

# توليد تلخيصات آلية دقيقة للنصوص الإخبارية الطويلة: دراسة في تحسين الكفاءة والدقة

اروى زوكار , اسراء حماديه , دنيا حوري , شيماء الموصلي

## السنة الخامسة , اختصاص الذكاء الصناعي , مشروع مادة معالجة اللغات الطبيعية

### (1) الملخص :

تم تنفيذ الدراسة وفق خطوات منهجية شاملة، بدءاً من المعالجة المسبقة للنصوص، والتي شملت إزالة الضوضاء وتحسين جودة البيانات، مروراً بتجزئة النصوص إلى جمل منفصلة لتسهيل التحليل، وحساب درجات أهمية الجمل باستخدام معايير محددة، وانتهاءً بإنشاء الملخصات عن طريق اختيار الجمل الأكثر أهمية. تم تقييم أداء النماذج المقترحة باستخدام مجموعة بيانات واسعة من النصوص الإخبارية، مع التركيز على قياس دقة التلخيص، والحفاظ على المعلومات الرئيسية، وكفاءة العملية. من المتوقع أن تسهم نتائج هذه الدراسة في تقديم رؤى قيمة حول فعالية النماذج المختارة، مما يساعد على تحسين أنظمة التلخيص الآلي وتوفير إطار عمل يُعتمد عليه في تطبيقات معالجة اللغة الطبيعية المستقبلية. كما تُبرز الدراسة أهمية الاختيار المدروس للنماذج بناءً على طبيعة المهمة والبيانات المتاحة، مما يفتح آفاقاً جديدة للبحث في هذا المجال

### (3) المراجعات الادبية:

تعد دراسة الفرق بين التلخيص الاستخراجي والتلخيص التوليدي خطوة أساسية لفهم آليات التلخيص الآلي وتحديد الطريقة الأنسب لكل تطبيق. في إطار هذه الدراسة، تمت مراجعة عدة دراسات شاملة حول التلخيص الاستخراجي، الذي يعتمد على اختيار الجمل الأكثر أهمية من النصوص الأصلية، دون تغيير في صياغة النصوص. والتلخيص التوليدي، الذي بدوره يركز على فهم النصوص وإعادة صياغة المعلومات لإنتاج ملخصات مترابطة ودقيقة من الناحية الدلالية. حسب ما ورد في دراسة ( Kimani, M, Kaur, G, & Mhd, M., 2024 ) حيث تناولت الدراسة آليات التلخيص التوليدي حيث يتم التركيز على إعادة صياغة النصوص باستخدام تقنيات معالجة اللغة الطبيعية المتقدمة. يتم تنفيذ العملية عبر مراحل متعددة تهدف إلى استخراج الكيانات المهمة وتحليل العلاقات بينها لتوليد ملخص مترابط ودقيق دلاليًا. تبدأ العملية بتنظيف النصوص وتقسيمها إلى كلمات باستخدام تقنيات التحليل اللغوي Tokenization، تليها استخراج الكيانات المسماة (NER) مثل الأشخاص، الأماكن، والتواريخ باستخدام أداة Stanford NER. بعد ذلك، يتم تحليل التبعية بين الكلمات والجمل باستخدام محلل Stanford لتحديد العلاقات

مع التزايد الهائل في حجم النصوص الإخبارية باللغة الإنجليزية، برزت الحاجة إلى أنظمة تلخيص آلية دقيقة وكفؤة. تهدف هذه الدراسة إلى مقارنة فعالية أسلوب التلخيص الخاضع للإشراف وغير الخاضع للإشراف في تحقيق التلخيص الاستخراجي للأخبار باستخدام تقنيات معالجة اللغة الطبيعية (NLP). تم إجراء التجارب على مجموعة بيانات واسعة من النصوص الإخبارية، حيث تم تقييم أداء النموذجين بناءً على معايير مثل دقة التلخيص، والحفاظ على المعلومات الرئيسية، وكفاءة العملية. أظهرت النتائج اختلافات ملحوظة بين الطريقتين، مما يبرز نقاط القوة والضعف لكل منهما. تقدم الدراسة رؤى مهمة لدعم اختيار الأسلوب الأنسب في تطبيقات التلخيص الآلي المختلفة كما اوضحت الدراسة تفوق التلخيص الخاضع للإشراف في التلخيص الاستخراجي.

### (2) المقدمة :

في ظل التزايد الهائل في حجم النصوص الإخبارية المنشورة باللغة الإنجليزية، أصبح من الصعب على القارئ العادي استخلاص المعلومات الأساسية من هذه الكميات الكبيرة من البيانات. هذا التحدي أدى إلى تنامي الحاجة إلى تطوير أنظمة تلخيص آلية قادرة على تقليل الوقت والجهد اللازمين لفهم المحتوى الإخباري بكفاءة. تُعد معالجة اللغة الطبيعية (NLP) أحد المجالات الواعدة التي توفر حلولاً تقنية متقدمة لمعالجة هذه المشكلة، حيث تعتمد على تطوير نماذج ذكية قادرة على تلخيص النصوص بشكل تلقائي. على الرغم من التقدم الكبير في تقنيات التلخيص الآلي، لا يزال هناك جدل حول فعالية الأساليب المختلفة المستخدمة في بناء هذه النماذج، وخاصةً بين أساليب التلخيص غير الخاضع للإشراف والتلخيص الخاضع للإشراف. في هذه الدراسة، تم اختيار أفضل النماذج من كلا النهجين لتقديم مقارنة شاملة ودقيقة. من جانب التلخيص غير الخاضع للإشراف، تم تطبيق نماذج رائدة مثل TextRank وLSA، والتي تعتمد على تحليل النصوص دون الحاجة إلى بيانات تدريب مسبقة. ومن جانب التلخيص الخاضع للإشراف، تم تطوير نموذج متقدم باستخدام أحدث التقنيات، حيث تم تدريبه على مجموعة بيانات معلمة بعناية لضمان دقة وفعالية التلخيص.

## الميزات الإحصائية Statistical Features

تكرار الكلمات - TF Term Frequency المتكررة تشير إلى أهمية الجملة.

معكوس تكرار المستند - Inverse Document Frequency

IDF الكلمات النادرة تُعد أكثر أهمية.

TF-IDF تقيس أهمية الكلمات من خلال الجمع بين التكرار TF

ومعكوس تكرار المستند IDF، مما يُركز على الكلمات ذات

الأهمية العالية ويقلل من تأثير الكلمات الشائعة.

حقبة الكلمات: تمثل الجمل كصفوف والتكرار كأعمدة.

البيانات الرقمية : تعتمد على حساب العدد الإجمالي للقيم

الرقمية في المستند، مما يعكس أهميتها الإحصائية.

الكلمات الموضوعية : تُعد الكلمات المتخصصة ذات الصلة

بمجال محدد أكثر أهمية.

الميزات الدلالية :

تضمين الكلمات : تُستخدم لتحديد معاني الكلمات وسياقها

باستخدام أدوات مثل Word2Vec و Gensim.

تشابه جيب التمام : يُستخدم لقياس التشابه بين الجمل بناءً على

متجهاتها، حيث يشير التشابه الأعلى إلى أهمية الجمل.

تشابه الموضوع : يقارن بين موضوع العنوان وموضوع الجمل

لتحديد الأهمية.

عبارات الإشارة : تُعتبر الجمل التي تحتوي على عبارات إرشادية في

البداية أكثر أهمية.

كما اشار الباحثون (Gupta et al (2024) إلى أنه يمكن أن يتم

التلخيص الاستخراجي بالتعلم الغير خاضع للإشراف بواسطة

خوارزميات مختلفة منها text Rank التي تمثل النص كرسوم

بيانية Graph Representation حيث تُعتبر كل جملة عقدة في

الرسم البياني، وترسم الروابط بين الجمل بناءً على درجة

التشابه بينها. يُحسب التشابه بين الجمل باستخدام تشابه جيب

التمام Cosine Similarity على تمثيل TF-IDF للجمل. الجمل

ذات التشابه العاليي تحصل على روابط أقوى في الرسم

البياني. تُطبق خوارزمية PageRank لحساب أهمية كل جملة في

الرسم البياني بناءً على قوة الروابط المتصلة بها وأهمية الجمل

المرتبطة. تُرتب الجمل بناءً على الدرجات التي تحصل عليها،

ويتم اختيار الجمل الأعلى أهمية لتكوين الملخص. و LSA والتي

تعتبر من أشهر الطرق الدلالية المستخدمة في التلخيص

الاستخراجي. تعتمد هذه الطريقة على تحليل العلاقات الكامنة

بين الكلمات والجمل في النصوص بناءً على تقنية رياضية تُعرف

ب تحليل القيم المفردة SVD - Value Decomposition

بين الكيانات المستخرجة. يتم استخدام هذه المعلومات لاستخراج الكلمات الرئيسية وتوليد جمل تلخص المعلومات الأساسية.

اما بالنسبة للتلخيص الاستخراجي:

تناولت الورقة البحثية للباحثين (Kimani et al (2024) آليات

التلخيص الاستخراجي، حيث ركزت على الخطوات الثلاث

الأساسية لهذا النهج. تمثيل النص المدخل بشكل وسيط: يتم

تحويل النصوص الخام إلى تمثيل ملائم للخوارزميات

المستخدمة، مثل تقسيم النص إلى كلمات أو جمل، أو تمثيله

كرسوم بيانية. حساب رتب أو درجات الجمل: تُطبق خوارزميات

لتحديد أهمية الجمل بناءً على معايير مثل التكرار TF، التكرار

العكسي للمستند IDF، الطول، الموقع، الأسماء الصحيحة،

وغيرها. إنشاء الملخص: يتم تكوين الملخص من خلال اختيار

الجمل ذات الرتب الأعلى مع إزالة التكرار.

كما استعرضت الدراسة طرق ترتيب الجمل في التلخيص

الاستخراجي، التي تنقسم إلى فئتين:

الطرق الإحصائية: تعتمد على تحليل النصوص بناءً على تكرار

الكلمات، طول الجمل، موقعها، أو تشابهها مع العنوان.

الطرق الدلالية: التي تُعنى بفهم المعنى العام للنصوص، إلا أن

التركيز في هذه الدراسة اقتصر على الطرق الإحصائية.

استعرضت دراسة قام بها الباحثون Gupta, P., Ngam S., &

Singh, R (2024) الميزات الشائعة المستخدمة في التلخيص

الاستخراجي، وركزت على تقسيم هذه الميزات إلى أربعة أنواع

رئيسية، مع إضافة ميزات تفصيلية لكل نوع:

الميزات اللغوية:

طول الجملة: الجمل الأطول غالباً ما تحتوي على معلومات أكثر.

موقع الجملة : الجمل في بداية أو نهاية الفقرات غالباً ما تكون

أكثر أهمية.

الكلمات المفتاحية: الجمل التي تحتوي على كلمات متكررة تُعتبر

ذات أهمية.

عبارات الإشارة: مثل "in conclusions" أو "although".

عبارة الاسم وعبرة الفعل: الجمل التي تحتوي على عبارات اسمية

وفعلية تُعد ذات أهمية أكبر.

الأسماء العلمية : تُحدد أهمية الجمل التي تحتوي على أسماء

علمية باستخدام أدوات تمييز أجزاء الكلام.

Singular حيث يتم تمثيل النص في مصفوفة المصطلحات -  
الجملة Term-Sentence Matrix يتم بناء مصفوفة حيث تمثل  
الصفوف الكلمات (أو المصطلحات) وتمثل الأعمدة الجملة، ويتم  
تعبئتها باستخدام قيم TF-IDF لكل كلمة في كل جملة. يُجرى  
تحليل القيم المفردة SVD

تستخدم القيم المفردة لتحديد الجملة الأكثر ارتباطاً بالمواضيع  
الدلالية الرئيسية. ترتب الجملة بناءً على مدى مساهمتها في  
الموضوعات الأكثر أهمية. يتم اختيار الجملة ذات القيم الأعلى  
لتكوين الملخص، بعد استخراج التلخيص يتبقى لدينا مرحلة  
التقييم ويمكن تقييم مدى جودة التلخيص المستخرج باستخدام

ال  
N-gram matching (ROUGE) وهو مقياس شائع لتقييم جودة  
الملخصات النصية، حيث يتم مقارنتها بملخصات مرجعية Gold  
Standard من خلال مطابقة الوحدات النصية. يستخدم  
مجموعة متنوعة من المقاييس لقياس مدى التشابه بين  
الملخص الآلي والملخص البشري .  
استعرضت دراسة (Uğkan, T., Karabulut, K (2024) التلخيص  
الاستخراجي باستخدام تعليم الآلة على عدة نماذج مختلفة  
SVM, logistic Regression, Random Forest ,K  
mean, XGBoost, حيث أظهرت النتائج أن logistic هي صاحبة  
أعلى نتائج باستخدام معيار ROUGE

#### 4 المنهجيات:

يعتمد هذا البحث على تطبيق إطار معالجة اللغة الطبيعية (NLP)  
لتحليل وتطوير طرق التلخيص المختلفة، مع التركيز على المقارنة  
بين التعليم الخاضع للإشراف والتعليم غير الخاضع للإشراف. تم  
استخدام مجموعة بيانات CNN/Daily Mail لتجريب واختبار نماذج  
التلخيص، حيث تحتوي هذه المجموعة على أكثر من 300,000 مقال  
إخباري باللغة الإنجليزية، تتضمن النصوص الإخبارية الأصلية  
والملخصات المكتوبة من قبل المؤلفين.

##### 4.1 مجموعة البيانات

مصدر البيانات: تم الحصول على البيانات من منصة Kaggle.  
حجم البيانات: تبلغ سعة المجموعة حوالي 1.37 جيجابايت،  
وتتضمن:

287,113 مقالاً في مجموعة التدريب.

13,368 مقالاً في مجموعة التحقق

11,490 مقالاً في مجموعة الاختبار.

لغة النصوص ونوعها: النصوص مكتوبة باللغة الإنجليزية

وتتعلق بالمقالات الإخبارية.

بنية البيانات: تحتوي كل عينة بيانات على:  
المقال: النص الإخباري الكامل.  
الملخص: أبرز النقاط الواردة في المقال.  
معرّف المقال: مُعرّف فريد للمقال.

#### 4.2 إجراءات البحث:

تم تقسيم العمل إلى خطوات رئيسية تُطبق على كلتا الطريقتين،  
التعليم الخاضع للإشراف والتعليم غير الخاضع للإشراف، وفق المراحل  
التالية:

##### المرحلة الأولى:

معالجة النصوص:

تشمل هذه المرحلة العمليات الأولية اللازمة لتحضير النصوص  
الخام للتلخيص تم العمل على الخطوات ذاتها الواردة في دراسة  
Gupta et al (2024). والتي تتضمن:

1. تنظيف النصوص:

1. إزالة المسافات الزائدة بين الكلمات.
2. حذف الرموز غير الضرورية.
3. إزالة علامات HTML من النصوص.
4. التخلص من الأحرف غير ASCII.
5. حذف المسافات الموجودة في بداية النصوص أو  
نهايتها.
6. تحويل النصوص إلى حروف صغيرة لتوحيد التنسيق.
7. إزالة الروابط الإلكترونية غير المهمة.
8. حذف الكلمات الشائعة (Stop Words) التي لا تضيف  
معنى دلاليًا.

2. تجزئة النصوص:

تقسيم النصوص إلى جمل باستخدام تقنيات التحليل اللغوي  
(Tokenization)، بحيث يمكن التعامل مع النص على مستوى الجملة  
ككيانات منفصلة.

##### المرحلة الثانية: استخراج الميزات

في هذه المرحلة، يتم استخراج الخصائص الرئيسية التي تساهم  
في عملية التلخيص. تختلف هذه المرحلة بناءً على الطريقة  
المستخدمة في التعليم، سواء كان التعليم خاضعاً للإشراف أو غير  
خاضع للإشراف

##### • التعليم غير الخاضع للإشراف:

في التعليم الغير خاضع للإشراف لا نحتاج إلى ذات الكم الهائل من  
التحليل المسبق للبيانات فلا يوجد لدي نموذج بحاجة الى التدريب  
فطريقة المعالجة هنا مختلفة لذلك قمنا بتطبيق اشهر خوارزميتين  
للتلخيص و تقييم نتائجهما .

TextRank: تعتمد هذه الخوارزمية على تحليل الروابط بين الجمل في النصوص. يتم قياس أهمية الجمل بناءً على ترابطها مع الجمل الأخرى باستخدام تقنية تشبه خوارزمية PageRank. Latent Semantic Analysis (LSA) : يتم تحليل العلاقات الدلالية بين الكلمات والجمل باستخدام تقنيات الجبر الخطي، مما يتيح تحديد الجمل التي تحمل المعنى الأساسي للنصوص .

- **التعليم الخاضع للإشراف:**

تم استخراج الميزات حسب الطرق المذكورة في دراسة Gupta et al (2024)

1. طول الجملة: يتم تقسيم النصوص إلى جمل، ثم يتم قياس طول كل جملة لتحديد أهميتها بناءً على كمية المعلومات التي تحتويها.
2. أول جملة وآخر جملة: يتم اختيار الجمل الأولى والأخيرة من النصوص بناءً على افتراض أن هذه الجمل غالباً ما تحتوي على معلومات رئيسية.
3. الجمل ذات الوزن الأعلى: يتم تحديد الجمل التي تحتوي على أسماء إشارة أو كلمات ذات دلالات مهمة بناءً على وزنها النسبي في النصوص

تمثل هذه المراحل الأساس في إعداد النصوص وتحليلها سواء باستخدام التعليم الخاضع للإشراف أو غير الخاضع للإشراف، بهدف تحقيق تلخيص دقيق وفعال.

#### ● المرحلة الثالثة:

بعد استخراج الميزات تم العمل على Baseline model يعتمد على خوارزمية Logistic Regression. وتم اختيار هذه الخوارزمية بناءً على دراسة Uçkan, T., & Karabulut, K. (2024) حيث يعمل على استخراج الجمل المهمة من مقالات بيانات تدريب باستخدام TF\_IDF، ثم تقييم النموذج على بيانات التحقق والاختبار يتم تقسيم كل مقال إلى جمل باستخدام خوارزمية تقسيم النصوص، مما يسمح بمعالجة الجمل بشكل فردي لكل جملة يتم تحديد ما إذا كانت مهمة (موجودة في الملخص) أو غير مهمة بناءً على المقارنة مع النص الملخص. يقوم النموذج بتقدير مدى أهمية كل جملة، وترتيب الجمل، واختيار أهم الجمل بناءً على النتيجة.

قمنا بدمج استخراج الميزات واعطاء كل ميزة وزن مع ال baseline لنرى هل هذا يعزز من قدرته على تحديد الجمل المهمة بشكل أفضل وبالتالي هل هذا يحسن النتيجة ام لا .

الخطوات التي قمنا بها :

1. إضافة ميزات إضافية :

- موضع الجملة: تمت إضافة ميزتين تشيران إلى موضع الجملة في المقال، وهما: is\_first : إذا كانت الجملة تقع في بداية المقال (قيمة 1) أو لا (قيمة 0). is\_last : إذا كانت الجملة تقع في نهاية المقال (قيمة 1) أو لا (قيمة 0).
  - تمت هذه الإضافة بناءً على دراسة Gupta et al (2024) أن الجمل الواقعة في بداية المقال أو نهايته غالباً ما تكون أكثر أهمية، حيث تميل إلى احتواء معلومات مُلخصة أو استنتاجية.
  - طول الجملة: تم حساب طول الجملة بعدد الكلمات كإحدى الميزات. الجمل الطويلة قد تحتوي على معلومات أكثر تفصيلاً وتكون أكثر أهمية في تشكيل الملخص.
  - عبارات الإشارة : تم تحديد مجموعة من العبارات الإشارية مثل ("however", "in conclusion", "for example") التي تُستخدم عادةً للإشارة إلى جمل مهمة أو انتقالية في النص. تم حساب وزن الجملة بناءً على عدد عبارات الإشارة التي تحتويها
3. تطبيق تحجيم البيانات Standard Scaling:
- تم تطبيق تحجيم البيانات باستخدام StandardScaler لضمان أن جميع الميزات لها نفس المقياس. يُعتبر تحجيم البيانات خطوة مهمة في معالجة البيانات قبل تمريرها إلى خوارزميات التعلم الآلي.
- تم تعديل أوزان كل ميزة للحصول على النتيجة الأمثل كما تم تطبيق الامر ذاته على خوارزمية Random forest التي ذكرت ضمن دراسة Uçkan, T et al(2024) لكن قمنا بإضافة استخراج الميزات الذي قمنا به وتعديل اوزان كل ميزة لنحدد هل هذا يحسن النتيجة أم لا .

#### المرحلة الرابعة:

تقييم أداء النموذج باستخدام مقياس ROUGE لتقييم جودة التلخيص التلقائي. وقد تم اختيار هذا المقياس بناءً على دراسة ROUGE-N: يقيس التشابه بناءً على n-grams نقوم بالتجريب مرة على 1 ومرة على 2 gram

ROUGE-L: يقيس التشابه بناءً على أطول تسلسل فرعي مشترك (Longest Common Subsequence - LCS) بين النصين. هذا المقياس يأخذ ترتيب الكلمات في الحسبان

## 5) النتائج:

في هذا القسم سنقوم باستعراض نتائج التعليم الخاضع للإشراف والغير خاضع للإشراف التي حصلنا عليها .

الغير خاضع للإشراف:

LSA	Text Rank	ROUGE
0.2543	0.2530	ROUGE-1
0.0952	0.1014	ROUGE-2
0.1672	0.2331	ROUGE-L

النتائج في المعيارين الاول والثاني كانتا متقاربتين ولكن في المعيار الثالث كان هنالك فرق كبير بين الخوارزميتين.

الخاضع للإشراف:

Random forest with fixed weight	Logistic Regression with weight	Baseline	ROUGE
0.2471	0.2468	0.3470	ROUGE-1
0.0977	0.0976	0.1604	ROUGE-2
0.1561	0.1558	0.2422	ROUGE-L

## 6) المناقشة:

نلاحظ في التعليم الغير خاضع للإشراف بأن خوارزمية Text Rank تفوق أداءها على خوارزمية ال LSA في المعيارين ROUGE-1 و ROUGE-2 بشكل ملحوظ بينما الفرق في معيار ROUGE-1 ايضا تفوقت ولكن بفارق بسيط كان صغير بينما في دراسة الباحثين (Gupta et al (2024) احرزت ال LSA تقدم واضح على ال TextRank حيث أن ال ROUGE-1 لل LSA 0.40483 بينما لل Text Rank 0.3644 كما نلاحظ في التعلم الخاضع للإشراف أن افضل نتائج كانت نتائج ال base line باستخدام خوارزمية ال logistic Regression كما في دراسة الباحثين (Uçkan, T., & Karabulut, K. (2024) حيث كانت الخوارزمية التي حققت أعلى نتيجة هي ال logistic إلا ان النتائج التي حازو هم عليها كانت أعلى حيث كانت قيمة ال ROUGE-1 0.6647 في حين أننا حصلنا على نتيجة 0.3470 وقد يعود هذا الاختلاف الى اختلاف الداتا سبت المستخدمة

أي أن عمليات استخراج الميزات لم تكن ذات فائدة فالنتيجة بدون هذه العمليات افضل على المعايير الثلاثة

في الختام نلاحظ ان نتائج التعليم الخاضع للإشراف احرز نتائج أعلى من التعليم الغير خاضع للإشراف بفارق كبير .

## المراجع:

- Chen, F., Zuq, H, Zhang, W, & Yu, F. (2020). Keywords Extraction Based on Word2Vec and TextRank. In Proceedings of the 2020 3rd International Conference on Big Data and Education (ICBDE'20), London, United Kingdom April 1-3, 2020. Association for Computing Machinery.
- Gupta, P., Ngam S, & Singh, R (2024). Automatic Extractive Text Summarization using Multiple Linguistic Features. *ACM Transactions on Asian and Low-Resource Language Information Processing*, Article No. ART, ACM DOI: <http://dx.doi.org/10.1145/3656471>
- Kirmani, M, Kaur, G, & Mohd, M (2024). Analysis of Abstractive and Extractive Summarization Methods. *Interdisciplinary Journal of E-Learning and Teaching (i-JET)*, 19, 1.
- Rathore, V. S, & Tavares, J. M R S (2023). *Emerging Trends in Expert Applications and Security - Proceedings of 2nd ICETEAS 2023, Volume 1*. Springer, 1, 253.
- Sreeja, B, Sreeja, R, Suchithra, G, Vaishak Reddy, C, Vaishnav, M, & Shafi, M D NewsZoom (focus on the important): Text summarization of News Articles based on named entity recognition using Spacy library.
- Uçkan, T., & Karabulut, K (2024). The effectiveness of machine learning algorithms in extractive text summarization: A comparative analysis of K-Means, Random Forest, GBM, Logistic Regression, and SVM. *Doğu Fen Bilimleri Dergisi / Journal of Natural & Applied Sciences of East*, 7(2), 77-91
- Wong, K-F, Wu, M, & Li, W (2008, August). Extractive Summarization Using Sentence Contextual Analysis. *The 22nd International Conference on Computational Linguistics (COLING'08)*, Manchester, United Kingdom