعد عن الحياة الاجتماعية واتجه للوحدة وقلت مؤلفاته، وأصبحت أكثر تعقيداً، وبالفعل مازالت أعماله حتى اليوم من أهم ما أنتجته الموسيقى الكلاسيكية العالمية، كما أنه أحدث الكثير من التغييرات في الموسيقى

الشكل (4) الملخص باستخدام الانطولوجيا (الدقة 0.57 ، الاستذكار 0.75 ومعامل F 0.65)

لودفيج فان بيتهوفن مؤلف موسيقي ألماني ولد عام 1770 م في مدينة بون، يعتبر من أبرز عباقرة الموسيقى في جميع العصور، وأبدع أعمالاً موسيقية خالدة، له الفضل الأعظم في تطوير الموسيقى الكلاسيكية، قدم أول عمل موسيقي وعمره 8 سنوات، تشمل مؤلفاته للأوركسترا تسعة سيمفونيات وخمس مقطوعات موسيقية على البيانو ومقطوعة على الكمان، كما ألّف العديد من المقطوعات الموسيقية كمقدمات للأوبرا، بدأ بيتهوفن

يفقد سمعه في الثلاثينيات من عمره إلا أن ذلك لم يؤثر على إنتاجه الذي ازداد في تلك الفترة وتميز بالإبداع، شهدت مدينة بون الألمانية ميلاد الفنان العبقري لودفج فان بيتهوفن في 16 ديسمبر عام 1770، وتم تعميده في 17 ديسمبر 1770، فنشرت أولى أعماله وهو في الثانية عشرمن عمره عام 1783 م، ثم زاد إنتاجه وذاع صيته كمؤلف موسيقي، عانى بيتهوفن كثيراً في حياته عائلياً وصحياً، فبالرغم من أن أباه هو معلمه الأول الذي وجه اهتمامه للموسيقى ولقنه العزف على البيانو والكمان، عاصمة الموسيقى في ذلك العصر، فهل كان التأليف الموسيقي هو نوع من أنواع العلاج والتغلب على المشاكل بالنسبة لبيتهوفن، فقد أرسله حاكم بون إلى فيينا، وهناك تتلمذ على يد هايدن ولكن بيتهوفن، وقد أسهمت كل هذه الدروس والاحتكاكات في تكوين شخصية بيتهوفن الفنية، وحاول أن يشق لنفسه طريق كعازف في عاصمة الموسيقى، بدأت إصابة بيتهوفن بصمم بسيط عام 1802، وأمضى حياته بلا زواج يرتبط بعلاقات عدة مع سيدات صغيرات، إلا أنه لم يتوقف عن الإنتاج الفني، امتنع عن العزف في الحفلات العامة، وبالفعل مازالت أعماله حتى اليوم من أهم ما أنتجته الموسيقى الكلاسيكية العالمية، وهما السيمفونية الخامسة والتاسعة، كما أنه أحدث الكثير من التغييرات في الموسيقى

الشكل (5) الملخص باستخدام التلخيص الاحصائي (الدقة 0.45 والاستذكار 0.6 ومعامل F 0.51)

لودفيج فان بيتهوفن مؤلف موسيقي ألماني ولد عام 1770 م في مدينة بون. وأبدع أعمالاً موسيقية خالدة. تشمل مؤلفاته للأوركسترا تسعة سيمفونيات وخمس مقطوعات موسيقية على البيانو ومقطوعة على الكمان. كما ألّف العديد من المقطوعات الموسيقية كمقدمات للأوبرا. بدأ بيتهوفن يفقد سمعه في الثلاثينيات من عمره إلا أن ذلك لم يؤثر على إنتاجه الذي ازداد في تلك الفترة وتميز بالإبداع. شهدت مدينة بون الألمانية ميلاد الفنان العبقري لودفج فان بيتهوفن في 16 ديسمبر عام 1770، فنشرت أولى أعماله وهو في الثانية عشرمن عمره عام 1783 م. ثم زاد إنتاجه وذاع صيته كمؤلف موسيقي. عانى بيتهوفن كثيراً في حياته، فبالرغم من أن أباه هو معلمه الأول الذي وجه اهتمامه للموسيقى ولقنه العزف على البيانو والكمان، كما أن والدته توفيت وهو في السابعة عشر من عمره بعد صراع طويل مع المرض، فهل كان التأليف الموسيقي هو نوع من أنواع العلاج والتغلب على المشاكل بالنسبة لبيتهوفن. فقد أرسله حاكم بون إلى فيينا، وهناك تتلمذ على يد هايدن. وعندما سافر هايدن إلى لندن، تحول بيتهوفن إلى معلمين آخرين مثل ساليري وشينك وألبريشتبيرجر. وقد أسهمت كل هذه الدروس والاحتكاكات في تكوين شخصية بيتهوفن الفنية. وحاول أن يشق لنفسه طريق كعازف في عاصمة الموسيقى، غناه هو أعماله الفنية المتميزة. فقد جاء إنتاجه الفني غزيراً حتى بعد إصابته بالصمم. بدأت إصابة بيتهوفن بصمم بسيط عام 1802، امتنع عن العزف في الحفلات العامة، وابتعد عن الحياة الاجتماعية واتجه للوحدة، وبالفعل مازالت أعماله حتى اليوم من أهم ما أنتجته الموسيقى الكلاسيكية العالمية. فجاءت رسالته إلى العالم كل البشر سيصبحون إخوة

الشكل (6) الملخص باستخدام لخصلي Lakhasly (الدقة 0.4 والاستذكار 0.5 ومعامل F 0.44)

الجدول (1) جمل التلخيص باعتماد الأنطولوجيا مع قيم أوزانها ونسب خصائصها

الجملة	وزن الجملة	موقع الجملة	التشابه مع العنوان	العبارات المفتاحية	طول الجملة	العبارات البارزة	العبارات الضعيفة
لودفيج فان بيتهوفن مؤلف موسيقي ألماني ولد عام 1770 م في مدينة بون	3.84	2.00	0.92	0.15	0.77	0.00	1.00
يعتبر من أبرز عباقرة الموسيقى في جميع العصور	2.49	2.00	0.08	0.14	0.27	0.00	1.00
وأبدع أعمالاً موسيقية خالدة	1.42	1.00	0.01	0.14	0.27	0.00	1.00
له الفضل الأعظم في تطوير الموسيقى الكلاسيكية	1.45	1.00	0.01	0.13	0.31	0.00	1.00
تشمل مؤلفاته للأوركسترا تسعة سيمفونيات وخمس مقطوعات موسيقية على البيانو ومقطوعة على الكمان	0.83	0.00	0.03	0.11	0.69	0.00	1.00
كما ألّف العديد من المقطوعات الموسيقية كمقدمات للأوبرا	0.59	0.00	0.01	0.20	0.38	0.00	1.00
بدأ بيتهوفن يفقد سمعه في الثلاثينيات من عمره إلا أن ذلك لم يؤثر على إنتاجه الذي ازداد في تلك الفترة وتميز بالإبداع	0.86	0.00	0.01	0.27	0.58	0.00	1.00
شهدت مدينة بون الألمانية ميلاد الفنان العبقري لودفج فان بيتهوفن في 16 ديسمبر عام 1770	2.94	0.00	1.86	0.08	1.00	0.00	1.00
فنشرت أولى أعماله وهو في الثانية عشرمن عمره عام 1783 م	0.64	0.00	0.03	0.07	0.54	0.00	1.00
ثم زاد إنتاجه وذاع صيته كمؤلف موسيقي	0.65	0.00	0.14	0.09	0.42	0.00	1.00
فبالرغم من أن أباه هو معلمه الأول الذي وجه اهتمامه للموسيقى ولقنه العزف على البيانو والكمان	0.72	0.00	0.01	0.06	0.65	0.00	1.00
عاصمة الموسيقى في ذلك العصر	0.65	0.00	0.00	0.50	0.15	0.00	1.00
فهل كان التأليف الموسيقي هو نوع من أنواع العلاج والتغلب على المشاكل بالنسبة لبيتهوفن	0.64	0.00	0.03	0.07	0.54	0.00	1.00
فقد أرسله حاكم بون إلى فيينا	0.76	0.00	0.03	0.50	0.23	0.00	1.00
وهناك تتلمذ على يد هايدن ولكن بيتهوفن	1.08	0.00	0.00	1.00	0.08	0.00	1.00
وقد أسهمت كل هذه الدروس والاحتكاكات في تكوين شخصية بيتهوفن الفنية	1.18	0.00	0.03	0.73	0.42	0.00	1.00
غناه هو أعماله الفنية المتميزة	0.56	0.00	0.00	0.33	0.23	0.00	1.00
بدأت إصابة بيتهوفن بصمم بسيط عام 1802	0.65	0.00	0.07	0.08	0.50	0.00	1.00
فبدأ في الانسحاب من الأوساط الفنية تدريجياً	0.73	0.00	0.00	0.50	0.23	0.00	1.00
وأمضى حياته بلا زواج يرتبط بعلاقات عدة مع سيدات صغيرات	0.68	0.00	0.00	0.06	0.62	0.00	1.00
إلا أنه لم يتوقف عن الإنتاج الفني	0.65	0.00	0.00	0.50	0.15	0.00	1.00
امتنع عن العزف في الحفلات العامة	0.68	0.00	0.03	0.50	0.15	0.00	1.00
وابتعد عن الحياة الاجتماعية واتجه للوحدة وقلت مؤلفاته	0.69	0.00	0.00	0.38	0.31	0.00	1.00
وأصبحت أكثر تعقيداً	0.58	0.00	0.00	0.50	0.08	0.00	1.00
وبالفعل مازالت أعماله حتى اليوم من أهم ما أنتجته الموسيقى الكلاسيكية العالمية	1.62	0.00	0.04	0.08	0.50	1.00	1.00

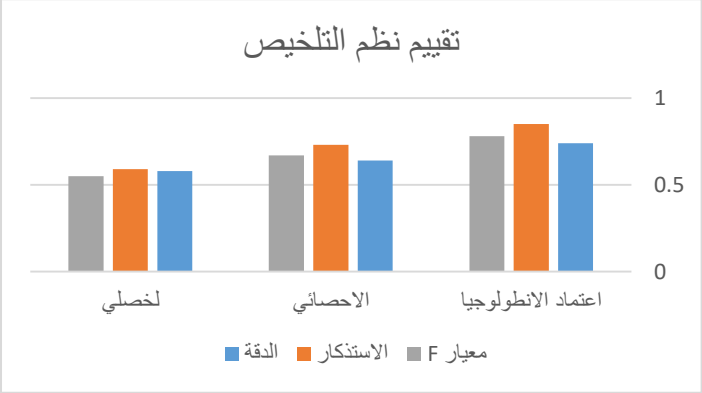
وهما السيمفونية الخامسة والتاسعة	0.79	0.00	0.00	0.60	0.19	0.00	1.00
كما أنه أحدث الكثير من التغييرات في الموسيقى	0.59	0.00	0.00	0.40	0.19	0.00	1.00

وبعد دراسة وتحليل نتائج التقييم على نصوص الملخصات المرجعية تبين أن متوسط الدقة والاستذكار ومعامل F في التلخيص باعتماد الأنطولوجيا المقترح هي: 0.74 و 0.85 و 0.78 على الترتيب. وكانت في التلخيص الإحصائي المقترح هي: 0.64 و 0.73 و 0.67 على الترتيب. وفي نظام تلخيص لخصلي هي: 0.58 و 0.59 و 0.55 على الترتيب. وهي تتفاوت حسب طول المستند.

وقد أوضحت النتائج كما هو مبين في الجدول (2) أن الدقة والاستذكار ومعامل F في التلخيص المعتمد على الأنطولوجيا أفضل من النظام الإحصائي المقارن معه وتتفوق عليه بنسب 0.1 و 0.12 و0.11 على الترتيب، وتفوق أيضاً على نتائج نظام لخصلي بنسب 0.16 و 0.26 و 0.23 على الترتيب. وتبين أيضاً أن النظامين المقترحين في هذا المشروع أفضل من نظام لخصلي، إلا أن نظام التلخيص المعتمد على الأنطولوجيا هو الأفضل. وهذا موضح في المخطط البياني في الشكل (7). باختصار، نظام التلخيص باعتماد الأنطولوجيا المقترح متفوق من حيث الدقة والاستذكار ومعامل F على طرق التلخيص الأخرى المقارن معها.

الجدول (2) مقارنة مقاييس الدقة والاسذكار والمعامل F لأنظمة التلخيص الثلاثة (التلخيص باعتماد الأنطولوجيا، والتلخيص الإحصائي، وتلخيص نظام لخصلي)

	الدقة	الاستذكار	معيار F
اعتماد الأنطولوجيا	0.74	0.85	0.78
الإحصائي	0.64	0.73	0.67
لخصلي	0.58	0.59	0.55



الشكل (7) مقارنة نتائج تقييم التلخيص باعتماد الأنطولوجيا مع التلخيص الإحصائي وتلخيص لخصلي.

الخاتمة:

يسهم مشروع التلخيص الآلي المقدم في هذا البحث إلى تحسين المحتوى الرقمي العربي على الانترنت وخدمة القارئ العربي واللغة العربية، وذلك من خلال استخدام الملخص الآلي المقترح لإيجاد ملخصات للنصوص العربية متناسقة وذات جودة مناسبة. وهو يفيد في تحسين عمليات استرجاع المعلومات بتوفير الوقت والجهد في الحصول على المعلومات الهامة. كما إن إيجاد أنطولوجيا عربية معتمدة على مصادر معرفية عربية غنية بكل جوانب المعرفة الإنسانية مثل الموسوعة العربية بتوسيع شبكة الكلمات العربية عمل هام ويسهم في دعم تطبيقات التي تعتمد معالجة اللغة العربية مثل الوب الدلالي والترجمة الآلية والتلخيص الآلي وغيرها.

الآفاق المستقبلية:

يمكن أن يستكمل نظام التلخيص باعتماد الأنطولوجيا المقترح في هذا البحث مستقبلاً بالتحول إلى اعتماد التلخيص التجريدي بدلاً من الاستخراجي بالاستفادة من التمثيل الدلالي لجمل المستند بالاستعانة بالأنطولوجيا العربية، كما يمكن أن تبحث الدراسات المستقبلية أيضاً بتحسين التلخيص المقترح بإلغاء التكرار باستخدام طرق تجميع الجمل المتشابهة وإضافة خصائص أخرى تسهم في تحسين أوزان الجمل التي تعكس أهمية تضمينها في الملخص النهائي.

Reference

Abdelkrime A, Djamel Eddine Z, Khaled Walid H. "Allsummarizer system at multiling multilingual single and multi-document summarization." *SIGDIAL 2015 Conference*. 2015. 237–244.

Abuobieda A., Salim N., Albaham A., Osman A., Kumar Y. "Text summarization features selection method using pseudo genetic-based model." *International conference on information retrieval knowledge management.* 2012. 193–197.

Al Maleh M., Desouki S.,. "Arabic text summarization using deep learning approach." *Journal of Big Data 7: 109*, 2020.

Al-Abdallah R.Z., Al-Taani A.T. "Arabic single-document text summarization using particle swarm optimization algorithm." *Procedia Computer Science 117*, 2017: 30–37.

Alami, N., Meknassi, M., Ouatik, S.A.,. "Ennahnahi N. Impact of stemming on arabic text summarization." *4th IEEE International Colloquium on Information Science and Technology (CiSt). IEEE.* 2016.

AL-Khawaldeh F, Samawi V. "Lexical cohesion and entailment based segmentation for arabic text summarization (lceas)." *World Comput Sci Inf Technol J 5(3):51–60*, 2015.

Al-Saleh A., Menail M. "Automatic arabic text summarization: a survey. Artif." *Intell. Rev. Arch. 45 (2)*, 2016: 203–234.

Al-Taani A.T., Maha M. "An extractive graph based Arabic text summarization approach." 2014.

Althobaiti M., Kruschwitz U., Poesio M. "Aranlp: A java-based library for the processing of arabic text." 2014.

AlZahrani A., Mathkour H., Ismail Abdalla H.. "Pso-based feature selection for Arabic text summarization." *J. UCS, 21(11)*, 2015: 1454–1469.

Ayedh A., Tan G., Alwesabi K., Rajeh H. "The effect of preprocessing on arabic document categorization.." *Algorithms 9 (2).*, 2016.

Azmi AM, Al-Thanyyan S. "A text summarizer for arabic." *Comput Speech Lang 26(4):260–273*, 2012.

Barzilay R., Elhadad, M. "Using lexical chains for text summarization." *Text Summarization.* 2015. 111–121.

Bawakid A. *Automatic documents summarization using ontology based methodologies.* Doctoral Thesis., University of Birmingham, 2011.

Baxendale P. "Man-made index for technical literature - an experiment." *IBM Journal of Research and Development*, 1958: vol. 2, no. 4.

Belguith L, Ellouze M. Maaloul M. Jaoua M. Jaoua F. Blache P. "Automatic summarization. In: Zitouni I (ed) Natural language processing of semitic languages, theory and applications of natural language processing." *Springer, Berlin*, 2014: 371–408.

Belkebir R., Guessoum A. "A supervised approach to arabic text summarization using adaboost.In:Rocha A, Correia AM, Costanzo S, Reis LP (eds) New contributions in information systems and technologies." *advances in intelligent systems and computing*, 2015: vol 353,227–236.

Edmundson H. P. "New Methods in Automatic Extracting." *ACM*, 1969: 264–285.

El-Beltagy S., Rafea A. "Kp-miner: a keyphrase extraction system for english and arabic documents.." *Inf Syst 34 (1), 132–144*, 2009.

El-Ghannam F, El-Shishtawy T. "Multi-topic multi-document summarizer." *Int J Comput Sci Inf Technol 5(6):77–90*, 2013.

Elghazaly T., Ibrahim A. "Arabic text summarization using rhetoricalstructure theory." *International Conference on INFOrmatics and Systems.* 2012.

El-Haj M. *Multi-document Arabic Text Summarisation Ph.D. thesis.* University of Essex, 2012.

El-Haj M., Kruschwitz Udo, Fox Ch. "Using mechanical turk to create a corpus of arabic summaries." 2010.

El-Haj M., Rayson P. "Using a keyness metric for single and multi document summarisation." *Proceedings of the MultiLing 2013 workshop on multilingual multi-document summarization.* Sofia, Bulgaria: Association for Computational Linguistics, 2013. 64-71.

El-Haj, M Kruschwitz U Fox. "Exploring clustering for multi-document arabic summarisation." *Salem M, Shaalan K, Oroumchian F, Shakery A, Khelalfa H (eds) Information retrieval technology, lecture notes in computer science, vol 7097. Springer, Berlin*, 2011: 550-561.

El-Shishtawy T.,, El-Ghannam F. "keyphrase based Arabic summarizer (KPAS)." *8th International Conference, IEEE.* informatics and systems (INFOS), 2012.

Fattah MA, Ren F. "Ga, mr, ffnn, pnn and gmm based models for automatic text summarization." *Comput Speech Lang 23(1):126–144*, 2009.

Fejer H., Omar N. "Automatic arabic text summarization using clustering and keyphrase extraction." *2014 International Conference on information technology and multimedia (ICIMU).* 2014. 293–298.

Fernando S., Stevenson M. *Mapping WordNet synsets to Wikipedia articles.* University of Sheffield, 2009.

Froud H, Lachkar A. Ouati kSA. "Arabic text summarization based on latent semantic analysis to enhance arabic documents clustering." *Int J Data Mining Knowl Manag Process (IJDKP) 3(1). doi:10.5121/ijdkp. 2013.3107*, 2013.

Gomaa W., Fahmy A. "A survey of text similarity approaches.." *Int. J. Computer Appl. 68 (13).*, 2013.

Gupta P., Pendluri V., Vats I. "Summarizing text by ranking text units according to shallow linguistic features.." *13th International conference on advanced communication technology,*. 2011. 1620–1625.

Gupta. V., Lehal S. L. "A Survey of Text Summarization Extractive Techniques." *Journal of emerging technologies in web intelligence*, 2010: 258-268.

Haboush A, Momani A. Al-Zoubi M. Tarazi M. "Arabic text summerization model using clustering techniques." *World Comput Sci Inf Technol J 2(2):62–67*, 2012.

Ibrahim A, Elghazaly T. "Rhetorical representation and vector representation in summarizing arabic text. Natural language processing and information systems." *lecture notes in computer science, vol 7934. Springer, Berlin*, 2013: 421–424.

Imam I, Hamouda A. Khalek HAA. "An ontology-based summarization system for arabic documents (ossad)." *Int J Comput Appl 74(17):38–43*, 2013.

John P. M. *Mapping WordNet Instances to Wikipedia.* Ireland: Insight Centre for Data Analytics National University of Ireland Galway, 2017.

Kanan R.A.S.G., Jaam J., Hailat E.,. "Stop-word removal algorithm for arabic language." *International Conference on Information and Communication Technologies: From Theory to Applications. IEEE.* 2004.

Keskes I. *Discourse analysis of arabic documents and application to automatic summarization.* Doctoral Thesis, University of Toulouse III-Paul Sabatier, 2015.

Lagrin S., Redjimi M., Azizi N. "Automatic arabic text summarization approaches." *Int. J. Computer Appl. 164 (5).*, 2017.

Lakshmi S., Deepthi K., Suresh C.,. "Text summarization basing on font and cue-phrase feature for a single document. Emerging ICT for Bridging the Future." *49th Annual Convention of the Computer Society of India CSI, vol. 2.* 2015. 537–542.

Litvak M., Vanetik N., Last M., Churkin E. "A multilingual text summarization tool. p.." *Museec:*, 2016: 73–78.

Lloret E., Ferrández O. Muñoz R. Palomar M. "A Text Summarization Approach Under the Influence of Textual Entailment." *5th International Workshop on Natural Language Processing and Cognitive Science (NLPCS 2008).* 2008. 22-31.

Maaloul M.H. *Approche Hybride pour le Résumé Automatique de Textes Application à la langue arabe.* Doctoral Thesis, France: Université de Provence-Aix-Marseille, 2012.

Meena Y., Gopalani D. (b). "Efficient voting-based extractive automatic text summarization using prominent feature set." *IETE J. Res. 62 (5).*, 2016.

Meena Y., Gopalani D. "Analysis of sentence scoring methods for extractive automatic text summarization." *International Conference on Information and Communication Technology for Competitive Strategies vol. 53.* 2014.

Moawad F., Aref M. "A Semantic Graph Reduction Approach for Abstractive Text Summarization." *The 2012 International Conference on Computer Engineering & Systems (ICCES'2012).* Cairo, Egypt, 2012. 27-29.

Mustafa M., Salah Eldeen A., Bani Ahmad S., Elfaki A. "A comparative survey on arabic stemming: approaches and challenges.." *Intell. Inform. Manage. 9 (2)*, 2017.

Nazari N., Mahdavi M. A. "A survey on Automatic Text Summarization." *Journal of AI and Data Mining*, 2019: 121-135.

Nobata C., Sekine S., Murata M., Uchimoto K., Utiyama M., Isahara H. "Sentence extraction system assembling multiple evidence." *Second NTCIR Workshop Meeting,.* 2009. p.2001.

Oufaida H, Nouali O. Blache P. "Minimum redundancy and maximum relevance for single and multidocument arabic text summarization." *J King Saud Univ Comput Inf Sci 26(4):450–461 special Issue on Arabic NLP*, 2014.

Baxendale P., "Machine-made index for technical literature: An experiment." *IBM J. Res. Dev. 2 (4), 354–361*, 1958.

Qaroush A., Abu Farha I., Ghanem W., Washaha M., Maali E. "An efficient single document Arabic text summarization using a combination of statistical and semantic features Journal of King Saud University – Computer and Information Scien." *Journal of King Saud University – Computer and Information Sciences 33*, 2021: 677-692.

Radev, Hovy E. McKeown K. "Introduction to the special issue on summarization." (Comput Linguist 28(4):399–408) 2002.

Salton G. *Automatic text processing: the transformation, analysis and retrieval of information by computer.* Addison-Wesley Longman, 1989.

Steinberger J., Jezek K. "Evaluation measures for text summarization." *Computing and Informatics, vol. 28, no. 2,*, 2009: 1001-1028.

Szymanski J., Korytkowski R. Krawczyk H. "Collaborative Approach to WordNet and Wikipedia Integration." *The Second International Conference on Advanced Collaborative Networks, Systems and Applications.* 2012.

Tanfouri I., Tlik G., JarrayF. " An automatic arabic text summarization system based on genetic algorithms,." *5th International Conference on AI in Computational Linguistics.* 2021. 195-202.

Thomas S., Beutenmüller C., de la Puente X., Remus R., Bordag S. "Exb text summarizer." *SIGDIAL 2015 Conference.* 2015. 260–269.

Turney P. "Learning algorithms for keyphrase extraction." *Inf. Retrieval 2 (4),*, 2000.

Yazan T., Jaradat A. Ahmad Al-Taani. "Hybrid-based Arabic Single-Document Text Summarization Approach Using Genatic Algorithm." *International conference on information and communication systems (ICICS).* 2016.