

الجمهورية العربية السورية
المعهد العالي للعلوم التطبيقية والتكنولوجيا
قسم النظم المعلوماتية
العام الدراسي 2023/2024

مشروع تخرج

أعد لنيل درجة الإجازة في اختصاص هندسة البرمجيات والذكاء الصنعي

تصميم نظام تحليل محتوى صفحات الويب معتمد على تقنيات التعلم العميق بهدف التصفح غير البصري لصفحات الويب

تقديم

لبنى إحسان الحلبي

إشراف

د. وسيم صافي

م. فراس عمران

ويبقى الفضل في هذا العمل . . .

إلى المثل الأعلى . . . إلى عزتي وقوتي . . . إلى سندي في الحياة وأماني . . . إلى النور الذي يضيء طريقي . . . إلى طمأنينتي . . . إلى من غرس في نفسي حب العلم . . . إلى مرشدني في حكمته . . .

أبي

إلى من بسمتها غايتها . . . إلى من تحت قدميها جنتي . . . إلى من بتضحيتها سهلت دربي . . . إلى من تمنى بجها وحنانها . . . إلى الدفء الذي يفك ضيقتي . . .

أمِي

إلى أغلى الناس على قلبي وسر بسمتي . . . إلى حنونتي . . . إلى بيت سري . . . إلى نور منزلنا . . .
إلى أخوتي . . .

لمى ولين

إلى كل من كان عوناً في دربي . . .

من كان لهm الأثر الأعمق في حياتي.

الشَّكْر

أتقدم بخالص الشكر والتقدير لكل من ساهم بإنجاز هذا العمل، وأخص بالذكر الأستاذة المشرفون: د. وسيم صافي و.م. فراس عمران الدين لولا دعمهما وتوجيهاتهما لما تم إنجاز أي جزء من هذا العمل.
كما أتقدم بالشكر لكل من علمني حرفًا وكان سبباً بوصولي إلى ما أنا عليه.

الخلاصة

صممت صفحات الويب للتفاعل البصري، إذ يعتمد الأشخاص عادة عند استخدام الإنترنت على الحاسة البصرية بشكل رئيسي. يستطيعون من خلال النظر تحديد المحتوى الرئيسي ضمن صفحة الويب ويتمكنون من الوصول إلى هدفهم ضمنها بسرعة عالية. هذا ما جعل الأشخاص ذوي الإعاقة البصرية يعانون عند استخدامهم للإنترنت، ويستهلكون الكثير من الوقت والجهد للوصول للمحتوى الفعلي لصفحة. نقدم من خلال هذا البحث نظاماً للتتصفح غير البصري، يوفر طريقة جديدة للتتصفح الويب، معتمدة على التقاطيع البصري وعنقدة محتويات صفحة الويب ثم استخراج المحتوى وتلخيصه لكل قطاع ضمن الصفحة وتقديمه للمستخدم على شكل صوت، مما يسمح للمستخدمين ذوي الإعاقة البصرية الوصول للمحتوى، كما يوفر أداة لتسهيل تصفح الصفحات ذات المحتوى الكثيف للمستخدمين جميعاً من خلال الاطلاع على تلخيص الفقرات قبل قراءتها بشكل كامل، الأمر الذي يوفر الوقت في الوصول إلى المعلومات المناسبة للمستخدمين، وقد أبدت الخوارزميات المقترنة للتقطيع ونماذج التلخيص أداءً جيداً وفقاً للمعايير المتبعة. وتم بناء تطبيق وب يؤمن عمل النظام ويقدم الملخص من كل قطاع للمستخدم بشكل صوتي.

Abstract

Web pages are designed for visual interaction, as people typically rely primarily on their sense of sight when using the internet. By looking, they can identify the main content on a web page and quickly reach their goal. This makes it challenging for visually impaired individuals to use the internet, as they spend a significant amount of time and effort trying to access the actual content of the page. In this research, we present a non-visual browsing system that offers a new method for web browsing, based on visual segmentation and clustering of web page contents, followed by extracting and summarizing the content of each segment and presenting it to the user in an audible form. This allows visually impaired users to access the content and provides a tool for all users to facilitate browsing of content-heavy pages by previewing summaries of paragraphs before reading them in full, which saves time in accessing relevant information. The proposed segmentation algorithms and summarization models have shown good performance according to established standards. Additionally, a web application has been developed to implement the system and provide users with audible summaries for each segment.

المحتويات

الفصل الأول: مقدمة وتعريف بالمشروع	1
- مقدمة	1.1
- التحديات المتعلقة بالوصول إلى محتوى صفحات الويب	2
- تقطيع صفحات الويب	3
- الهدف من المشروع	4
- تقطيع صفحات الويب	4
- استخراج وتلخيص النصوص	4
- تحليل الصور وتوصيفها	4
- المتطلبات	5
- المتطلبات الوظيفية	5
- المتطلبات غير الوظيفية	6
- الخطة الزمنية	6
الفصل الثاني: الدراسة النظرية	9
- مقدمة	10
- Machine Learning	10
- المنهجيات العامة لتعلم الآلة	1.2.2
- Learning	11
- Clustering	11
- التعلم العميق	13
- Yolo (You Only Look Once)	14
-1.3.2	

14	SAM (Segment Anything Model) -2.3.2
19	SAM2 (Segment Anything in Images and Videos) -3.3.2
Vision-Language Models (VLM) and Pre-training (VLP) with	-4.3.2
20	ViT
21	Large-Language Models (LLM) -5.3.2 - نماذج اللغة الكبيرة
4- معالجة الصورة (Mathematical morphology and Edge detection)	
25	1.4.2 - التحليل الرياضي للأشكال
27	2.4.2 - كشف الحواف
28	الفصل الثالث: دراسة المرجعية
29	1.3 - مقدمة
29	2.3 - الأبحاث المتعلقة بتنقية صفحات الويب
29	1.2.3 - المنهجيات المستخدمة في تنقية صفحات الويب
31	2.2.3 - الطرق التي تم وضعها لتنقية صفحات الويب
37	2.2.3 - المنهجيات المستخدمة في تنقية صفحات الويب بالاعتماد على التعلم العميق
37	3.2.3 -أحدث التقييمات الموجودة للمنهجيات المستخدمة في تنقية صفحات الويب
42	3.3 - الأبحاث المتعلقة بتلخيص النصوص
الفصل الرابع: مجموعات البيانات	
45	1.4 - مقدمة
46	2.4 - مجموعات البيانات المستخدمة لتقسيم الخوارزميات المقترنة
46	1.2.4 - مجموعة بيانات DSpopular
46	2.2.4 - مجموعة بيانات DSrandom

47 4.3 - مجموعات البيانات المستخدمة لتدريب النماذج المستخدمة
47 1.3.4 - مجموعة بيانات COCO: Common Objects in Context
48 2.3.4 - مجموعة بيانات EdinburghNLP/Xsum
48 3.3.4 - مجموعة بيانات abisee/cnn_dailymail
49 الفصل الخامس: تصميم مكونات النظام
50 1.5 - مقدمة
50 2.5 - الخوارزمية المقترنة لتقطيع صفحات الويب باستخدام الطرق التقليدية
56 3.5 - الخوارزميات المقترنة للتنبؤ بعدد القطاعات الموجودة بصفحة الويب
57 1.3.5 - اعتماداً على تحليل المحتوى النصي
60 2.3.5 - اعتماداً على تحليل الصورة
66 4.5 - الخوارزمية المقترنة لتقطيع صفحات الويب باستخدام تقنيات التعلم العميق
68 5.5 - التقنيات المتتبعة لاستخراج المعلومات من قطع صفحات الويب
68 1.5.5 - تلخيص النصوص
70 2.5.5 - التعرف على محتوى قطع الصور الناتجة عن التقطيع
72 الفصل السادس: تنجز النظام
73 1.6 - مقدمة
73 2.6 - بيئة العمل والمكتبات المستخدمة
73 1.2.6 Selenium
74 2.2.6 BeautifulSoup
74 3.2.6 Asyncio
74 4.2.6 Puppeteer

74 TfifdVectorizer -5.2.6
74 NLTK (Natural Language Toolkit) -6.2.6
75 Numpy (Numeric Python) -7.2.6
75 Pandas -8.2.6
75 Matplotlib -9.2.6
75 Visual Studio Code (VS Code) -10.2.6
75 بيئة Google Colab -11.2.6
76 Requests -12.2.6
76 Scikit-learn -13.2.6
76 Torch -14.2.6
76 PIL -15.2.6
77 Scipy -16.2.6
77 Imutils -17.2.6
77 Transformers -18.2.6
78 gTTS (Google Text-to-Speech) -19.2.6
78 Pytesseract -20.2.6
78 SpaCy -21.2.6
79 3.6 التنفيذ العملي
79 1.3.6 - الخوارزمية المقترنة لقطع صفحات الويب باستخدام الطرق التقليدية
86 2.3.6 - الخوارزميات المتبعة للتنبؤ بعدد القطاعات الموجودة بصفحة الويب (اعتماداً على تحليل النصوص)
87 3.3.6 - الخوارزميات المتبعة للتنبؤ بعدد القطاعات الموجودة بصفحة الويب (اعتماداً على تحليل الصورة)

88	4.3.6 - الخوارزمية المقترنة لقطع صفحات الويب باستخدام تقنيات التعلم العميق
96	5.3.6 - التقنيات المتباينة لاستخراج المعلومات من قطع صفحات الويب
103	6.3.6 - قسم الويب.....
106	الفصل السابع: الخاتمة والآفاق المستقبلية
107	1.7 - الخاتمة والفائدة المكتسبة
107	2.7 - الصعوبات
107	3.7 - التوسيع والآفاق المستقبلية.....
108	الملحق آ: اختبارات وتجارب
112	الملحق ب: تفاصيل التحسينات.....
123	المراجع

قائمة الأشكال

الشكل 1-أعداد مستخدمي الإنترن特 في السنوات الأخيرة.	2
الشكل 2- مخطط غانت لتوضيح خطة العمل المتبعة خلال فترة العمل على المشروع.	8
الشكل 3 -مثال على خرج النموذج عند إعطاء مستطيل كتوجيه للقطع.	16
الشكل 4- بعض الصور من مجموعة بيانات تقطيع أي شيء.	18
الشكل 5- ناتج تطبيق النموذج دون أي موجهات (نقط، مربعات).	19
الشكل 6- بنية نموذج BART	24
الشكل 7-مثال على التأكيل والتمدد.	26
الشكل 8-مثال على تحسين التجزئة بالتشابه الدلالي.	32
الشكل 9-خطوات الطريقة المقسمة على مراحلتين.	33
الشكل 10-خوارزمية إعادة تجميع الكتل.	35
الشكل 11- مثال على عملية التقطيع.	36
الشكل 12-مصفوفة الالتباس.	38
الشكل 13 ...	41
الشكل 14	41
الشكل 15	42
الشكل 16-سحابة للألفاظ الأكثر تكراراً.	43
الشكل 17-تكرار الأطوال المختلفة للملخصات.	44
الشكل 18-بنيةمجموعات البيانات.	47
الشكل 19- دمج قطاعين.	52
الشكل 20-بنية صفحة الويب الأساسية.	53
الشكل 21-بنية صفحة الويب بعد المرحلة الأولى.	54
الشكل 22-بنية صفحة الويب بعد المرحلة الثانية.	54
الشكل 23-بنية صفحة الويب بعد المرحلة الثالثة.	55
الشكل 24-بنية صفحة الويب بعد المرحلة الرابعة.	55

الشكل 25- بنية صفحة الويب بعد المرحلة الخامسة.....	55
الشكل 26 -نتيجة التقطيع باختيار عدد أمثل للعناقيد.....	56
الشكل 27-نتيجة التقطيع باختيار عدد أقل من العدد الأمثل للعناقيد.....	56
الشكل 28 -نتيجة التقطيع باختيار عدد أكبر من العدد الأمثل للعناقيد.....	57
الشكل 29-إيجاد المسافة بين الكائن والكائنات الأخرى.....	58
الشكل 30-إيجاد الحواف في الصورة باستخدام خوارزمية canny	61
الشكل 31-إيجاد Contours	62
الشكل 32-إيجاد Centroids في الصورة باستخدام خوارزمية moments	63
الشكل 33- العناقيد الناتجة ومرائزها.....	63
الشكل 34-صفحة الويب والنقط المميزة بها دخل لنموذج sam	64
الشكل 35- مقارنة بين النماذج الأشهر في التلخيص.....	70
الشكل 36- مثال لوصف النموذج المستخدم لصورة.....	71
الشكل 37- مثال للمعلومات المحقونة.....	80
الشكل 38- بنية الرماز المستخدم لتنجيز الخوارزمية.....	81
الشكل 39- ناتج تطبيق الخوارزمية على موقع LightOJ	81
الشكل 40 - ناتج تطبيق الخوارزمية على عدة مواقع.....	82
الشكل 41- ناتج تطبيق الخوارزمية على موقع W3Cs	83
الشكل 42- بنية الرماز المستخدم لتنجيز الخوارزمية.....	85
الشكل 43- عدد القطع المتبع به لمجموعة من الموقع.....	87
الشكل 44- عدد القطع المتبع به لمجموعة من الموقع.....	88
الشكل 45- بنية الرماز المستخدم لتنجيز الخوارزمية.....	91
الشكل 46- خطوات تطبيق الخوارزمية على جزء من صفحة وب.....	92
- بنية الرماز المستخدم لتنجيز الخوارزمية.47 الشكل	95
الشكل 48- بنية الرماز المستخدم لتنجيز الخوارزمية.....	96
الشكل 49- بنية الرماز المستخدم للتطبيق.....	104
الشكل 50- الواجهة الأساسية.....	104

الشكل 51- معلومات إضافية مطلوبة من المستخدم عند اختيار الطريقة التقليدية.....	105
الشكل 52- النتائج النهائية.....	105
الشكل 53- خطوات المعالجة بعد تطبيق النموذج.....	108
الشكل 54-نتيجة تطبيق sam2.....	110
الشكل 55 -مزيد من النتائج.....	111
الشكل 56- هيكل تنسيق الانتباه.....	115
الشكل 57-تكيف مكانی محسن.....	117
الشكل 58-الشبكة الحسنة.....	117
الشكل 59-الاتصالات المتداخلة المضافة.....	117
الشكل 60-دمج المربعات المؤثرة.....	120
الشكل 61-تقنية نقل التعلم.....	122

قائمة الجداول

جدول 1 - المهام المنجزة خلال الفترة الزمنية للمشروع.....	8
جدول 2 - نتائج تقييم الخوارزميات على مجموعة البيانات Popular	40
جدول 3 - نتائج تقييم الخوارزميات على مجموعة البيانات Random	40
جدول 4 - نتائج تقييم الخوارزميات على مجموعة البيانات Snew	40
جدول 5 - محصلة النتائج النهائية Total	41
جدول 6 - نتائج تقييم النماذج المستخدمة.....	43
جدول 7 - نتائج تقييم الخوارزمية المستخدمة على مجموعة بيانات Popular	85
جدول 8 - نتائج تقييم الخوارزمية المستخدمة على مجموعة بيانات Random	85
جدول 9 - نتائج تقييم الخوارزمية المستخدمة على مجموعة بيانات Random	96

الاختصارات

	المصطلح باللغة الإنكليزية	المصطلح باللغة العربية
ALIGN	A Large-scale Image and Noisy-text	صورة واسعة النطاق ونص مشوش
ARI	Adjusted Rand Index	مؤشر راند المعدل
BART	Bidirectional Auto-Regressive Transformer	المحول ثنائي الاتجاه الترجمي التلقائي
BERT	Bidirectional Encoder Representation from Transformers	تمثيل الترميز ثنائي الاتجاه من المحوّلات
BLIP	Bootstrapping Language-Image Pre-training	التمهيد للتدريب المسبق للغة والصورة
BOM	Bag Of Models	حقيقة النماذج
CLIP	Contrastive Language–Image Pretraining	التدريب المسبق التباعي للغة والصورة
CMS	Content Management Systems	نظم إدارة المحتوى
CNN	Convolutional Neural Network	الشبكة العصبية التلافية
COCO	Common Objects in Context	الكائنات الشائعة في السياق
CWI	Centrum Wiskunde & Informatica	مركز الرياضيات والمعلوماتية
DBSCAN	Density-Based Spatial Clustering of Applications with Noise	التجميع المكاني المعتمد على الكثافة للتطبيقات مع ضجيج
DOC	Degree of Coherence	درجة التماسك
DOM	Document Object Model	نموذج كائن مستند
GPT	Generative Pre-trained Transformer	المحول التوليدی المدرب مسبقاً
GLM	Gestalt Layer Merging	دمج الطبقات بأسلوب الجشطالت
GMM	Gaussian Mixture Models	نماذج المزيج الغاوسي
GPU	Graphics Processing Unit	وحدة معالجة الرسوميات
HEPS	Heading-based Page Segmentation	تجزئة الصفحة على أساس العنوان
HTML	Hypertext Markup Language	لغة ترميز النصوص الشعبية
HTTP	Hypertext Transfer Protocol	بروتوكول نقل النص التشعبي
ICT	Information and Communication	تكنولوجيا المعلومات والاتصالات

	المصطلح باللغة الإنكليزية	المصطلح باللغة العربية
IDF	Inverse Document Frequency	معكوس تكرار المستند
ITU	International Telecommunication Union	الاتحاد الدولي للاتصالات
LLM	Large-Language Models	نماذج اللغة الكبيرة
MTV	Model-Template-View	نموذج-عرض- قالب
NER	Named Entity Recognition	التعرف على الكيانات المسماة
NHD	Normalized Hausdorff Distance	مسافة هوسدورف
NMS	Non-Maximum Suppression	قيم القمع غير القصوى
OPTICS	Ordering Points To Identify the Clustering Structure	ترتيب النقاط لتحديد هيكل التجميع
PDOC	Preferred Degree of Coherence	درجة التماسك المفضلة
PEGASUs	Pre-training with Extracted Gap-Sentences for Abstractive Summarization	التدريب المسبق باستخدام جمل الفجوة المستخرجة للتلخيص التجريدي
ROUGE	Recall-Oriented Understudy for Gisting Evaluation	مقياس التذكر الموجه لتقدير التلخيص
SAM	Segment Anything Model	نموذج تقسيم أي شيء
SPP	Spatial Pyramid Pooling	تحميم الهرم المكاني
T5	Text-to-Text Transfer Transformer	محول تحويل النص إلى النص
TF	Term Frequency	تكرار المصطلح
TPU	Tensor Processing Unit	وحدة معالجة الموترات
VIPS	Vision-based Page Segmentation algorithm	تجزئة الصفحة على أساس الرؤية
ViT	Vision Transformer	المحول البصري
VLM	Vision-Language Models	نماذج اللغة والرؤية
VLP	Vision Language Pre-training	التدريب المسبق على اللغة والرؤية
W3c	World Wide Web Consortium	الاتحاد شبكة الويب العالمية
WAI	Web Accessibility Initiative	مبادرة الوصول الى الويب
WCAG	Web Content Accessibility Guideline.	إرشادات الوصول إلى محتوى الويب
YOLO	You Only Look Once	أنت تنظر مرة واحدة فقط

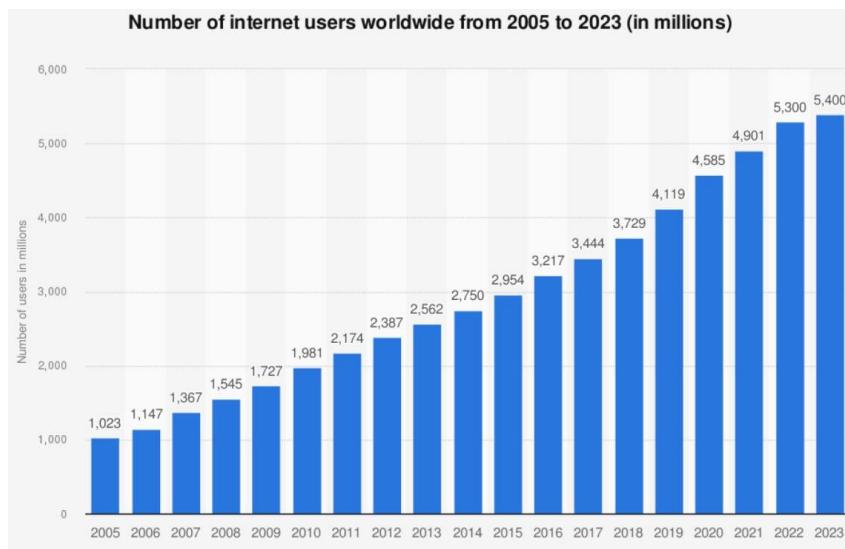
الفصل الأول

مقدمة وتعريف بالمشروع

نقدم في هذا الفصل مقدمة عامة عن المشروع وأبرز النقاط التي دفعتنا للعمل على هذا المشروع، ونبين فكرة المشروع وأهميته والأهداف المرجوة منه والمتطلبات والخطوة الزمنية له.

1.1- مقدمة

في ظل تزايد أهمية الإنترت في حياة الإنسان اليومية وفي مختلف المجالات، أصبحت صفحات الويب مصدرًا أساسياً للمعلومات لعدد متزايد من المستخدمين، فيعد وسيلة للتعلم من خلال الدورات التدريبية الإلكترونية والمحاضرات، كما يعد وسيلة معتمدة من قبل العديد للتسوق عن بعد والشراء من مختلف البلدان في أي وقت دون أي جهد، كما يوفر فرص عمل للأفراد القاطنين في أماكن تتزايد بها البطالة. حيث يمكنهم من إيجاد فرص عمل في شركات بعيدة ويسهل التواصل والتنسيق بين فرق العمل، إضافة لكونه وسيلة معتمدة من قبل الحكومات لتقديم الخدمات للأفراد كتجديد الوثائق الشخصية ودفع الفواتير والضرائب بأقل جهد ممكن، وغيرها من الخدمات التي جعلت منه وسيلة لا يمكن الاستغناء عنها في حياتنا اليومية. كما بينت الدراسة التي أجرتها وكالة الأمم المتحدة المتخصصة في مجال المعلومات والاتصالات ICT (Information and communication technology) [1]، أن أعداد المستخدمين متزايد سنويًا كما هو موضح في الشكل 1، وهي نتيجة طبيعية لاعتماد الناس المتزايد على الإنترت كمصدر رئيسي للمعلومات والخدمات.



الشكل 1-أعداد مستخدمي الإنترت في السنوات الأخيرة.

رغم فوائد الإنترت العديدة وسهولة الوصول إلى كميات هائلة من المعلومات عبر الإنترت، إلا أن طبيعة هذه الفوائد لا تأتي دون تحديات، لعل أبرزها طبيعة المعلومات غير المنظمة والمتنوعة بالإضافة إلى التصاميم المختلفة والعشوانية لصفحات الويب. إذ تنسم معظم صفحات الويب بتصاميم فريدة تتباين بشكل كبير من موقع إلى آخر، مما يجعل من

الصعب تطبيق أسلوب موحد لاستخراج المعلومات منها، الأمر الذي يجعل من الحصول على المعلومات بدقة وفعاليةً أمراً صعباً، وخاصةً من قبل الأفراد الذين يعانون من ضعف في البصر أو فقدان كامل للبصر. إذ يتطلب الحصول على المعلومات من صفحات الويب ذات التصميمات المتباعدة أدوات وتقنيات متقدمة لتسهيل التصفح والوصول إلى المحتوى. هذا ما خلق تحديات عديدة ترتبط بالوصول إلى محتويات صفحات الويب الموجودة.

2.1- التحديات المتعلقة بالوصول إلى محتوى صفحات الويب

وضعت منظمة W3C (World Wide Web Consortium) [2] إرشادات لجعل صفحات الويب متوافرة وسهلة الوصول لنذوي الاحتياجات الخاصة، بما في ذلك الأشخاص ذوي الإعاقة البصرية. من أبرز مبادرات W3C في هذا المجال هي "مبادرة الوصول إلى الويب" (Web Accessibility Initiative)، التي تعمل على إصدار إرشادات الوصول إلى محتوى الويب (WCAG - Web Content Accessibility Guidelines) [3]. هذه الإرشادات توفر توصيات مفصلة لمطوري الويب حول كيفية جعل المحتوى الرقمي أكثر سهولة للوصول. تتضمن WCAG مجموعة من المبادئ مثل توفير بدائل نصية للصور، وتوفير تسميات ذات معنى لغوي للروابط والعناصر التفاعلية، وضمان أن يكون المحتوى قابلاً للوصول من خلال لوحة المفاتيح وحدها، تسعى إلى ضمان أن يكون كل محتوى الويب متاحاً وقابلاً للاستخدام من قبل الجميع، بغض النظر عن القدرات البدنية أو الحسية. بتوفير بدائل نصية للصور، يمكن للمستخدمين المكفوفين فهم محتوى الصور من خلال وصف نصي دقيق. وبإضافة تسميات واضحة للروابط، يمكن لبرامج قراءة الشاشة (screen reader) تقديم وصف دقيق للرابط، مما يسهل على المستخدمين معرفة وجهتهم عند النقر عليه.

إن استخدام برامج قراءة الشاشة قد لا يكون سهلاً دائماً بسبب الصعوبات في تحديد المحتويات الهامة والمتعلقة بالموضوع الرئيسي دون الحاجة لقراءة محتويات كثيرة غير مرتبطة، غالباً ما يمر المستخدمون المكفوفون بكميات كبيرة من المحتوى غير الفعلي قبل أن يجدوا ما يحتاجون إليه، مما يجعلهم يعانون من الحمل الزائد للمعلومات. هذه الصعوبات تتبع من طبيعة صفحات الويب المتنوعة والمعقدة، حيث يجري ترتيب المحتوى بشكل متغير ومعتمد على التصميم البصري بشكل كبير، إذ لا تراعي هذه التقنيات حقيقة أنَّ محتوى الويب منظم ضمن كيانات دلالية. هذا التنظيم يجعل من الصعب على المستخدمين معرفة الكيانات الموجودة في الصفحة وحدودها. من خلال القراءة التسلسلية، تتدخل محتويات الكيانات في بعضها البعض، مما قد يؤدي إلى إرباك المستخدم وعدم قدرته على فهم المحتوى بشكل كامل. بعبارة أخرى، التداخل في القراءة يجعل من الصعب على المستخدمين المكفوفين استيعاب السياق الكامل للصفحة أو التنقل بكفاءة بين الأجزاء المختلفة من المحتوى. بالإضافة إلى ذلك، قد تكون الروابط والعناصر التفاعلية غير موصوفة بشكل كافٍ أو منطقى، مما يعقد عملية التصفح لنذوي الإعاقة البصرية.

كما أنّ تزايد استخدام نظم إدارة المحتوى CMS (Content Management Systems) التي تسمح للمستخدمين بإنشاء وإدارة وتعديل المحتوى الرقمي دون الحاجة إلى معرفة متخصصة مثل البرمجة، من خلال واجهات رسومية سهلة الاستخدام، يتسمى لنظم إدارة المحتوى إنشاء صفحات الويب المطلوبة وتنسيقها بالكامل بدون الحاجة إلى استخدام اللغات المتقدمة مثل HTML (Hypertext Markup Language) و CSS (Cascading Style Sheets) هذا النهج يسهل على الأفراد والشركات إنشاء موقع الويب بسرعة وكفاءة، مما يتيح لهم التركيز بشكل أكبر على تجربة المستخدم وجودة المحتوى بدلاً من التفاصيل الفنية الدقيقة للتصميم والتنفيذ.

في هذا السياق، يأتي البحث والتطوير الحديث في مجال تحليل محتوى صفحات الويب باستخدام تقنيات التعلم العميق، وتحسين التطبيقات التي يمكنها تحليل الصفحة بالكامل، تحديد الكيانات الدلالية، وتقديم ملخصات دقيقة للمحتوى الهام، مما يسهل على المستخدمين وخاصة المكفوفين الوصول إلى المعلومات الضرورية دون المرور عبر كميات كبيرة من المحتوى غير المهم.

3.1- الهدف من المشروع

يهدف هذا المشروع إلى تحسين القدرة على استخراج وتنظيم المحتوى من صفحات الويب باستخدام تقنيات التعلم العميق، مما يعزز دعم نظم التصفح غير البصري. يتمثل هدف المشروع في تحسين تجربة التصفح للأشخاص ذوي الاحتياجات الخاصة من خلال تطبيق خطوات محددة تشمل:

1.3.1- تقطيع صفحات الويب

تحليل وتقطيع صفحات الويب إلى قطع صغيرة ومتابطة تتناسب مع تصميم الصفحة، مما يسهل عملية معالجة المحتوى بشكل أكثر دقة.

2.3.1- استخراج وتلخيص النصوص

استخراج النصوص من كل قطعة وتلخيصها ثم تحويل هذا التلخيص إلى صيغة صوتية، مما يتيح للأشخاص ذوي الإعاقة البصرية الوصول للمعلومات بشكل فعال.

3.3.1- تحليل الصور وتوصيفها

التقاط صور لكل قطعة من الصفحة، ثم استخدام نماذج التعلم العميق لتوصيف هذه الصور، يتم تحويل هذا التوصيف إلى الصيغة الصوتية أيضاً لتمكين المستخدمين من فهم المحتوى البصري.

يسعى هذا المشروع إلى تحسين تجربة تصفح الإنترنت للأفراد خاصةً ذوي الإعاقة البصرية من خلال تقطيع صفحات الويب إلى قطع منظمة وتقديم تلخيصات صوتية، مما يسهل الوصول السريع إلى المعلومات الرئيسية دون الحاجة لتصفح نصوص طويلة وغير ملائمة. يعزز المشروع الشمولية الرقمية يجعل المحتوى الإلكتروني أكثر قابلية للوصول، ويضمن تحقيق المساواة في الوصول إلى المعلومات والخدمات الرقمية لجميع الأفراد، بما في ذلك ذوي الاحتياجات الخاصة. من خلال تطبيق الإرشادات والتقييمات الحديثة في تحليل وتقديم محتوى الويب، يمكننا تحسين تجربة الإنترنت بشكل عام، مما يساهم في تعزيز التفاعل الفعال، ويضمن أن يستفيد جميع فئات المجتمع من الوصول المتساوي إلى المعلومات والخدمات.

4.1- المتطلبات

نسعى في هذا المشروع للحصول على أداة مساعدة نظم التصفح غير البصري من خلال استخراج المعلومات من صفحات الويب بينها المختلفة، تقسم متطلبات المشروع إلى متطلبات وظيفية وغير وظيفية، نجدها فيما يلي:

1.4.1- المتطلبات الوظيفية

يتحقق النظام مجموعة من المتطلبات الوظيفية وهي:

1. دراسة مرجعية للأبحاث المتعلقة بتنقية صفحات الويب بالطرق التقليدية بما في ذلك الخوارزمية المقترنة والتي تعتمد على شجرة DOM (Document Object Model).
2. تقديم تحسينات على الخوارزمية السابقة وخصوصاً فيما يتعلق بعدد القطاعات المطلوب، وذلك لتجنب حالات عدم معرفة المستخدم ببنية صفحة الويب المراد تقطيعها.
3. دراسة مرجعية للأبحاث المتعلقة بتنقية صفحات الويب باستخدام تقنيات التعلم العميق واقتراح خوارزمية تقطيع بالاعتماد على تقنيات التعلم العميق الموجودة.
4. استخراج النصوص الموجودة في كل قطاع.
5. استخراج الصور الناتجة للحصول على معلومات حول المحتوى الصوري في صفحة الويب.

7. تلخيص النصوص الناتجة.
8. تحويل النصوص وملخصاتها وتوصيف الصور إلى الصيغة الصوتية لتمكين المستخدمين ضعاف البصر من الوصول للمعلومات.
9. تغليف الخوارزميات السابقة بتطبيق وب لإجراء المكاملة مع نظم الداعمة لضعف البصر.
10. يتبع تطبيق الويب للمستخدم الاختيار بين الطرق المختلفة لتقطيع صفحات الويب.
11. يتبع تطبيق الويب للمستخدم الاختيار بين طرق التلخيص المختلفة (استخراجي، تحريري).
12. في حال تم اختيار طريقة التقطيع وفق الخوارزمية التي تعتمد على الطرق التقليدية، يقترح النظام للمستخدم عدد القطع المتوقع بعماً للطريقة المعتمدة على تحليل النصوص، ووفقاً للطريقة المعتمدة على تحليل الصورة، ويقدم إمكانية الاختيار بين الرقمين أو ادخال رقم ثالث للمستخدم.
13. في حال تم اختيار طريقة التقطيع وفق الخوارزمية التي تعتمد على الطرق التقليدية، يقدم النظام للمستخدم حرية اختيار driver الذي يستخدمه لاستخراج المعلومات من شجرة ال DOM.

2.4.1- المتطلبات غير الوظيفية

1. القدرة على التعامل مع الصفحات الكثيفة المحتوى وتقطيعها بشكل صحيح والتعامل مع المحتوى الضخم.
2. قابلية التوسيع وإضافة طرق جديدة للتقطيع والتلخيص والتوصيف بسهولة.
3. يجب أن يكون النظام موثوقاً وأمن.
4. يجب أن تنسجم الواجهات بالسهولة والوضوح وأن تكون تفاعلية.
5. يجب أن يكون النظام متوفراً في أي وقت.

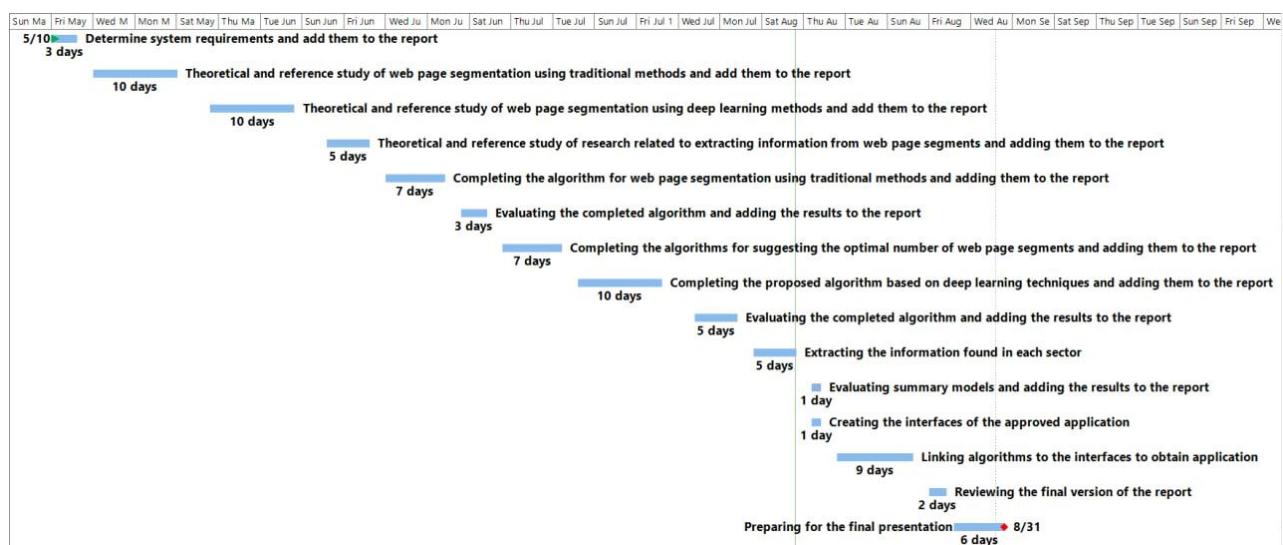
5.1- الخطة الزمنية

يوضح الجدول التالي المهام المنجزة وفق الفترات المرفقة:

المهمة	المدة	تاريخ	تاريخ	نهاية	بداية
	بالأيام				
تحديد متطلبات النظام وإضافتها للتقرير	3	5/10	5/12		
الدراسة النظرية والمرجعية لقطع صفحات الويب بالطرق التقليدية	10	5/15	5/24		
الدراسة النظرية والمرجعية لقطع صفحات الويب باستخدام طرق التعلم العميق وإضافتها للتقرير	10	5/29	6/7		
الدراسة النظرية والمرجعية للأبحاث المتعلقة باستخراج المعلومات من قطع صفحات الويب وإضافتها للتقرير	5	6/12	6/16		
تجزيز خوارزمية تقطيع صفحات الويب بالطرق التقليدية وإضافتها للتقرير	7	6/19	6/25		
تقييم الخوارزمية المنجزة وإضافة النتائج للتقرير	3	6/28	6/30		
تجزيز خوارزمي اقتراح العدد الأمثل لقطع صفحات الويب بالاعتماد على المحتوى النصي وعلى معالجة الصورة وإضافتها للتقرير	7	7/3	7/9		
تجزيز الخوارزمية المقترحة لقطع صفحات الويب بالاعتماد على تقنيات التعلم العميق وإضافتها للتقرير	10	7/12	7/21		
تقييم الخوارزمية المنجزة وإضافة النتائج للتقرير	5	7/26	7/30		
استخراج المعلومات الموجودة في كل قطاع	5	8/2	8/6		
تقييم نماذج التلخيص على قطع صفحات الويب وإضافة النتائج للتقرير	1	8/9	8/9		
إنشاء واجهات التطبيق المعتمد	1	8/9	8/9		

8/20	8/12	9	ربط الخوارزميات المنجزة سابقاً مع الواجهات التي تم الحصول عليها للوصول إلى التطبيق المساعد في نظم التصفح غير البصري وإضافة طريقة التنجيز للتقرير
8/24	8/23	2	تدقيق النسخة النهائية للتقرير
8/31	8/26	6	التحضير للعرض النهائي

كما يمكن عرض الخطة وفق مخطط غانت



الشكل 2 - مخطط غانت لتوضيح خطة العمل المتبعة خلال فترة العمل على المشروع

الفصل الثاني

الدراسة النظرية

نقدم في هذا الفصل تعريفاً للمفاهيم النظرية والتقنيات المعتمدة في هذا المشروع وأبرز تكنيات التعلم الآلي المستخدمة في تقطيع صفحات الويب وأبرز تكنيات التعلم العميق المستخدمة في استخراج المعلومات من الصور وتلخيص النصوص وبعض التعريف والأدبيات حول المصطلحات والمفاهيم التي سترد خلال الدراسة.

1.2 - مقدمة

في هذا الفصل، نقدم تعريفات ومفاهيم أساسية تتعلق بالتقنيات النظرية والعملية التي تُستخدم في مشروعنا. نبدأ بتوسيع أساسيات التعلم الآلي وأبرز تقنياته، بما في ذلك العنقدة، التي تساعد في تنظيم وتصنيف البيانات. ثم نتناول {YOLO(You Only Look Once) ,SAM(Segment Model) مثل Anything Model) ,SAM2 ,VLM (Vision-Language Models) ,LLM (Large-Language Model) {BART (Bidirectional and Auto-Regressive Transformers) ,T5 (Text-to-Text Transfer Transformer) ,Pegasus (Pre-training with Extracted Gap-Sentences for Abstractive Summarization)}، وكيفية استخدامها في معالجة وتقطيع الصور واستخراج المعلومات من النصوص. أخيراً، نستعرض مفهوم معالجة الصور بما في ذلك Mathematical Morphology, Edge Detection، وهي تقنيات تُستخدم لتحليل المكونات في الصورة وتحسين الصورة باتباع أساليب رياضية.

2.2 - تعلم الآلة

تعلم الآلة يمثل فرعاً من فروع الذكاء الصنعي، وهو تقنية تُستخدم لتحليل البيانات بشكل أساسي. يشير التعلم الآلي إلى مجموعة الأدوات والمفاهيم والمنهجيات التي تُستخدم لبرمجة الحواسيب، مما يُمكّنها من استخلاص المعرفة من البيانات المتاحة واستخدامها لتحليل بيانات جديدة بأقل قدر من التدخل البشري.

مع تزايد حجم وتنوع البيانات المتاحة، أصبح تعلم الآلة ضرورياً بشكل متزايد، حيث توفر تقنيات تكلفتها نسبياً منخفضة وقدرات حسابية عالية، مما يساعد على إنشاء نماذج قادرة على تحليل بيانات معقدة وكبيرة، وإنتاج نتائج دقيقة بسرعة كبيرة مع تقليل الأخطاء إلى أدنى حد ممكن.

تعتمد الشركات العديدة التي تعامل مع كميات كبيرة من البيانات على تعلم الآلة لتحسين كفاءتها والحصول على ميزة تنافسية. ويُلاحظ استخدام تعلم الآلة في مجموعة واسعة من المجالات مثل الخدمات المالية، والحكومة، والرعاية الصحية، وقطاعات الطاقة والنقل، مما يعزز الأداء ويسهل استخدام الموارد بشكل فعال.

إذ أن تعلم الآلة يعتبر ركيزة أساسية في تطبيقات الذكاء الصنعي، حيث يمكنه تمثيل وتحليل وفهم البيانات بشكل تلقائي، مما يسهم في تحسين العمليات واتخاذ القرارات الذكية في مختلف المجالات.

1.2.2- المنهجيات العامة لتعلم الآلة

General Methodologies of Machine Learning

يمكن تصنيف منهجيات تعلم الآلة وفقاً لآلية التدريب، ومن أبرز هذه المنهجيات:

1. التعلم بإشراف (Supervised Learning)

يتم تدريب الخوارزمية على بيانات مصنفة، حيث تكون قيم الخرج معروفة. تُستخدم هذه القيم لتوجيه الخوارزمية في تعلم العلاقات بين المدخلات والخرجات. من الأمثلة الشائعة على ذلك مسألة التصنيف[4] .

2. التعلم غير الخاضع لإشراف (Unsupervised Learning)

تتعلم الخوارزمية من بيانات غير مصنفة دون أن تكون قيم الخرج معروفة. تُترك للخوارزمية مهمة تحديد الأنماط والسمات المشتركة بين البيانات. مسألة العنقدة (clustering) هي مثال شائع على هذا النوع [5] .

3. التعلم المشرف عليه جزئياً (Semi-Supervised Learning)

يجمع بين التعلم بإشراف والتعلم غير الخاضع لإشراف، حيث تكون بعض أمثلة التدريب مصنفة والبعض الآخر غير مصنف. يساعد هذا النهج الخوارزمية على التعلم من كمية أقل من البيانات المصنفة.

4. التعلم بالتعزيز (Reinforcement Learning)

يتفاعل النظام مع البيئة المحيطة به، ويتعلم من خلال الحصول على تغذية راجعة (feedback) بناءً على أفعاله. يُستخدم هذا النهج في المهام التي تتطلب اتخاذ قرارات لتحقيق أهداف محددة. [6].

2.2.2- العنقدة

هي واحدة من أهم التقنيات في تعلم الآلة، وخصوصاً في التعلم غير الخاضع للإشراف (Unsupervised Learning) يتمثل المهدى الأساسي من العنقدة في تجميع مجموعة من البيانات بحيث تكون البيانات في نفس المجموعة (العنقود) متباقة إلى حد كبير مع بعضها البعض و مختلفة عن البيانات في المجموعات الأخرى.

1.2.2.2- أشكال العنقدة

1. عنقدة التجزئة (Partitioning Clustering)

تقوم هذه الطريقة بقطيع البيانات إلى مجموعات أو عناقيد غير متداخلة. من أشهر الخوارزميات في هذا النوع:

- خوارزمية k-means تقوم بقطيع البيانات إلى k عنقود، حيث k هو عدد العناقيد المحدد مسبقاً.

تبدأ الخوارزمية بتحديد مراكز أولية للعناقيد (عادةً يتم اختيارها عشوائياً). بعد ذلك، تُخصص كل نقطة بيانات إلى أقرب مركز عنقود (وفقاً لمسافة الإقليلية مثلاً) ثم تقوم بتحديث هذه المراكز من خلال حساب المتوسط الجديد لكل عنقود. بشكل متكرر إلى أن تقارب مراكز العناقيد وتسقى [7].

- خوارزمية k-medoids مشابهة لخوارزمية k-means ولكن بدلاً من استخدام المتوسطات لاختيار مراكز العناقيد، يتم اختيار النقاط التي تمثل مراكز العناقيد بشكل عشوائي من البيانات الفعلية (لا يتم إيجاد نقاط من خارج البيانات) [8].

2. العنقدة الهرمية (Hierarchical Clustering)

تقوم هذه الطريقة ببناء تسلسل هرمي من العناقيد. هناك نوعان رئيسيان:

- الطريقة التراكمية (Agglomerative) تبدأ كل نقطة كعنقود منفصل وتندمج العناقيد بشكل تدريجي حتى تجتمع كل النقاط في عنقود واحد كبير.

- الطريقة التقطعية (Divisive) تبدأ بعنقود واحد يحتوي على كل النقاط ثم تقسم العناقيد بشكل تدريجي حتى تصبح كل نقطة عنقوداً منفصلاً [9].

3. عنقدة قائمة على الكثافة (Density-Based Clustering)

تركز هذه الطرق على تحديد العناقيد على أساس كثافة البيانات. من أشهر الخوارزميات:

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) -
تحدد العناقيد كمناطق ذات الكثافة العالية، وتعتبر النقاط في المناطق ذات الكثافة المنخفضة نقاط شاذة.

OPTICS (Ordering Points to Identify the Clustering Structure) -
ولكنها تعالج بعض مشاكله مثل تحديد العناقيد ذات الأشكال المختلفة والكثافات المختلفة [10].

4. عنقدة قائمة على النموذج (Model-Based Clustering)

تفترض هذه الطرق أن البيانات تنشأ من مزيج من التوزيعات الاحتمالية. تقوم الخوارزميات بتقدير معلمات هذه التوزيعات لتحديد العناقيد. من أشهر الخوارزميات:

Gaussian Mixture Models (GMMs) -
ويُمكن أن تكون مزيج من توزيعات غاوسيّة لكل توزيع وزن معين (Gaussian distribution)
حسب أهميته، وتستخدم لتقدير معلمات التوزيعات المختلفة. [11]

تعد العنقدة من الأدوات المهمة في تحليل البيانات وتعلم الآلة، حيث توفر وسيلة لفهم البنية الكامنة في البيانات بشكل غير موجه. على الرغم من التحديات المرتبطة بها، تظل تقنية ذات قيمة عالية في العديد من المجالات والتطبيقات.

3.2- التعلم العميق Deep Learning

التعلم العميق هو فرع من فروع التعلم الآلي يعتمد على الشبكات العصبية الصناعية ذات الطبقات المتعددة (الشبكات العميقية) لمعالجة البيانات واستخلاص الأنماط منها. يعتبر التعلم العميق تطوراً طبيعياً للتعلم الآلي ويتميز بقدرته على التعلم من كميات هائلة من البيانات بشكل غير مُراقب أو شبه مُراقب.

يعتمد التعلم العميق على الشبكات العصبية العميقية التي تتكون من طبقات متعددة من الخلايا العصبية (neurons) كل طبقة تقوم بتحويل بياناتها إلى شكل أكثر تحريراً قبل تمريرها إلى الطبقة التالية. مع تقدم البيانات عبر الطبقات، تتعلم الشبكة قنوات أكثر تعقيداً وعمقاً للبيانات، مما يسمح لها بفهم الأنماط وال العلاقات المعقدة. ويتميز بقدرته على:

- **التعرف على الأنماط:** التعلم العميق قادر على التعرف على الأنماط المخفية في البيانات، سواء كانت صوراً أو نصوصاً أو إشارات صوتية.

- **التعلم التلقائي:** يمكن للتعلم العميق تحسين أدائه تلقائياً مع مرور الوقت، كلما زادت كمية البيانات المتوفرة له.

- **التعليم:** يساعد التعلم العميق في تعليم المعرفة المكتسبة من البيانات على حالات جديدة، مما يجعل النماذج قادرة على التعامل مع سيناريوهات غير مسبوقة.

- الدقة والفعالية: يوفر التعلم العميق دقة عالية في التنبؤ واتخاذ القرارات، مما يجعله مناسباً لتطبيقات حرجية تتطلب مستوى عالياً من الدقة.

من أحدث الشبكات التي ظهرت في الآونة الأخيرة في مجال تقطيع الصور وصفحات الويب، وفي مجال تلخيص النصوص واستخراج المعلومات من الصور:

Yolo (You Only Look Once) - 1.3.2

هو نموذج شبكة عصبية عميق مخصص لأغراض كشف وتحديد الأجسام في الصور والفيديو. تم تقديم YOLO لأول مرة في عام 2015، ومنذ ذلك الحين، تطورت هذه التقنية بشكل كبير وأصبحت واحدة من أكثر الأساليب شيوعاً وفعالية في هذا المجال. تعد خوارزمية YOLO من الخوارزميات السريعة التي تعمل على تحديد الأغراض في الوقت الفعلي (real time). وهي من خوارزميات الكشف ذات المرحلة الواحدة (تحديد موقع الأجسام وتصنيفها معاً) إذ تستخدم عملية أساسية واحدة فقط أثناء عملية الفحص، مما يمكن أن يقلل بشكل كبير من التعقيد الحسابي واستخدام الذاكرة و يجعلها تعمل بشكل أسرع. على الرغم من أنها ليست دقيقة مثل خوارزمية الكشف على مرحلتين، إلا أنها تساعده على تحقيق طريقة كشف أسرع وأكثر في الوقت الفعلي من خلال تبسيط العملية [12].

SAM (Segment Anything Model) - 2.3.2

تقنية "تقطيع أي شيء" (Segment Anything) هي من الابتكارات الحديثة التي تسعى إلى تحسين أداء مهام تقطيع الصور في مجال رؤية الحاسوب. هذه التقنية تعني فصل الصورة إلى أجزاء متميزة، وتعد خطوة أساسية في تطبيقات عديدة مثل التعرف على الأشياء والتحليل الطبي والرؤية الروبوتية. باستخدام نموذج "تقطيع أي شيء" (SAM)، يمكن تنفيذ تقطيع دقيق وفعال للصور باستخدام مختلف أنواع المطالبات مثل النقاط، المربعات، وحتى النصوص الحرة (النصوص المدخلة من قبل المستخدم كإيجاد القطط الموجودة في الصورة فيكون النص الحر هنا "قطة")، مما يجعلها تقنية متعددة الاستخدامات وقابلة للتطبيق في مجموعة واسعة من التطبيقات.

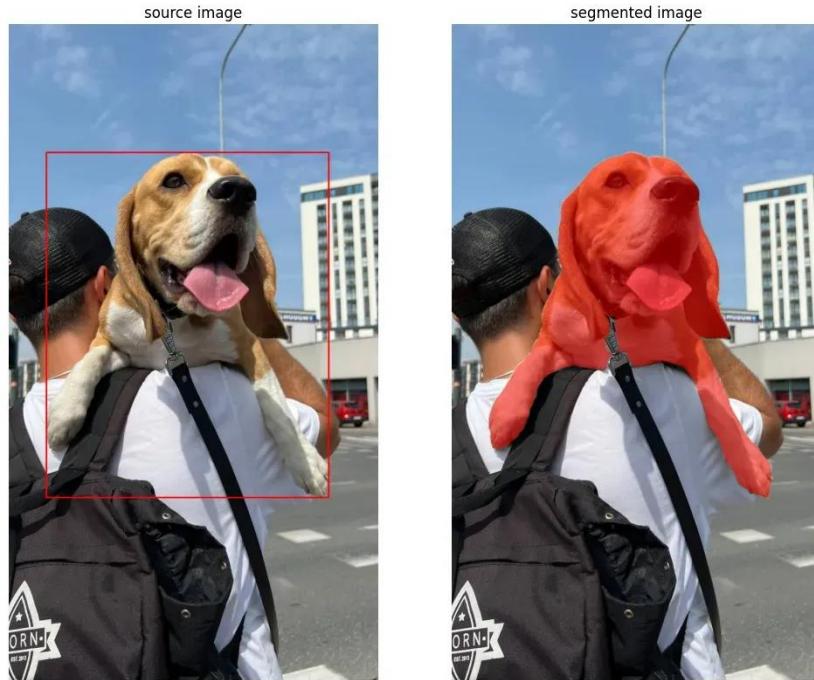
يجدر بالذكر أنَّ النماذج اللغوية الكبيرة التي جرى تدريبيها مسبقاً علىمجموعات بيانات ضخمة أحدثت ثورة في مجال معالجة اللغة الطبيعية بفضل قدرتها على التعلم دون الحاجة إلى تدريب مسبق كبير أو بتدريب قليل. تعرف هذه النماذج باسم "النماذج الأساسية"، وهي تستطيع تعميم المهام والبيانات التي لم تكن معروفة أثناء التدريب، باستخدام تقنيات مثل تصميم النصوص لتحفيز النموذج على إنتاج استجابات نصية صالحة للمهام المختلفة.

تستخدم هذه النماذج الأساسية أيضًا في مجال رؤية الحاسوب، مثل نموذج CLIP (Contrastive Language-Image Pretraining) [13] الذي يربط بين النصوص والصور باستخدام التعلم التباعي (تمييز أوجه التشابه والاختلاف بين عينات البيانات). تمكن هذه النماذج من التعميم على مفاهيم بصرية جديدة باستخدام تصميم النصوص. يتم استخدام هذه النماذج في مهام متنوعة مثل توليد الصور والتعرف على الأشياء.

المدارك من نموذج SAM هو تطوير نموذج أساسى كالنماذج السابقة التي أحدثت ثورة في مجال معالجة اللغة ثم مجال الرؤية الحاسوبية، وذلك لقطع الصور باستخدام نهج شامل يشمل "المهمة، النموذج، والبيانات". يتم بداية العمل بتحديد "مهمة قطع الصور" التي تدعم التعرف بدقة على الحدود الفاصلة بين الكائنات والخلفية، وكذلك التمييز بين الكائنات المختلفة داخل الصورة. بالإضافة إلى ذلك، يتبع على "النموذج" أن يكون قادرًا على استخدام "تصميم النصوص" لإنتاج أقنية قطع الصور بفعالية في الوقت الفعلي. وأخيرًا، يعتمد تدريب النموذج على مجموعة بيانات كبيرة ومتنوعة، ويتم استخدام "محرك البيانات" الذي يعمل على جمع بيانات جديدة أو تحسين جودة البيانات الموجودة مما يحسن من النموذج بشكل مستمر.

• مهمة التجزئة القابلة للتوجيه (Promptable Segmentation Task)

تسعى لتطوير مهمة جديدة لقطع الصور مستوحة من نماذج معالجة اللغة الطبيعية الحديثة. تهدف هذه المهمة، التي نسميها "القطع القابل للتوجيه"، إلى إنتاج قناع تجزئة صالح استجابة لأى مطالبة تجزئة، سواء كانت مجموعة من النقاط، المربعات، أو النصوص الحرة كما هو موضح في الشكل 3. تتيح هذه المرونة تطبيق النموذج على مجموعة واسعة من مهام التجزئة دون الحاجة إلى تدريب مسبق لكل مهمة. يجب أن يكون النموذج قادرًا على إنتاج قناع قطع معقول حتى عندما تكون التوجيهات غامضة. يمكن للمستخدمين ببساطة تقديم مطالبة مناسبة والحصول على نتائج دقة فورًا، مما يسرع العمليات ويقلل الجهد المطلوب. استخدمت هذه المهمة لتدريب النموذج بشكل مسبق وحل مهام قطع الصور المستقبلية باستخدام تقنيات توجيه النصوص.



الشكل 3 - مثال على خرج النموذج عند إعطاء مستطيل كتوجيهه للتقطيع.

• نموذج تقطيع أي شيء(SAM)

نموذج "تقطيع أي شيء" (SAM) هو نموذج جديد مصمم لتحقيق مهمة التقطيع القابل للتوجيه واستخدامه في التطبيقات الواقعية. يتكون SAM من ثلاثة مكونات رئيسية:

1. مشفر الصورة: يحول الصور المدخلة إلى تمثيلات مميرة تُستخدم في أجزاء النموذج الأخرى.
2. مشفر التوجيهات: يحول المطالبات المدخلة (مثل النقاط، الصناديق، النصوص، أو الأقنعة) إلى تمثيلات تتوافق مع تمثيلات الصور.
3. مفكك القناع: وحدة خفيفة الوزن تجمع بين تمثيلات الصورة والتوجيهات لإنتاج أقنعة التجزئة النهائية.

يتميز تصميم SAM بالكفاءة العالية والدعم التفاعلي في الوقت الفعلي، مما يتبع التعامل مع العموم في المطالبات المختلفة وإنتاج أقنعة صالحة متعددة لنفس التوجيه. النموذج مصمم لدعم التوجيهات المرنة، الاستجابة في الوقت الفعلي، والتعامل مع التوجيهات الغامضة بفعالية.

يتيح التصميم المفصول لمشفر الصورة والتوجيهات وفكك القناع إعادة استخدام تضمين الصورة مع توجيهات مختلفة، مما يوفر الوقت والموارد. بمجرد الحصول على تضمين الصورة، يمكن للنموذج توقع قناع التقطيع في حوالي 50 ملي ثانية في متصفح الويب.

يركز SAM على التوجيهات مثل النقاط، الصناديق، والأقنعة، ويقدم نتائج أولية باستخدام التوجيهات النصية الحرة (النصوص المدخلة من قبل المستخدم). بفضل تصميمه البسيط والفعال، يوفر النموذج تجربة تقطيع صور تفاعلية وسريعة باستخدام توجيهات مختلفة، مع القدرة على التعامل مع الغموض من خلال توقع أقنعة متعددة لنفس التوجيه.

• محرك البيانات (Data Engine)

هو استراتيجية متقدمة لتطوير نموذج "تقطيع أي شيء" (SAM) من خلال جمع مجموعة بيانات ضخمة ومتنوعة تضم أكثر من مليار قناع بجزئه على 11 مليون صورة. تتألف هذه الاستراتيجية من ثلاثة مراحل:

1. المرحلة الأولى - المساعدة اليدوية: يقوم المعلقون البشريون بإضافة التعليقات التوضيحية

بمساعدة أدوات مساعدة، مما يشبه أعداد تقطيع تفاعلي تقليدي.

2. المرحلة الثانية - شبه الآوتوماتيكية: يتم استخدام نماذج تعلم الآلة للمساعدة في التعليقات

التوضيحية مع إشراف بشري جزئي. يمكن لـ SAM تلقائياً توليد أقنعة لبعض الكائنات بناءً على توجيهات محددة، بينما يرکز المعلقون على وضع الأقنعة لبقية الكائنات، مما يزيد من تنوع البيانات.

3. المرحلة الثالثة - الآوتوماتيكية بالكامل: تقوم النماذج المدرية بتوليد التعليقات التوضيحية

دون تدخل بشري من خلال توجيه SAM بشبكة منتظمة من النقاط الأمامية، مما يؤدي إلى إنتاج حوالي 100 قناع عالي الجودة لكل صورة في المتوسط.

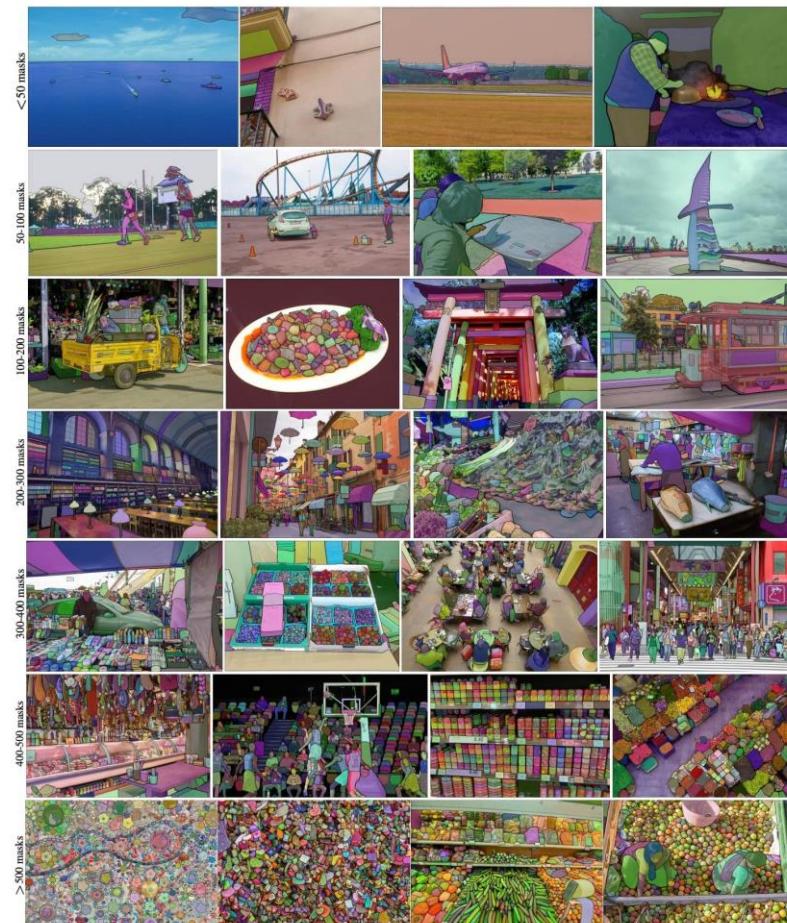
هذا النهج يضمن الحصول على مجموعة بيانات كبيرة ومتنوعة وعالية الجودة، مما يدعم تدريب نماذج SAM لتكون دقيقة وفعالة وقدرة على التعميم على توزيعات البيانات الجديدة بشكل قوي.

• مجموعة بيانات تقطيع أي شيء (SA-1B)

هي نتاج محرك البيانات، تحتوي على أكثر من مليار قناع بجزئه موزعة على 11 مليون صورة مرخصة ومحافظة على الخصوصية. تعد هذه الجموعة الأكبر بمراحل مقارنة بأي مجموعة بيانات أخرى للتجزئة المتاحة حالياً، مما يوفر موارد ضخمة لتدريب وتحسين نماذج تعلم الآلة في مجالات التجزئة المتعددة يوضح الشكل 4 بعض الأمثلة على صور مجموعة البيانات. جرى جمع SA-1B بشكل تلقائي باستخدام أحد التكنولوجيات، وقد

جرى التحقق من جودة وتنوع الأقنعة بشكل مكثف. يطلع الباحثون إلى أن تكون SA-1B مصدرًا مهمًا للبحث المستقبلي في تطوير خاذج تقطيع الصور عالية الأداء والشمولي.

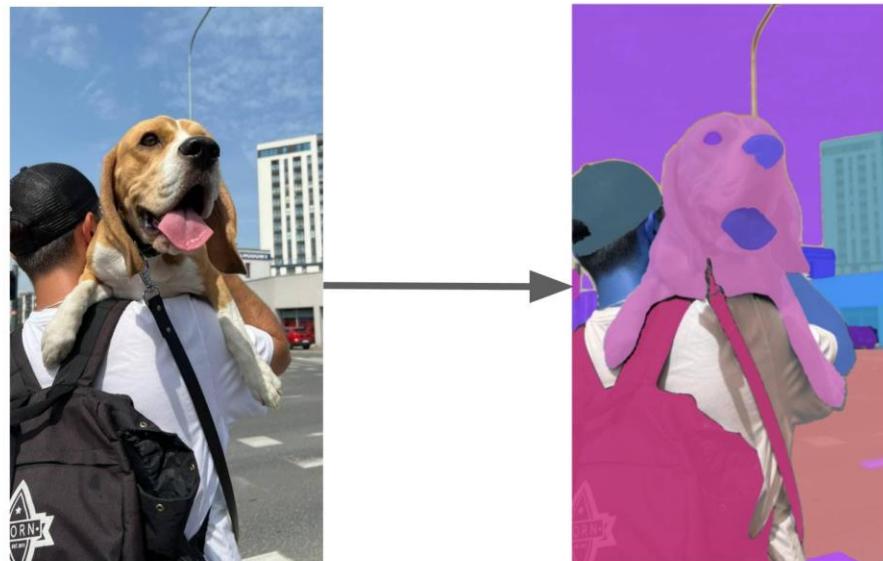
تقنية تقطيع أي شيء (Segment Anything) تمثل نقلة نوعية في مجال معالجة الصور ورؤيه الحاسوب، حيث توفر أدوات قوية ومرنة لتحقيق تجزئة دقيقة وفعالة باستخدام مجموعة متنوعة من المطالبات. من خلال الجمع بين تقنيات التشفير المتقدمة ومحرك البيانات الضخم، يمكن لنموذج SAM تقديم أداء عالي الجودة في مجموعة واسعة من التطبيقات، مما يفتح آفاقاً جديداً للتحسين والابتكار في هذا المجال. [14]



الشكل 4- بعض الصور من مجموعة بيانات تقطيع أي شيء.

SAM2 (Segment Anything in Images and Videos) -3.3.2

وهو نسخة محسنة من نموذج SAM أصدر في نهاية الشهر السابع من سنة 2024، يعمل على التجزئة المرئية القابلة للتوجيه في الصور ومقاطع الفيديو من خلال توفير إمكانية ادخال النقط المميزة من خلال النقرات أو المربعات المحددة مما يعزز تجربة المستخدم، يدمج 2 SAM تجزئة الفيديو والصورة، ويعامل الصور كمقاطع فيديو أحادية الإطار. يوفر SAM 2 المدرب أداءً قويًا عبر مجموعة واسعة من المهام، في تجزئة الصور، يكون نموذج أكثر دقة وأسرع بقدر 6 مرات من نموذج (SAM)، إذ جرى تحقيق تحسينات كبيرة في مجال تجزئة الصور من خلال الاستفادة من البنية المتقدمة ومجموعة البيانات التدريبية الأكبر، وهذا يؤدي إلى تحديد أكثر دقة للأشياء داخل الصور، إذ تدمج النماذج المتقدمة المعلومات السياقية من وحدات البكسل والإطارات الحبيطة، مما يحسن فهم حدود الأشياء وال العلاقات داخل المشاهد المعقدة، إضافة إلى أنه يمكن تعديمه عبر مجالات مختلفة على سبيل المثال، التصوير الطبي، وصور الأقمار الصناعية، والمشاهد اليومية، تساهم هذه التحسينات بشكل جماعي في التطور المستمر لجزئية الصور، مما يجعلها مكوناً أساسياً في التطورات والتطبيقات التكنولوجية المختلفة، حيث يوفر هذا النموذج قدرات محسنة لقطع أي كائن في الوقت الفعلي عبر كل من الصور ومقاطع الفيديو[15]، (تم تطبيق النموذج على عدد من صفحات الويب للتحقق من إمكانية الاستفادة منه في التنفيذ العملي، تعرض النتائج التي تم الحصول عليها في الملحق آ).



الشكل 5- ناتج تطبيق النموذج دون أي موجهات (نقط، مربعات، نصوص حرة).

Vision-Language Models (VLM) and Pre-training (VLP) with -4.3.2 ViT

جرى تصميم نماذج الرؤية واللغة (VLM) لفهم وإنشاء محتوى يدمج المعلومات المرئية والنصية، مما يجعلها ضرورية لمجموعة متنوعة من المهام متعددة الوسائط. تُعتبر هذه النماذج حيوية لتحسين أداء التطبيقات التي تحتاج إلى معالجة وتفسير البيانات المتعددة الوسائط، مثل ترجمة الصور، التي تتطلب من النموذج تفسير محتوى الصورة ثم تحويله إلى نصوص بلغة أخرى، والإجابة على الأسئلة المرئية (VQA)، حيث يقوم النموذج بالإجابة على الأسئلة المتعلقة بمحتوى الصورة. كما تلعب دوراً أساسياً في استرجاع الصورة والنص، حيث يستخدم النص أو الصورة كاستعلام لاسترجاع معلومات مرتبطة بها من مجموعة بيانات كبيرة.

يشير التدريب المسبق للرؤبة واللغة (VLP) إلى عملية تدريب النماذج علىمجموعات بيانات ضخمة تحتوي على أزواج من الصورة والنص لتحسين أدائها في المهام اللاحقة. يتضمن هذا التدريب تعلم العلاقات بين الصور والنصوص ذات الصلة، مما يعزز قدرة النموذج على فهم السياق والتفسير الدقيق للبيانات المتعددة الوسائط. يتم ذلك باستخدام إطار عمل VLP مثل CLIP و ALIGN (A Large-scale Image and Noisy-text) الذي تستفيد من التفاعل بين البيانات المرئية والأوصاف النصية لتدريب نماذج قوية تلتقط دلالات وتفاصيل دقيقة.

Vision Transformer (ViT) هو بنية نموذجية تطبق تقنيات المحول، المصممة في الأصل لمعالجة اللغة الطبيعية، على بيانات الصور، إذ يقسم ViT صورة الإدخال إلى أقسام بحجم ثابت على سبيل المثال 16x16، بعدها يجري معالجة كل قسم بمفرده كمتوجه وذلك باستخدام آليات الانتباه الذاتي متعدد الرؤوس لالتقاط العلاقة بين البقع المختلفة من الصورة مما يسمح للنموذج بالتركيز على الأجزاء ذات الصلة من الصورة سياقياً، تساعد هذه الآلية النموذج على فهم العلاقات والميزات المكانية داخل الصورة، يجري استخدام الترميز الموضعي لأن المحولات لا تحتوي على مفهوم مدمج للترتيب، فيقوم ViT بإضافة تصميمات موضعية إلى متجهات التصحيح للاحتفاظ بالمعلومات المكانية، إذ أنه أظهر أداءً تنافسيًا أو متفوقاً مقارنة بالشبكات العصبية التلaffيفية التقليدية (Convolutional Neural Networks) (CNNs) في مهام الرؤبة المختلفة، وخاصة عند تدريبه علىمجموعات بيانات كبيرة، كما أنه جرى تصميم ViT للعمل جنباً إلى جنب مع محول نصي، مما يمكن النموذج من التعامل مع البيانات المرئية والنصية بشكل فعال، و التقاط العلاقات المعقدة داخل الصور يساعد على تعزيز فهم النموذج للسياقات المرئية، وهو أمر بالغ الأهمية لمهام مثل ترجمة الصور واسترجاعها.[16]

5.3.2 نماذج اللغة الكبيرة (LLM)

لقد أحدثت نماذج اللغة الكبيرة تحولاً كبيراً في مجال تلخيص النصوص، وخاصة في مجالات التلخيص التجزيوي والاستخراجي. و تستفيد هذه النماذج، مثل GPT-3 و GPT-4، من كميات هائلة من البيانات المدرية مسبقاً لفهم وإنشاء نص يشبه النص البشري. ويستند هيكلها في المقام الأول إلى نموذج Transformer، الذي يستخدم آليات الانتباه الذاتي لمعالجة تسلسلات الإدخال بشكل فعال. و تسمح هذه القدرة لنماذج اللغة الكبيرة بفهم الأنماط اللغوية المعقدة وال العلاقات الدلالية والفروق الدقيقة السياقية، مما يمكنها من إنتاج ملخصات عالية الجودة تنافس غالباً المحتوى الذي ينتجه الإنسان.

نظراً لاستخدام التلخيص في استخراج المعلومات من صفحات الويب وتسهيل عملية التصفح، يمكن استخدام نماذج اللغة الكبيرة في كل من النهجين الاستخراجي والتجزيوي، حيث أن التلخيص هو عملية تستخدم لتكثيف جزء من النص إلى نسخة أقصر، مع الحفاظ على المعلومات الأساسية والمعنى. هناك طريقتان أساسيتان للتلخيص: الاستخراج والتجزيد.

▪ **تلخيص الاستخراج Extractive summarization:** يتضمن تحديد واستخراج الجمل أو العبارات أو الأقسام الأكثر أهمية مباشرة من النص الأصلي. والمهدف هو إنشاء ملخص باستخدام عبارات أو جمل دقيقة من المصدر. هذه الطريقة مباشرة نسبياً لأنها تعتمد على محتوى النص الأصلي دون إنشاء محتوى جديد، ويشمل الخطوات التالية:

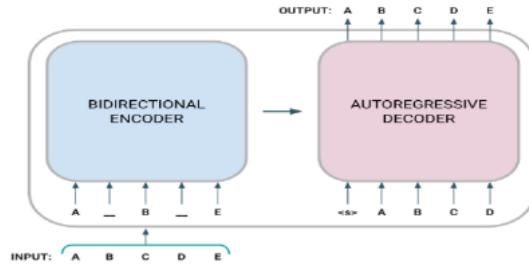
1. المعالجة المسبقة Preprocessing: تنقية النص ومعالجته مسبقاً لإزالة المعلومات غير ذات الصلة.
2. تقطيع الجملة Sentence Segmentation: تقطيع النص إلى جمل فردية.
3. استخراج السمات Feature Extraction: تحديد سمات الجمل، مثل تكرار المصطلح TF (Term Frequency)، وتكرار المستند العكسي IDF (Inverse Document Frequency)، وطول الجملة، والموضع في النص.
4. إعطاء الأهمية Scoring: استخدام الخوارزميات لتسجيل الجمل بناءً على أهميتها. تشمل الطرق الشائعة:
 - TF-IDF: تسجيل الجمل بناءً على عدد مرات ظهور المصطلحات الرئيسية.
 - الطرق القائمة على الرسم البياني: تقوم تقنيات مثل TextRank [17] ببناء رسم بياني للجمل وتصنيفها بناءً على أهميتها في الرسم البياني.
5. الاختيار Selection: اختيار الجمل الأعلى درجة لتضمينها في الملخص.

- **تلخيص التجريد Abstractive Summarization**: يتضمن إنشاء ملخص قد يستخدم عبارات وجمل جديدة، وإعادة صياغة المحتوى الأصلي. تهدف هذه الطريقة إلى فهم المعنى الأساسي للنص ثم إنتاج نسخة أقصر تقلل هذا المعنى بشكل أكثر إيجاراً، ويشمل الخطوات التالية:
 1. المعالجة المسبقية Preprocessing: تنقية النص ومعالجته مسبقاً على غرار التلخيص الاستخراجي.
 2. فهم النص Text Understanding: استخدام تقنيات فهم اللغة الطبيعية لفهم المفاهيم وال العلاقات الرئيسية في النص. يمكن أن يتضمن هذا التحليل الدلالي وفهم السياق.
 3. إنشاء المحتوى Content Generation: إنشاء نص جديد ينقل نفس المعنى مثل النص الأصلي. يتم ذلك عادةً باستخدام نماذج تعتمد على:
 - تسلسل إلى تسلسل: نماذج تقوم بترميز النص المدخل وفك تشفيره إلى ملخص.
 - المحوّلات Transformers: تعمل النماذج المتقدمة مثل BERT (Bidirectional Encoder) (تمثيلات المشفر ثنائي الاتجاه من المحوّلات)، أو GPT (Generative Pre-trained Transformer) (المحول المدرب مسبقاً التوليدية)، أو T5 على إنشاء ملخصات أكثر شبهاً بالبشر.
 - التقييم Evaluation: يجري استخدام مقاييس مثل ROUGE (Recall-Oriented Evaluation Understudy for Gisting) مقاييس يستخدم لتقييم جودة النصوص الناتجة عن النظم الآلية من خلال مقارنة مدى تطابقها مع النصوص المرجعية.
 - ROUGE-1 يقيس مدى تطابق الكلمات الفردية بين النصوص الناتجة والنصوص المرجعية، مما يعكس مدى القدرة على استرجاع المعلومات الأساسية.
 - ROUGE-2 يقيّم تطابق ثنائيات الكلمات (bigrams) بين النصوص، مما يوفر رؤية حول جودة الترابط بين الكلمات.
 - ROUGE-L يقيس أطول تسلسل متتابع من الكلمات المتطابقة بين النصوص، مما يعكس قدرة النظام على الحفاظ على التسلسل والتركيب اللغوي.
- تساعد هذه المقاييس في فهم مدى قرب النصوص الناتجة من النصوص المرجعية في التطبيقات اللغوية مثل التلخيص والترجمة.

نجد أن التلخيص التجريدي أكثر تعقيداً لأنه يتطلب إنشاء جمل جديدة تعيد صياغة النص الأصلي وتكلّفه. تُستخدم تقنيات مثل التعلم من التسلسل إلى التسلسل، حيث يقوم المشفّر بمعالجة النص المدخل ويقوم المفكّك بإنشاء الملخص، بشكل شائع. توضح نماذج مثل BART و T5 هذا النهج، حيث تستخدم استراتيجيات إزالة الضوضاء والتدريب المسبق لتعزيز أدائها في إنشاء ملخصات سلسة ومفيدة. وعلاوة على ذلك، يمكن لبرامج التعلم العميق أن تتضمن تعليمات المستخدم لتخفيض الملخصات لاحتياجات محددة، مما يجعلها أدوات متعددة الاستخدامات لتطبيقات مختلفة.[20] فيما يلي مجموعة من أشهر نماذج التلخيص.

1.5.3.2 - نموذج (BART (Bidirectional and Auto-Regressive Transformers)

(المحولات ثنائية الاتجاه والتراجعية التلقائية) هو نموذج قائم على المحولات، جرى تطويره بواسطة Facebook AI Research. وهو يجمع بين مبادئ كل من BERT و GPT يجري تدريب BART باستخدام نجح يسمى "التدريب المشوه" (Denoising Autoencoder). يتضمن هذا الأسلوب إدخال نصوص مشوهة (مثل حذف أجزاء من النص) وطلب من النموذج تصحيح هذه النصوص إلى شكلها الأصلي. هذه الطريقة تساعده على تعلم كيفية توليد نصوص جديدة من نصوص مشوهة. مما يجعله فعالاً لمهام معالجة اللغة الطبيعية المختلفة، بما في ذلك إنشاء النص والتلخيص والترجمة. يعمل BART كنموذج تسلسل إلى تسلسل (seq2seq)، مستفيداً من مشفّر ثنائي الاتجاه لفهم السياق وفك تشفير انحداري تلقائي لتوليد نص متماسك، يجري تدريب النموذج مسبقاً باستخدام نجح مشفّر تلقائي لإزالة الضوضاء، حيث يتم إخفاء أجزاء من النص المدخل أو إتلافها، ويتعلم النموذج إعادة بناء النص الأصلي. تسمح طريقة التدريب الفريدة هذه لـ BART بالتفوق في فهم اللغة وتولیدها، مما يجعلها مفيدة بشكل خاص للمهام التي تتطلب مستويات عالية من التجريد والإبداع، مثل التلخيص وتوليد الحوار. يشير نموذج Large إلى الإصدار الموسع من نموذج BART الأصلي، والذي يحتوي على المزيد من المعلومات والطبقات، مما يعزز قدرته على التعلم منمجموعات البيانات المعقّدة. يعمل هذا الهيكل الأكبر على تحسين الأداء في معايير مختلفة، وخاصة في المهام التي تتضمن إنشاء مخرجات نصية أطول وأكثر دقة. وقد ثبت أن نموذج BART الكبير يتفوق على النماذج الأصغر في العديد من السيناريوهات بسبب قدرته على التقاط أنماط أكثر تعقيداً في البيانات، وتكون فاعليته واضحة في مهام مثل التلخيص، حيث يمكنه إنتاج ملخصات ليست موجزة فحسب، بل تحفظ أيضاً بالمعلومات المهمة من المواد المصدرية. بالإضافة إلى ذلك، يميل النموذج الأكبر إلى التعميم بشكل أفضل عبر مجالات مختلفة، مما يجعله الخيار المفضل للتطبيقات التي تتطلب إنشاء نص عالي الجودة، إذ تميز هذه النماذج بقدرها على توليد ملخصات موجزة من المستندات الطويلة.[21]



الشكل 6 - بنية نموذج BART

2.5.3.2- نموذج T5 (Text-to-Text Transfer Transformer)

محول نقل النص إلى النص، هو نموذج معالجة لغة طبيعي يعتمد على فكرة رئيسية: كل مهمة يمكن تمثيلها كتحويل نص إلى نص. هذا يعني أن T5 لا يميز بين أنواع مختلفة من المهام مثل الترجمة، التلخيص، الإجابة على الأسئلة، أو تصنيف النصوص، بل يعامل جميعها كمشاكل تحويل نصوص، يجري تقديم كل من المدخلات (النصوص الأصلية) والمخرجات (النصوص المستهدفة) كسلسل نصية. على سبيل المثال، إذا كانت المهمة هي تلخيص نص، فإن المدخل سيكون النص الكامل، والمخرج سيكون الملخص. بما أن T5 يعامل كل المهام كمشاكل تحويل نص إلى نص، فإنه لا يستخدم نفس بنية النموذج لجميع هذه المهام. هذا يساعد في تبسيط تصميم النموذج ويجعله أكثر مرونة، لأنه لا يحتاج إلى تخصيص بنية مختلفة لكل مهمة. أثناء تدريب T5، يتم استخدام طريقة تسمى "ملء الفراغات". في هذه الطريقة، يتم إخفاء أجزاء معينة من النص (يعني أن أجزاء من النص تكون غير مرئية للنموذج) ويطلب من النموذج توقع الأجزاء المخفية بناءً على السياق الحيط بها.[20]

3.5.3.2- نموذج PEGASUS (Pre-training with Extracted Gap-Sentences for Abstractive Summarization)

التدريب المسبق باستخدام جمل الفراغات المستخرجة للتلخيص التجريدي خصيصاً لمهام التلخيص. وهو يقدم هدفاً جديداً للتدريب المسبق يتضمن إخفاء الجمل المهمة من مستند الإدخال وتدريب النموذج لتوليد هذه الجمل المقنعة كمخرجات وتسمى طريقة "الجمل الفجوية". من خلال محاكاة عملية التلخيص أثناء التدريب المسبق، يتعلم PEGASUS تحديد المعلومات المهمة وإنشاء ملخصات موجزة بشكل فعال. يتيح هذا النهج لـ PEGASUS التفوق في إنتاج ملخصات عالية الجودة تشبه البشر، وخاصة في السيناريوهات حيث يكون فهم جوهر النص أمراً بالغ الأهمية.[20]

تمثل هذه النماذج معاً تقدماً كبيراً في مجال التلخيص التجريدي، حيث تعرض استراتيجيات مبتكرة لإنشاء ملخصات متماسكة ومتناهية للسياق. يسلط تطويرها الضوء على التطور المستمر لـ LLMs وتأثيرها على تعزيز فهم اللغة الطبيعية وتوليدها.

4.2- معالجة الصورة (Mathematical morphology and Edge detection)

في مجال معالجة الصور، تعتبر كل من الـ Mathematical Morphology التحليل الرياضي للأشكال و Edge Detection كشف الحواف من الأساسيات المهمة التي تلعب دوراً حيوياً في تحليل وفهم الصور الرقمية.

1.4.2- التحليل الرياضي للأشكال

تركز على تحليل شكل وهيكل الأجسام داخل الصور. يعتمد هذا المجال على تطبيق عمليات رياضية تُعد ضرورية لتحليل الصور ومعالجتها. تشمل العمليات الأساسية في الـ Mathematical Morphology التأكيل (Erosion)، التمدد (Dilation)، الفتح (Opening)، والإغلاق (Closing). تُستخدم هذه العمليات لتعديل قيم البكسل بناءً على شكل وحجم العنصر الميكلبي، وهو عبارة عن قناع ثانوي ذو شكل معين ونقطة مرجعية معينة.

1. عملية التأكيل (Erosion): تقلل من حجم الكائنات في الصورة، وتعمل عن طريق إزالة البكسلات من حواف الكائنات، يتحرك عنصر هيكلبي (structuring element) أو قالب صغير عبر الصورة، وإذا كانت جميع بكسلاط الكائن التي تغطيها البكسلاط في العنصر الميكلبي تساوي البكسلاط المطابقة في الصورة الأصلية، تبقى البكسلاط الموجودة في مركز العنصر الميكلبي. وإلا، يجري إزالة تلك البكسلاط، تستخدم لتقليل الضوضاء وفصل الكائنات المتصلة.

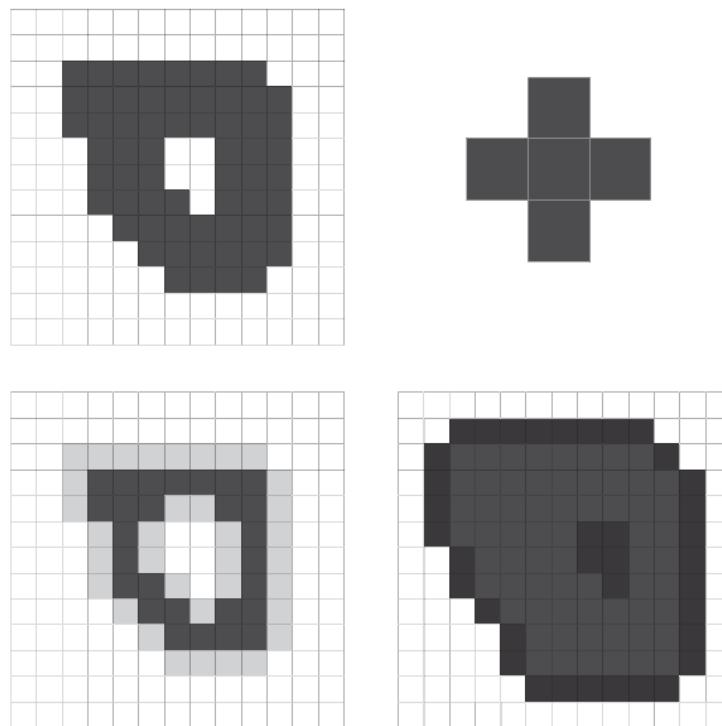
2. عملية التمدد (Dilation): تزيد من حجم الكائنات في الصورة عن طريق إضافة بكسلاط إلى حواف الكائنات، يتحرك عنصر هيكلبي عبر الصورة، وإذا كان أي من البكسلاط التي يعطيها العنصر الميكلبي يتواافق مع بكسلاط الكائن في الصورة الأصلية، تضاف البكسلاط المركزية، يستخدم ملء الفجوات الصغيرة، ووصل الكائنات المتقطعة.

3. عملية الفتح (Opening): عبارة عن تطبيق التأكيل متبوعاً بالتمدد. يتم استخدام هذه العملية لإزالة الضوضاء الصغيرة دون تغيير كبير في حجم وشكل الكائنات الكبيرة، إذ يتم أولاً تأكيل الكائنات مما يقلل

من حجمها، ثم يجري تمددها لإعادة الحجم الأصلي تقريباً ولكن بدون الضوضاء الصغيرة، ثم تتم عملية إزالة الضوضاء الصغيرة وتنعيم حواف الكائنات.

4. عملية الأغلاق (Closing): عبارة عن تطبيق التمدد متبعاً بالتأكل، يتم استخدام هذه العملية ملء الفجوات الصغيرة في الكائنات وتوصيل الكائنات القريبة، إذ يتم اولاً تمديد الكائنات، مما يزيد من حجمها وملء الفجوات، ثم يتم تأكلاً لها لإعادة الحجم الأصلي تقريباً ولكن مع ملء الفجوات، تستخدم ملء الفجوات الصغيرة وتوصيل الكائنات المتقطعة.

تُستخدم هذه العمليات في مهام متعددة مثل تقليل الضوضاء، الكشف عن الحواف، تقطيع الأجسام، وتحليل الأشكال. نشأت Mathematical Morphology في أوائل السبعينيات، ومنذ ذلك الحين تطورت في العديد من الحالات، خاصة في معالجة الصور. في الـ Binary Morphology، تعالج الصور الثنائية التي تحتوي فقط على قيمتين للبيكسل، عادةً 1 للأجسام و 0 للخلفية. تُنفذ العمليات المورفولوجية الثنائية عن طريق تحريك العنصر الميكلبي عبر الصورة الثنائية المستهدفة.



الشكل 7-مثال على التأكل والتمدد.

يوضح الشكل السابق تطبيق التأكّل والتتمدد على الشكل في الزاوية العليا اليسرى، ويشكل العنصر الهيكلي الشكل في الزاوية العليا اليمنى، الصورة المتأكّلة في أسفل يسار الشكل حيث تمثل البكسولات الرمادية الفاتحة النقاط التي تم إلغاؤها بسبب التأكّل، الصورة الممتدة موجودة في أسفل بيمين الشكل، حيث تمثل البكسولات السوداء النقاط المضافة بواسطة التمدد.

2.4.2- كشف الحواف Edge detection

هي عملية حيوية في معالجة الصور، حيث تركز على الكشف عن حدود بين مناطق ذات تباين عالي في السطوع. يتضمن هذا الكشف تحديد وجود وموقع الحواف في الصورة، وهي معلومات أساسية لفهم وتفسير المحتوى البصري. بدأت أهمية الكشف عن الحواف في أوائل السبعينيات، وقد تطورت الأساليب المستخدمة في هذا المجال لتشمل الكاشفات القائمة على التدرج والكاشفات القائمة على لاپلاسيان.

- **الكاشفات القائمة على التدرج:** تعتمد على تطبيق المشتقفات الأولى للصورة لتحديد التغيرات في السطوع. من الأمثلة الشائعة لهذه الكاشفات روبرتس (Roberts)، سوبيل (Sobel)، بريويت (Prewitt)، وشار (Scharr).
- **الكاشفات القائمة على لاپلاسيان:** تستخدم المشتقفات الثانية لتحديد التغيرات في التدرج، مثل LoG (Laplacian of Gaussian) وكاني (Canny)، الذي يعتبر من أكثر الكاشفات دقة وشهرة في تحديد الحواف بفضل قدرته على التقليل من الضوضاء وتحديد الحواف بدقة عالية.

تلعب كل من عمليات Edge Detection و Mathematical Morphology دوراً مهماً في تحسين جودة وتحليل الصور. بينما تُركز الـ Mathematical Morphology على تحليل الأشكال والبني في الصور، تتركز عملية الكشف عن الحواف على تحديد الحدود بين المناطق المختلفة في الصورة، مما يساهم في فهم أعمق للمحتوى البصري وتحسين تطبيقات معالجة الصور [22].

الفصل الثالث

الدراسة المرجعية

نقدم في هذا الفصل دراسة مرجعية لهذا البحث تتناول الأبحاث المتعلقة بأساليب تقطيع صفحات الويب والأبحاث المتعلقة بتقييم خوارزميات تقطيع صفحات الويب، إضافةً لأحدث الأبحاث المتعلقة بتلخيص النصوص والنماذج المستخدمة لذلك.

1.3- مقدمة

في عصر المعلومات الرقمية المتسارع، باتت الحاجة إلى التعامل الفعال مع الكم الهائل من البيانات والنصوص على الويب أمراً لا غنى عنه. يُعتبر تقطيع صفحات الويب وتلخيص النصوص من أبرز المجالات التي تشهد تقدماً ملحوظاً في الأبحاث التقنية. يهدف تقطيع صفحات الويب إلى استخراج المعلومات ذات الصلة من موقع الإنترن特 بطرق منتظمة، مما يسهم في تسهيل الوصول إلى البيانات وتحليلها. بينما يسعى تلخيص النصوص إلى تقديم ملخصات دقيقة وفعالة لحتوى النصوص الطويلة والمعقدة، مما يوفر الوقت والجهد في استيعاب المعلومات الأساسية. في هذا المقطع، سنستعرض أحد الأبحاث والابتكارات في هذين المجالين.

2.3- الأبحاث المتعلقة بتقطيع صفحات الويب

يعد تقطيع صفحات الويب عملية أساسية لاستخراج المعلومات وجلب المحتوى، حيث يتم تقسيم الصفحات إلى أقسام أصغر لفهم هيكلها واستخراج البيانات بفعالية. هناك عدة طرق لتحقيق ذلك، مثل خوارزمية VIPS التي تعتمد على قواعد تكرارية للتقطط الميزات البصرية والهيكلية، لكنها قد لا تغطي جميع أنماط تصميم الصفحات. يقترح استخدام خوارزميات قائمة على النصوص لاسترجاع الأقسام النصية، ولكنها تتجاهل الكائنات غير النصية مثل الصور والجداول. كما تُستخدم أيضاً خوارزميات الرسم البياني والنهج الهجين لتحسين عملية التقطيع.

1.2.3- المنهجيات المستخدمة في تقطيع صفحات الويب

يوجد العديد من المنهجيات التي تستند إلى مبادئ مختلفة في تقطيع صفحات الويب ذكر منها:

1.1.2.3- النهج المستند إلى شجرة DOM

يتضمن تجزئة الويب المستندة إلى DOM تقطيع محتوى صفحة الويب إلى أجزاء ذات معنى بناءً على بنية نموذج كائن المستند (DOM) الخاص بها. تعتبر هذه العملية ضرورية لمهام مثل استخراج المحتوى، وتحمييع الويب، وتحليل بنية صفحة الويب. في البداية، يعد فهم بنية DOM أمراً ضرورياً، وهو ما يمكن القيام به باستخدام أدوات مطورة المتصفح. يأتي بعد ذلك تحديد معايير التجزئة، مثل علامات HTML أو الفئات أو المعرفات. يمكن إجراء التقطيع إما يدوياً عن طريق فحص صفحة الويب أو برمجياً باستخدام مكتبات معالجة DOM مثل BeautifulSoup أو Cheerio [23]. بمجرد إنشاء قواعد التجزئة، يتبع التنفيذ، حيث يتم تقطيع صفحة الويب إلى أجزاء بناءً على الطريقة المختارة. من الضروري التتحقق من دقة التجزئة وضبط المعايير أو التنفيذ حسب الحاجة. تضمن هذه العملية التكرارية التجزئة الفعالة، مما يتيح المزيد من المعالجة والتحليل لحتوى صفحة الويب.

2.1.2.3- النهج المستند إلى تحليل النصوص

تتضمن تجزئة الويب المستندة إلى النص تقسيم محتوى صفحة الويب إلى أقسام أو شرائح مميزة اعتماداً على المعلومات النصية، ويتم ذلك وفقاً للخطوات التالية:

1. تحليل بنية HTML: تبدأ العملية بتحليل بنية HTML لصفحة لتحديد العناصر النصية ذات الصلة مثل العناوين، الفقرات، والعلامات الدلالية.
2. معالجة النص: يتم استخدام تقنيات معالجة اللغة الطبيعية (Natural Language Processing) مثل الترميز (Tokenization)، التعرف على الكيانات المسماة NER (Named Entity Recognition)، استخراج الكلمات الرئيسية (Keyword Extraction)، لفهم المحتوى بشكل أعمق.
3. تجميع النصوص: تُستخدم الخوارزميات لتجميع أجزاء النص ذات الصلة معاً، وذلك باستخدام ميزات مثل مقاييس التشابه (Topic Modeling)، مذكرة الموضوع (Similarity Metrics).
4. استخدامات المحتوى المجزأ: يمكن الاستفادة من المحتوى المجزأ في عدة أغراض مثل: تحسين فهرسة محركات البحث، تعزيز تجربة المستخدم من خلال تنظيم المحتوى، تسهيل مهام استرجاع المعلومات مثل التلخيص أو تحليل المشاعر.

3.1.2.3- النهج المستند إلى تحليل الصور

يتطلب هذا النهج الحصول على مجموعة بيانات من لقطات شاشة صفحات الويب لتدريب نموذج التجزئة، والذي يستخدم عادةً بنيات التعلم العميق مثل الشبكات العصبية التالافية (CNNs). معالجة الصور مسبقاً عن طريق تغيير حجمها وتطبيعها وزيادتها لتعزيز تعميم النموذج. بعد ذلك، يتم تدريب نموذج CNN على مجموعة البيانات هذه لتعلم كيفية تقطيع صفحات الويب إلى مناطق متميزة بناءً على الميزات المرئية مثل الألوان والأنسجة والتخطيطات. بمجرد التدريب، يتم تطبيق النموذج على لقطات شاشة صفحة الويب الجديدة للتجزئة، واستخراج المناطق المقابلة للرؤوس والتدليلات والأشرطة الجانبية والمحتوى الرئيسي والعناصر الأخرى.

4.1.2.3- النهج المستند إلى الرسم البياني Graph

تقسيل بيانى لصفحة الويب، حيث تمثل العقد عناصر مثل كتل النص أو الصور أو الروابط، وتشير الحواف إلى العلاقات مثل القرب أو التشابه الدلالي. تقوم الخوارزميات بعد ذلك بتحليل هذا الرسم البياني لتحديد الأجزاء المتماسكة، وتحصى العناصر المرتبطة بشكل وثيق أو المرتبطة لغوياً معاً.

5.1.2.3- النهج الهجين

يجمع النهج الهجين بين العناصر المستندة إلى DOM وخصائص النصوص وال العلاقات المكانية والبصرية للصفحة، كما يضمن ترابطًا دلائليًّا للقطع ويستخدم خوارزميات تقطيع متعددة لتحقيق هذا الهدف.

2.2.3- الطرق التي تم وضعها لتقطيع صفحات الويب

إذ تم اتباع المنهجيات السابقة لتنفيذ طرق متعددة منها:

1.2.2.3- طريقة BOM (Bag of Models)

يستخدم BoM مجموعة من السمات الميكيلية والمرئية والمنطقية لتحليل صفحات الويب بشكل شامل (استخدام نماذج تعلم آلي تستخدم السمات السابقة للتعلم).

2.2.2.3- طريقة VIPS (Vision-based Page Segmentation algorithm)

تُعد واحدة من أبرز الخوارزميات في مجال تقطيع صفحات الويب، حيث تقوم بمحاكاة طريقة فهم العين البشرية لبنية صفحة الويب. بدايةً تقوم الخوارزمية باستخدام تقنيات الرؤية الحاسوبية لتحليل التصميم المرئي للصفحة، ثم إنشاء شجرة هرمية من الأقسام استناداً إلى شجرة DOM للصفحة، ثم تقسم المقاطع المستطيلة باستخدام ما يعرف بدرجة التماسك (DoC) (Degree of Coherence)، والتي يتم حسابها باستخدام أسماء العلامات وألوان الخلفية والفوائل المرئية. تتكرر عملية التقطيع هذه حتى الوصول إلى قيمة DoC أكبر من قيمة PDoC (Preferred Degree of Coherence). باستخدام هذه الطريقة، يمكن تحقيق تقطيع دقيق لصفحات الويب استناداً إلى Coherence جميع الكتل. خصائصها البصرية، مما يساعده في فهم وتحليل البنية الداخلية للمحتوى بشكل أكثر دقة وفعالية.[24]

3.2.2.3- طريقة SegBlock

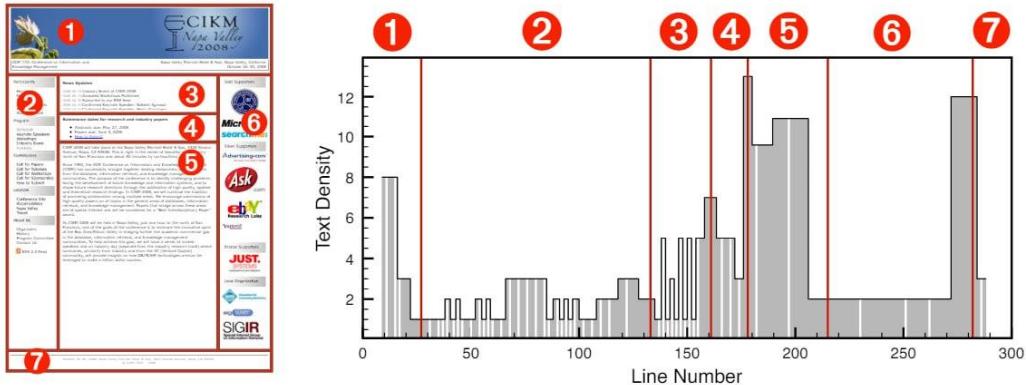
تجمع هذه الخوارزمية بين خوارزميات الرؤية الحاسوبية والتحليل الدلالي ومقاييس ملاءمة المحتوى لتقطيع صفحات الويب بشكل فعال، إذ يدمج بين السمات المرئية والمنطقية لإجراء ذلك.

4.2.2.3- طريقة Semantic-Block

تستمد الإلهام من مبادئ قانون Gestalt، مثل القرب والتتشابه والاستمرارية والإغلاق على مجموعة العناصر ذات الصلة على صفحة الويب في كتل ذات معنى، أي تقطيع محتوى صفحة الويب تبعاً للعلاقات بين العناصر.

5.2.2.3 طريقة Fusion-Block

تتضمن التقسيم على أساس تشابه النص الدلالي، يقوم Fusion-Block بتحسين التجزئة من خلال تحليل التشابه الدلالي للنص كما هو موضح في الشكل 8 داخل الكتل أو المناطق المحددة، قد تتضمن عملية التحسين هذه مقارنة المحتوى النصي باستخدام تقنيات معالجة اللغة الطبيعية مثل تضمين الكلمات أو مقاييس التشابه الدلالي [25].



الشكل 8-مثال على تحسين التجزئة بالتشابه الدلالي.

6.2.2.3 طريقة HEPS (Heading-based Page Segmentation)

تعتمد هذه المنهجية على استخدام العناوين لتحديد المقاطع داخل الصفحة، حيث تُعتبر العناوين بارزة بصرياً وتصرف موضوع المقطع. يعتمد القارئ البشري عادةً على العناوين لفهم الهيكل الدلالي للصفحة، وذلك بسبب الخصائص التالية:

- تظهر العناوين في بداية الكتل.
- تُعطى أنماط مرئية بارزة.
- تشتراك العناوين من نفس المستوى في نفس النمط المرئي.
- تُعطى عناوين المستويات الأعلى أنماط مرئية أكثر بروزاً.

لا تعتمد خوارزمية HEPS فقط على علامات العناوين HTML، حيث أنَّ العناوين تُعرف بشكل متكرر بوسائل أخرى، كما أن علامات العناوين غالباً ما تُستخدم لأغراض غير العنوان. بدلاً من ذلك، تحدد الخوارزمية العناوين والمقاطع المقابلة لها بناءً على عدة قواعد، تعتمد على موضعها في شجرة DOM، واسم العلامة، وحجم الخط والوزن.

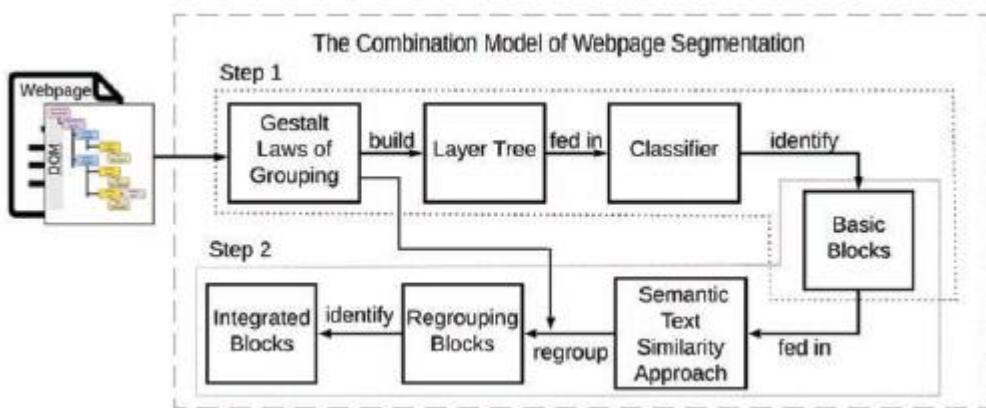
تببدأ الخوارزمية بتحديد العنوانين المرشحة باستخدام العقد النصية والصور، ثم تحدد الكتل المقابلة لها. بعد ذلك، تنشئ تجزئة هرمية بناءً على الكتل المحددة.

من المشاكل الرئيسية لهذه الأساليب أنَّ عدد المجموعات يتم تحديده تلقائياً، وبالتالي يمكن أن يختلف اختلافاً كبيراً من صفحة إلى أخرى. بالإضافة إلى ذلك، قد تبقى بعض العناصر في الصفحة غير مجمعة [26].

7.2.2.3 طريقة Integrated-Block

يدمج أساليب متعددة، بما في ذلك تحليل بنية DOM، والتجزئة المرئية، ومقاييس التشابه المستندة إلى النص. يستفيد من البنية الهرمية لنموذج كائن المستند (DOM) لفهم تنظيم عناصر صفحة الويب، يتم استخدام تقنيات التجزئة المرئية لتحديد المناطق أو الكتل المميزة بصرياً على صفحة الويب، بالإضافة إلى ذلك، يقوم بتحليل المحتوى النصي داخل هذه الكتل باستخدام مقاييس التشابه لتحديد أجزاء المحتوى ذات الصلة لغويًا، من خلال الجمع بين هذه التقنيات، يهدف إلى تحقيق تقطيع أكثر شمولاً ودقة لمحظى صفة الويب لمختلف التطبيقات [27].

إذ تم هذه الطريقة على مرحلتين أساسيتين الأولى تحدد الكتل الأساسية باستخدام قوانين Gestalt للتجميع ثم في المرحلة الثانية استخدام التشابه الدلالي لإعادة تجميع الكتل الأساسية كما هو موضح في الشكل 9



الشكل 9- خطوات الطريقة المقسمة على مرحلتين.

- الخطوة الأولى الكتل الأساسية:

في البداية جرى القيام بقطع صفة الويب إلى كتل أساسية عن طريق محاكاة الفهم البشري ووضع الكتل في شجرة layer tree لتمثيل صفحة الويب، تعد لقطات الشاشة (الصور) وأشجار DOM طريقتين رئيسيتين لتمثيل صفحة وب لتقدير التشابه البصري.

ت تكون عقد شجرة الطبقات من الكتل الأساسية الموضحة في الشكل السابق، عن طريق دمج الكتل المترابطة باستخدام قوانين تجميع Gestalt Layer Merging (GLM) في البداية تأخذ شجرة DOM ثم يتم تحويلها إلى شجرة الطبقة بواسطة منشئ الشجرة، إذ يتم إنشاء عقد الشجرة بالتزامن مع بناء تلك الشجرة وبدأ من إضافة العقدة الجذرية إلى شجرة الطبقة ثم يتم تنفيذها بشكل متكرر حتى يتم استخراج جميع عناصر DOM المرئية واضافتها لشجرة الطبقة.

بعد إنشاء شجرة الطبقة وإزالة التناقضات الهرمية بين تمثيل شجرة DOM والتخطيط المرئي لصفحة الويب، الناشئة بسبب قواعد CSS التي تسبب مشكلة في معاذة شجرة DOM بشكل غير صحيح مقابل التسلسل الهرمي مقابل التسلسل الهرمي المرئي، إضافة لإزالة العناصر غير المرئية أي تكون شجرة الطبقات حاوية على عناصر شجرة DOM المرئية فقط، كما يتم دمج هذه العناصر في مجموعات منفصلة تبعاً لمعانيها، تشرح قوانين Gestalt كيفية بناء كل كتلة من شجرة الطبقات ويقوم المترجم بتفسير القوانين الخاصة بالتجميم وهي 6 قوانين على الشكل التالي:

1. البساطة: البشر يميلون لتصنيف الكائنات لأبسط الوحدات في صفحة الويب.
2. الاغلاق: يشير هذا القانون إلى أن البشر يميلون لإدراك الأشكال غير المكتملة على أنها أشكال كاملة.
3. التقرير: تجميع الأشياء المتقاربة، ويتم تعريف المسافة وفقاً للقانون هاوستورف Normalized Hausdorff Distance حيث يعطى بالعلاقة التالية لحساب المسافة بين مجموعتين A,B

$$NHD(A, B) = \frac{1}{2} \left(\frac{\| \sup_{a \in A} \inf_{b \in B} \|a - b\|}{|A|} + \frac{\| \sup_{b \in B} \inf_{a \in A} \|b - a\|}{|B|} \right) \quad (1)$$

4. التشابه: البشر ينظرون للأشياء المتشابهة على أنها مجموعة واحدة، لحساب التشابه يمكن استخدام مؤشر Structural Similarity Index (SSIM) ومكوناته:

السطوع (Luminance) يقارن متوسط السطوع بين الصورتين.

التبابين (Contrast) يقارن تباين السطوع في الصورتين.

البنية (Structure) يقارن الأنماط أو البنية في الصورتين.

5. الاستمرارية: البشر يميلون إلى الحكم على العناصر الموجودة على صفحة الويب على أنها مرتبطة.
6. المصير المشترك: الناس يميلون إلى اعتبار العناصر التي لها نفس اتجاه الحركة مرتبطة ببعضها البعض.

وفقاً لهذه القوانين الستة، يمكن للنموذج تحديد العناصر التي سيتم تقسيمها إلى مجموعة أم لا. تتضمن كل مجموعة (كتلة أساسية) نتائج ستة قوانين مدمجة.

- الخطوة الثانية الكتل المتكاملة:

تتم مقارنة ميزات الكتلة مثل أنماط النص والعرض والارتفاع واللون بغض النظر عن التحليل الدلالي. وبالتالي، قد يتم تقسيم الكائنات إلى كتل مختلفة، على الرغم من أنها تحتوي على نص مرتبط دلالي

- **تمثيل النص:** الهدف هو تمثيل المستندات النصية الغير منظمة عددياً، في شكل منظم يمكن معالجته وتحليله في تقنيات NLP المختلفة، تقسم طرق تمثيل النص لقسمين: قسم تمثيل السياق وقسم التمثيل غير السياق. لتحقيق ذلك تم استخدام BERT, SBERT

- **التحليل الدلالي للنص:** يتم فصل الفقرات التي تحتوي على مواضيع مشابهة إلى كتل مختلفة لأنها تحتوي على نص بتنسيقات مختلفة مثل أحجام الخطوط المختلفة وألوان الخطوط وما إلى ذلك. يعالج هذا النهج المقترن هذه المشكلات باستخدام التحليل الدلالي وإعادة تجميع الكتل المتاثرة ذات الصلة في كتلة متكاملة تحتوي على نص أطول، في الخطوة الأولى يتمأخذ القوانين الست التي تم ذكرها سابقاً لتحديد الكتل الأساسية، في الخطوة الثانية يتم مقارنة التشابه النصي للكتل التي تحتوي على محتوى نصي، ويتم إعادة التجميع الدلالية وفق الخوارزمية الموضحة بالشكل 10.

ALGORITHM 1: Semantic Regrouping Algorithm

Input: Two Neighboring Basic Blocks (B_1, B_2), Text Difference Limit t

Output: Integrated Block

Begin

If $\text{Text_Similarity}^*(B_1, B_2) > t$

$\text{Integrated_Block} \leftarrow \text{Regroup}^{**}(B_1, B_2);$

else

Break;

End

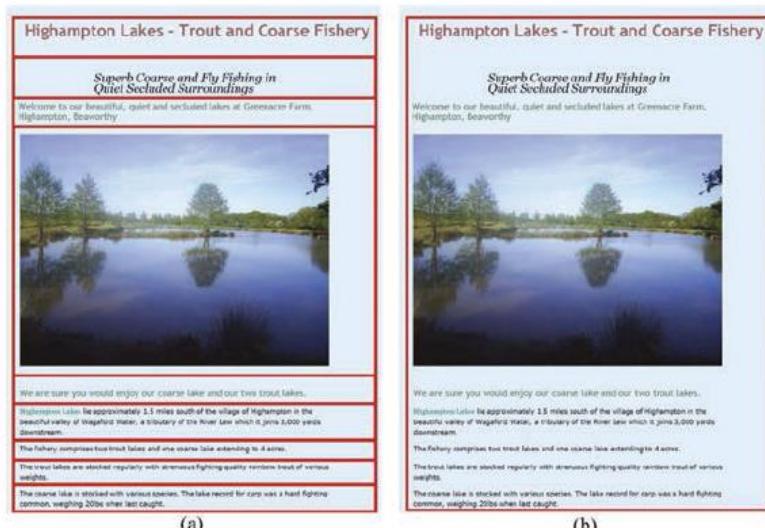
Return $\text{Integrated_Block};$

الشكل 10-خوارزمية إعادة تجميع الكتل.

يكون الدخل في هذه الخوارزمية الكتل الأساسية المجاورة إضافة لعتبة t للتشابه، تقوم الخوارزمية باستخدام قوانين Gestalt للتجميع (القرب والاستمرارية)، كما استخدم SBERT لتقدير التشابه النصي للكتل الأساسية.

مثال على ذلك:

لقطع جزء صفحة الويب (www.fishdevon.co.uk) دلليًا، يحدد هذا النموذج أولاً الكتل الأساسية وفقًا لقوانين التجميع Gestalt ليتسع الشكل a، في الخطوة الثانية من النموذج، يتم حساب التشابه الدلالي للكتل الأساسية المستخرجة ويعكس تجميع الكتل المرتبطة بالنص الدلالي وفقًا لقوانين Gestalt الخاصة بالقرب والاستمرارية كما هو موضح في الخوارزمية السابقة. يقوم النموذج بإعادة تجميع هذه الكتل الأساسية ككتلة واحدة متکاملة block نظرًا لأن حد اختلاف النص لهذه الكتل الأساسية أكبر من t فيتسع الشكل b كما هو موضح في الشكل 11 [26].



الشكل 11 – مثال على عملية التقطيع.

مع الأخذ بعين الاعتبار المصطلحات التالية:

- **السمات المنطقية (logical features):** تشير إلى تنظيم المحتوى، بما في ذلك العناوين والفقرات والقوائم والعناصر الدلالية الأخرى التي تنقل المعنى.

- **السمات الهيكلية (structural features):** تشمل على عناصر مثل علامات HTML وبنيتها المترادفة وال العلاقات بين العناصر المختلفة في الصفحة.
- **السمات المرئية (visual features):** تشمل جوانب مثل الألوان والخطوط والتخطيط والعناصر الرسومية الموجودة على صفحة الويب.
- **سمات المحتوى (content features):** أهمية المحتوى النصي ومحظى الوسائط الموجود على صفحة الويب.
- **قانون Gestalt:** يصف كيفية إدراك البشر للعناصر البصرية وتنظيمها في أنماط وهيكل متماسكة .[27]

2.2.3- المنهجيات المستخدمة في تقطيع صفحات الويب بالاعتماد على التعلم العميق

إن تطبيق أساليب الكشف عن الكائنات القائمة على التعلم العميق على مهام تجزئة صفحة الويب له إمكانات كبيرة. يتم إدخال خوارزميات اكتشاف الكائنات في تجزئة صفحة الويب عن طريق تحويل مهمة تجزئة صفحة الويب إلى مهمة اكتشاف الكائنات. استناداً إلى نموذج YOLOv5 [28]، تم إجراء تحسين لمهام تجزئة صفحات الويب من جوانب متعددة، مثل بنية الشبكة، ووظيفة الخسارة، وطريقة ما بعد المعالجة، وطريقة التدريب.

1.2.2.3- استخدام YOLOv5 لتجزئة صفحات الويب

تم اختيار نموذجاً شائعاً للكشف عن الكائنات يسمى YOLOv5 ثم القيام بإجراء العديد من التحسينات لجعله أكثر ملاءمة لتجزئة صفحات الويب [29]، ولكن النموذج المدرب لم يكن متاحاً للاستخدام، وعملية التعديل على النموذج YOLOv5 لم تكن ممكناً وذلك بسبب البنية المعقّدة له وعدم توافر الإمكانيات الالزامية للتعدّيلات التي جرى القيام بها، وهي موضحة بالتفصيل في الملحق بـ. كما تم استخدام النموذج YOLOv5 وتجربته على صفحات الويب، وكانت النتائج غير مفيدة في استخلاص المعلومات بشكل صحيح من صفحة الويب، كما هو موضح في الملحق آ.

3.2.3- أحدث التقييمات الموجودة لمنهجيات المستخدمة في تقطيع صفحات الويب

تم اقتراح تقييم عام 2022 للطريقة Integrated-Block إضافة لخمس طرائق أخرى على ثلاثة مجموعات بيانات [26] ، وتم استخدام أربعة مقاييس precision, recall, F-1 score, and the Adjusted Rand Index .(ARI)

1.3.2.3-مجموعات البيانات

تُستخدم مجموعات البيانات هذه لتقطيع محتوى صفحات الويب وفقًا للحكم البشريين (يتم تحديد الكتل الدلالية يدوياً).

- DSpopular مكونة من 70 صفحة رئيسة لموقع الويب الشهير، تتضمن ثلاث نسخ من كل صفحة
 - صفحة HTML الأساسية
 - نسخة متسلسلة من DOM بعد تحميل جميع الموارد الخارجية
 - صفحة DOM مع كتل دلالية مصنفة يدوياً
- DSrandom مكونة من 82 صفحة رئيسة لموقع الويب عشوائية وتحوي ثلاث نسخ كالسابقة.
- DSnew مكونة من 50 صفحة ويب رئيسة وهي غير متاحة لل العامة

2.3.2.3-المنهجيات التي قمت المقارنة عليها

VIPS	-
BoM	-
.SegBlock	-
Semantic-Block	-
Fusion-Block	-
Integrated-Block	-

3.3.2.3-المقاييس التي تم اعتمادها

حساب المقاييس تم استخدام مصفوفة الالتباس (Confusion matrix) وهي أداة تُستخدم بشكل شائع لقياس أداء نماذج التعلم الآلي، خاصة في مشكلة تصنيف البيانات إلى فئات مختلفة. تلخص مصفوفة الالتباس نتائج أداء النموذج عن طريق عرض عدد التنبؤات الصحيحة والخاطئة لكل فئة. في حالة مشكلة التصنيف بين فئتين (Binary)، تكون المصفوفة الالتباس على الشكل التالي:

		True Class	
		Positive	Negative
Predicted Class	Positive	TP	FP
	Negative	FN	TN

الشكل 12-مصفوفة الالتباس.

حيث تعرف الحالياً بالتالي:

- True Positive (TP): عدد النماذج التي تم تصنيفها بشكل صحيح كإيجابية.
- True Negative (TN): عدد النماذج التي تم تصنيفها بشكل صحيح كسلبية.
- False Positive (FP): عدد النماذج التي تم تصنيفها خطأ كإيجابية (خطأ نوع أول).
- False Negative (FN): عدد النماذج التي تم تصنيفها خطأ كسلبية (خطأ نوع ثان).

للتحقق من دقة طريقة التجزئة المحسوبة بواسطة النهج المقترن وخوارزميات المقارنة الخمسة الأخرى، يتم استخدام

مقاييس التقييم الأربع التالية:

- Precision نسبة الكتل المجزأة بشكل صحيح على الكتل المجزأة بواسطة الخوارزمية وتعطى عبارته بالشكل التالي:

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

حيث أن TP يشير إلى كتلتين متشابهتين تم تحديدهما بشكل صحيح على انهما متشابهان، بينما يشير FP تحديد كتلتين مختلفتين على انهما متشابهتين بشكل غير صحيح.

- Recall نسبة الكتل المجزأة بشكل صحيح على الكتل التي تم الحصول عليها بدويا بواسطة البشر (الحقيقة الأساسية) وتعطى عبارته بالشكل التالي

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

حيث يشير FN تحديد كتلتين متشابهتين على انهما مختلفتين بشكل غير صحيح.

- F-1 score مقياس يجمع بين المقاييس السابقتين على الشكل التالي:

$$F1\ score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (4)$$

- ARI (ARI) هو مقياس يستخدم لتحديد مدى الاتفاق بين مجموعتين من المتجزئات (أو الكتل)، مثل الكتل المجزأة والكتل الحقيقة الأساسية، على مجموعة معينة من البيانات

$$ARI = \frac{\sum_{ij} \binom{n_{ij}}{2} - [\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}] / \binom{n}{2}}{\frac{1}{2} [\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2}] - [\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}] / \binom{n}{2}} \quad (5)$$

لنفترض لدينا مجموعة من n العناصر هي $\{O_1, O_2, \dots, O_n\} = S$ ونفترض أن $\{x_1, x_2, \dots, x_r\} = X$ وأن لدينا $\{y_1, y_2, \dots, y_s\} = Y$ تجزئان مختلفان (كتلتان لهذه العناصر)، يحتوي الجدول على n_{ij} عدد العناصر المشتركة بين التجزئة x_i في X والتجزئة y_j في Y ، ويكون a_i هو مجموع العناصر في التجزئة x_i و b_j هو مجموع العناصر في التجزئة y_j . وتتراوح قيمته بين 0 و1، حيث تكون اتفاق المجموعتين يمثل الرقم 1 والعكس للقيمة 0 وتسمى كتل عشوائية. (يوجد المزيد من التفاصيل حول القانون) [30].

4.3.2.3-النتائج

يمثل الجدول التالي نتائج المقاييس الأربع للخوارزميات السبعة على كل مجموعة بيانات، حيث أن الحقل Total يدل على نتائج تطبيقمجموعات البيانات الثلاث المدججة، ويلاحظ أن طريقة Integrated-block تحقق تحسناً من حيث جميع المقاييس المعطاة، كما نجد أن صفحات الويب في DSrandom تميل إلى أن تكون أقل تعقيداً (مع محتوى أقل) بدلاً من DSnew وDSpopular بسبب دقتها العالية.

Algo	Precision	Recall	F1-Score	ARI
BOM	30.5%	26.1%	28.1%	0.452
VIPS	23.7%	26.2%	24.9%	0.420
SegBlock	38.1%	40.2%	39.1%	0.530
Semantic-Block	40.3%	43.4%	41.8%	0.532
Fusion-Block	44.7%	54.1%	48.9%	0.598
Integrated-Block	51.7%	61.7%	52.6%	0.632

جدول 2 - نتائج تقييم الخوارزميات على مجموعة البيانات Popular

Algo	Precision	Recall	F1-Score	ARI
BOM	30.8%	33.0%	31.8%	0.473
VIPS	27.8%	26.4%	27.1%	0.371
SegBlock	41.9%	44.8%	43.3%	0.531
Semantic-Block	43.3%	53.6%	48.0%	0.549
Fusion-Block	49.3%	61.2%	54.6%	0.610
Integrated-Block	56.6%	67.5%	61.5%	0.718

جدول 3 - نتائج تقييم الخوارزميات على مجموعة البيانات Random

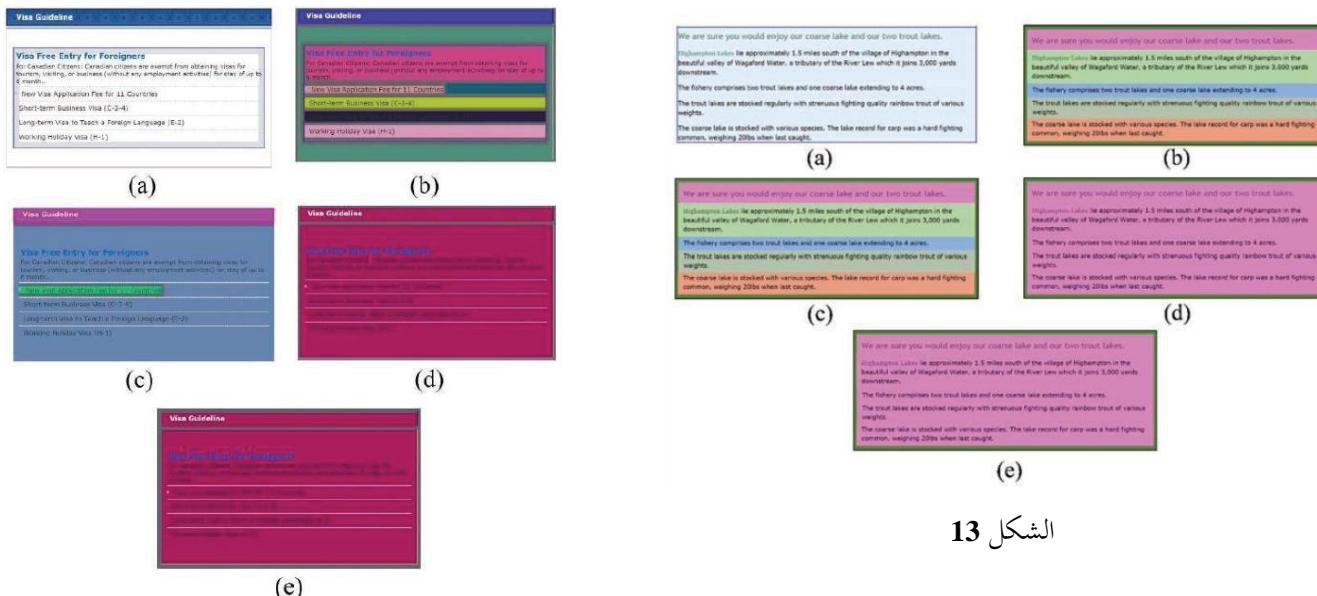
Algo	Precision	Recall	F1-Score	ARI
BOM	30.5%	21.7%	25.3%	0.426
VIPS	20.7%	24.4%	22.4%	0.415
SegBlock	39.2%	38.9%	39.0%	0.464
Semantic-Block	40.6%	41.7%	41.1%	0.483
Fusion-Block	45.8%	48.2%	47.0%	0.533
Integrated-Block	52.9%	58.7%	55.4%	0.619

جدول 4 - نتائج تقييم الخوارزميات على مجموعة البيانات Snew

Algo	Precision	Recall	F1-Score	ARI
BOM	31.4%	27.9%	29.5%	0.450
VIPS	24.7%	25.8%	25.2%	0.405
SegBlock	39.6%	42.4%	40.9%	0.514
Semantic-Block	41.2%	46.7%	43.8%	0.536
Fusion-Block	46.3%	54.9%	50.2%	0.583
Integrated-Block	53.4%	63.2%	57.8%	0.660

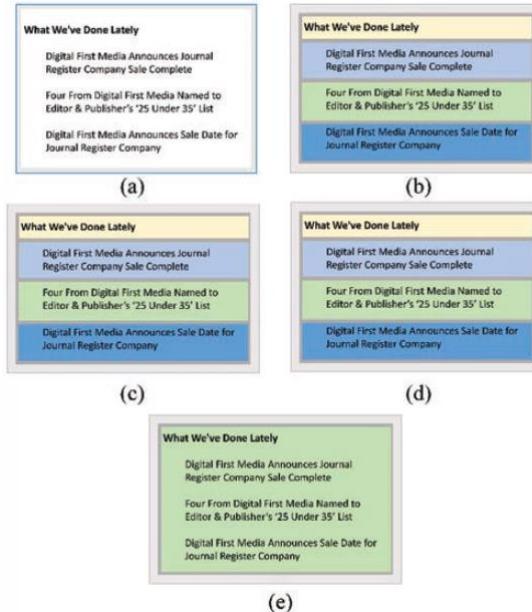
جدول 5 - محصلة النتائج النهائية .Total

تمثل الأشكال الثلاث التالية نتيجة تجزئة صفحة وب باستخدام اربع طرق مختلفة، تمثل الأشكال الفرعية a الأشكال الأساسية التي تم تصنيفها يدوياً، أما الشكل b يمثل الكتل الناتجة من تطبيق خوارزمية SegBlock والشكل c يمثل الكتل الناتجة عن تطبيق الخوارزمية Semantic-Block اللتان لا تقومان باستخدام التشابه الدلالي في التقطيع، أما الشكل d يمثل التجزئة باستخدام خوارزمية Fusion-Block والشكل e لخوارزمية Integrated-Block، نلاحظ ان في الشكل 15 تقوم طريقة Fusion-Block بقطع الفقرة إلى اربع كتل منفصلة بينما تكون هذه الكتل مرتبطة لغويًا كما نجد ان الطريقة المقترنة من قبلهما قد جمعت هذه العناصر في كتلة واحدة.



الشكل 14

الشكل 13



الشكل 15

3.3- الأبحاث المتعلقة بتلخيص النصوص

لاستخراج المعلومات بشكل فعال من المحتوى المتعلق بالأعمال، وفي عصر يسود فيه التحميل الزائد للمعلومات، يمكن أن توجه نتائج هذه الدراسة اختيار ونشر نماذج تلخيص النصوص لتعزيز عمليات استخراج المعلومات، مما يسهل في نهاية المطاف اتخاذ القرارات ونشر المعلومات بكفاءة أكبر في مجال الأعمال.

يوضح البحث المنهجية المستخدمة لتلخيص مقالات الأخبار باستخدام ثلاثة نماذج متقدمة لتلخيص النصوص: BART و T5 و PEGASUS [31]. وتنقسم المنهجية إلى عدة خطوات رئيسية:

1. **اكتساب مجموعة البيانات ومعالجتها مسبقاً:** تتضمن هذه الخطوة جمع مجموعة بيانات من المقالات التجارية من BBC News [32]. وتتضمن مجموعة البيانات المقالات وملخصاتها الحالية. ويؤكد المؤلفون على أهمية عملية التنظيف الشاملة لضمان ملاءمة البيانات للتحليل.

2. **نهاية النموذج:** هنا، يتم تثبيت الحزم اللازمة، ويتم استيراد النماذج من مكتبة Hugging Face. يتم تهيئة كل نموذج (BART و T5 و PEGASUS) باستخدام برنامج الترميز الخاص به، والذي يقوم بإعداد النص المدخل للمعالجة.

- نموذج BART: يتم تحية نموذج BART باستخدام نموذج Facebook/bart-large-CNN المدرب مسبقاً لتوليد الملخصات.
 - نموذج T5: يتم تحية نموذج T5 باستخدام الإصدار t5-small، والذي يستخدم أيضاً أدوات التجزئة لمعالجة الإدخال مسبقاً.
 - نموذج PEGASUS: يتم استيراد نموذج PEGASUS باستخدام النموذج المدرب مسبقاً `google/pegasus-cnn_dailymail`.

3. مهمة تلخيص النص: يتم إدخال مجموعة البيانات المنظفة في النماذج، والتي تولد ملخصات جديدة للمقالات. يتم فهرسة هذه الملخصات المولدة لسهولة الوصول إليها ومقارنتها بالملخصات الموجودة.

4. مقاييس التقييم: يتم تقييم أداء نماذج التلخيص باستخدام العديد من المقاييس منها ROUGE-1.

5. تحليل البيانات الاستكشافي: يختتم القسم بتحليل المقالات الملخصة، بما في ذلك التمثيلات المرئية مثل السحب والفوظية والمخططات البيانية لفهم توزيع أطوال الملخصات وخصائص المقالات.

أبو زيد النتائج:

التقييم باستخدام ROUGE-1:

models	scores
BART	0.308
T5	0.353
PEGASUS	0.245

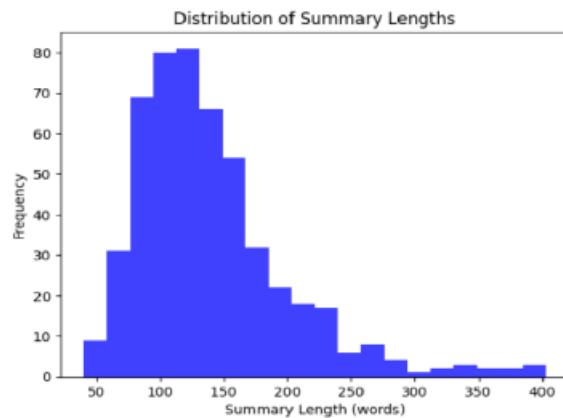
جدول 6 - نتائج تقييم النماذج المستخدمة.

السحب اللغوية للكلمات الأكثر تكراراً:



الشكل 16- سحابة للألفاظ الأكثر تكراراً.

أطوال الملخصات:



الشكل 17- تكرار الأطوال المختلفة للملخصات.

نوصي باستخدام نموذج T5 لتلخيص مقالات الأخبار نظرًا لأدائه العالي في ROUGE-1

الفصل الرابع

مجموعات البيانات

نبين في هذا الفصل مجموعات البيانات التي تم استخدامها في تقسيم تقطيع صفحات الويب، ثم مجموعات البيانات المستخدمة في تدريب التماذج التي تم الاستعانة بها لاستخراج المعلومات من قطع صفحات الويب الناتجة.

1.4 - مقدمة

تُعد مجموعات البيانات أحد الأركان الأساسية لأي مشروع بحثي أو تطبيقي، حيث تشكل الأساس الذي يتم بناء التحليلات والنماذج عليه. في هذا الفصل، سنستعرض بالتفصيل مجموعات البيانات التي تم استخدامها في المشروع.

2.4 - مجموعات البيانات المستخدمة لتقدير الخوارزميات المقترنة

تم استخدام مجموعة بيانات وهي:

DSpopular - مجموعة بيانات 1.2.4

تألف مجموعة البيانات هذه من 70 صفحة رئيسة لموقع الويب الشهير، مجموعة البيانات هذه تضم صفحات الويب مقسمة وفقاً للحكام البشريين (يتم تحديد الكتل الدلالية يدوياً)، وتتضمن ثلاثة نسخ من كل صفحة:[33]

1. صفحة HTML الأساسية
2. نسخة متسلسلة من DOM بعد تحميل جميع الموارد الخارجية
3. صفحة DOM مع كتل دلالية مصنفة يدوياً
4. اللغات المستخدمة في هذه الصفحات

CSS 53.4%	-
JavaScript 43.6%	-
Classic ASP 3.0%	-

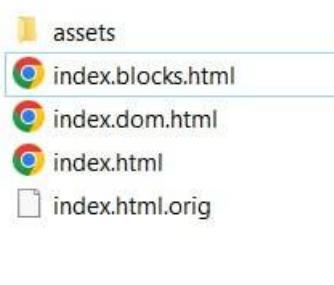
DSrandom - مجموعة بيانات 2.2.4

تألف من 82 صفحة رئيسة لموقع الويب عشوائية، مجموعة البيانات هذه تضم صفحات الويب مقسمة وفقاً للحكام البشريين (يتم تحديد الكتل الدلالية يدوياً)، وتتضمن ثلاثة نسخ كمجموعة البيانات السابقة[34]

1. اللغات المستخدمة في هذه الصفحات

CSS 46.9%	-
JavaScript 35.7%	-
PHP 17.4%	-

تم استخدامهما في الوراق البحثية [26] و[27] لتقدير الطرق التقليدية والمقارنة بينها، تضمان المحتويات التي تم ذكرها سابقاً كما هو موضح بالشكل 18، حيث تحوي الصفحة index.blockd.html على لقطة شجرة DOM مطلوب بحيث يجري التقاطع بإعطاء العنصر الذي يشكل كتلة ضمن شجرة ال DOM مسمى data-block والتي لها القيم 1 أو 2، مما يدل على أن الكتلة التي لها data-block=1 أي أنها كتلة أم ولها أبناء (كتل داخلها)، أما الكتل التي تحمل data-block=2 دل ذلك على أن الكتلة هي كتلة فرعية ضمن كتلة أم.



```

▶ <div class="headerBgContainer">::</div>
▼ <div data-block-type="Header" data-block="1" class="headerbg">
  ▼ <div class="header">
    ▼ <div class="upperheader">
      ▶ <div class="show_logos">::</div>
      ▼ <a href="index.html" name="lpos=nav[header_homepage]&lid=[Logo]">
          ::</div>
      ▶ <script type="text/javascript">::</script>
    </div>
  ▶ <div id="sectionnav">::</div>
  </div>
</div>
▶ <div id="subHeadWidgets" class="abc_wide">::</div>

```

الشكل 18 - بنية مجموعات البيانات.

3.4- مجموعات البيانات المستخدمة لتدريب النماذج المستخدمة

فيما يلي المجموعات التي تم تدريب النماذج المستخدمة لاستخراج المعلومات من قطع صفحات الويب:

1.3.4- مجموعة بيانات COCO: Common Objects in Context

هي مجموعة بيانات واسعة النطاق ومستخدمة على نطاق واسع في مجال الرؤية الحاسوبية والتعلم الآلي. تم إنشاؤها لتعزيز تطوير نماذج اكتشاف الأشياء وتقطيعها وترجمتها، تحوي على أكثر من 200000 صورة بأكثر من 80 فئة من مثلاً الأشخاص والحيوانات والمركبات والأشياء المنزلية، تتضمن مجموعة البيانات تعليمات توضيحية مفصلة مثل أقنعة تجزئة على مستوى البكسل لحدود الأشياء، و النقاط الرئيسية من أجسام البشر إضافة لنص وصفي لكل صورة، كما أنها مفيدة لمهام ترجمة الصور، تم تصميم هذه المجموعة لاكتشاف وتجزئة الكائنات وإنشاء أوصاف نصية للصور، تحوي

على فئات أقل من مجموعات البيانات الأخرى (imagenet) إلا أنها ت Hoy على مزيد من مثيلات كل فئة [35]، وهي مجموعة ماحية للعامة، وتعد موردا أساسيا في تطوير تقنيات اكتشاف الكائنات وتقسيعها وفهم الصور.

EdinburghNLP/Xsum - 2.3.4- مجموعة بيانات

هي جزء من مجموعة XSum (Extreme Summarization) الأكبر، والتي تم تصميمها لمهمة التلخيص الشديد للمقالات الإخبارية. تم تقديم مجموعة بيانات XSum في الأصل بواسطة مجموعة Edinburgh NLP لمعايرة النماذج الخاصة بإنشاء ملخصات موجزة للغاية من المقالات الإخبارية، تكون مجموعة البيانات من مقالات إخبارية مقتنة بملخصات من جملة واحدة، وتحتوي على آلاف المقالات والملخصات، تُستخدم عادةً لتدريب وتقدير نماذج تلخيص النصوص، وخاصة تلك التي تركز على إنشاء ملخصات قصيرة للغاية ومؤثرة من نصوص أطول [36].

abisee/cnn_dailymail - 3.3.4- مجموعة بيانات

مصممة خصيصاً لتقدير النماذج الخاصة بتلخيص المقالات الإخبارية، تكون مجموعة البيانات من مقالات إخبارية من CNN و Daily Mail، إلى جانب ملخصات كتبها بشر، تحتوي مجموعة بيانات CNN/Daily Mail على حوالي 300000 مقال، مما يجعلها مجموعة بيانات كبيرة لتدريب وتقدير نماذج التلخيص، إذ أنها تعد مورداً شاملًا لتلخيص الأخبار، حيث توفر حجماً كبيراً من البيانات لتدريب وتقدير نماذج التلخيص. وتحلها غطتها الواسعة للموضوعات والملخصات التفصيلية أداة قيمة لتطوير قدرات تقنيات تلخيص النصوص [37].

الفصل الخامس

تصميم مكونات النظام

نقدم في هذا الفصل تصميم مكونات النظام، ومراحل بناءه، نبدأ بشرح الخوارزميات المقترنة لقطع صفحات الويب، ثم نقدم شرحاً عن طريقة استخراج المعلومات من قطع صفحات الويب مع تبرير استخدامها.

1.5 - مقدمة

في الفصول السابقة، قمنا بتوضيح المهد الرئيسي لهذا البحث، والذي يتمثل في معالجة المشاكل التي يواجهها ضعاف البصر عند تصفح الويب. كما استعرضنا بعض الأبحاث والخوارزميات المقترحة لحل هذه المشاكل. في هذا الفصل، سنعرض الطرق المقترحة للعمل على تقطيع صفحات الويب واستخراج المعلومات منها، والتي تعتمد على تشخيص قطع صفحات الويب وإيجاد توصيف لقطع صفحات الويب باستخدام تقنيات التعلم العميق.

2.5 - الخوارزمية المقترحة لتقطيع صفحات الويب باستخدام الطرق التقليدية

باعتبار أنَّ كل قسم ضمن صفحة الويب يتكون عملياً من مجموعة من العقد المجاورة، سواءً أفقياً أو عمودياً، قمنا في هذا البحث باقتراح خوارزمية جديدة لتقطيع صفحات الويب تستند إلى هذا المفهوم. تعتمد الخوارزمية بشكل رئيسي على المساحة التي تشغليها عقد الـ DOM ضمن صفحة الويب، والمسافة بين هذه العقد، بالإضافة إلى بعض الميزات المرئية لكل عقدة.

تبعد هذه الخوارزمية نحجاً من أعلى إلى أسفل (Top-Down) أو من أسفل إلى أعلى (Bottom-Up) بناءً على عدد الأقسام التي يدخلها المستخدم، والتي تمثل المدخل الوحيد لهذه الخوارزمية. تمر الخوارزمية بعدة مراحل لإنتاج الأقسام المطلوبة لكل صفحة، وهي:

- حقن الميزات المرئية لكل عقدة ضمن شجرة الـ DOM في الصفحة المراد تقطيعها:

بدايةً، نحصل على شجرة الـ DOM لصفحة الويب، ثم نقوم بحقن إحداثيات نقطة البداية، أي إحداثيات الزاوية اليسرى العليا للعقدة (top, left)، بالإضافة إلى طول وعرض ومساحة كل عقدة من الشجرة، وإضافة الميزات المرئية للعقدة (معلومات الـ CSS). كما نضيف المسار لكل عقدة ضمن الشجرة لاستخدامها في المراحل اللاحقة.

- فلترة العقد:

في هذه المرحلة، نقوم بإجراء فلترة وتصفية للعقد التي تكون مساحتها معروفة، أو تقع خارج إطار الصفحة، أو تمتلك أمطاً يجعلها غير ظاهرة ضمن الصفحة؛ كي لا تؤثر على نتائج التقطيع في المراحل اللاحقة.

- التقطيع والتجميع:

وهي المرحلة الأساسية لعمل هذه الخوارزمية، حيث يتم تجميع الصفحة ضمن عدد القطاعات التي حددها المستخدم. نعرف كل قطاع بالبنية التالية والتي تضم: أسماء العلامات (Tags) التي يحتويها القطاع، وإحداثيات نقطة بداية القطاع أي إحداثيات الرواية اليسرى له (left, top)، وطول القطاع وعرضه ومساحته، وإحداثيات مركز القطاع المتمثلة بـ $(\text{left} + \text{width}) / 2$, $(\text{top} + \text{height}) / 2$ ، والنص الذي يحتويه كل قطاع، بالإضافة إلى المسار لكل عقدة ضمن القطاع.

يوضح شبيه الرماز (pseudo code) التالي خطوات هذه المرحلة:

```

N_Clusters = The number of clusters entered by the user
List_Clusters = children's of body node
N_Iterations = 1
Repeat till N_Iterations = Max_Iterations
    If N_Clusters = number of clusters in List_Clusters
        ▪ end the algorithm
    Else if N_Clusters < The number of clusters in List_Clusters
        ▪ Find the smallest cluster of List_Clusters and group it with its
          nearest neighbors' clusters
        ▪ Add the resulting cluster to the List_Clusters, and delete the two
          clusters that were grouped
        ▪ N_Iterations = N_Iterations +1
    Else if N_Clusters > The number of clusters in List_Clusters
        ▪ Find the largest cluster of List_Clusters and split it into its direct
          children's
        ▪ Add the resulting clusters to the List_Clusters, and Delete the cluster
          that has been divided
  
```

مراحل الخوارزمية

1. في البداية، يتم أخذ الأبناء المباشرين فقط لعقدة ال`body` من شجرة الـ DOM، ويتم وضعهم ضمن قائمة، واعتبار كل منهم قطاعاً. هذه القطاعات تمثل الحالة الابتدائية للخوارزمية.
2. مقارنة عدد القطاعات في القائمة مع العدد المدخل من قبل المستخدم:
 - **الحالة الأولى:** إذا كان عدد القطاعات في القائمة يساوي عدد القطاعات المدخلة من قبل المستخدم، تتوقف الخوارزمية هنا. تكون القطاعات الناتجة هي الأبناء المباشرين لعقدة ال`body`.

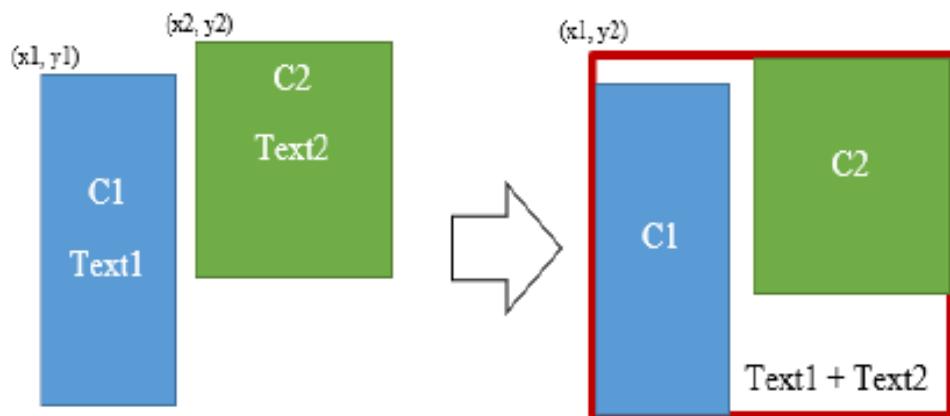
- **الحالة الثانية:** إذا كان عدد القطاعات في القائمة أكبر من عدد القطاعات المدخلة، يتم البحث ضمن القطاعات عن أصغر قطاع من حيث المساحة باستخدام المعلومات التي تم حفظها سابقاً في المرحلة الأولى. ثم يتم حساب المسافة بين هذا القطاع والقطاعات المجاورة له، ثم يتم دمج القطاع الأصغر مع أقرب قطاع له واعتبارهما قطاعاً واحداً. تم حساب المسافة باستخدام المسافة الإقليدية بين مركز القطاع الأول (الأصغر من حيث المساحة) ومركز القطاع الثاني (كل قطاع من القطاعات المجاورة)، باستخدام العلاقة التالية:

$$Distance = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (6)$$

حيث (x_1, y_1) تمثل احداثيات مركز القطاع الأول، (x_2, y_2) تمثل احداثيات مركز القطاع المجاور، يتم تعديل القطاعات تدريجياً للوصول إلى العدد المطلوب من القطاعات الذي حدده المستخدم.

في حال كان لأكثر من قطاع نفس المساحة الصغرى، يجري حساب بعد كل قطاع عن القطاعات المجاورة له، ثم نقوم باختيار أول قطاع يحقق أصغر مسافة عن مجاوراته.

تعديل قائمة القطاعات لدينا بحيث نحذف القطاعين اللذين حررنا بهما، ونضيف القطاع الذي يعبر عن ناتج دمج هذين القطاعين أي $(\min(left_{c1}, left_{c2}), \min(top_{c1}, top_{c2}))$ ، وطول القطاع الجديد هو مجموع طولي القطاعين المدمجين باستثناء الجزء المشترك بينهما، وكذلك نحسب العرض، والنط هو نصي القطاعين المدمجين.



الشكل 19 - دمج قطاعين.

- **الحالة الثالثة:** إذا كان عدد القطاعات في القائمة أصغر من العدد المطلوب، هذه الحالة، يتم تقطيع أكبر قطاع في القائمة إلى عدة قطاعات أصغر كما يلي:

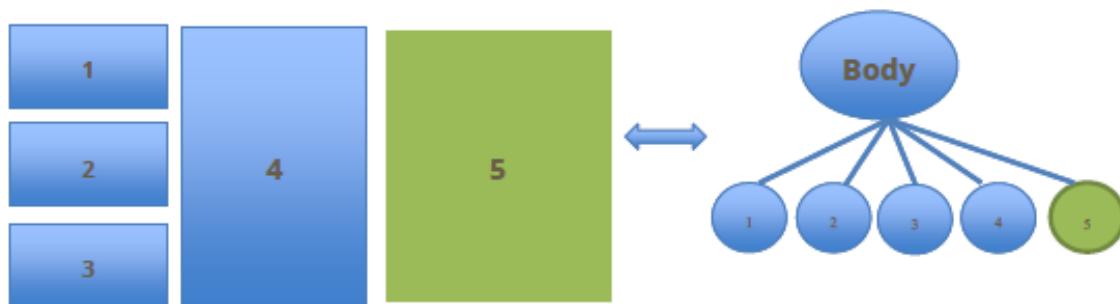
- الحصول على أكبر قطاع: يتم تحديد أكبر قطاع من حيث المساحة في القائمة.
- تقطيع القطاع:
 - إذا كان لهذا القطاع أبناء مباشرين، يتم تقطيعه عبرأخذ أبنائه المباشرين.
 - يتم تعديل قائمة القطاعات بإضافة الأبناء المباشرين للقطاع الأكبر وحذف القطاع الذي جرى تجزئته.
 - في حال لم يكن للقطاع الأكبر أبناء مباشرين، يتم الانتقال إلى القطاع التالي في المساحة وتقطيعه.
 - إذا تساوى أكثر من قطاع في المساحة الكبرى، يتم حساب عدد الأبناء المباشرين لكل منهم.
 - يتم اختيار القطاع الذي يمتلك العدد الأكبر من الأبناء المباشرين وبحري عليه التجزئة.

الحالات الثلاث السابقة تمثل دورة واحدة للخوارزمية، يتم تكرار هذه المرحلة حتى يتحقق أحد الشرطين:

1. يصبح عدد القطاعات المدخلة مساوياً لعدد القطاعات الناتجة.
2. تتحقق الخوارزمية العدد الأقصى للدورات الذي يتم تحديده لضمان توقفها بعد عدد معين من التكرارات.

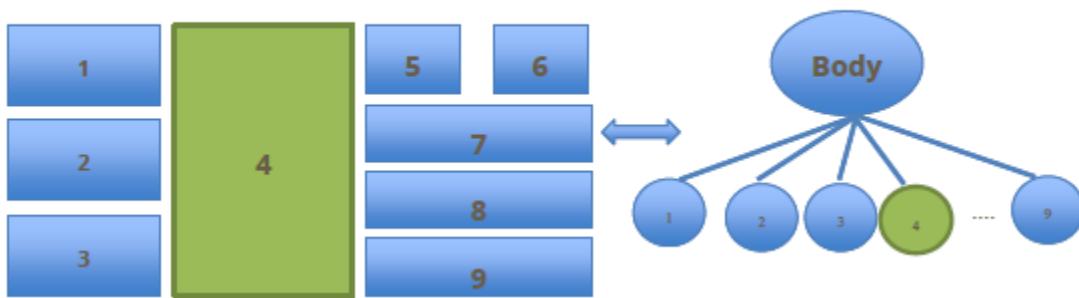
بعد الانتهاء من هذه المرحلة نكون قد قسمنا صفحة الويب إلى عدد القطاعات المطلوبة وحصلنا على مساحة كل قطاع والنص الذي يحتويه من أجل مراحل وأجزاء النظام التالية.

توضح الأشكال التالية مراحل سير الخوارزمية وفق المثال التالي:



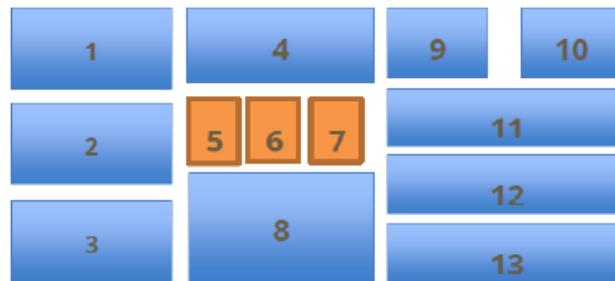
الشكل 20- بنية صفحة الويب الأساسية.

في حال أراد المستخدم تقطيع الصفحة لـ 10 قطاعات، وكانت بنية الصفحة كما هو موضح بالشكل 20 حيث العقدة body تحوي على 5 أبناء مباشرين، يكون لدينا عدد القطاعات المطلوب أكبر من عدد قطاعات القائمة، وبالتالي سنبحث عن أكبر قطاع وهو القطاع 5 وبحزته إلى أبناء مباشرين، وعدهم 5 أبناء فنحصل على الشكل التالي:



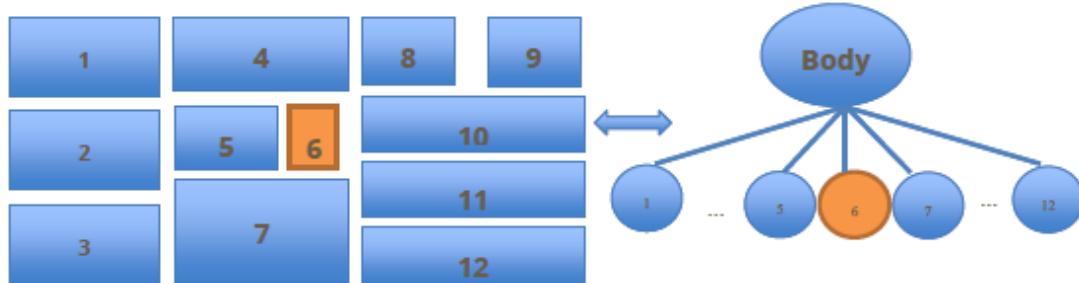
الشكل 21-بنية صفحة الويب بعد المرحلة الأولى.

كذلك نتج لدينا عدد عناقيد في القائمة مساوي لـ 9 وهو أصغر من العدد المطلوب، لذلك سنلجمأ للبحث عن أكبر قطاع فنجد أنه القطاع 4 وبحزته إلى أبناءه المباشرين وعدهم 5 فنحصل على التالي:



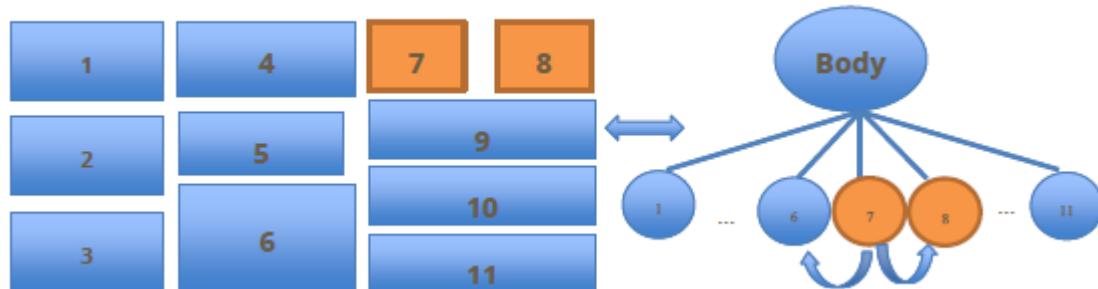
الشكل 22-بنية صفحة الويب بعد المرحلة الثانية.

أصبح لدينا 13 قطاع سنجد القطاع الأصغر ضمن القطاعات الموجودة وندمجه مع أقرب قطاع له فنجد أن هناك ثلاثة قطاعات 5, 6, 7، متساوية بالمساحة الصغرى لختار القطاعين الأقرب لبعضهما وهما 5, 6 ليصبح ناتج الدمج كما هو موضح في الشكل التالي:



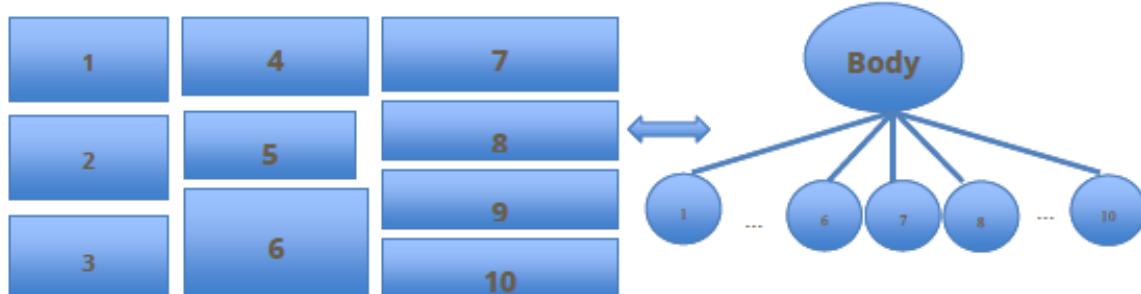
الشكل 23-بنية صفحة الويب بعد المرحلة الثالثة.

أصبح لدينا عدد العناقيد في القائمة مساوياً ل 12، سنبحث عن القطاع الأصغر ضمن القطاعات الموجودة وندمجه مع أقرب قطاع له، وهو القطاع 6.



الشكل 24-بنية صفحة الويب بعد المرحلة الرابعة.

أصبح لدينا 11 قطاع في القائمة، نبحث عن أصغر قطاع وهو القطاع 7 وندمجه مع أقرب قطاع له وهو القطاع 8 فحصل على التالي:



الشكل 25-بنية صفحة الويب بعد المرحلة الخامسة.

أصبح لدينا 10 قطاعات في القائمة وهو العدد المساوي لعدد القطاعات المطلوب من قبل المستخدم ونكون بذلك حصلنا على النتيجة النهائية للتقطيع.

3.5- الخوارزميات المقترنة للتنبؤ بعدد القطاعات الموجودة بصفحة الويب

إحدى التحديات الرئيسية في عملية العنقدة هي تحديد العدد الأمثل للعنقائد، إذ تعد خطوة حاسمة لأنها يؤثر بشكل مباشر على جودة النتائج والتحليلات اللاحقة. إذا كان عدد العنقائد قليلاً جداً، فقد يتم دمج مجموعات بيانات غير متجانسة معًا، مما يؤدي إلى فقدان التفاصيل الدقيقة والمعلومات المهمة. على الجانب الآخر، إذا كان عدد العنقائد كبيراً جداً، فقد يؤدي ذلك إلى تقطيع البيانات إلى مجموعات صغيرة غير ذات مغزى، مما يصعب تفسير النتائج ويقلل من فعالية التحليل، مما يجعله معامل حساس بالنسبة لعمل الخوارزمية التي تم اقتراحتها سابقاً في تقطيع صفحات الويب.

The screenshot shows the LightOJ Contest Hosting Pricing page. At the top, there is a navigation bar with links for Home, Problems, Compete, Ranks, Pricing, a search bar, and sign-up/login buttons. Below the navigation, a header reads "LightOJ Contest Hosting Pricing". A note says "Please send us an email on admins@lightoj.com to contact about the premium contest hosting" and "** Academic institution will get 15% discount on pricing". There are four plan boxes:

- Small**
 - 50 contestants (Max)
 - No support during contest
 - No guarantee the website will be up during the contest
 - Price: Free
- Medium**
 - 200 contestants (Max)
 - A LightOJ admin will be available in first 3 hours
 - Premium support to create problems and managing contests
 - 99.95% Uptime
 - Price: 200 USD
- Large**
 - 500 contestants (Max)
 - A LightOJ admin will be available in first 3 hours
 - Premium support to create problems and managing contests
 - 99.95% Uptime
 - Price: 400 USD
- Extra large**
 - If you want to host a contest with participants size more than 500, please reach us out to get pricing quote.

الشكل 27- نتيجة التقطيع باختيار عدد أمثل للعنقائد.

The screenshot shows the LightOJ Contest Hosting Pricing page. At the top, there is a navigation bar with links for Home, Problems, Compete, Ranks, Pricing, a search bar, and sign-up/login buttons. Below the navigation, a header reads "LightOJ Contest Hosting Pricing". A note says "Please send us an email on admins@lightoj.com to contact about the premium contest hosting" and "** Academic institution will get 15% discount on pricing". There are four plan boxes:

- Small**
 - 50 contestants (Max)
 - No support during contest
 - No guarantee the website will be up during the contest
 - Price: Free
- Medium**
 - 200 contestants (Max)
 - A LightOJ admin will be available in first 3 hours
 - Premium support to create problems and managing contests
 - 99.95% Uptime
 - Price: 200 USD
- Large**
 - 500 contestants (Max)
 - A LightOJ admin will be available in first 3 hours
 - Premium support to create problems and managing contests
 - 99.95% Uptime
 - Price: 400 USD
- Extra large**
 - If you want to host a contest with participants size more than 500, please reach us out to get pricing quote.

الشكل 26- نتيجة التقطيع باختيار عدد أقل من العدد الأمثل للعنقائد.



الشكل 28- نتائج التقاطع باختيار عدد أكبر من العدد الأمثل للعقايد.

1.3.5- اعتماداً على تحليل المحتوى النصي

تهدف الطريقة لإيجاد العدد الأمثل للعقايد في صفحة الويب بالاعتماد على المعلومات النصية الموجودة في الصفحة.

1.1.3.5- تحليل الصور الظلية Silhouette Analysis

تقيس هذه الطريقة مدى تقارب الكائن وتطابقه مع مجموعته مقارنة بالمجموعات الأخرى، إذ تعطي كل كائن قيمة ضمن المجال [1, -1]، تشير القيمة الأعلى إلى أنَّ الكائن مطابق جيداً لمجموعته الحالية ومعاير للمجموعات الأخرى، وتدلل القيمة السالبة على العكس.

آلية حساب درجة الصورة الظلية للكائن (نص في حالة صفحات الويب) في مجموعة مصنفة ضمن عنقودين:

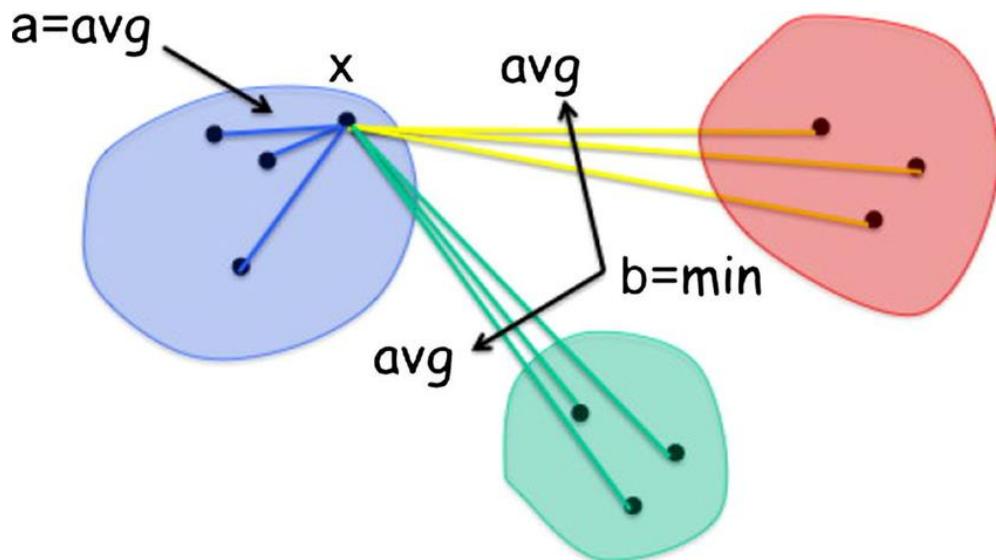
- في البداية سيتم تعريف C_n, C_m كعنقودين منشأين من قبل خوارزمية العنقدة.
- ليكن O_i العنصر الذي ينتمي للعنقود C_m والذي علاقته بالعنقودين المعروفين C_n, C_m سيتم تحليلها.
- نفترض a متوسط المسافة المغيرة عن البعد بين O_i والعناصر الأخرى في نفس العنقود C_m .
- نفترض b متوسط مسافة ما بين العنصر O_i والعناصر في العقايد الأخرى C_n (في حال وجود أكثر من عنقود نأخذ أصغر المتوسطات).

- تشير القيم الصغيرة للمعامل a إلى أن الكائنات متشابهة داخل الكتلة.
 - تشير القيم الكبيرة للمعامل b بين المجموعات إلى انفصال واضح بين المجموعات.
 - تكون جودة التجميع أعلى ما يمكن عند جعل المسافات داخل المجموعة صغيرة وخارجها كبيرة ولتحقيق هذا المهد تم تعريف درجة الصورة الظلية في العلاقة 9.
- وتعطى بالعلاقة:

$$a(o_i) = \frac{1}{|C_M|-1} \sum_{O_j \in C_M, O_j \neq o_i} d(O_i, O_j) \quad (7)$$

$$b(o_i) = \min_{C_N} \neq C_M \frac{1}{|C_N|} \sum_{O_j \in C_N} d(O_i, O_j) \quad (8)$$

$$\text{silhouette}(o_i) = \frac{b(o_i) - a(o_i)}{\max(b(o_i), a(o_i))} \quad (9)$$



الشكل 29 - إيجاد المسافة بين الكائن والكائنات الأخرى.

بعد حساب درجات الصورة الظلية لجميع الكائنات، يتم الحصول على متوسط درجة الصورة الظلية، مما يشير إلى الجودة الشاملة والمثالية لعدد المجموعات وذلك بأخذ المعادلة التالية:

$$silhouette(C_i) = \frac{1}{|C_i|} \sum_{O_j \in C_i} silhouette(O_j) \quad (10)$$

$$silhouette(C) = \frac{1}{m} \sum_{i=1}^m silhouette(C_i) \quad (11)$$

بعد حساب قيمة الصورة الظلية لحالات مختلفة للتقطيع، يتمأخذ عدد العناقيد الموفق لأعلى درجة صورة ظلية، إذ يجري تطبيق الخوارزمية السابقة على قطع النصوص المستخرجة من صفحات الويب، وتم العنقدة عدة مرات وفي كل مرة تم عملية حساب الصورة الظلية ويتم اعتماد العدد الموفق لأعلى صورة ظلية [38].

يوضح شبه الرمaz (pseudo code) التالي خطوات هذه الخوارزمية:

URL = The URL of the webpage to be segmented, provided by the user.

Content = Fetch the webpage content from the given **URL**.

Extracted_Segments = Parse the content to obtain text from tags such as `<h1>`, `<h2>`, `<h3>`, `<p>`, etc.

Preprocess Text:

Repeat until **Extracted_Segments** array is empty:

- **Remove Extra Whitespace**: Replace multiple whitespace characters with a single space.
- **Remove Non-Word Characters**: Remove characters that are not letters or spaces.
- **Remove Digits**: Eliminate numeric characters from the text.
- **Convert to Lowercase**: Convert all text to lowercase.
- **Remove Common Stop Words**: Remove common stop words (e.g., "and", "the") using a stop words list.

Feature Extraction:

- Convert the preprocessed text data into numerical features using TF-IDF (Term Frequency-Inverse Document Frequency) vectorization.
//This process transforms the text data into a matrix of TF-IDF features.

Determine Optimal Clusters:

- **Silhouette_Scores_List**: A list to store silhouette scores for different numbers of clusters.
- **Range_of_Clusters**: Initialize a range of cluster numbers (e.g., from 2 to 200 or the number of texts - 1).

Repeat till all numbers in **Range_of_Clusters** tested:

IF the number of texts is sufficient for the number of clusters:

- Apply KMeans clustering with the current number of clusters.
- Compute the silhouette score to evaluate the clustering quality.
- Store the silhouette score to **Silhouette_Scores_List**

Else if the number of texts is insufficient for the number of clusters:

- Store a very low score (e.g., -infinity) to **Silhouette_Scores_List**

- **Result** = the number of clusters with the highest silhouette score as the optimal number of clusters.

2.3.5- اعتماداً على تحليل الصورة

تهدف الطريقة التالية لإيجاد العدد الأمثل للعنقיד في صفحة الويب بالاعتماد على معالجة الصورة، ويتم ذلك وفق المراحل التالية:

1.2.3.5- معالجة الصورة

- بدايةً تم استخدام مرشح غاوسي Gaussian Blur وهو نوع من مرشحات التسгيم يستخدم لتقليل الضوضاء والتفاصيل الصغيرة في الصورة حيث يتم تطبيقه عن طريق تمرير مرشح على الصورة بقيم مأخوذة من توزيع غاوسي، إذ يتم حساب توزيع غاوسي لكل بكسل في الصورة استناداً إلى البكسلات المجاورة، وتكون القيمة الجديدة للبكسل عبارة عن متوسط موزون للبكسلات المجاورة، تكون القيم الأبعد لها وزن أقل.
- ثم يتم استخدام خوارزمية Canny لتحديد الحواف Edge Detection. تهدف لتحديد الحواف البارزة والأكثر أهمية في الصورة، وذلك وفقاً للخطوات التالية:
 - بدايةً يتم ذلك بحساب التدرج Gradient في الصورة في الاتجاهين الأفقي والرأسي باستخدام مرشحات مثل Sobel والتي تعطى بالعلاقة:

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad (12)$$

$$G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \quad (13)$$

- ثم يتم تطبيق جداء التلاطف بين المرشحات السابقة والصورة المنعمة والمتحول للتدرج الرمادي كما هو موضح بالعلاقة:

$$G_x = I * G_x \quad (14)$$

$$G_y = I * G_y \quad (15)$$

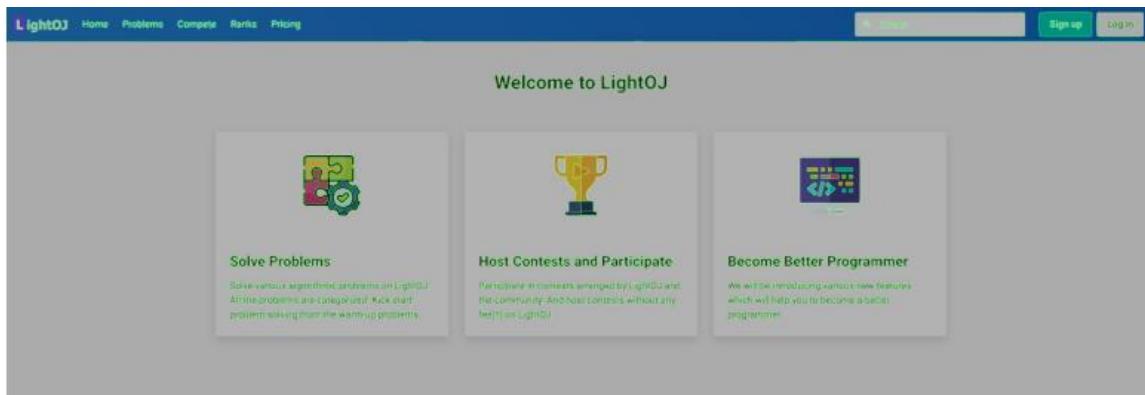
- ثم حساب قيمة التدرج المعطاة في العلاقة التالية:

$$G = \sqrt{{G_x}^2 + {G_y}^2} \quad (16)$$

$$\theta = \arctan\left(\frac{G_y}{G_x}\right) \quad (17)$$

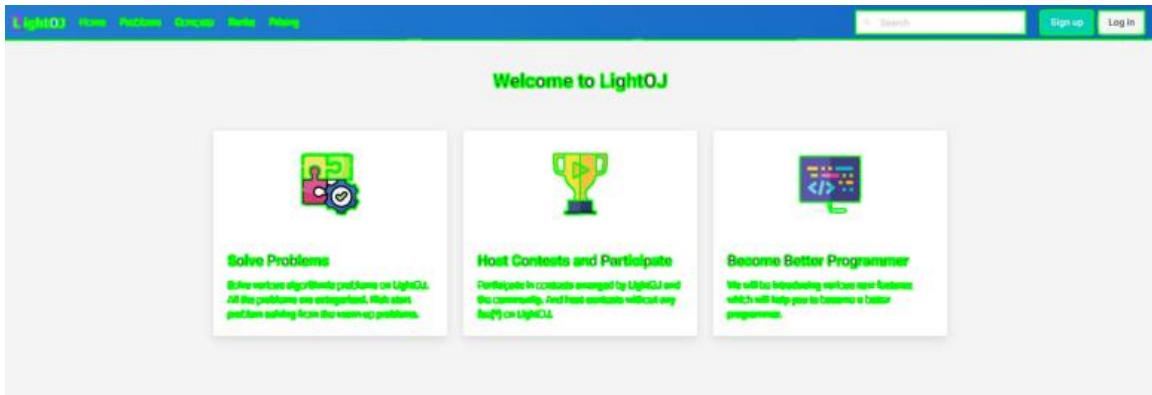
إذ يعطي التدريج فكرة عن مدى تغير شدة الصورة في كل نقطة وهو مؤشر عن وجود حافة.

- ثم استخدام Non-Maximum Suppression لإزالة الاستجابات غير المحلية أي النقاط التي ليست بالضرورة جزءاً من الحافة القوية، وذلك من خلال مقارنة مقدار التدرج في كل بكسل في الصورة مع مقادير التدرج للبكسلات المجاورة في اتجاه التدرج، فإذا كان مقدار التدرج في البكسل أقل من أحد البكسلات المجاورة يتم تعين مقدار التدرج في هذا البكسل للصفر مما يساعد في تنحيف الحواف، فمثلاً إذا كانت زاوية التدرج مساوية ل 45 درجة يتم مقارنة مقدار التدرج مع البكسلات الموجودة في الاتجاه 45 أي بشكل قطري، إذا كان مقدار التدرج في البكسل الحالي أكبر من مقادير التدرج لجميع البكسلات المجاورة يتم الاحتفاظ به.
- ثم استخدام العتبات المزدوجة لتحديد الحواف القوية والضعيفة إذ تم استخدام القيمتين 50 للعتبة السفلية و 150 للعتبة العليا، ثم تصنيف الحواف إلى حواف قوية إذا كانت قيمتها أكبر من العتبة العليا وحواف ضعيفة إذا كانت قيمتها بين العتبتين والحواف التي تكون قوتها أقل من العتبة السفلية يتم تجاهلها.
- ثم يتم تعقب الحواف باستخدام التحليل التواصلي Edge Tracking by Hysteresis والذي يقوم بتتبع الحواف الضعيفة والتأكد مما إذا كانت متصلة بحواف قوية في حال اتصالها يتم اعتبارها جزءاً من الحافة وإلا سيتم تجاهلها.



الشكل 30- إيجاد الحواف في الصورة باستخدام خوارزمية `canny`.

- بعد اكتشاف الحواف يتم استخدام خوارزمية `findContours` لتتبع الحدود الخارجية للأشكال في الصورة، يتم تخزين كل `Contour` كسلسلة من النقاط (الأحداثيات) التي تمثل الحدود الخارجية للકائنات في الصورة، مما يساعد في فهم شكل وهيكل الكائنات داخل الصورة. فيتم استخراج `Contours` من الحواف المكتشفة في هذه المرحلة.



.الشكل 31-إيجاد **Contours**

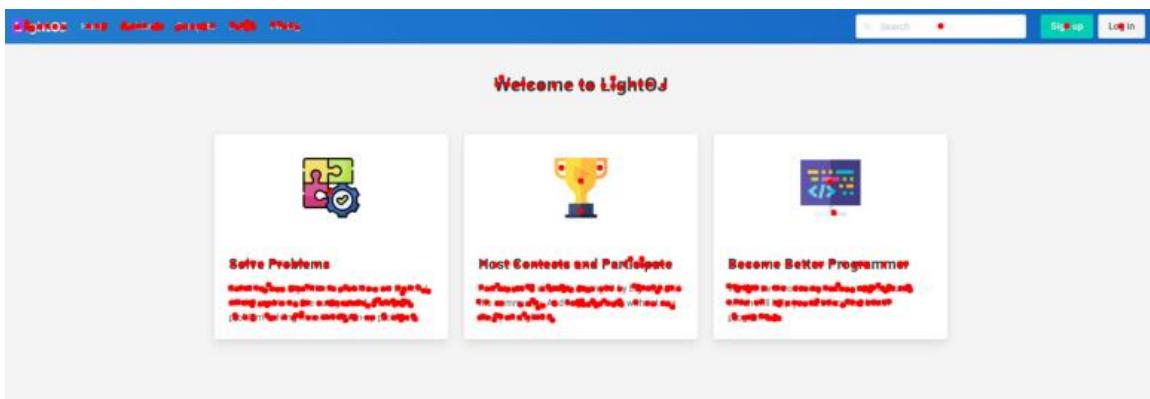
2.2.3.5- استخراج الخصائص والسمات

في هذه المرحلة تم حساب ورسم مراكز الكتل باستخدام مفهوم لحظات الصورة وهي عبارة عن واصفات توفر معلومات قيمة حول التوزيع المكاني للكثافة في الصورة. وهي عبارة عن قيم رقمية يتم حسابها من شدة البكسل في الصورة، وتستخدم لوصف الشكل والحجم والاتجاه والخصائص الهندسية الأخرى للكائنات داخل الصورة. يمكن حساب لحظات الصورة Moments من أنواع مختلفة من الكائنات، مثل الخطوط العريضة المغلقة أو مجموعات من النقاط، وتحليل الأشكال، Contours، وتستخدم على نطاق واسع في رؤية الكمبيوتر لمهام مثل التعرف على الكائنات، وتحليل الأشكال، وتسجيل الصور. من خلال التقاط الميزات المهمة للصورة، تلعب اللحظات دوراً حاسماً في تطبيقات معالجة الصور وتحليلها [39]. يتم من خلال حساب النقطة المركزية من النقاط Contours في الصورة وذلك لتحديد موقع الدقيقة للكائنات في الصورة ليتخرج ما يسمى ب Centroids، يتم حساب لحظات الصورة حول Contour بالعلاقة التالية:

$$C_x = \frac{M_{10}}{M_{00}} \quad (18)$$

$$C_y = \frac{M_{01}}{M_{00}} \quad (19)$$

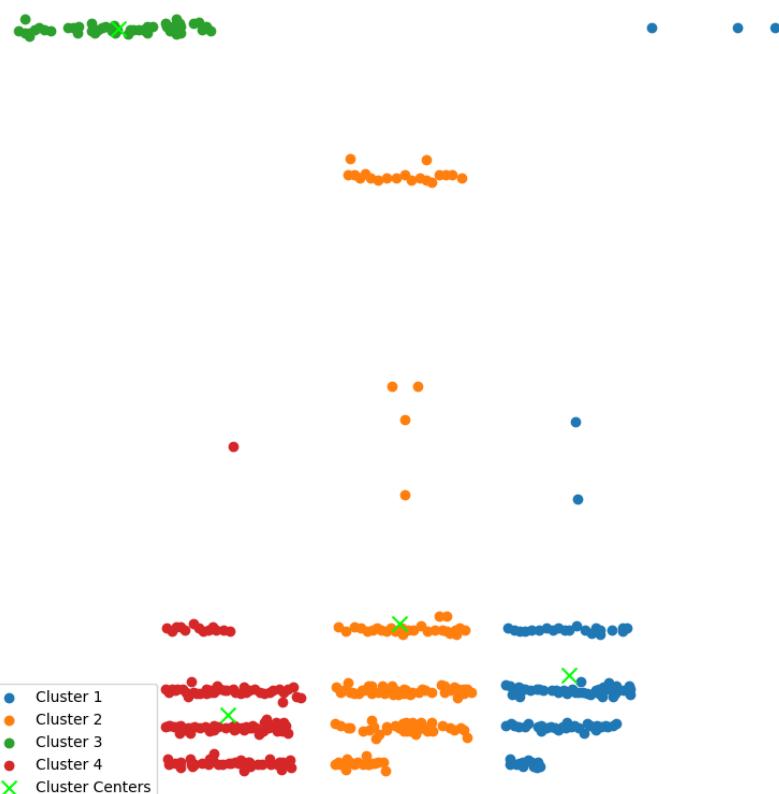
حيث تمثل حيث M_{00} هي اللحظة الصفرية (التي تمثل المساحة)، و M_{10} و M_{01} هما اللحظات الأولى في المحورين X و Y على التوالي.



الشكل 32- إيجاد Centroids في الصورة باستخدام خوارزمية moments.

3.2.3.5- التجميع العنقودي

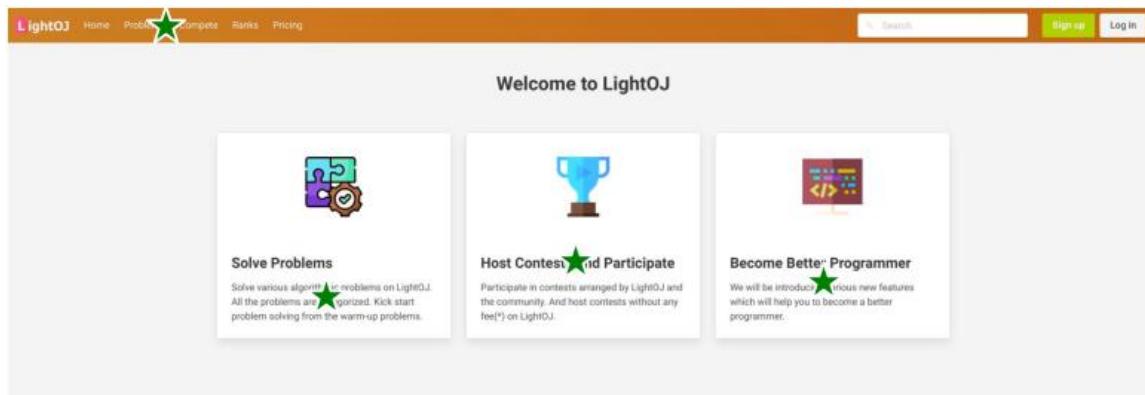
تم استخدام خوارزمية DBSCAN (Density-Based Spatial Clustering of Applications with Noise) التي تكون متقاربة بكثافة ضمن عنقود واحد، وتعتبر النقاط البعيدة كنقاط شاذة .(Outliers)



الشكل 33- العناقيد الناتجة ومركزها.

4.2.3.5 - التعلم العميق وتجزئة الصورة باستخدام نموذج تقطيع أي شيء SAM

تم استخدام هذا النموذج لتجزئة والذي يتطلب عدداً من المدخلات وهي صورة لصفحة الويب المراد تجزئتها، إضافة لمجموعة من النقاط التي تحدد المناطق المميزة في الصورة والتي تشكل مراكز العناقيد التي تم إيجادها في الفقرة السابقة، ليعطي قناع يقسم الصفحة لأجزاء مختلفة.



الشكل 34-صفحة الويب والنقط المميزة بها دخل لنموذج sam

كما يوضح شبه الرمaz (pseudo code) التالي خطوات هذه الخوارزمية:

URL = The URL of the webpage to be segmented, provided by the user.

Image = Fetch screen shot to the webpage from the given **URL** using apiflash.

Process Image:

- Convert **Image** to grayscale.
- Apply Gaussian Blur on gray image.
- Detect edges using the Canny edge detection algorithm.
- **Edges** = binary image with detected edges.
- **Contours** = list of detected contours found by **Find_Contours (Edges)** method.

Feature Extraction:

- **Centroids**: an empty list
- **Min_area**: minimum area threshold for contour consideration

Repeat till all contour in **Contours**:

If the area of contour is greater than **Min_area**:

- **M** = The moments of the contour.
- Calculate the centroid coordinates (cx, cy):
If $M["m00"] \neq 0$:
 - $cx = \text{int}(M["m10"] / M["m00"]);$
 - $cy = \text{int}(M["m01"] / M["m00"]);$**Centroids** $\leftarrow (cx, cy);$

Cluster Aggregation:

- **Eps** (float, default 10): The maximum distance between samples for them to be considered in the same neighborhood.
- **Min_Samples**: (int, default 2): The minimum number of samples required in a neighborhood for a point to be considered a core point.
- **db**: Initialize DBSCAN with **Eps** and **Min_Samples**
- **db.fit (Centroids)** ;
- **Labels** = **db.Labels**
- **Unique_Labels** = Set of unique labels from **Labels**
- **Clusters_Centers**: an empty list for centers.
Repeat till all **Labels** in **Unique_Labels** processed:
 - **Center**: The mean of coordinates for centroids belonging to the current label
 - **Clusters_Centers** \leftarrow **Center**

SAM:

- **Mask**: a numpy array representing the binary mask of the segmented image generated by sam model.
- **Labeled_Mask, Num_Pieces** = **label (Mask)** where **Labeled_Mask**: A label matrix where each connected component in the mask is assigned a unique label and
- **Num_Pieces** The number of unique labels (components) found in the mask.
- **Area_Threshold** = 2000;
- Repeat till all component in **Labeled_Mask** processed:
 - IF the area of the component is less than **Area_Threshold**:
 - Set all pixels in that component to background (0):
 - **Component_Mask** = (**Labeled_Mask** == component)
 - **Labeled_Mask [Component_Mask]** = 0
 - **Filtered_Mask** = (**Labeled_Mask** > 0)
 - Re-label the filtered mask to count the remaining clusters:
 , **Number_Of_Clusters**= **label (Filtered_Mask)**

4.5 الخوارزمية المقترنة لتقطيع صفحات الويب باستخدام تقنيات التعلم

العميق

تم تقديم مقترن خوارزمية لتقطيع صفحات الويب بالاعتماد على تقنيات التعلم العميق والنمذج الموجودة إضافةً لتقنيات معالجة الصورة، وذلك وفقاً للخطوات التالية:

1. في البداية يتم تطبيق مرحلة معالجة الصورة ومرحلة استخراج الخصائص والسمات ومرحلة التجميع العنقودي من الفقرة 2.3.5 وذلك لإيجاد النقط الأساسية من الصورة المراد تقطيعها.
2. يتم تحويل هذه الصورة إلى صورة ثنائية وذلك للحصول على صورة من البكسلات البيضاء والسوداء فقط مما يساعد في إزالة التفاصيل غير المهمة والمضوضاء مما يجعل المياكل والأشكال أكثر بروزاً، وتعزيز التباين بين المناطق الحاوية على النصوص مما يجعل الحدود أوضح وأكثر تحديداً وبالتالي يحسن دقة استخراج المراكز ومن ثم التجميع، إضافةً لنتائج تقطيع أفضل.
3. يتم استخدام نموذج التعلم العميق SAM للحصول على تقطيع للصورة الثنائية بالإضافة للنقط المميزة التي تم استخراجها في الخطوة الأولى، ليتتج القناع الممثل لعملية التقطيع والذي يتم معالجته للتخلص من القطع الصغيرة والمهملة المساحة.
4. يتم تطبيق مجموعة من العمليات Mathematical Morphology لمعالجة القناع الناتج، وهي عمليات تطبق على الصورة بشكلها الثنائي، إذ تستخدم العمليات السابقة لاستخراج الميزات ومعالجة الكائنات بناءً على شكلها الهندسي وفي نظم التعرف على الأنماط.
5. تطبيق عمليات معالجة الصورة الناتجة (القناع بعد ملء الفجوات الموجودة به) وتقطيعها إلى فقرات باستخدام المربعات المحيطة وذلك من خلال:
 - الكشف عن الحواف الموجودة (إيجاد حواضف كل قطعة ضمن القناع الناتج)، وذلك باستخدام مرشح Canny ثم خوارزمية Gaussian.
 - إيجاد الحدود الخارجية contours للأقسام في القناع وذلك بالاعتماد على الحواف.
 - إيجاد Bounding Boxes للحدود الناتجة لكل قطعة من القناع.
 - تجميع المربعات الناتجة والتي تتقارب من بعضها بأقل من حد معين من المسافة.
 - تصفية المربعات الداخلية، إزالة المربعات التي تقع بالكامل داخل مربعات أخرى أكبر منها.
 - دمج المربعات المتداخلة مع بعضها البعض.

بعد تطبيق العمليات السابقة نجد تحسين لتحديد القطع الموجودة، يوضح شبيه الرماز التالي خطوات هذه الخوارزمية:

URL = The URL of the webpage to be segmented, provided by the user.

Image = Fetch screen shot to the webpage from the given **URL** using apiflash.

Process Image: Same steps in previous algorithm.

Feature Extraction: Same steps in previous algorithm.

Cluster Aggregation: Same steps in previous algorithm.

Image Preprocessing:

- Convert **Image** to grayscale.
- Compute adaptive threshold.
- **Binary_Image:** comparing each pixel value in the grayscale image to the adaptive threshold and setting the pixel value to either 0 or 255.

SAM Segmentation:

- Setting the **Binary_Image** for the model predictor and obtaining image embeddings.
- Inputs are organized into a dictionary that is passed to the ONNX runtime.
- **Mask:** a numpy array representing the binary mask of the segmented image generated by model.
- **Labeled_Mask, Num_Pieces** = label (**Mask**) where **Labeled_Mask:** A label matrix where each connected component in the mask is assigned a unique label and
- **Num_Pieces** The number of unique labels (components) found in the mask.
- **Area_Threshold** = 2000;

Repeat till all component in **Labeled_Mask** processed:

IF the area of the component is less than **Area_Threshold**:

- Set all pixels in that component to background (0):
 - **Component_Mask** = (**Labeled_Mask** == component)
 - **Labeled_Mask [Component_Mask]** = 0
- **Filtered_Mask** = (**Labeled_Mask** > 0)

Mathematical Morphology:

- **Masked_Image:** **Image** + **Filtered_Mask**.
- Convert the **Masked_Image** to grayscale.
- Compute adaptive threshold using **Gray_Image**.
- Create **Binary_Image** by thresholding **Gray_Image** using adaptive threshold.
- **Opened_Image:** Apply morphological opening on **Binary_Image** using **kernel_open**.
- **Closed_Image:** Apply morphological closing on **Opened_Image** using **kernel_close**.
- **Inverted_Image:** Invert colors of **Closed_Image**.
- Get height (h) and width (w) of **Inverted_Image**.
- Create a mask of size (h+2, w+2) with zeros for flood filling.
- Perform flood fill on **Flood_Filled_Image** using the mask.
- Invert colors of **Flood_Filled_Image**.
- **Mathematicl_Image:** Combine **Closed_Image** with **Flood_Filled_Image** inverted using bitwise OR

Image Postprocessing:

- Convert **Mathematical_Image** grayscale.
- Apply Gaussian blur to the grayscale image.
- Detect edges in blurred_image using Canny edge detection.
- Find contours in the edge-detected image.
- Calculate bounding boxes for each contour.
- Convert bounding boxes to array.
- Cluster bounding boxes based on distance threshold.
- Group bounding boxes by clusters and merge them.
- Filter out inner boxes.
- **Final_Boxes:** Merge overlapping boxes.
- Draw **Final_Boxes**.

5.5- التقنيات المتبعة لاستخراج المعلومات من قطع صفحات الويب

بعد القيام بتنقطيع صفحة الويب المراد تصفحها سواءً بالطرق التقليدية أو بالاعتماد على تقنيات التعلم العميق، فإننا نقوم بأخذ المحتوى النصي وتلخيصه وتحويله لصوت، ويمكن الاستفادة من تقنيات التعلم العميق في تلخيص النصوص الناتجة إضافةً إلى القدرة على معالجة قطع الصور الناتجة عن التنقطيع باستخدام معالجة الصورة، حيث يتم معالجة قطع الصورة لأخذ شرح عن المحتوى الموجود في كل صورة، وذلك لتوصيف الصور الموجودة في صفحات الويب.

1.5.5- تلخيص النصوص

كما تم التوضيح في الفصول السابقة، أنَّ التلخيص يقسم إلى تلخيص تجزيدي وتلخيص استخراجي، تم استخدام نوعي التلخيص وجعل النظام يقدم حرية الاختيار للمستخدم بين النوعين.

1.1.5.5- التلخيص الاستخراجي

تم تلخيص النصوص باستخدام مكتبة Spacy، إذ تعمل على التلخيص بناءً على أهمية الجمل، ويتم ذلك باتباع الخطوات التالية:

1. **تحميل النموذج:** تحميل نموذج مدرب مسبقاً من مكتبة Spacy والذي يحوي على أدوات لتحليل النصوص مثل تصنيف الكلمات واستخراج الجمل، إذ تم استخدام نموذج en_core_web_sm، وهو نموذج مدرب على نصوص وب مكتوبة (مدونات والأخبار والتعليقات).

2. معالجة النصوص: وذلك باستخدام nlp، وتحويله لوثيقة مما يمكن مكتبة Spacy من تحليل النص واستخراج المعلومات منه، إذ يحوي على معلومات حول الجمل والكلمات والعلامات التحوية POS Tags، بالإضافة للبيانات المسممة NER.

3. استخراج الجمل: يتم استخراج الجمل من كائن Doc إذ يتم تخزينها في قائمة sentences.

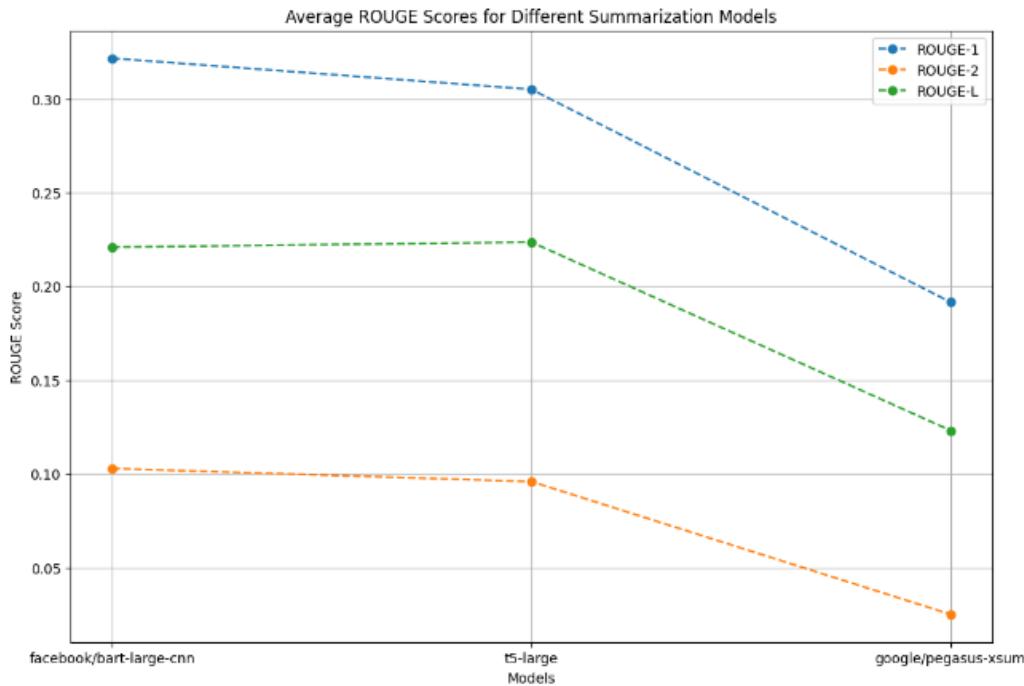
4. تحديد درجات الأهمية: يتم حساب "الأهمية" كل جملة بناءً على عدد الكلمات غير الشائعة (STOP_WORDS) فيها. الكلمات غير الشائعة هي الكلمات التي يتم تجاهلها عادة في تحليل النصوص مثل "the" ، "is" ، "and" ، وما إلى ذلك، إذ يتم حساب عدد هذه الكلمات غير الشائعة في كل جملة، وكلما زاد عدد الكلمات غير الشائعة في الجملة، تعتبر الجملة أقل أهمية. العكس صحيح: كلما قلَّ عدد الكلمات غير الشائعة، تكون الجملة أكثر أهمية.

5. ترتيب الجمل: وذلك بناءً على درجات الأهمية المحسوبة. إذ تكون الجمل التي تحتوي على عدد أقل من الكلمات غير الشائعة في أعلى القائمة.

6. إنشاء ملخصات: اختيار الجمل التي تحمل أهمية عالية من القائمة المرتبة وتوحيدتها في ملخص واحد .

2.1.5.5- التلخيص التجريدي

يوجد العديد من النماذج المستخدمة في مهمة التلخيص التجريدي، تم اختيار أشهر ثلات نماذج في هذا المجال، ولاعتماد أحدها تم استخدام طريقة التقييم المستخدمة في الأبحاث الأخيرة وهي من خلال الاعتماد على مقياس ROUGE إذ تم تحميل النموذج T5, BART, PEGASUS واستخدام نصوص لـ 15 قطعة من صفحات الويب مع تلخيصها (يدوياً)، لإجراء مقارنة أكثر دقة من إعطاء نص وحيد مع تلخيصه، إذ يتمأخذ قيمة المتوسط لكل معيار ROUGE من بين المعايير التالية ROUGE-1, ROUGE-2, ROUGE-L على الأمثلة الـ 15، وكانت النتائج على الشكل التالي:



الشكل 35- مقارنة بين النماذج الأشهر في التلخيص.

نلاحظ مما سبق أن نموذج bart-larg-cnn هو الذي يعطي أعلى النتائج على جميع المقاييس، لذلك سيتم استخدام النسخة الأنسع منه وهو نموذج sshleifer/distilbart-CNN-12-6 وهو النسخة الأصغر والأنسع من نموذج bart، إذ تم فيه تقليل عدد الطبقات من 12 إلى 6 وعدد المعلمات من 406 إلى 306 مليون، مما جعل الاداء أسرع بقدر 1.68 مقارنة بنموذج bart-larg-cnn، مع الحفاظ على الاداء نفسه في مهام تلخيص CNN/dailyMail، مما ادى إلى الحصول على نموذج محسن من حيث السرعة والحجم المنخفض، تم تدريسه على مجموعة بيانات [40] abisee/cnn_dailymail ومجموعة بيانات edinburghNLP/xsum.

2.5.5- التعرف على محتوى قطع الصور الناتجة عن التقاطع

يوجد العديد من النماذج التي تعمل على توصيف الصور الموجودة، ولتحقيق ذلك تم استخدام نموذج salesforce/blip-blip image-captioning-base على بنية المحولات المطورة والتي اظهرت أداءً استثنائيًّا عبر مختلف مهام البرمجة اللغوية العصبية ورؤبة الكمبيوتر ومن خلال تدريسيه على كميات هائلة من أزواج التسميات التوضيحية للصور، إذ تم التدريب على مجموعة بيانات coco مع استخدام بنية قاعدة ViT.

Extracted Descriptive Text: a woman with arabic calligraphy



الشكل 36- مثال لتصويف المموج المستخدم بصورة.

3.5.5 تحويل النصوص وتصويف الصور إلى صوت

استخدام مكتبات Python:

gTTS (Google Text-to-Speech) لتحويل النصوص إلى صوت لتوفير الوصول للأشخاص ذوي الاعاقة البصرية إلى المحتوى.

خلاصة: تتضمن التقنيات المتعددة لاستخراج المعلومات من صفحات الويب تقطيع النصوص والصور، تلخيص النصوص باستخدام نماذج مثل Distil BART، التعرف على محتوى الصور عبر نماذج مثل BLIP (Bootstrapping) كل من هذه التقنيات تساهم في تحسين معالجة المعلومات وتجربة المستخدم في التفاعل مع المحتوى الرقمي.

الفصل السادس

تجيز النظام

نقدم في هذا الفصل مراحل تطوير وتجيز النظام، ونعرض بدأياًً توصيفاً لبيئة العمل التي تم تجيز النظام ضمنها، وأدوات التطوير المستخدمة، ثم نشرح مراحل التنفيذ العملي لمكونات النظام بالإضافة إلى تطبيق الوب.

1.6 - مقدمة

في هذا الفصل، سنستعرض تطوير نظام برمجي مخصص لتمكين المستخدمين ضعاف البصر من تصفح صفحات الويب بشكل أكثر فعالية. يرتكز هذا النظام على مجموعة الخوارزميات المصممة لتقطيع صفحات الويب إلى أجزاء يمكن معالجتها بسهولة بواسطة تقنيات التعلم العميق المتعددة. سنقوم بتغطية كيفية تصميم وتنفيذ هذه الخوارزميات التي تم ذكرها في الفصل السابق، وكيفية إنشاء تطبيق ويب يغلف تلك الخوارزميات لتمكين المستخدم من إعطاء عنوان الصفحة المراد استخراج المعلومات منها، وإعطاء المستخدم إمكانية اختيار الطرق والتقنيات التي يريد استخدامها في عملية استخراج المعلومات.

2.6 - بيئة العمل والمكتبات المستخدمة

تعتبر لغة البرمجة بايثون من أبسط اللغات وأكثرها قوة وجذباً للمطورين. هي لغة برمجة عالية المستوى ومفتوحة المصدر، تميز بسهولة الكتابة والقراءة مقارنة باللغات الأخرى، وتستخدم أسلوب البرمجة الكائنية Object Oriented Programming، حيث يحتاج البرنامج المكتوب بها إلى تفسير باستخدام مفسر قبل التنفيذ. تم إنشاء هذه اللغة في مركز (CWI) Centrum Wiskunde & Informatica في مدينة أمستردام في أواخر ثمانينيات القرن الماضي. ووفقاً لـ IEEE Spectrum لعام 2021، صنفت بايثون كلغة البرمجة الأكثر نمواً وتطوراً [41] ، وذلك لأنها تتيح للمطورين إضافة تعديلات ومكتبات بسهولة وسلامة كبيرة. تكتسب بايثون قوتها من احتوائها على مجموعة كبيرة من المكتبات التي تسهل العمل في مجال تعلم الآلة، حيث يمكن من خلال هذه المكتبات بناء شبكات عصبية بمحظوظ طبقاتها وأنواعها. هذه الميزات والمكتبات التي تقدمها لغة البرمجة بايثون هي التي دفعتنا لاختيارها لإنجاز هذا العمل. فيما يلي تعريف بأهم المكتبات، الأدوات، وبيئات العمل المستخدمة في هذا المشروع.

Selenium - 1.2.6

هي أداة مفتوحة المصدر تستخد لآلة التفاعل مع متصفحات الويب، مما يسمح بالتحكم والتشغيل التلقائي لهذه المتصفحات، واستخراج المحتوى والبيانات من موقع الويب المختلفة. تتخصص Selenium بشكل رئيسي في إجراء اختبارات لصفحات الويب، حيث تقدم إطار عمل متكملاً لفحص واختبار نظم الويب [42] تكمن أهميتها في أنها تحاكي سلوك المستخدم الحقيقي لموقع الويب، حيث تعمل عن طريق تشغيل المتصفح آلياً، وتحميل الموقع، ومن ثم

إجراء عمليات مختلفة ومتعددة عليه، مثل تبعة البيانات في حقول معينة، وقراءة محتوى صفحات الموقع، وإيجاد عناصر هذه الصفحات، وإمكانية التنقل بينها والعديد من العمليات الأخرى.

Beautiful Soup -2.2.6

هي مكتبة مفتوحة المصدر تستخدم لتحليل الملفات التي تعتمد في هيكلها على DOM، مثل مستندات HTML و XML. يمكن استخدامها لاستخراج البيانات من صفحات الويب، حيث تقوم بإنشاء شجرة تحليل للصفحات المطلوبة، مما يجعل من الممكن الوصول إلى كل عنصر ضمن هذه الصفحة وأخذ البيانات المطلوبة منها [43].

Asyncio -3.2.6

هي مكتبة مدمجة في Python تُستخدم لإدارة البرمجة غير المتزامنة (Asynchronous Programming) وإدارتها. تم تقديمها في Python 3.3 وازدادت أهميتها في الإصدارات اللاحقة. تتيح المكتبة للمطوريين كتابة برامج تعامل مع عمليات غير متزامنة أو متعددة المهام بطريقة أكثر كفاءة.

Puppeteer -4.2.6

هي مكتبة في بايثون تُستخدم للتفاعل مع متصفحات الويب بشكل برمجي. تقوم بتوفير واجهة برمجية تتيح للمطوريين التحكم في متصفح Chromium أو Google Chrome عبر برمجة Python. يمكن استخدام Puppeteer لأتمتة المهام مثل اختبار الواجهات، تجميع البيانات، والتفاعل مع التطبيقات الويب.

TfidfVectorizer -5.2.6

هو أداة في مكتبة Scikit-learn تستخدم لتحويل النصوص إلى مصفوفة من القيم العددية باستخدام تقنية TF-IDF. تقوم بحساب مدى أهمية كل كلمة في النصوص بناءً على ترددتها في كل مستند وتوزيعها في كل المستندات. هذه العملية تساعد في استخراج الأنماط والمعلومات الهامة من البيانات النصية، مما يجعلها مفيدة في تحليل النصوص وفهم محتوياتها بشكل أفضل [44].

NLTK (Natural Language Toolkit) -6.2.6

هي مجموعة من البرامج لمعالجة اللغات الطبيعية الرمزية والإحصائية، وتعود من المكتبات الأساسية للتعامل مع البيانات النصية وتحليلها. توفر حزم برمجية سهلة الاستخدام تدعم مجالاً كبيراً من المعالجة النصية مثل التصنيف، والتترميز،

والتحليل. تهدف NLTK إلى دعم البحث في معالجة اللغات الطبيعية أو المجالات ذات الصلة بها، بما في ذلك استرجاع المعلومات، والتعلم الآلي [45].

Numpy (Numeric Python) -7.2.6

هي حزمة للحسابات العلمية ضمن لغة Python. توفر إمكانية إجراء حسابات سريعة وفعالة للمصفوفات متعددة الأبعاد، وتتوفر أيضًا أدوات لقراءة وكتابة البيانات المعتمدة على هذه المصفوفات. بالإضافة لما سبق، توفر Numpy تعليقًا للبيانات لتسهيل نقلها بين الخوارزميات المختلفة. تعد أكثر فعالية لتخزين وإجراء العمليات على البيانات الرقمية مقارنة ببني المعطيات الأخرى المعرفة ضمن لغة Python [46].

Pandas -8.2.6

هي حزمة مفتوحة المصدر تستخدم على نطاق واسع في علوم البيانات وتحليلها. تتصف بسرعة الأداء والمونة وسهولة الاستخدام، وتتوفر هيكل البيانات والعمليات اللازمة لمعالجة الجداول العددية والسلسل الرممية. تُستخدم في مهام التعلم الآلي لأنها توفر أدوات لفهرسة مجموعات البيانات، وقراءتها، وكتابتها ضمن هيكل البيانات الموجودة في الذاكرة بتنسيقات مختلفة [47].

Matplotlib -9.2.6

هي مكتبة تصميم رسوميات خاصة بلغة Python. يمكن من خلالها توليد رسوم بيانية وأشكال خاصة بالبيانات. وتدعم الرسم التفاعلي وغير التفاعلي، ويمكن من خلالها حفظ الصور والرسوم البيانية. تمكن المطوريين من رسم العديد من المخططات المختلفة، مثل الرسوم الخطية، والرسوم العمودية، والرسوم الدائرية، والمخططات [48].

Visual Studio Code (VS Code) -10.2.6

هي بيئة تطوير متكاملة (IDE) مفتوحة المصدر تم تطويرها بواسطة مايكروسوفت. تُعد واحدة من أكثر بيئة التطوير شيوعًا بين المطوريين بفضل مرونتها ودعمها للعديد من لغات البرمجة. يُعد VS Code خيارًا ممتازًا لمشاريع تعلم الآلة وتحليل البيانات بفضل دعمه القوي لـ Python وعدد كبير من الإضافات المتعلقة بهذا المجال، مثل Jupyter Notebooks، والتي تسمح للمطوريين بكتابة وتشغيل أكواد تحليل البيانات وتعلم الآلة داخل البيئة نفسها.

Google Colab -11.2.6

هي بيئة عمل برمجية لغة Python تعمل في المتصفح، تقدمها Google بشكل مجاني محدود كخدمة سحابية. بحيث يمكن إجراء كمية محددة من الحسابات (تدريب الشبكات العصبية) عن بعد باستخدام أجهزة ذات مقدرات حاسوبية عالية. تُعد مناسبة بشكل خاص لتعلم الآلي وتحليل البيانات، حيث توفر إمكانية تشغيل الأكواد البرمجية مباشرة في السحابة مع دعم لتقنيات مثل GPU (Graphics Processing Unit) و TPU (Tensor Processing Unit) لتسريع العمليات الحاسوبية.

Requests -12.2.6

تُستخدم لإرسال طلبات HTTP (Hypertext Transfer Protocol) بسهولة وبساطة، مما يتيح الحصول على البيانات من الإنترنت أو إرسالها إلى الخوادم. تميز المكتبة بواجهة سهلة الاستخدام وتدعم إدارة الجلسات، والتعامل مع ملفات الكوكيز، وإرسال البيانات بتنسيقات مختلفة مثل JSON ورفع الملفات. هي أداة قوية ومفضلة للتعامل مع بروتوكول HTTP في تطبيقات بايثون.

Scikit-learn -13.2.6

هي مكتبة برمجية مفتوحة المصدر للتعلم الآلي في Python. تقدم مجموعة واسعة من الأدوات والخوارزميات للتعلم الآلي، بما في ذلك التصنيف، والتجميع، والتحليل النمطي، والتعلم الشبكي، وغيرها من التقنيات الهامة في مجالات البيانات الكبيرة وعلم البيانات. تميز بسهولة الاستخدام والقدرة على التكامل مع بيئة Python، مما يجعلها خياراً شائعاً بين المطورين والمهندسين الذين يعملون في مجال التحليل البياني والتعلم الآلي.^[49]

Torch -14.2.6

مكتبة Torch هي إطار عمل للتعلم العميق والحوسبة العلمية، تم تطويرها بواسطة Facebook's AI Research (FAIR) Lab. تقدم Torch بنية قوية وفعالة لبناء وتدريب الشبكات العصبية الصناعية. تميز Torch بدعم قوي للتباين التلقائي، مما يسهل عملية تطوير النماذج وتعديلها بسهولة. تتيح Torch أيضاً استخدام وحدات معالجة الرسومات (GPUs) لتسريع عمليات التدريب والتحسين. بالإضافة إلى ذلك، تتمتع Torch بمجتمع نشط من المطورين والباحثين، مما يسهم في تحديث وتطوير المكتبة بشكل مستمر وتقديم دعم فني شامل.

PIL -15.2.6

مكتبة PIL (Python Imaging Library) هي مكتبة برمجية للتعامل مع الصور في لغة Python. تتيح PIL للمطورين قدرات متقدمة لمعالجة الصور بسهولة وفعالية، مثل فتح وحفظ الصور بتنسيقات مختلفة، وتغيير حجم

الصور، وتعديل الألوان، وإضافة الفلاتر، والعديد من العمليات الأخرى، تأتي PIL بجموعة واسعة من الدوال والأدوات التي تساعد في التلاعب بالصور، سواء كانت رقمية أو ماسحة ضوئياً. على الرغم من أنها مكتبة قديمة إلى حد ما، إلا أنها لا تزال تستخدم بشكل واسع بسبب سهولة استخدامها وموتها في التعامل مع مجموعة متنوعة من العمليات الصورية.

Scipy -16.2.6

هي مكتبة برمجية مفتوحة المصدر مخصصة للحوسبة العلمية والهندسية في Python. تعتبر Scipy جزءاً لا يتجزأ من البيئة العلمية في Python بجانب مكتبات مثل NumPy وMatplotlib، تقدم Scipy مجموعة واسعة من الوظائف والأدوات لحل مشاكل علمية معقدة، مثل التكامل العددي، وحل المعادلات التفاضلية، والتحسين، ومعالجة الإشارات، وتحليل البيانات الإحصائية، وغيرها، تعتمد Scipy على NumPy للعمل مع بيانات الأرقام الصحيحة والمصفوفات بشكل فعال، تتميز Scipy بسهولة الاستخدام والقدرة على التكامل الوثيق مع مكتبات أخرى في البيئة العلمية بما في ذلك Matplotlib لرسم البيانات وPandas لتحليل البيانات. بفضل مجتمعها النشط والمطوريين الملتزمين، تحصل Scipy على تحديثات منتظمة ودعم تقني مستمر لتلبية احتياجات الباحثين والمهندسين في مجالات العلوم والهندسة.[50]

Imutils -17.2.6

هي مكتبة Python خفيفة الوزن تسهل عمليات معالجة الصور، وتبسيط استخدام مكتبة OpenCV. توفر imutils وظائف مريحة للتغيير حجم الصور، تدويرها، قصها، ومعالجتها بشكل عام دون الحاجة إلى كتابة الكثير من الكود. تساعد هذه المكتبة المطوريين في تنفيذ المهام الروتينية في معالجة الصور بسرعة وكفاءة. بفضل سهولة الاستخدام والتكامل السلس مع OpenCV، أصبحت imutils خياراً شائعاً بين المطوريين في مجالات مثل التعلم الآلي والرؤية الحاسوبية.

Transformers -18.2.6

مكتبة Transformers هي واحدة من أشهر المكتبات في مجال معالجة اللغة الطبيعية باستخدام نماذج الذكاء الصنعي. توفر المكتبة مجموعة واسعة من النماذج المدرية مسبقاً، مثل BERT وGPT، وتدعم مجموعة متنوعة من التطبيقات مثل الترجمة، التحليل العاطفي، وتوليد النصوص. من أبرز ميزاتها واجهة "pipeline" التي تسهل استخدام النماذج لأغراض محددة بدون الحاجة إلى الغوص في تفاصيل الكود. تشمل المكتبة مكونات متعددة مثل:

BlipProcessor : لتحويل البيانات النصية والصور إلى تنسيق قابل للاستخدام في نموذج **BlipForConditionalGeneration** ، الذي يستخدم لتوصيف الصور أو الإجابة على الأسئلة.

BART Tokenizer : لتحويل النصوص إلى رموز قابلة للمعالجة بواسطة نموذج **BartForConditionalGeneration** ، الذي يستخدم في توليد النصوص وتصحيحها.

T5 Tokenizer : لتحويل النصوص إلى رموز يمكن معالجتها بواسطة نموذج **T5ForConditionalGeneration** ، الذي يستخدم في المهام المتنوعة مثل الترجمة وتلخيص النصوص.

Pegasus Tokenizer : لتحويل النصوص إلى رموز يتم معالجتها بواسطة نموذج **PegasusForConditionalGeneration** ، الذي يستخدم في توليد النصوص وتلخيصها.

تُعد مكتبة Transformers أداة قوية ومرنة لمجموعة واسعة من مهام معالجة اللغة الطبيعية، وتتوفر دعماً متميزاً لها مهام متنوعة بفضل النماذج والتقنيات المتقدمة التي تشملها [51].

gTTS (Google Text-to-Speech) -19.2.6

تتيح تحويل النصوص إلى صوت باستخدام API، تعد أداة مفيدة لإنشاء ملفات صوتية من النصوص وتتميز بسهولة استخدامها، إضافة لكونها تدعم العديد من اللغات.

Pytesseract -20.2.6

هي واجهة برمجية (API) لـ Tesseract OCR (Optical Character Recognition) ، وهو محرك متقدم لتحويل النصوص المطبوعة أو المكتوبة يدوياً إلى نصوص قابلة للتحرير عبر التعرف البصري على الحروف . توفر pytesseract طريقة سهلة لدمج Tesseract OCR مع برامج Python ، مما يجعل من السهل استخراج النصوص من الصور.

SpaCy -21.2.6

مكتبة SpaCy هي أداة مفتوحة المصدر لمعالجة اللغة الطبيعية بلغة Python، وتعُرف بسرعتها وكفاءتها في تحليل النصوص. توفر نماذج عالية الجودة تدعم العديد من اللغات وتتميز بدقة في تحليل النصوص وتصنيف الكيانات المسماة. [52]

3.6 التنفيذ العملي

بالاعتماد على بيئة العمل والادوات السابقة، تم تنجيز مكونات النظام وفق المراحل التالية:

1.3.6 الخوارزمية المقترنة لتقسيط صفحات الويب باستخدام الطرق التقليدية

يتم بدايةً الحصول على صفحة الويب باستخدام webdriver الذي توفره مكتبة selenium، بناءً على الرابط المدخل من قبل المستخدم، بعد ذلك يتم القيام بعدة عمليات لتطبيق الخوارزمية وإنجاز التقسيط:

1. حقن المعلومات داخل الرمaz المصدرri لصفحة الويب

لتنفيذ عملية إضافة السمتين "data-bbox" و "data-cleaned" و "data-style" إلى كل علامة (tag) داخل الرمز المصدرri للصفحة، سنحتاج إلى استخدام مكتبة Selenium لاستخراج عناصر الصفحة وإضافة السمتين المطلوبتين. يمكن تحقيق ذلك باتباع الخطوات التالية:

- **تحميل الصفحة باستخدام Selenium:** Selenium نستخدم Selenium لتحميل الصفحة والحصول على الكائنات التي تمثل العناصر المختلفة في الصفحة.
- **إضافة السمة "data-bbox":** ستحسب موقع العلامة ومساحتها (بداية العلامة، ارتفاعها، عرضها، ومساحتها)، ثم نضيف السمة "data-bbox" إلى كل علامة.
- **إضافة السمة "data-xpath":** سنستخدم Selenium لحساب المسار الكامل (XPath) لكل علامة ونضيف هذه المعلومات كسمة "data-xpath" للعلامة.
- **إضافة السمة "data-style":** استخراج الأنماط النمطية (CSS styles) المطبقة على كل عنصر، وإضافتها كقيمة للسمة.
- **إضافة السمة "data-cleaned":** لتنفيذ عملية فلترة العلامات التي ليس لها تأثير مئوي داخل صفحة الويب، سنقوم بتحديد العلامات التي تستوفي أحد الشروط التالية:
 - مساحة العلامة تساوي الصفر: نحتاج للتحقق مما إذا كانت مساحة العلامة صفرًا باستخدام سمة "data-bbox" التي تحدد موقع العلامة ومساحتها.
 - وقوع العلامة خارج الصفحة: يجب التتحقق من موقع العلامة ضمن صفحة الويب لمعرفة ما إذا كانت خارج الحدود الظاهرة للصفحة.

- العلامة مخفية: يجب التحقق مما إذا كانت العلامة مخفية باستخدام الخصائص CSS المعروفة

hidden: true, display: none, visibility: hidden مثل

كما هو موضح بالشكل:

```
StackExchange.gps.init(false);
;
pt>
ipt id="noscript-css" data-bbox="0 673.5999755859375 0 0" data-style="accent-color:auto;align-content:normal;align-
    data-xpath="/html/head/noscript" data-cleaned="true"><style>body , s-topbar{margin-top:1.9em}</style></noscript>
t src="https://accounts.google.com/gsi/client" async="" data-bbox="0 673.5999755859375 0 0" data-style="accent-col-
netrust-consent-sdk #onetrust-banner-sdk {background-color: #252627; }
#onetrust-consent-sdk #onetrust-policy-title,
```

الشكل 37 - مثال للمعلومات المخونة.

2. تطبيق الخوارزمية

في البداية نقوم بأخذ جميع محتويات الصفحة من الرماد المصدري باستخدام webdriver وبعدها مكتبة Beautiful

Soup، مما يمكننا من الوصول إلى أي علامة عن طريق نوعها أو مسارها.

بعد ذلك نقوم بتطبيق التقاطع ويكون الخرج في هذه المرحلة هو ملف json، يحتوي على القطاعات التي جرى

تقاطع الصفحة إليها ومعلومات كل قطاع، حيث يحوي كل قطاع ضمه على عدة معلومات وهي:

• **TagName**: نوع العلامات التي يضمها القطاع.

• **X, Y**: يمثلان إحداثيات نقطة بداية القطاع.

• **Width**: يمثل عرض القطاع.

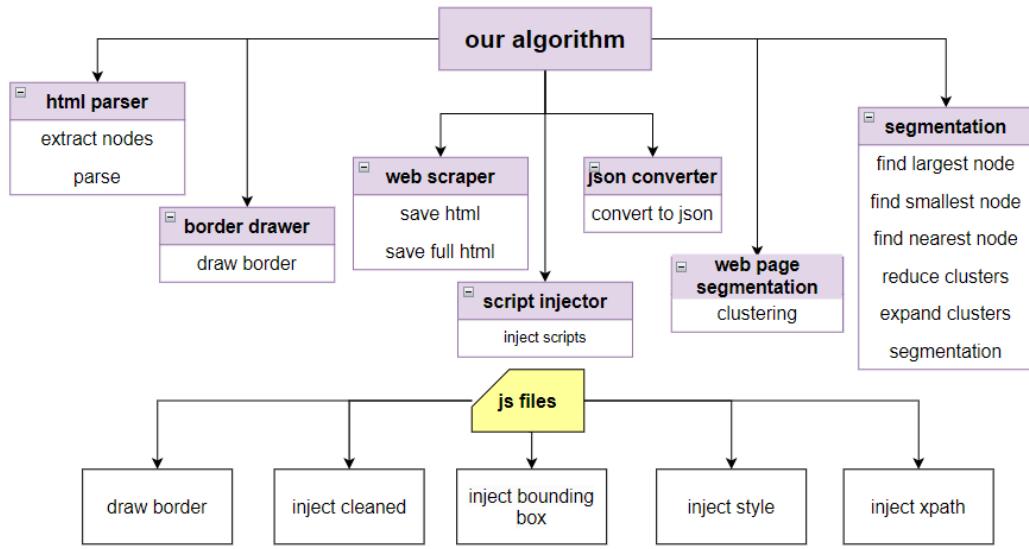
• **Height**: يمثل ارتفاع القطاع.

• **XCenter, YCenter**: يمثلان إحداثيات مركز القطاع.

• **Text**: يحوي النص الذي يحتويه القطاع.

• **XPath**: يمثل المسار إلى عقدة القطاع.

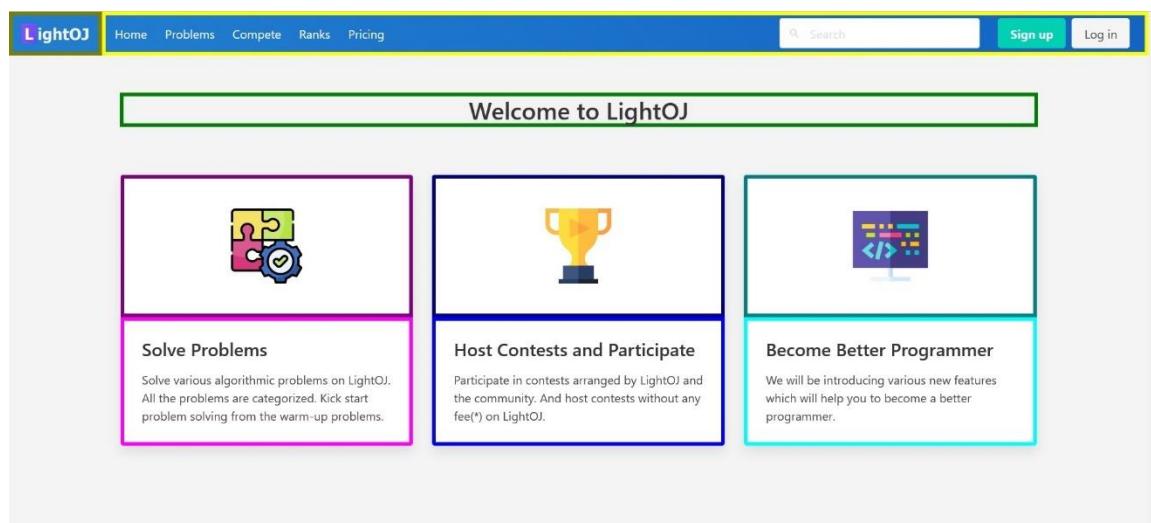
تم تصميم الكود وفق المخطط التالي:



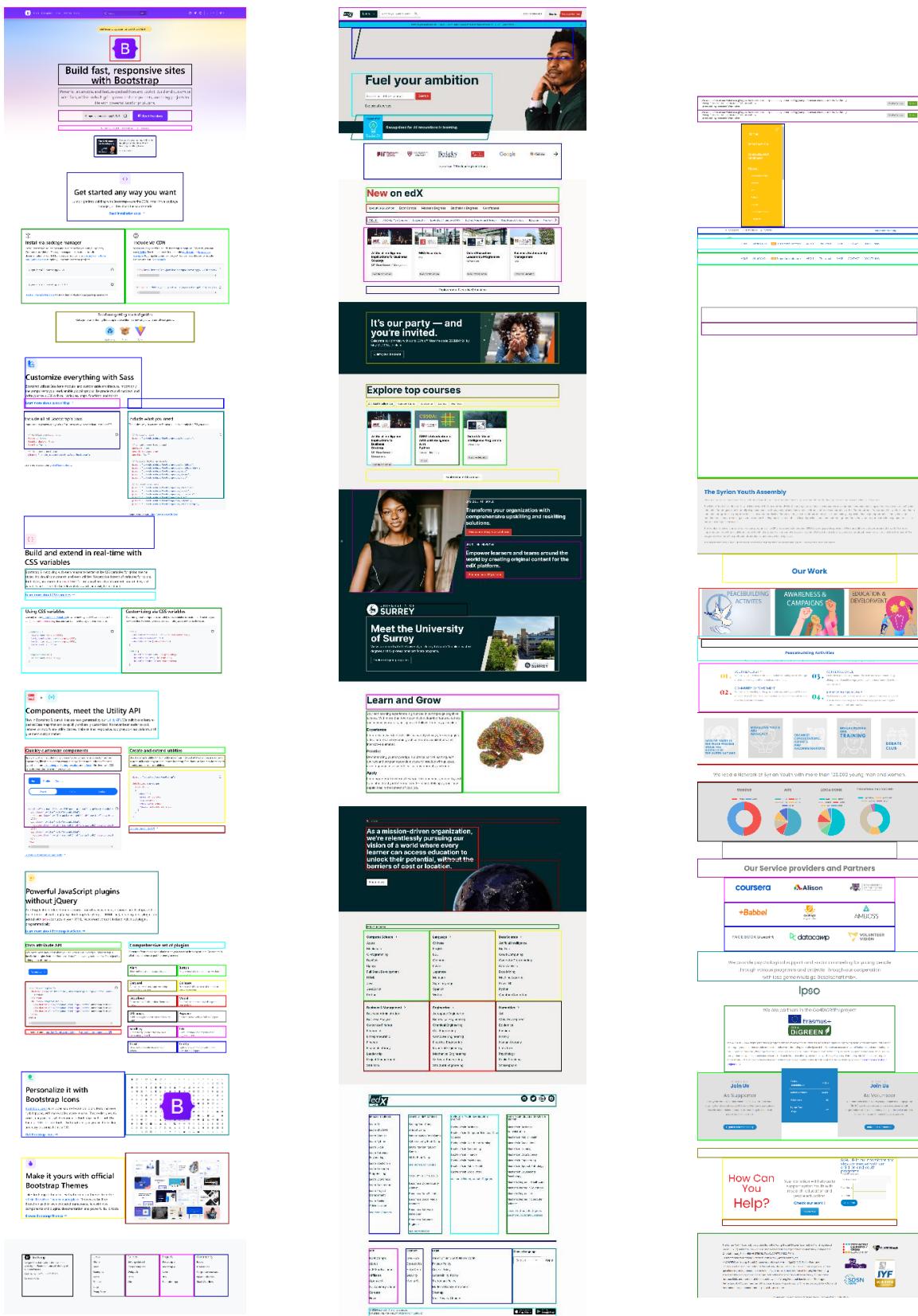
الشكل 38- بنية الرمaz المستخدم لتنجيز الخوارزمية.

فيما يلي تقاطع بعض الصفحات الشهيرة باستخدام الخوارزمية المقترنة

(LightOJ, edX, Bootstrap, W3Cs, Syrian youth org)



الشكل 39- ناتج تطبيق الخوارزمية على موقع LightOJ



الشكل 40 - ناتج تطبيق الخوارزمية على عدة مواقع.

日本語 | ページ | 関連中文版

W3C Standards Groups Get involved Resources News and events About Sign in

Making the Web work

The World Wide Web Consortium (W3C) develops standards and guidelines to help everyone build a web based on the principles of accessibility, internationalization, privacy and security.

[Read more about W3C](#)

Working with stakeholders of the Web

A range of organizations join the World Wide Web Consortium as Members to work with us to drive the direction of core web technologies and exchange ideas with industry and research leaders. We rotate randomly a few of our Member organizations' logos underneath.

The Web Consortium and its members, with help from the public and the web community, focus on a range of business ecosystems that the Web transforms, including E-commerce, Media & Entertainment, Publishing and several other areas.

Discover how W3C supports these organizations

Web standards

Web standards are the building blocks of a consistent digitally connected world. They are implemented in browsers, blogs, search engines, and other software that power our experience on the Web.

W3C is an international community where [Members](#), full-time staff, and the public work together to develop web standards.

[Learn more about Web standards](#)

Get involved

W3C works at the nexus of core technology, industry needs, and societal needs. Everyone can get involved with the work we do.

There are many ways individuals and organizations can participate in the Web Consortium to advance web standardization.

[Ways to get involved](#)

Latest news

Latest entries from across our News, Press Releases or Blog:

- First Public Working Draft: Controller Documents 1.0**
A controller document is a set of data that specifies one or more relationships between a controller and a set of data, such as a set of public cryptographic keys.
[News](#)
- Authorized Translation of WCAG 2.2 in Catalan**
W3C published the Authorized Catalan Translation of Web Content Accessibility Guidelines (WCAG) 2.2, Directrius per a l'accessibilitat del contingut web (WCAG) 2.2.
[News](#)
- Updated W3C Recommendation: Media Queries Level 3**
The CSS Working Group has republished Media Queries Level 3 as an updated W3C Recommendation.
[News](#)
- RDF Dataset Canonicalization is a W3C Recommendation**
The RDF Dataset Canonicalization and Hash Working Group published RDF Dataset Canonicalization as a W3C Recommendation.
[News](#)

Home Contact Help Donate Legal & Policies Corporation System Status
Copyright © 2024 World Wide Web Consortium. W3C® liability, trademark and permissive license rules apply.

الشكل 41- ناتج تطبيق الخوارزمية على موقع W3Cs

3. تقييم الخوارزمية المقترنة

تم تقييم الخوارزمية المقترنة بالاعتماد على مجموعة البيانات DSpopular, DSrandom، والذى يوفر تقييم الخوارزمية وفق ثلات معايير مختلفة وهى:

- المحتوى النصي للكتلة (text).
- المسار وفق شجرة الـ DOM (xpath).
- إحداثيات بداية الكتلة مع طولها وعرضها (x, y, width, height).

إذ يتم حساب مصفوفة الالتباس Confusion matrix باعتبار أنَّ:

• **TP**: عدد الكتل المتطابقة في المجموعتين (يحصل التطابق في حال تساوي المسارين أو في حال تساوي المحتوى النصي).

• **TN**: جميع الكتل الغير موجودة في المجموعتين (0 دوماً).

• **FP**: عدد الكتل الموجودة في خوارزميتنا وغير موجودة في الكتل المخزنة بمجموعة البيانات.

• **FN**: عدد القطع الموجودة في مجموعة البيانات وغير موجودة في خوارزميتنا.

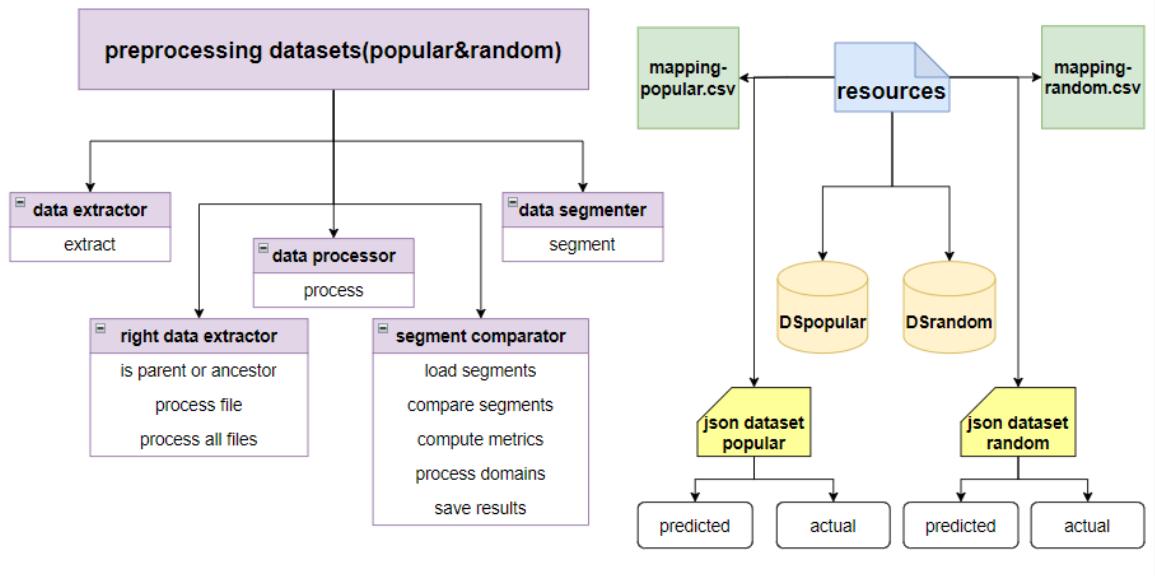
وبعد حساب:

- | | |
|-----------|---|
| Recall | • |
| Precision | • |
| F1score | • |

إذ تم استخدام 74 صفحة من random اضافة ل 60 صفحة من popular وذلك من خلال القيام بالخطوات التالية:

1. استخراج معلومات القطع الموجودة ضمن قاعدة المعطيات الموافقة للخرج الموصف للخوارزمية المقترنة كما تم توصيفه بفقرة تطبيق الخوارزمية من هذا الفصل.
2. التخلص من الكتل المتداخلة (كتلة داخل كتلة)، وأخذ الكتل الداخلية فقط.
3. تقطيع الصفحة المعطاة باستخدام الخوارزمية المقترنة وإيجاد الخرج تبعاً للصيغة المحددة.
4. التقييم على أساس الكتل الصحيحة والمشتركة في القسمين.

تم القيام بما سبق باستخدام البنية التالية:



الشكل 42- بنية الرمaz المستخدم لتنجيز الخوارزمية.

وكانت النتائج بالمقارنة مع نتائج الخوارزميات التقليدية عندما تم تقييمها على مجموعة البيانات السابقتين، على الشكل التالي:

Algo	Precision	Recall	F1-Score
BOM	30.5%	26.1%	28.1%
VIPS	23.7%	26.2%	24.9%
SegBlock	38.1%	40.2%	39.1%
Semantic-Block	40.3%	43.4%	41.8%
Fusion-Block	44.7%	54.1%	48.9%
Integrated-Block	51.7%	61.7%	52.6%
Our Algorithm	38.8%	41.3%	40%

جدول 7 - نتائج تقييم الخوارزمية المستخدمة على مجموعة بيانات Popular.

Algo	Precision	Recall	F1-Score
BOM	30.8%	33.0%	31.8%
VIPS	27.8%	26.4%	27.1%
SegBlock	41.9%	44.8%	43.3%
Semantic-Block	43.3%	53.6%	48.0%
Fusion-Block	49.3%	61.2%	54.6%
Integrated-Block	56.6%	67.5%	61.5%
Our Algorithm	57.6%	47.9%	52.3%

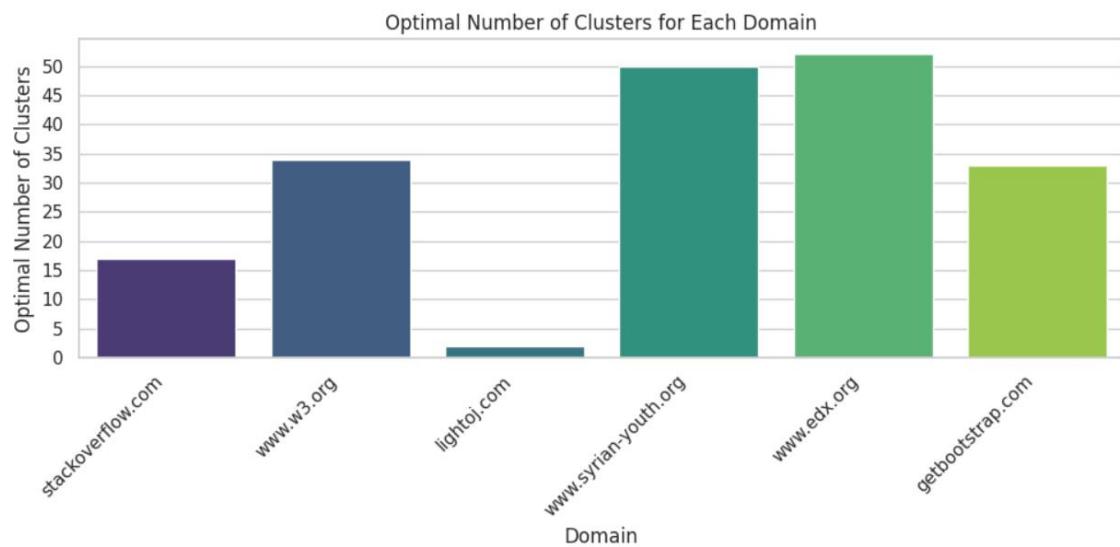
جدول 8 - نتائج تقييم الخوارزمية المستخدمة على مجموعة بيانات Random.

2.3.6- الخوارزميات المتبعة للتتبؤ بعدد القطاعات الموجودة بصفحة الويب (اعتماداً على تحليل النصوص)

تم تنجيز طريقة تحليل الصور الظلية المذكورة في الفصل السابق من خلال استخدام عدة صفوف وهي:

- URLContentExtractor: وهو الصنف المسؤول عن استخراج المحتوى النصي لصفحة الويب من خلال جلب محتوى صفحة الويب باستخدام مكتبة requests ثم تحليل هذا المحتوى باستخدام مكتبة BeautifulSoup واستخراج النصوص من العناصر المحددة وإضافتها للقائمة texts وارجاع هذه القائمة.
- TextProcessor: يُستخدم لتحضير النصوص الخام عن طريق تنظيفها وتحويلها إلى شكل يمكن استخدامه بسهولة في التحليل النصي أو نماذج التعلم الآلي. تشمل الخطوات إزالة الفراغات الزائدة، الأرقام، تحويل الأحرف إلى صغيرة، إزالة كلمات التوقف، وتطبيق التجذير على الكلمات.
- DomainClusterer: يستخدم الصفين السابقين ثم يقوم بتحويل النصوص إلى ميزات رقمية باستخدام TfidfVectorizer مما يسمح بتمثيلها بشكل يمكن استخدامه في نمذجة التجميع، حيث يتم استخراج السمات الأساسية المشتركة بجميع الفقرات المستخرجة وإعطاء كل سمة مقدار للأهمية وفق مقياس tf-idf ضمن كل فقرة موجودة، مما يشكل مصفوفة ثنائية أعمدتها هي السمات المستخرجة وأسطرها هي الفقرات النصية إذ تشكل القيمة الداخلية في الخلية $z_{i,j}$ هي أهمية السمة z ضمن الفقرة i ، ليتم تمثيل كل فقرة بشعاع من القيم، بعد ذلك يتم عقدة الفقرات تبعاً لشعاع كل فقرة، ثم يقوم بحساب درجة الصورة الظلية لقطعات صفحة الويب المقطعة باستخدام KMeans لعدة مرات في كل مرة يتم اعتماد عدد قطع مختلف بدءاً من استخدام عنقودين وصولاً لاستخدام العدد الاعظمي الممكن وجوده في صفحات الويب، ثم أخذ العدد الموافق لقيمة العظمى لدرجة الصورة الظلية المحسوبة.

تم تطبيق الخوارزمية السابقة على عدد من الصفحات التي تم تقطيعها في الخوارزمية المقترنة السابقة والتي تظهر في الشكل التالي:



الشكل 43- عدد القطع المتبع به لمجموعة من المواقع.

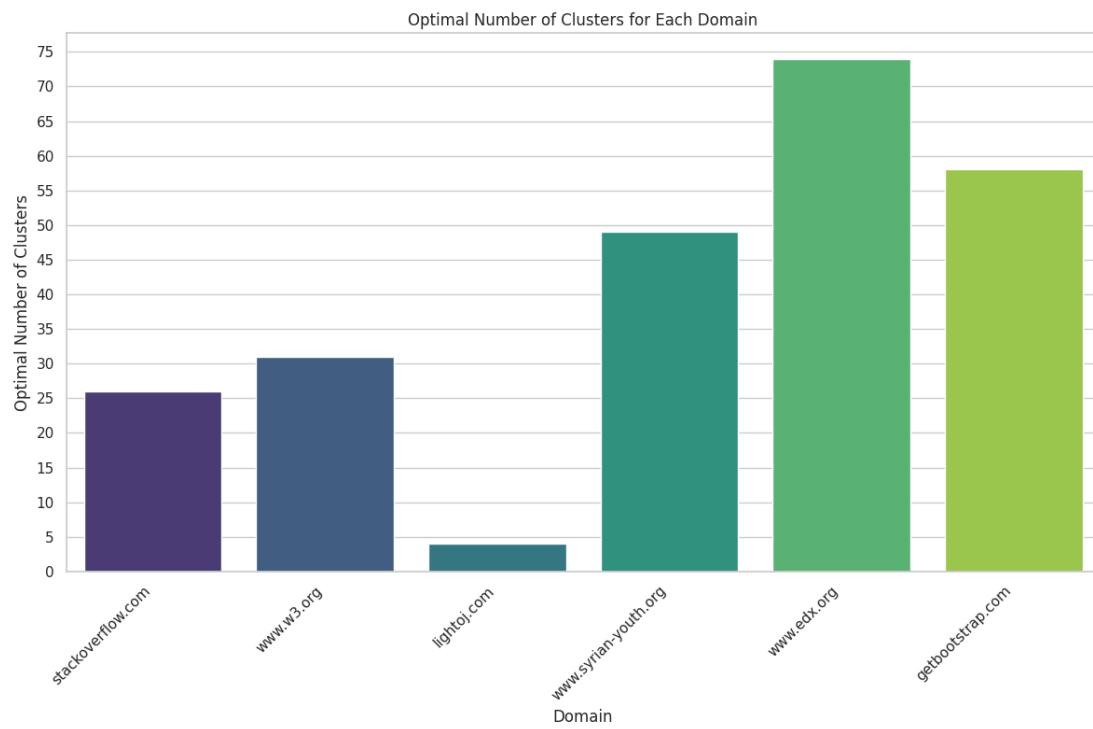
3.3.6- الخوارزميات المتبعة للتنبؤ بعدد القطاعات الموجودة بصفحة الويب (اعتماداً على تحليل الصورة)

تم تتجيز طريقة تحليل الصورة المذكورة في الفصل السابق من خلال استخدام عدة صنوف وهي:

- ScreenshotFetcher: الهدف هو جلب لقطات شاشة لصفحات الويب وذلك عن طريق مفتاح API المقدم وعنوان URL للصفحة المراد التقاط لقطة شاشة لها، إذ يتم إرسال طلب GET إلى رابط API باستخدام مكتبة requests.
- EdgeDetector: يهدف إلى الكشف عن الحواف في الصورة والعنود على الأشكال وتحديد مراكزها، إذ يقوم بدايةً بكشف حواف الصورة ثم استخراج الأشكال contours من الحواف المكتشفة ثم استخراج مراكز الأشكال الموجودة.
- Clustering: يتم في هذا الصدد تنفيذ خوارزمية DBSCAN، والحصول على العناقيد الموجودة ثم حساب مركز كل عنقود وارجاع قيمها.

- **SamSegmentation**: خص لقطع الصور باستخدام نموذج SAM ويشمل عمليات التحميل والتصدير للنموذج بتنسيق ONNX وتنفيذ عمليات التقطيع بناءً على مراكز العناقيد، ليتخرج قناع عملية التقطيع وتحسب المجموعات المتبقية بعد تصفية المجموعات الصغيرة بناءً على عتبة المساحة ويكون العدد الناتج هو العدد المتوقع لعدد العناقيد اللازمة لقطع صفحة الويب.
- **MainProcessor**: يأخذ عناصر التحكم الرئيسية كمتغيرات داخلية، مثل رابط الصفحة(url) ، مفتاح الواجهة (api_key) ، ونقطة التحقق (checkpoint) للنموذج، ليقوم بتنسيق عملية متكاملة لمعالجة صورة.

تم تطبيق الخوارزمية السابقة على عدد من الصفحات التي تم تقطيعها في الخوارزمية المقترنة السابقة والتي تظهر في الشكل التالي:



الشكل 44- عدد القطع المتبع به لمجموعة من المواقع.

4.3.6- الخوارزمية المقترنة لقطع صفحات الويب باستخدام تقنيات التعلم العميق

يتم تنظيم خطوات الخوارزمية باستخدام الصف MainProcessor وفق عدة خطوات وهي على الشكل التالي:

- الخطوة الأولى: أخذ لقطة شاشة لصفحة الويب المراد تقطيعها وعرضها والاحتفاظ بنسخة من اللقطة كصورة مرجعية، يتم ذلك باستخدام الصف Screenshot Fetcher وفق التالي:

- يتم أخذ API key مفتاح API من قبل المستخدم لتكوين الرابط من المفتاح وعنوان url لصفحة الويب المراد تقطيعها.
- إرسال طلب get إلى API وذلك باستخدام مكتبة requests.
- التحقق من الاستجابة الناجحة وفقاً ل status code.
- تحميل الصورة من محتوى الاستجابة باستخدام PIL لمعالجة وتحميل الصور.
- تحويل الصورة إلى مصفوفة NumPy للتعامل مع الصورة كمصفوفة.
- تحويل الصورة من تنسيق BGR إلى RGB باستخدام مكتبة OpenCV.
- ثم إعادة الصورة في تنسيق RGB كمصفوفة NumPy.

الخطوة الثانية: الكشف عن الحواف واستخراج المتجهات المركزية للأشكال الموجودة في صورة باستخدام الصيغ EdgeDetector وفق التالي:

- تحويل لقطة الشاشة لصفحة المراد تقطيعها إلى التدرج الرمادي.
- تطبيق مرشح غاوسي للتخلص من الضجيج.
- الكشف عن الحواف باستخدام خوارزمية Canny.
- العثور على المتجهات contours للصورة الناتجة مع الحواف باستخدام مكتبة cv2.
- استخراج المتجهات المركزية centroids للأشكال التي مساحتها أكبر من حد أدنى يتم تعينه (10).

الخطوة الثالثة: عملية تجميع النقط الناتجة عن الخطوة السابقة باستخدام خوارزمية DBSCAN للحصول على المميز في الصورة من خلال الصيغ Clustering وفق التالي:

- تحويل قائمة المتجهات المركزية إلى مصفوفة NumPy.
- تنفيذ التجميع باستخدام خوارزمية DBSCAN لكشف المناطق الحاوية على كثافة عالية.
- استخراج عدد العنقود الناتجة من عملية العنقدة السابقة.
- حساب مركز كل عنقود عن طريق أخذ متوسط إحداثيات نقط كل عنقود.

الخطوة الرابعة: القيام بتحويل الصورة المراد تقطيعها إلى صورة ثنائية بناءً على عتبة تكيفية مما يسهل التعرف على الأشكال والتعرف على المعلومات الهامة.

الخطوة الخامسة: إدخال الصورة بشكلها الثنائي، بالإضافة إلى النقاط المميزة وهي مراكز العناقيد الناتجة كدخل لنموذج SAM، وذلك باستخدام الصف SamSegmentation لتم عملية التقطيع كما يلي:

- تهيئة النموذج: نقوم بتحميل نقطة التحقق للنموذج 'vit_h' وتعيينه للاستخدام على وحدة المعالجة المركزية (CPU). كما نجهز متنبي SAM وجلسة تشغيل ONNX.
- تحميل نموذج ONNX: نقوم بتحميل نموذج ONNX جلسة التشغيل لتكون جاهزة للاستخدام.
- تصدير النموذج إلى تنسيق ONNX: نصدر النموذج الحالي إلى تنسيق ONNX باستخدام مدخلات وهمية لتحديد الأبعاد الديناميكية للنموذج.
- تجزئة الصورة: نقوم بتجزئة الصورة باستخدام نموذج SAM مع مراكز العناقيد المقدمة. نحدد مناطق معينة في الصورة بناءً على هذه المراكز.
- تصنيف القناع: نقوم بتصنيف القناع الناتج إلى مكونات متصلة، مما يساعدنا على التعرف على المناطق المختلفة في الصورة.
- إزالة المكونات الصغيرة: نزيل المكونات الصغيرة بناءً على عتبة المساحة المحددة، وذلك لتصفيية الأجزاء غير المهمة في القناع، وحساب عدد القطع المتبقية.
- إرجاع القناع المصفى: نعيد القناع المصفى وعدد القطع المتبقية بعد إزالة المكونات الصغيرة.
- عرض القناع الناتج: نعرض القناع الناتج عن عملية التصفية لتبين النتائج بشكل واضح.

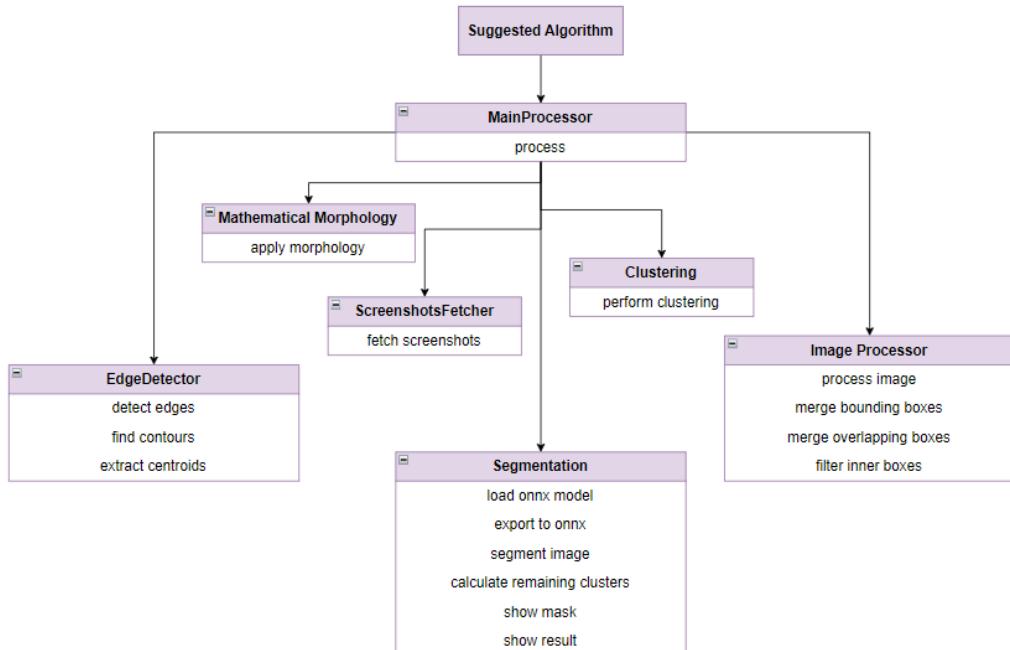
الخطوة السادسة: تطبيق العمليات المورفولوجية الرياضية على صورة الخرج الناتجة من الخطوة السابقة، وذلك لملئ الفجوات وإزالة الضوضاء، وذلك باستخدام الصف MathematicalMorphology، وذلك وفق التالي:

- تحويل الصورة إلى تدرجات الرمادي لتسهيل المعالجة.
- استخدام عتبة تكيفية لتحويل الصورة إلى صورة ثنائية حيث يتم تقطيع الصورة إلى كتل صغيرة ثم تحديد عتبة لكل كتلة بناءً على الإضاءة المحلية.
- إنشاء نواة للعمليات المورفولوجية، نواة للتقريب (close) بحجم (5x5) وأخرى للتوسيع (open) بحجم (3x3).
- تطبيق عمليات الفتح والغلق المورفولوجية على الصورة الثنائية: الفتح (open) لإزالة الضوضاء الصغيرة، الغلق (close) لملء الفجوات الصغيرة.

- عكس ألوان الصورة ثم تطبيق تقنية الملء بالفيض، وذلك من خلال عكس الألوان لتصبح الفجوات بيضاء ثم استخدام الملء للتخلص من الفجوات، ثم إعادة الصورة لشكلها الطبيعي من خلال عكس الألوان.
- دمج الصورة المغلقة الأصلية مع الصورة المملوأة بالفيض ملء جميع الفجوات.

الخطوة السابعة: معالجة الصورة الناتجة مع القناع المملوء للعثور على المربعات المحاطة النهائية وذلك من خلال الصنف وفق التالي: Image processor

- يتم تحويل صورة القناع إلى الرمادي لتسهيل المعالجة، ثم يتم تطبيق تمويه Gaussian لتقليل الضوضاء.
- يتم استخدام خوارزمية Canny لاكتشاف الحواف في الصورة، ثم يتم إيجاد الخطوط المحاطة في الصورة بناءً على الحواف المكتشفة. يتم حساب المربعات المحاطة لكل خط محيط وتحويلها إلى مصفوفة.
- يتم تجميع المربعات المحاطة بناءً على مسافة معينة باستخدام دالة fclusterdata فعندما تكون المسافة أقل من المسافة الحدية يتم الدمج.
- تصفية المربعات التي تكون داخل مربعات أخرى بناءً على معيار الانحراف المسموح به.
- دمج المربعات التي تتدافق فيما بينها إلى صندوق واحد أكبر.
- رسم المربعات الناتجة على الصورة الأصلية (لقطة الشاشة لصفحة الويب).



الشكل 45- بنية الرماز المستخدم لتنجيز الخوارزمية.

إعطاء مثال عن جزء من صفحة ويب للمراحل السابقة (تتوفر الصور للكامل صفحة الويب في الملحق آ):

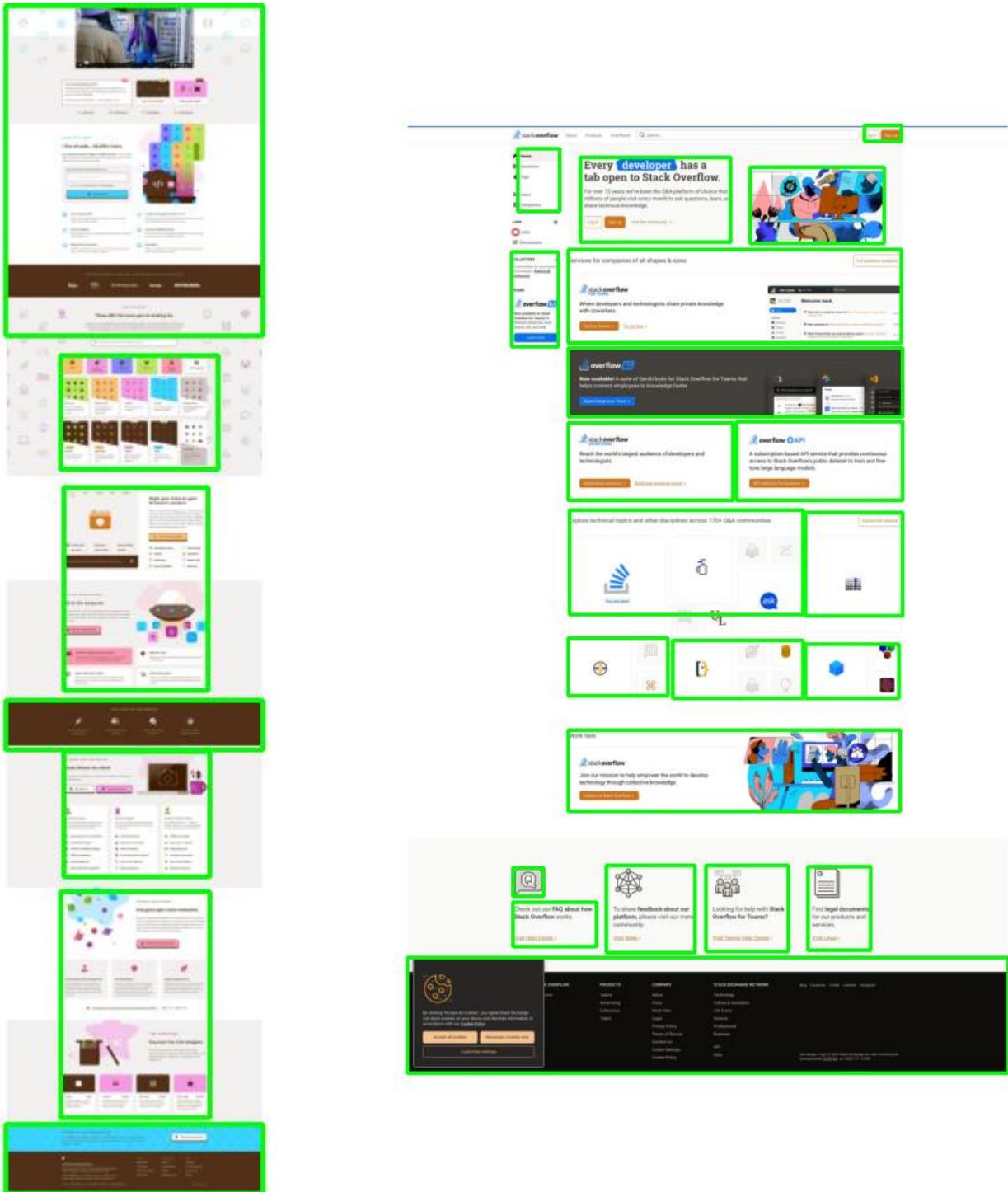


الشكل 46- خطوات تطبيق الخوارزمية على جزء من صفحة ويب.

أبرز النتائج:

The collage consists of several screenshots from the W3C website, each framed by a green border. The images include:

- The homepage featuring a large video player with the text "Bring your goals into focus".
- A news article titled "Working with stakeholders of the Web" featuring a graphic of people in wheelchairs.
- A banner for "W3C TPAC 2024 & W3C@30" with a blue background and the text "23-27 SEPTEMBER HYBRID MEETING Aachen, CA, USA & online".
- A section titled "Web standards" showing a pile of colorful 3D cubes.
- A section titled "Get involved" showing a person working at a desk with many electronic devices.
- A sidebar with a "Latest news" section containing links to "Draft Note: Undo Gap Analysis", "Draft Note for Web Review: Guidance on Applying WCAG 2.1 to Non-Web Information and Communications Technologies (WCAG2ICT)", "Internationalization (I18N): enabling access to a Web for All", and "Updated W3C Recommendation: CSS Containment Module Level 1".
- A footer with links to "About", "Participate", "Join", "Events", "News", "Resources", "About Us", "Contact Us", and "Feedback".

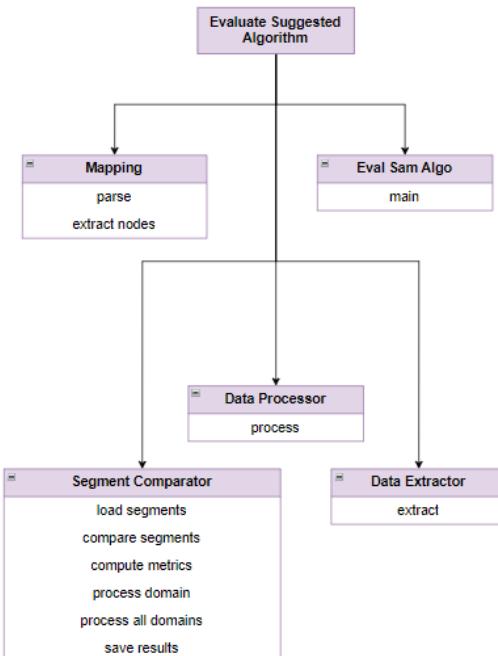


تقييم الخوارزمية:

تم التقييم باستخدام مجموعة البيانات random وذلك وفق الخطوات التالية:

1. تم أخذ لقطات شاشة لكل صفحة لتشكيل دخل للخوارزمية المقترحة.
2. تطبيق الخوارزمية المقترحة على مجموعة الصور الناتجة واستخراج معلومات القطع المرسومة (إحداثيات الزاوية العليا اليسرى بالإضافة لعرض وطول كل مستطيل).
3. استخراج معلومات ملفات json للتقسيم المقترن (إحداثيات الزاوية العليا اليسرى، العرض والطول، المساحة وإحداثيات المركز، المحتوى النصي، المسار ضمن شجرة DOM)، وذلك من خلال تحليل شجرة ال DOM وأخذ العقد الموقعة للمعلومات التي تم استخراجها في الخطوة السابقة.
4. المقارنة بين ملفات ال json المرجعية والملفات المستخرجة وفقاً للمعايير المقترنة في تقييم الخوارزمية السابقة لتمكن من القيام بعملية المقارنة.

البنية المستخدمة في التقييم:



الشكل 47 – بنية الرمaz المستخدم لتنجيز الخوارزمية.

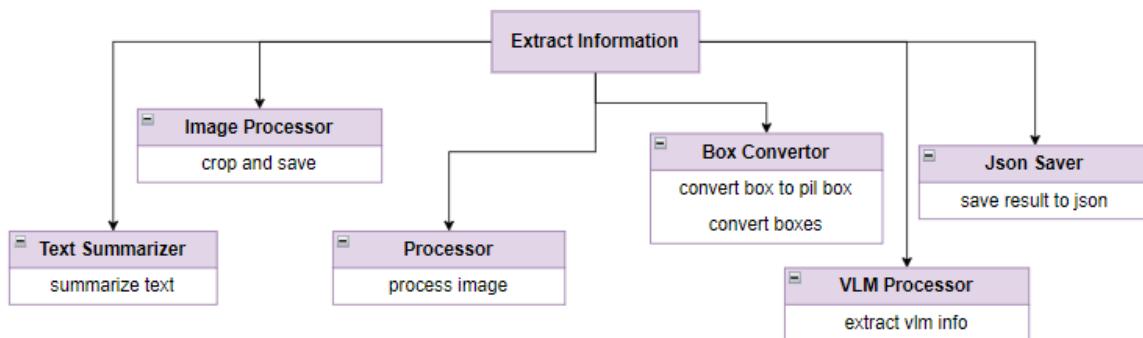
إضافةً لاستخدام بنية موارد شبيهة بنية الموارد المستخدمة في تقييم الخوارزمية السابقة، وكانت النتائج على الشكل التالي:

Algo	Precision	Recall	F1-Score
BOM	30.8%	33.0%	31.8%
VIPS	27.8%	26.4%	27.1%
SegBlock	41.9%	44.8%	43.3%
Semantic-Block	43.3%	53.6%	48.0%
Fusion-Block	49.3%	61.2%	54.6%
Integrated-Block	56.6%	67.5%	61.5%
Our Algorithm	57.6%	47.9%	52.3%
Suggested Algo	47.3%	0.1%	0.15%

جدول 9- نتائج تقييم الخوارزمية المستخدمة على مجموعة بيانات Random.

5.3.6- التقنيات المتّبعة لاستخراج المعلومات من قطع صفحات الويب

بعد القيام بتنقطيع صفحات الويب، تمّ أخذ النصوص الموجودة في كل قطعة (إما باستخدام شجرة DOM أو باستخدام OCR لقراءة النصوص الموجودة في كل قسم من الصورة)، ثم تلخيصها ثم القيام بتحويلها لصوت، كما تمّ أخذ قطع الصور الناتجة ومعالجتها بنموذج مدرب مسبقاً لاستخراج توصيف لكل قطعة موجودة وبعد استخراج التوصيف جرى تحويله لصوت أيضاً، إذ يتم الاستفادة من ذلك بوصف المحتوى الصوري الموجود في صفحة الويب، تم تنجيز ما سبق وفق البنية التالية:



الشكل 48- بنية الرمaz المستخدم لتنجيز الخوارزمية.

يوضح الشكل السابق البنية العامة للرماز المستخدم في تنظيم ما سبق والتي يتم تنظيم خطواتها باستخدام الصنف Processor وفق عدة خطوات وهي على الشكل التالي:

الخطوة الأولى: تحويل تنسيق الصناديق من (x, y, w, h) حيث تمثل الشائبة y , x , إحداثيات الزاوية العليا اليسرى، وتمثل w العرض و h الارتفاع، إلى تنسيق $[(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4)]$ إحداثيات الزوايا الأربع لكل صندوق موجود.

الخطوة الثانية: معالجة كل صندوق من خلال استخراج النص منه إما من ملف ال json الناتج عن التقطيع بالطرق التقليدية، أو باستخدام pytesseract لاستخراج النصوص من الصور.

الخطوة الثالثة: قص الصورة وذلك من خلال استخراج الموارف من الزوايا الأربع لكل صندوق $(top, bottom, left, right)$ إلى القطع المقسمة وحفظ القطع الناتجة.

الخطوة الرابعة: استخراج المعلومات من الصور باستخدام نماذج توصيف الصور.

الخطوة الخامسة: تحويل النصوص المستخرجة وتلخيصها، والنصوص الناتجة عن توصيف الصور إلى صوت.

الخطوة السادسة: حفظ النص الأصلي والنص الملخص والنص الناتج عن التوصيف بالإضافة لمسارات الملفات التي تحوي على الملفات الصوتية الموافقة لكل نص، ومسار الملف الحاوي على الصور المقطعة في ملف json.

أبرز النتائج:

1

Your GhatGPT AI Assistant Anywhere Chat, write, read, summarize, and create effortlessly with Sider, powered by ChatGPT and other leading AI models. Easily access @ GPT-40, % Claude 3.5, 4 Gemini1.5 and more AI models in one place with Sider!

summary

ChatGPT and other leading AI models in one place with Sider! Your GhatGPT AlophobicAssistant Anywhere . Chat, write, read, summarize, and create effortlessly

1

Sider

Your ChatGPT AI Assistant Anywhere

Chat, write, read, summarize, and create effortlessly with Sider, powered by ChatGPT and other leading AI models.

Easily access GPT-4o, Claude 3.5, Gemini 1.5 and more AI models in one place with Sider!

Add to Chrome Now - It's Free

4.9 out of 5.0 | 40K+ Reviews | 5M+ Users

Also available on

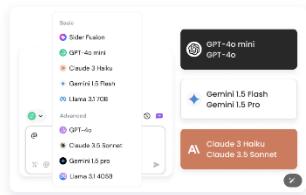
One Stop AI Assistant

All-in-one AI Chatbot

Sider integrates iconic AI models into a single chatbot, like GPT-4o mini, GPT-4o, Google Gemini and Claude. Get ready to explore the distinct capabilities of these different AI!

Furthermore, Sider now fully supports group chats with AI bots, allowing you to compare the variances among various bots.

[Install Now →](#)



2

Time-saving AI Writer

Generating articles, poems, thesis, email replies, and comments can be incredibly easy with the help of AI.

Previously, creating an article, encompassing topic selection, outline formation, content creation, and optimization, could take hours to days. But now, with Sider, it could be well finished in minutes or even seconds!

[Install Now →](#)

All Mighty AI Reader

Sider offers content reading for documents, website pages, PDFs and videos. Choose features from translation, summarization, quiz, rewriting, and break free from mundane reading experiences!

That's not all. Sider also unlocks the door to your private knowledge base. Your uploaded documents and pages are ready to provide personal assistance in your work and life.

[Install Now →](#)

3

AI PDF Summarizer & Translator

Sider helps you quickly understand your PDF documents by creating concise summaries, saving you time and highlighting key points. Perfect for anyone needing to digest information quickly.

Translate your PDF documents with our AI-powered tool. Upload your files and receive accurate translations in your desired language while preserving the original format. Ideal for business communications, travel documents, or studying foreign languages, making global interaction more accessible.

[Install Now →](#)

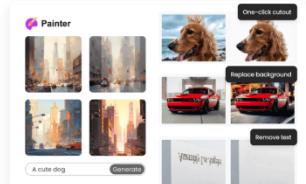
Text-to-image AI Painter

With Sider, anyone can transform plain words or images into stunning arts, based on the latest Stable Diffusion.

Unlike any other AI picture program, such as Midjourney, Sider ignites your imagination by a suite of pre-trained styles that cater to over 95% usage scenarios. For odds of prompts, Sider understands that customization is key, and allows you to effortlessly tailor your AI-generated masterpieces.

[Install Now →](#)

4



5

6

2

All-in-one AI Chatbot Sider integrates iconic AI models into a single chatbot, like GPT-40 mini, GPT-40, Google Gemini and Claude. Get ready to explore the distinct capabilities of these different AIs! Furthermore, Sider now fully supports group chats with AI bots, allowing you to compare the variances among various bots. [InstallNow >](#)

summary

Sider integrates iconic AI models into a single chatbot, like GPT-40 mini, Google Gemini and Claude . Get ready to explore the distinct capabilities of these different AIs .

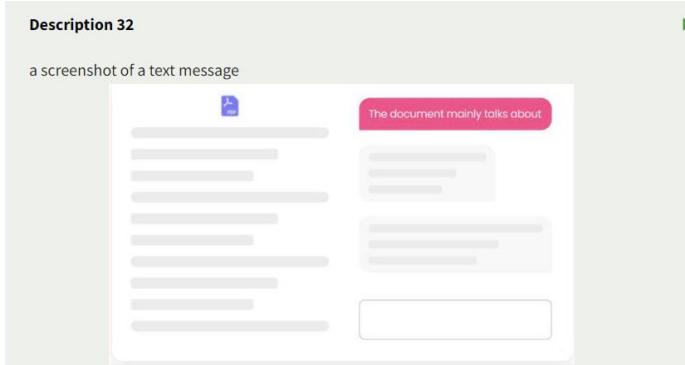
3

All Mighty AI Reader Sider offers content reading for documents, website pages, PDFs and videos. Choose features from translation, summarization, quiz, rewriting, and break free from mundane reading experiences! That's not all. Sider also unlocks the door to your private knowledge base Your uploaded documents and pages are ready to provide personal assistance in your work and life. [InstallNow >](#)

summary

Sider offers content reading for documents, website pages, PDFs and videos . Choose features from translation, summarization, quiz, rewriting and rewriting . Sider unlocks the door to your private knowledge base .

4



5

Text-to-image AI Painter With Sider, anyone can transform plain words or images into stunning arts, based on the latest Stable Diffusion. Unlike any other AI picture program, such as Midjourney, Sider ignites your imagination by a suite of pre-trained styles that cater to over 95% usage scenarios. For dabs of prompts, Sider understands that customization is key, and allows you to effortlessly tailor your AI-generated masterpieces. [InstallNow >](#)

summary

Sider lets anyone transform plain words or images into stunning arts . Pre-trained styles cater to over 95% usage scenarios .

6

AI PDF Summarizer & Translator Sider helps you quickly understand your PDF documents by creating concise summaries, saving you time and highlighting key points. Perfect for anyone needing to digest information quickly. Translate your PDF documents with our AI-powered tool. Upload your files and receive accurate translations in your desired language while preserving the original format. Ideal for business communications, travel documents, or studying foreign languages, making global interaction more accessible.

summary

AI PDF Summarizer & Translator helps you quickly understand your PDF documents by creating precise summaries and highlighting key points . Perfect for anyone needing to digest information quickly .

Home > Our history

Our history



1

MAMM is a space that questions traditions and has a proactive attitude towards the defining issues of contemporary society. We make available to the public different ways of seeing and interpreting the world, offering meaningful experiences around art.

More than a physical space, we are a place where heritage, art production, the public and the institution come together, becoming a platform of knowledge and expertise that impacts today's society.

We work to be a global and at the same time local Museum, which promotes encounter, work, production and research among multiple professions of contemporary creation and fosters the development of subjectivities. These dimensions are mixed and nourished to be exhibited and disseminated to society through various exhibitions each year.



2

Our beginnings



Arrival from River City



Expansion



Museum today



Historical events



Alberto Sierra: his story and legacy

Buy your tickets to the Museum

Visit us Tuesday to Friday: 11:00 am to 7:00 pm or Saturday, Sunday and holidays: 11:00 am to 6:00 pm

[Get to know the rates](#)



1

the building is made of concrete and has a triangular shape



2

a black and white photo of a building with stairs



Nuestros inicios

3

a black and white photo of a crowd of people



4

the building is made of concrete and has a large square shape

Compra tus entradas al Museo

Visítanos de martes a viernes: 10:00 a.m. a 7:00 p.m., o sábado, domingo y festivos:
10:00 a.m. a 10:00 p.m.

[Conoce las tarifas](#)



6.3.6- قسم الويب

تم تغليف الخوارزميات السابقة بتطبيق ويب، يمكن استخدامه في نظم التصفح غير البصري لتنق剔 واستخراج المعلومات من صفحات الويب وذلك باستخدام Django.

• تصميم النظام

تم استخدام نمط MTV (Model-Template-View) وهو تصميم معماري يفصل المكونات المختلفة لتطبيق الويب إلى ثلاثة أجزاء رئيسية: النموذج Model، القالب Template، والعرض view.

- View العرض: هو القسم الذي يتعامل مع دورة request/response الطلب/الاستجابة،

إذ يعالج إدخال المستخدم وتنفيذ منطق العمل، وإرجاع الاستجابة المناسبة.

- Template القالب: هو القسم المسؤول عن طبقة الواجهات الخاصة بالتطبيق، وتحديد كيفية

عرض البيانات المرسلة من العروض views للمستخدم.

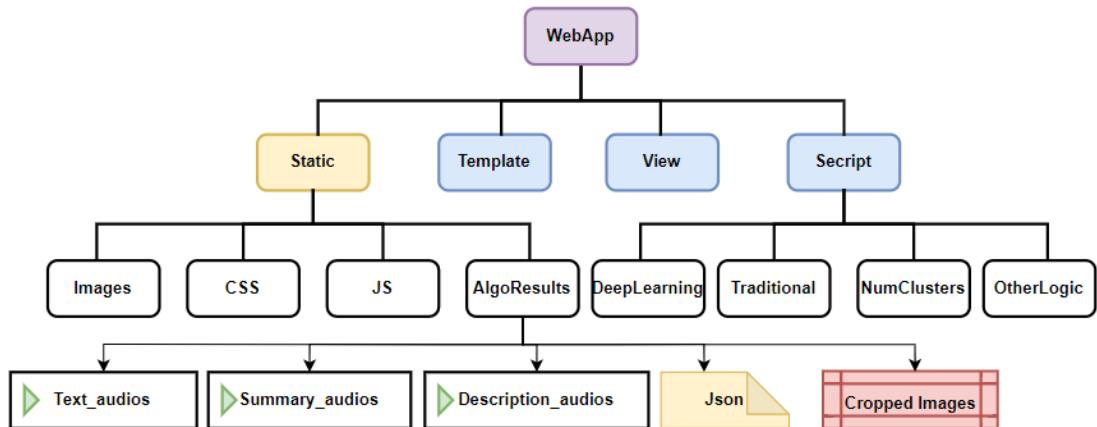
- Model النموذج: مسؤول عن منطق العمل ومعالجة البيانات.

• تنजير النظام

1. View: يتم استخدام مجموعة من التوابع التي تعمل على تلقي الطلبات من المستخدم وتنفيذ scripts المناسب للطلب.

2. Template: وهي مجموعة من ثلاث ملفات .html.

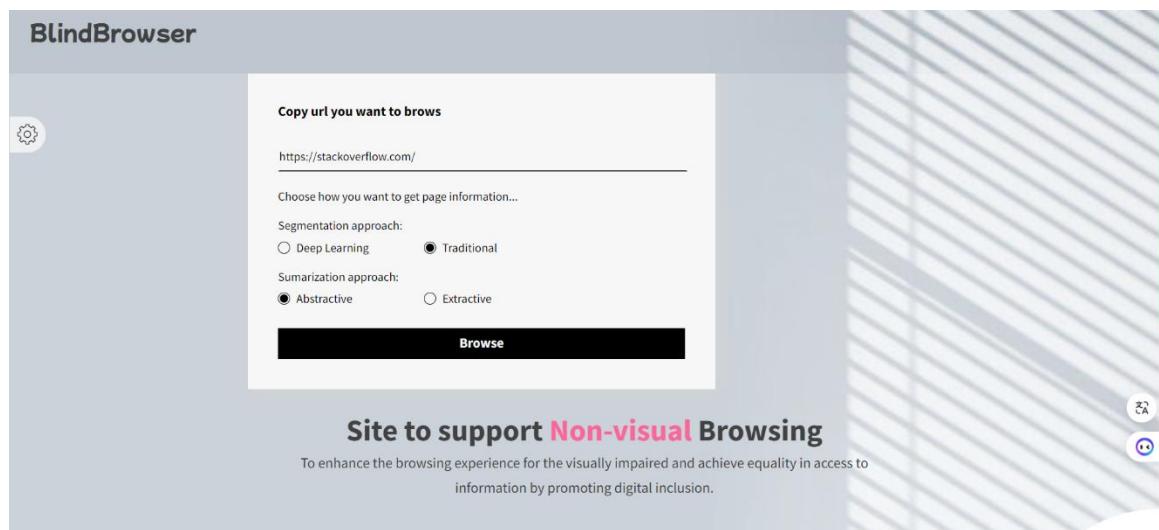
3. Model: تم التعديل على بنية models لعدم الحاجة لاستخدام قاعدة بيانات، إذ تم تضمين منطق التطبيق بمجلد scripts، والذي يتضمن على الخوارزميات السابقة كل منها منفصل عن الآخر، وذلك لتسهيل عملية إضافة خوارزميات جديدة في حال تم إصدار نماذج مدربة لتنق剔 صفحات الويب أو لاستخراج المعلومات منها، كما تم التننجيز بحيث يمكن فصل أي خوارزمية واستخدامها في تطبيق آخر، وبالرغم من غياب قاعدة البيانات إلا أننا قد حافظنا على مبدأ الفصل لمنطق العمل عن الطبقات الأخرى في النموذج (template, view).



الشكل 49- بنية الرمざ المستخدم للتطبيق.

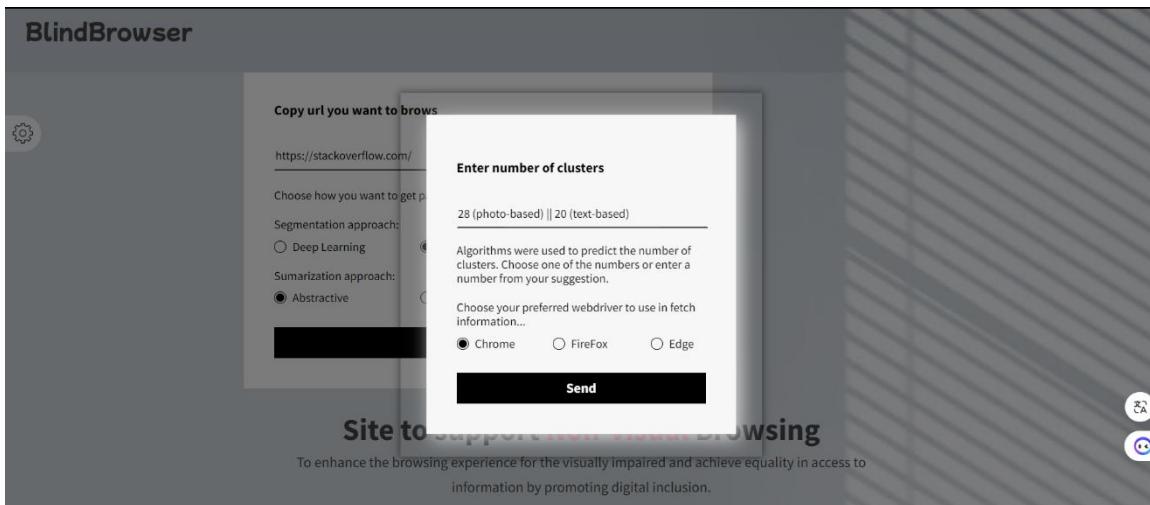
• الواجهات

في البداية يقوم المستخدم بإدخال عنوان الصفحة المراد تصفحها، ثم يقوم باختيار طريقة التقاطع إما بالطريقة التقليدية أو بالطريقة التي تستخدم تقنيات التعلم العميق في التقاطع، كما يجب على المستخدم أن يدخل طريقة التلخيص المراد استخدامها (تجريدي، استخراجي)، كما يوضح الشكل 50.



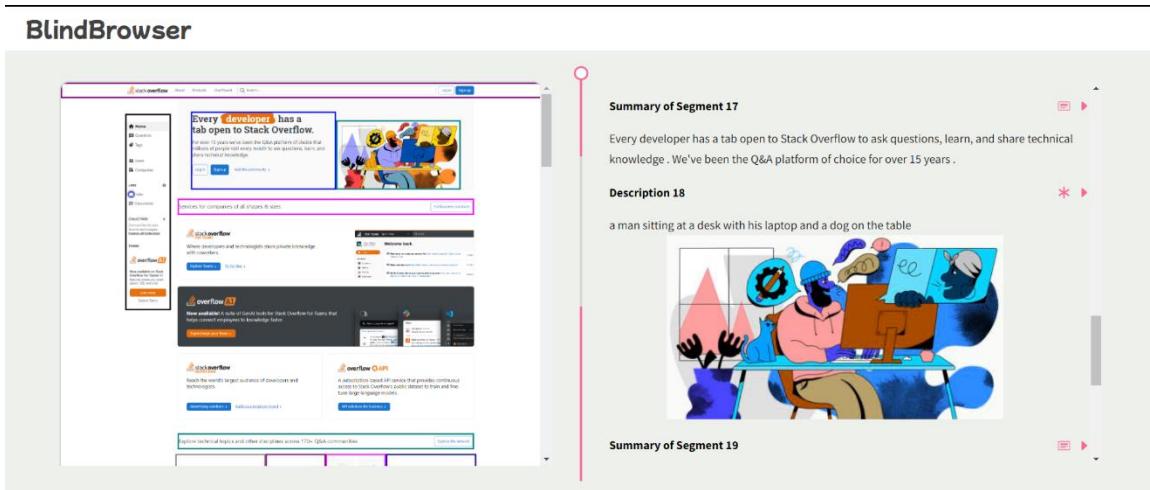
الشكل 50- الواجهة الأساسية.

في حال قام المستخدم باختيار الطريقة التقليدية في التقطيع، يقوم النظام باقتراح عدد العناوين المعتمد على تحليل المحتوى النصي وعدد العناوين المعتمد على تحليل الصورة، ليقوم المستخدم باختيار العدد المناسب (عدد مقترن بناءً على النص أو عدد مقترن بناءً على الصورة أو عدد يقترحه هو)، كما يتوجب عليه اختيار webdriver الذي يريد استخدامه لاستخراج معلومات الصفحة كما هو موضح في الشكل 51.



الشكل 51- المعلومات الإضافية المطلوبة من المستخدم عند اختيار الطريقة التقليدية.

ثم يعرض نتائج التقطيع بالإضافة للنصوص المستخرجة من القطع والصوت الموافق لها، إضافةً لتلخيص تلك النصوص مع الصوت الموافق لها، بالإضافة للتوضيف الناتج عن نماذج توصيف لقطع الصور الناتجة والصوت المرافق لذلك التوضيف كما هو موضح في الشكل 52.



الشكل 52- النتائج النهائية.

الفصل السابع

الخاتمة والآفاق المستقبلية

نقدم في هذا الفصل الفائدة المكتسبة من المشروع، وأبرز الصعوبات التي واجهتنا خلاله، وفي النهاية الآفاق المستقبلية للمشروع.

1.7- الخاتمة والفائدة المكتسبة

تم في هذا المشروع بناء نظام مساعد لتقاطيع صفحات الويب واستخراج المعلومات من القطع الناتجة وذلك بالاعتماد على العديد من التقنيات الخاصة بتعلم الآلة والتعلم العميق ومعالجة الصورة، والتعرف على التقنيات المختلفة لتقدير الخوارزميات المقترحة، كما تم التعامل مع العديد من المكتبات التي تقدمها لغة Python.

2.7- الصعوبات

أبرز الصعوبات التي تمت مواجهتها:

- كثرة الخوارزميات التي يجب التعرف عليها في تقاطيع صفحات الويب.
- صعوبة التكامل بين بيئة عمل Colab وتطبيق الويب المنجز باستخدام Django.
- صعوبة في إيجادمجموعات البيانات الملائمة، حيث أن مجموعة البيانات التي تم استخدامها لا تحتوي على صور لصفحات الويب وهو دخل الخوارزمية الثانية بتقاطيع صفحات الويب، إضافةً لعدم توافرمجموعات بيانات لتلخيص قطع صفحات الويب.

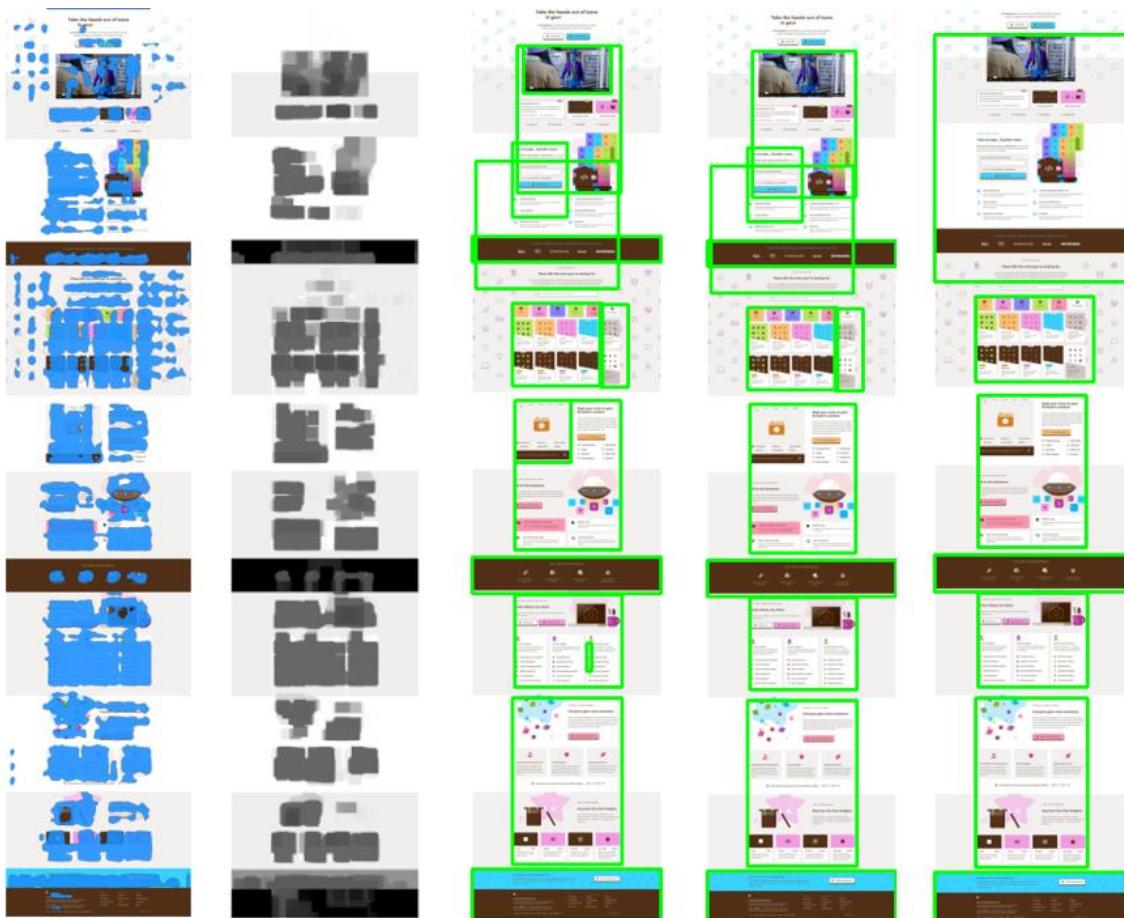
3.7- التوسيع والآفاق المستقبلية

تم تنجيز النظام السابق مع الأخذ بعين الاعتبار إمكانية إصدار نماذج خاصة بتقاطيع صفحات الويب فيما بعد من قبل الشركات الكبيرة، الأمر الذي يسمح بإضافة أي طريقة جديدة للنظام، كما يمكن الدمج بين الخوارزميتين المقترحتين لتقاطيع صفحات الويب لإعطاء نتيجة أفضل للتقطيع.

الملحق آ

اختبارات وتجارب

- نتائج الخوارزمية المقترحة في الفقرة 4.2.6:

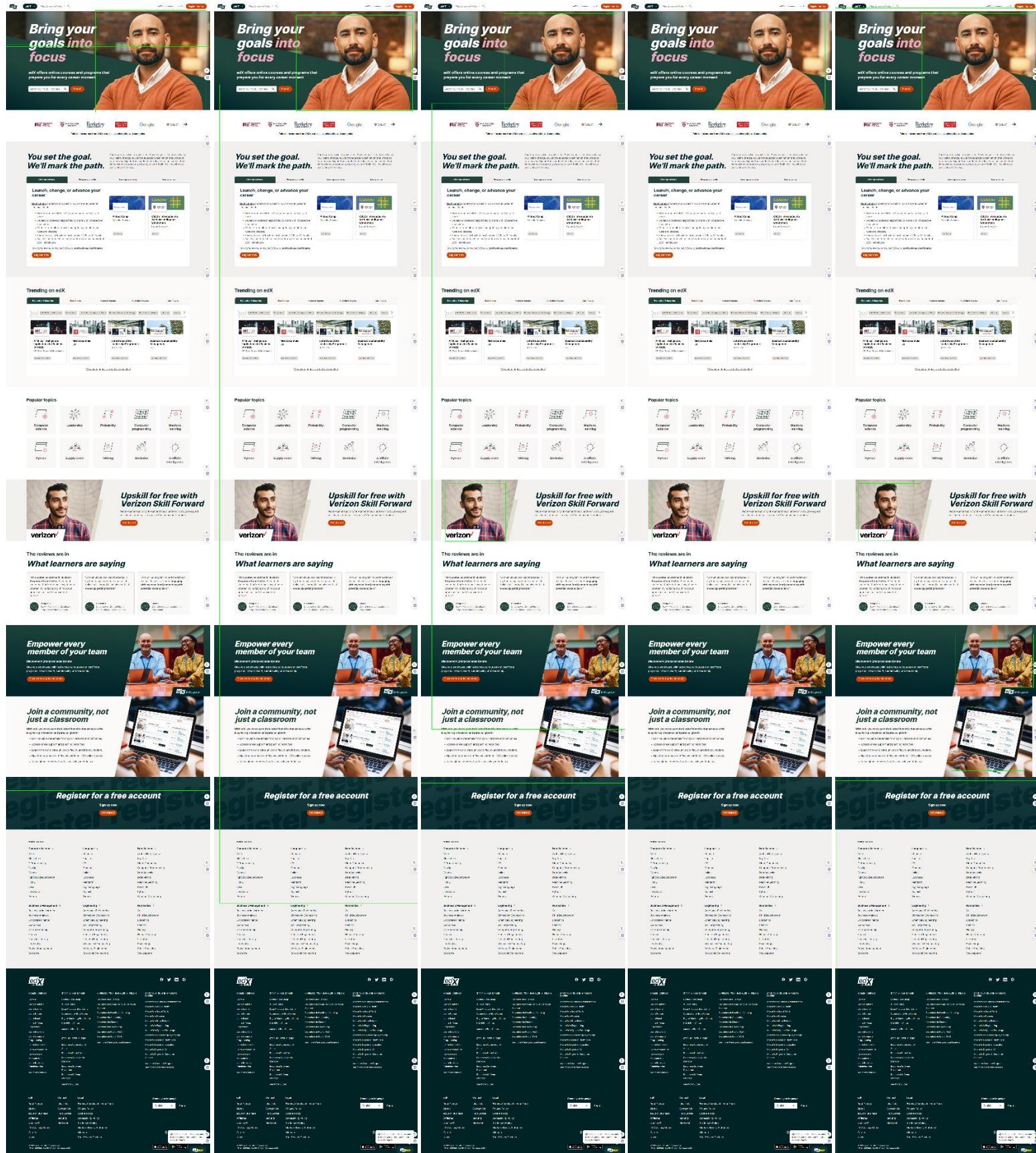


الشكل 53- خطوات المعالجة بعد تطبيق النموذج.

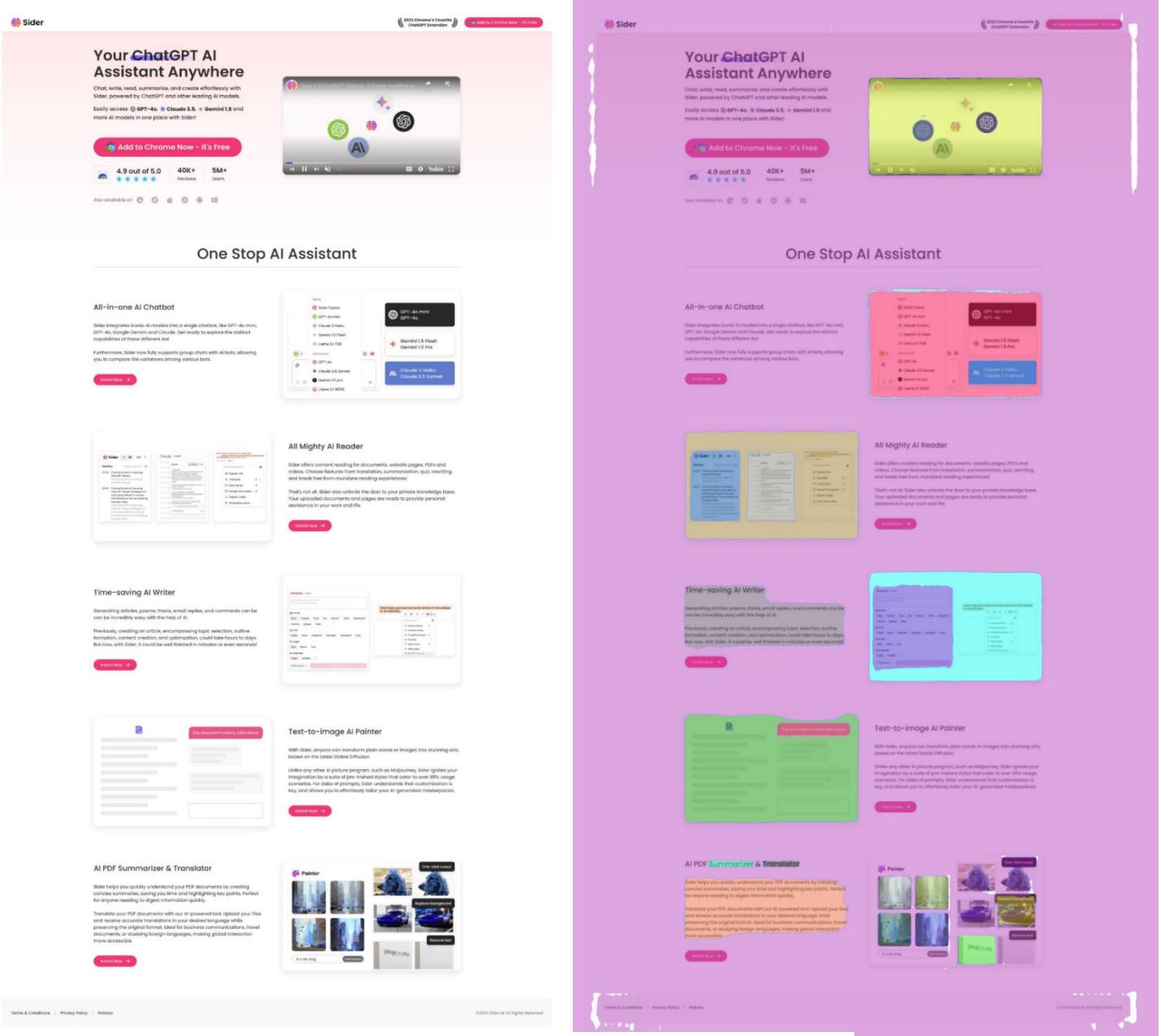
- نتائج تطبيق نماذج yolov5 المختلفة على صفحة وب:

وهي بالترتيب من اليسار إلى اليمين

yolov5n, yolov5s, yolov5m, yolov5l, yolov5x



• نتائج تطبيق نموذج SAM2 على عدد من صفحات الويب:



الشكل 54-نتيجة تطبيق sam2.

(نلاحظ أنه يقوم بتحديد التفاصيل الموجودة في الصور داخل صفحة الويب ويقوم بتقطيع محتويات تلك الصور أكثر من كونه قادر على استخراج قطع المحتوى الموجودة في الصفحة كل كمًا هو مبين في المثال السابق، فيما يلي مزيد من النتائج).

B

Build fast, responsive sites with Bootstrap

Download Bootstrap, the most advanced front-end toolkit. Build fast, responsive, and highly performant web applications with powerful front-end engines.

Get started any way you want

Install via package manager
Include via CSS

Customize everything with SASS

Build and extend in real-time with CSS variables

Components, meet the Utility API

Powerful JavaScript plugins without jQuery

Personalize it with Bootstrap Icons

Make it yours with official Bootstrap Themes

W3C

Standards Groups Get involved Resources News and events About

Making the Web work

The World Wide Web Consortium (W3C) develops standards and guidelines to help everyone build a web based on the principles of accessibility, internationalization, privacy and security.

Working with stakeholders of the Web

A range of organizations join the World Wide Web Consortium as Members to work with us to drive the direction of core web technologies and exchange ideas with industry and research leaders. We invite randomly a few of our Member organizations' logos below.

The Web Consortium and its members, with help from the public and the web community, focus on a range of business ecosystems that the Web transforms, including E-commerce, Media & Entertainment, Publishing, and several other areas.

HUAWEI, ensysnet, @combook, Access Australia, LEVEL ACCESS, NAVY FEDERAL Credit Union

W3C TPAC 2024 & W3C@30

1 October 2024 is W3C's 30th anniversary. We will celebrate at our annual TPAC, our three decades, advances in the Web, the impact of our standards on the world, imagine a bright future, and enjoy interactive discussions with our community.

TPAC, our major annual event, gathers our community for thought-provoking discussions and coordinated work to advance the mission of the group, innovative challenging technical and social issues for the Web itself.

TPAC brings together W3C Members and technical groups, the Board of Directors, the Advisory Board, and the Technical Architecture Group.

Information about W3C@30 & TPAC

Web standards

Web standards are the building blocks of a consistent digitally connected world. They are implemented in browsers, blogs, search engines, and other software that power our experience on the Web.

W3C is an international community where standards, full-time staff, and the public work together to develop web standards.

Get involved

W3C works at the cross of core technology, industry needs, and societal needs. Everyone can get involved with the work we do. There are many ways individuals and organizations can participate in the Web Consortium to advance web standardization.

Way to get involved

Latest news

Nine Draft Notes published by the Internationalization Working Group

Third-party cookies have got to go

The W3C Technical Architecture Group explains how third-party cookie serving does not respect privacy and why they must be removed from the web. This blog post introduces the latest bid finding. Third-party cookies must be removed.

First Public Working Draft: CSS Scroll Snap Module Level 2

This module contains rules for scroll snap serving and scrolling behavior with "snap position".

Eight Draft Notes published by the Internationalization Working Group

The Internationalization Working Group has published the Mongolian, Persian, Pashto, Punjabi, Devanagari, Bengali, and Cyclic Script Resources, as well as the Text Analysis, as group Draft Notes.

Header

Footer

Bring your goals into focus

You set the goal. We'll mark the path.

Launch, change, or advance your career

Trending on edX

Popular topics

Information Technology, Data Science, Machine Learning, Computer Science, Programming, Analytics, Future, Business, Marketing, Communication, Writing, Design, Metrics

Upskill for free with Verizon Skill Forward

The reviews are in What learners are saying

Empower every member of your team

Join a community, not just a classroom

Register for a free account

Header

Footer

الملحق ب

تفاصيل التحسينات

▪ تحسين بنية الشبكة لتجزئة صفحة الويب

التعديل الأول: يستخدم نموذج YOLOv5 شبكة CSPDarknet53 باعتبارها الشبكة الأساسية لاستخراج الميزات ويجمع بين PANet و SPP لدمج الميزات. كما يستخدم أيضًا ثلاث شبكات تنبؤ بأحجام مختلفة، وهي 52x52، و 26x26، و 13x13، لاكتشاف الأجسام الكبيرة والمتوسطة والصغيرة على التوالي.

• شبكة CSPDarknet53

1. Darknet: هو بيئة لشبكة عصبية مفتوحة المصدر مكتوب بلغة C و CUDA. تم تحسينه لتحقيق الأداء العالي ويعمل كأساس لنماذج YOLO.

2. Darknet53: هذه نسخة محسنة من بنية Darknet الأصلية المستخدمة في YOLOv3. يحتوي على 53 طبقة تلافيفية، ومن هنا جاء اسم Darknet53.

3. CSPNet (Cross-Stage Partial Network): عبارة عن بنية مصممة لتحسين القدرة على التعلم وتقليل التكلفة الحسابية. من خلال تقطيع خريطة المعلم للطبقة الأساسية إلى جزأين ثم دمجهما من خلال تسلسل هرمي عبر المراحل، تساعد CSPNet في تحسين تدفق التدرج وتقليل العمليات الحسابية وتحسين الدقة.

4. CSPDarknet53: يجمع بين فوائد Darknet53 و CSPNet، مما يجعله العمود الفقري الفعال لاستخراج الميزات في YOLOv5.

• دمج الميزات: SPP، PANet

1. PANet (Path Aggregation Network): يتم استخدام PANet لتعزيز تدفق المعلومات عبر مستويات مختلفة من هرم الميزات. فهو يساعد في التقاط مستويات مختلفة من المعلومات الدلالية، وهو أمر بالغ الأهمية للكشف عن الكائنات ذات الأحجام المختلفة، تقدم PANet اتصالات إضافية لإنشاء تعزيز المسار من الأسفل إلى الأعلى، مما يساعد في الحفاظ على الميزات ذات المستوى المنخفض مع دمج المعلومات الدلالية عالية المستوى.

2. SPP (Spatial Pyramid Pooling): يقدم طبقة تجميع تجمع الميزات بمقاييس مختلفة (على سبيل المثال، 1×1 ، 3×3 ، 5×5 ، وما إلى ذلك) وتقوم بتسلسلها. يتيح ذلك للشبكة التقاط المعلومات بمقاييس متعددة دون الحاجة إلى تغيير حجم الصورة المدخلة، يساعد في زيادة المجال الاستقبالي، مما يسمح للشبكة بفهم التسلسل الهرمي المكاني وبالتالي تحسين اكتشاف الأشياء على مستويات مختلفة.

- شبكات التنبؤ (اكتشاف متعدد المقاييس): يستخدم YOLOv5 ثلات شبكات تنبؤ بأحجام مختلفة

لاكتشاف الكائنات بمقاييس مختلفة:

1. كائنات كبيرة (شبكة 13×13): خريطة المعلم 13×13 مسؤولة عن اكتشاف الأجسام الكبيرة.

تحتوي هذه الشبكة على مجال استقبال أكبر، مما يجعلها مناسبة لالتقاط كائنات كبيرة في الصورة.

2. الكائنات المتوسطة (شبكة مقاس 26×26): يتم استخدام خريطة المعلم مقاس 26×26 للكشف

عن الأجسام متوسطة الحجم. فهو يوفر التوازن بين حجم المجال الاستقبالي والدقة المكانية، مما يجعله فعالاً بالنسبة للأجسام المتوسطة.

3. الكائنات الصغيرة (شبكة مقاس 52×52): خريطة المعلم مقاس 52×52 مخصصة للأشياء

الصغيرة. تحتوي هذه الشبكة على مجال استقبال أصغر ولكن دقة مكانية أعلى، وهو أمر ضروري للكشف عن الأجسام الصغيرة.

آلية العمل:

1. يستخرج CSPDarknet53 ميزات غنية من الصورة المدخلة، ويلتقط مستويات مختلفة من التفاصيل والدلائل.

2. دمج الميزات: يتم تمرير الميزات المستخرجة عبر PANet و SPP لتحسين تمثيلها. تساعد PANet في تجميع الميزات عبر طبقات مختلفة، بينما تميز مجموعات SPP بمقاييس مختلفة.

3. التنبؤ متعدد المقاييس: يتم بعد ذلك إدخال الميزات المدمجة في ثلاث شبكات تنبؤ منفصلة، تتوافق كل منها مع حجم شبكة مختلف (13×13 ، 26×26 ، 52×52). تنبأ كل خلية شبكة

بالمربعات المحيطة ودرجات الموضوعية واحتمالات فئة الكائنات.

التعديل الثاني: Attention module وحدة الاهتمام. لتحسين أداء خوارزمية YOLOv5 في مهام تجزئة صفحة الويب، أضفنا وحدة تنسيق الاهتمام [53] إلى نهاية الإخراج لوحدة C3 في الشبكة الأساسية لـ YOLOv5.

تهدف هذه الآلية إلى تحسين اهتمام الشبكة بكتل المحتوى المستهدف ذات الاهتمام بإدخال بيانات صفحة الويب والتركيز على المعلومات المفيدة.

- وحدة C3 هي اختصار لـ Cross-Stage Partial3، وهي نوع من الكتل المستخدمة في YOLOv5 لتحسين استخراج الميزات وأداء الشبكة. تتضمن هذه الوحدة مجموعة من الطبقات التلافيفية، تسوية الدفعات، ووظائف التنشيط. الرقم "3" يشير إلى التكوين أو الاختلاف ضمن نجح CSP.

تعمل وحدة C3 على تقطيع خريطة ميزات الإدخال إلى قسمين. يمر جزء واحد عبر عدة طبقات تلافيفية، بينما قد يتم تجاوز الجزء الآخر أو يخضع لعدد أقل من التحولات. بعد ذلك، يتم دمج هذين الجزأين من خلال طبقات تلافيفية إضافية، مما يحقق توازنًا بين الكفاءة الحسابية وثراء الميزات.

• وحدة تنسيق الانتباه: Coordinate Attention Module

- Coordinate Attention: تعامل وحدة تنسيق الاهتمام على تعزيز قدرة الشبكة على التركيز على المناطق المهمة من خلال تضمين المعلومات الموضعية في آلية انتباه القناة، على عكس آليات الانتباه التقليدية التي تركز بشكل أساسي على تبعيات القناة، يلتقط تنسيق الانتباه معلومات القناة والمعلومات المكانية، تقوم هذه الوحدة بتقطيع خريطة ميزات الإدخال على طول الأبعاد المكانية وترميز معلومات الإحداثيات في خرائط الانتباه الناتجة.

- هيكل وعمل تنسيق الاهتمام: Structure and Working of Coordinate Attention

.1 Splitting and Pooling: يتم تقطيع خريطة ميزات الإدخال X إلى فرعين: أحدهما للبعد الأفقي والآخر للبعد الرأسى، في التجميع الأفقي يتم تطبيق عملية التجميع العام الأفقي لإنتاج واصف F_h يحتوي على معلومات الارتفاع. أما في التجميع العمودي: يتم تطبيق عملية التجميع العام الرأسى لإنتاج واصف F_v يحتوي على معلومات العرض، ويمثل التجميع الأفقي والعمودي (X avg pool, Y avg pool) الطبقة الثانية في الشكل.

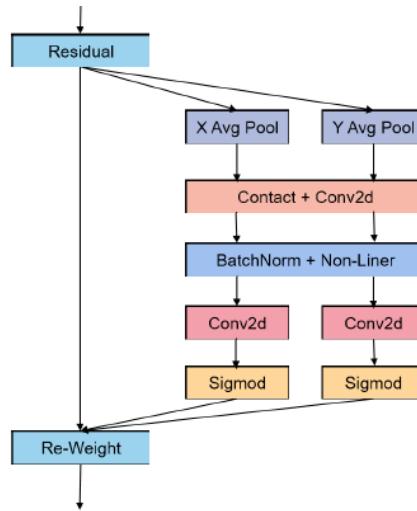
.2 Transformation

- المشترك: يتم تمرير كل من F_h و F_v من خلال Perceptron متعدد الطبقات MLP مشترك لتضمين المعلومات الموضعية.
- التسلسل وإعادة التشكيل: يتم بعد ذلك تسلسل مخرجات MLP لكلا الاتجاهين وإعادة تشكيلها مرة أخرى إلى حجم خريطة الميزات الأصلية.

(Contact + conv2d => batchNorm + Non-linear => conv2d, conv2d) ○
الطبقات الثالثة والرابعة والخامسة من الشكل.

3. Attention Generation: تُستخدم خرائط الميزات المتسلسلة لإنشاء أوزان الانتباه لكل قناة، معأخذ المعلومات الموضعية فياعتبار من خلالتابع sigmoid وهو الطبقة السادسة في الشكل.

4. Reweighting: تتم إعادة وزن خريطة الميزات الأصلية من خلال خرائط الاهتمام هذه، مما يعزز المناطق المهمة ويحسن التركيز على كتل المحتوى المهمة وهي الطبقة الأخيرة بالشكل 56.



الشكل 56 - هيكل تنسيق الانتباه.

Adding Coordinate Attention to YOLOv5

1. في YOLOv5، تتم إضافة وحدة تنسيق الانتباه إلى طرف الإخراج لوحدة C3 داخل الشبكة الأساسية.

2. من خلال دمج تنسيق الاهتمام، يمكن للشبكة التمييز بشكل أفضل والتأكيد على كتل المحتوى المستهدف في بيانات صفحة الويب المدخلة. يعد هذا مفيداً بشكل خاص لمهام تجزئة صفحات الويب حيث يعد تحديد أقسام أو عناصر معينة والتركيز عليها (مثل الصور أو الكتل النصية أو الأزرار) أمرًا بالغ الأهمية.

التعديل الثالث: Activation function optimization تحسين وظيفة التنشيط: بتحسين قدرة الشبكة على استخراج الميزات الدلالية المكانية في صفحات الويب عن طريق استبدال وظيفة تنشيط SiLU بوظيفة تنشيط FReLU [54] في طبقات التلافييف. تعمل وظيفة FReLU على توسيع وظيفة PReLU/ReLU عن طريق

إضافة حالة مكانية لإنشاء وظيفة تنشيط ثنائية الأبعاد، وهي مصممة لمهام المعالجة المرئية ويمكنها تحسين كفاءة هذه المهام بشكل كبير.

1. ReLU: وهو أحد وظائف التنشيط الأكثر استخداماً في التعلم العميق، يتميز ببساطة وسهولة التنفيذ، إضافة لمشكلة موت بعض الخلايا في الشبكة مما يجعل خرجها صفرى لجميع المدخلات الأمر الذي يجعل التدرج غير نشط ويعطى بالعلاقة:

$$ReLU(x) = \max(0, x) \quad (20)$$

2. PReLU: يعالج مشكلة موت بعض الخلايا حيث يضيف معلمة alpha وهي معلمة فائقة يتم تعلمها، لكن زيادة معلمة إضافية بالتعلم تعنى زيادة التعقيد الحسابي ويعطى بالعلاقة:

$$PReLU(x) = \max(0, x) + \alpha \min(0, x) \quad (21)$$

3. SiLU: مستمرة وقابلة للتفاضل في كل مكان بما في ذلك حوالي الصفر، تدرجه السلس يعطى نتيجة أفضل من ReLU، ولكنه أكثر تكلفة من الناحية الحسابية منه، ويعطى بالعلاقة:

$$SiLU(x) = x \cdot \sigma(x); \quad \sigma(x) = \frac{1}{1+e^{-x}} \quad (22)$$

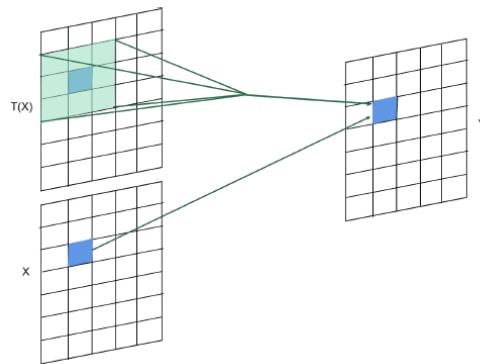
4. FReLU: وهي وظيفة تنشيط حديثة مصممة لتوفير المرونة في ReLU، ويتم ذلك بناء على المعلومات المكانية، كما يضيف معلمة إضافية للتعلم، ويعطى بالعلاقة:

$$FReLU(x) = \max(0, x) + max(0, Y(x)) \quad (23)$$

أو تعطى بالعلاقة:

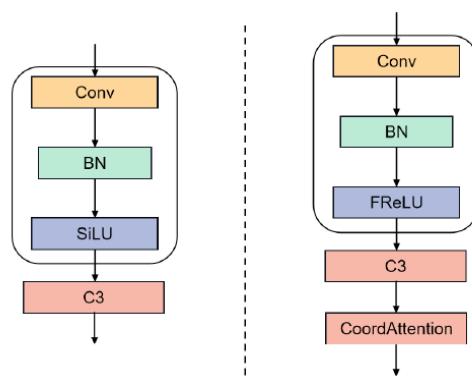
$$y = \max(x, T(x)) \quad (24)$$

إذ أنَّ استبدال وظيفة التنشيط SiLU بوظيفة التنشيط FReLU في طبقات التلافييف يعطي تكيف مكاني محسن، زيادة قوة التمثيل، تحسين عملية التعلم في العمق، والتكيف الأفضل مع تأثيرات التجميع المكاني. هذه التحسينات يمكن أن تؤدي إلى نموذج أكثر دقة وفعالية في معالجة البيانات المرئية المعقدة كما يوضح الشكل 57.



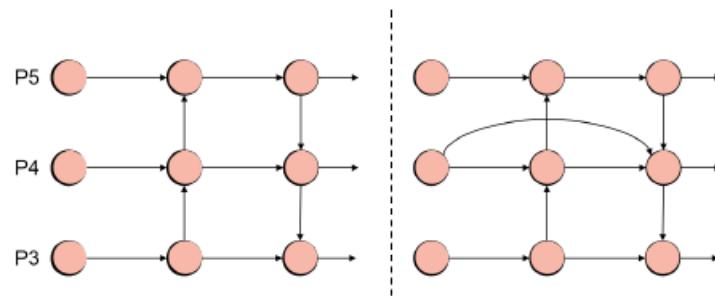
الشكل 57- تكيف مكاني محسن.

لتصبح الشبكة بعد التعديلات السابقة على الشكل 58



الشكل 58- الشبكة المحسنة.

التعديل الرابع: تم استخدام PANet المحسنة بدلاً من PANet، والتي تميز بإضافة الاتصالات المتداخلة skip connections في طبقة الإخراج ومنع تدهور الشبكة وتقليل فقدان المعلومات [55] كما هو موضح في الشكل 59



الشكل 59- الاتصالات المتداخلة المضافة.

▪ تحسين تابع الخسارة لتجزئة صفحة الويب

لتحسين أداء نموذج YOLOv5، استبدلنا وظيفة الخسارة الأصلية، CIoU، بوظيفة خسارة جديدة تسمى EIoU في نموذج YOLO-WS المقترن:

1. CIoU (Complete Intersection over Union) تُستخدم في نماذج اكتشاف الكائنات مثل YOLOv5.

يتكون من ثلاثة أجزاء رئيسية:

- خسارة التداخل: يقيس مدى تداخل المربع المحيط المتوقع مع المربع المحيط بالحقيقة الفعلية.
- فقدان المسافة المركزية: يقيس المسافة بين النقاط المركزية للمربع المتوقع ومربع الحقيقة الأرضي.
- خسارة نسبة العرض إلى الارتفاع: تأخذ في الاعتبار الاختلاف في الشكل (نسبة العرض والإرتفاع) بين المربع المتوقع ومربع الحقيقة الأرضي.

2. EIoU (Enhanced Intersection over Union) هو نسخة محسنة من CIoU مع بعض التغييرات الرئيسية:

- خسارة العرض والارتفاع: بدلاً من مجرد النظر إلى نسبة العرض إلى الارتفاع، تعمل EIoU بشكل مباشر على تقليل الفرق بين عرض وارتفاع المربع المتوقع ومربع الحقيقة. يساعد هذا النموذج على تعلم التنبؤ بحجم الكائنات بشكل أكثر دقة وسرعة.
- خسارة التداخل وخسارة المسافة المركزية: هذه الأجزاء هي نفسها الموجودة في CIoU، مما يضمن تداخل المربع المتوقع بشكل جيد مع مربع الحقيقة الفعلية ويتم توسطيه بشكل صحيح.
- الخسارة البؤرية: يضيف EIoU شيئاً يسمى الخسارة البؤرية أعلى CIoU. يساعد فقدان التركيز على معالجة مشكلة عدم توازن العينة. في عملية اكتشاف الكائنات، غالباً ما يكون هناك العديد من الكائنات التي يسهل اكتشافها (مثل الكائنات الكبيرة ذات الحدود الواضحة) وعدد أقل من الكائنات التي يصعب اكتشافها (مثل الكائنات الصغيرة أو المتداخلة). يقلل فقدان التركيز البؤري من وزن الكائنات التي يسهل اكتشافها ويركز بشكل أكبر على الكائنات التي يصعب اكتشافها، مما يجعل النموذج أفضل في تحديد الكائنات الصعبة وتعطى صيغتها

بالعلاقة:

$$FL(p_t) = -(1 - p_t)^{\gamma} \log(p_t) \quad (25)$$

$$p_t = \begin{cases} p, & y = 1 \\ 1 - p, & \text{otherwise} \end{cases} \quad (26)$$

كما يتم حساب دالة الخسارة EIoU وفق المعادلة التالية:

$$L_{EIoU} = L_{IoU} + L_{dis} + L_{asp} \quad (27)$$

$$L_{EIoU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{(w^c)^2 + (h^c)^2} + \frac{\rho^2(w, w^{gt})}{(w^c)^2} + \frac{\rho^2(h, h^{gt})}{(h^c)^2}$$

$$L_{EIoU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{(w^c)^2 + (h^c)^2} + \frac{\rho^2(w, w^{gt})}{(w^c)^2} + \frac{\rho^2(h, h^{gt})}{(h^c)^2}$$

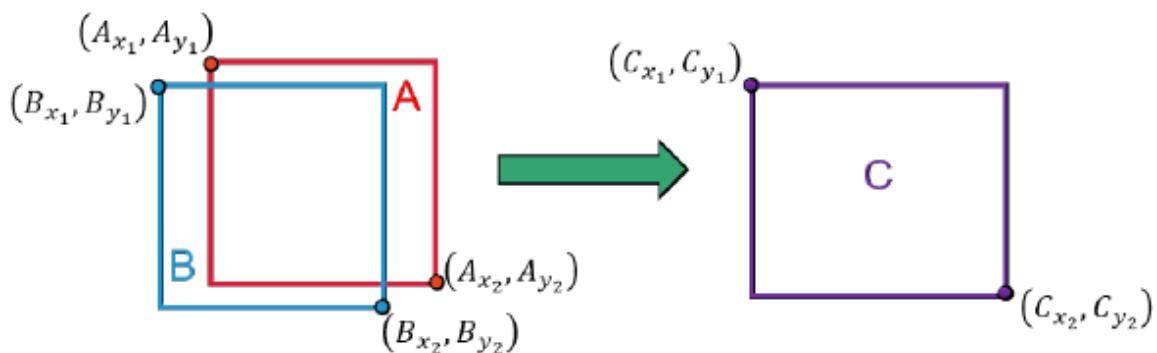
حيث تدل w^c, h^c عن عرض وارتفاع المستطيل الأصغر الذي يغطي كلا المربعين الفعلي والمتوقع، وتعبر المسافة ρ عن المسافة الأقلية و IOU نسبة التداخل بين المربع التبؤى والمربع الحقيقي، اما b فهو المربع التبؤى

- تحسين طريقة ما بعد المعالجة
- في عملية معالجة النهاية، تستخدم خوارزمية YOLOv5 بشكل افتراضي خوارزمية **Non-Maximum Suppression (NMS)** لإزالة المربعات التبؤية المكررة. تقوم خوارزمية NMS بمقارنة المربع التبؤى الذي يحصل على أعلى درجة مع المربعات التبؤية الأخرى بناءً على نسبة التداخل (IoU) وتزيل تلك التي لديها ثقة منخفضة وتدخل عالي مع المربع الأعلى درجة.
- رغم أن NMS يمكنها إزالة المربعات التبؤية المكررة بفعالية، إلا أن لديها بعض العيوب إذ تقوم NMS بإزالة المربعات التي لديها تداخل أكبر من العتبة المحددة مباشرة، مما يعني أنه إذا كان هناك كائن في المنطقة المتدخلة، فإن المربع التبؤى قد يُزال، وبالتالي تقل دقة الكشف.

- من أجل تحسين اختيار المربعات التنبؤية تم استخدام طريقة دمج المربعات بالوزن (Weighted Box Box Fusion) والتي تنص على:

- عدم إهمال أي مربعات تنبؤية: بخلاف NMS التي تزيل بعض المربعات التنبؤية، طريقة دمج المربعات بالوزن لا تخلص من أي مربع.
- توزيع وزن لكل مربع تنبؤي متداخل: تقوم الطريقة بتعيين وزن لكل مربع تنبؤي متداخل.
- دمج المناطق التنبؤية: بدلاً من التخلص من المربعات المتداخلة، يتم دمج المناطق التي تغطيها هذه المربعات للحصول على مربع تنبؤي جديد وأكثر دقة. يتم هذا الدمج عن طريق حساب الأوزان المعطاة لكل مربع تنبؤي [56].

ما يسمح بأخذ جميع التنبؤات بعين الاعتبار مما يحسن الدقة ويقلل التحيز في التنبؤ.



الشكل 60- دمج المربعات الموزنة.

- 1- في البداية، يتم تحديد المربعات المرشحة المتداخلة بناءً على نسبة التداخل (IoU) المربعات التي لديها أعلى IoU من العتبة المحددة تعتبر متداخلة.
- 2- تعين وزن لكل مربع مرشح بناء على درجة الثقة.
- 3- يتم الحصول على الصندوق المرشح الجديد C عن طريق حساب الإحداثيات المرجحة للصندوقين المرشحين A و B ، وأوزانهما على الشكل التالي:

$$C_{x_1} = \frac{A_{x_1} \cdot A_w + B_{x_1} \cdot B_w}{A_w + B_w} \quad (28)$$

$$C_{y_1} = \frac{A_{y_1} \cdot A_w + B_{y_1} \cdot B_w}{A_w + B_w} \quad (29)$$

$$C_{x_2} = \frac{A_{x_2} \cdot A_w + B_{x_2} \cdot B_w}{A_w + B_w} \quad (30)$$

$$C_{y_2} = \frac{A_{y_2} \cdot A_w + B_{y_2} \cdot B_w}{A_w + B_w} \quad (31)$$

4- ويتم أيضًا حساب وزن الصندوق المرشح C عن طريق حساب المتوسط.

$$C_w = \frac{A_w + B_w}{2} \quad (32)$$

5- يمكن أيضًا للصندوق المرشح المدمج C أن يستمر في المشاركة في الاندماج اللاحق وتشكيل سلسلة من الاندماج

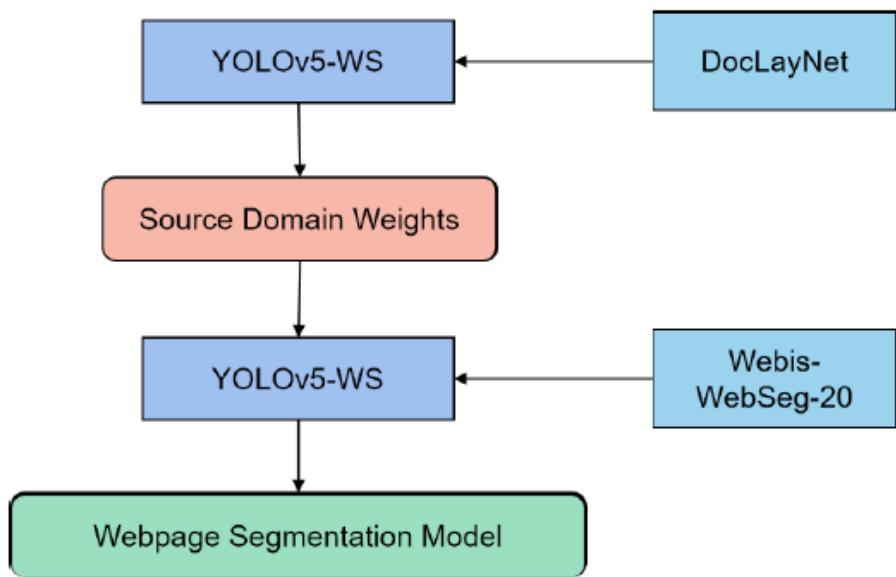
▪ نقل التعلم Transfer learning

نظرًا لصغر حجممجموعات البيانات المستخدمة في مهام تقطيع صفحات الويب، مثل Webis-WebSeg-20، يمكن أن يكون نقل المعرفة من مجالات مشابهة مفيدًا في تحسين فعالية وسرعة تدريب هذه المهام. إحدى هذه المجالات هي تحليل تحظيط المستندات، والتي يتم دراستها باستخدام مجموعة بيانات DocLayNet. تتكون هذه المجموعة بشكل رئيسي من الكتب والصحف والمجلات، وتحتوي على ميزات مثل النصوص الرئيسية، العنوانين، والرسوم التوضيحية، والتي تشبه جزئياً ميزات عرض صفحات الويب.

خطوات التدريب:

1. يجري تدريب نموذج باستخدام بيانات DocLayNet للحصول على أوزان النموذج.
2. نقل الأوزان: نستخدم هذه الأوزان كنقط ابداية لشبكة نموذج تقطيع صفحات الويب كما يتم تحميل بعض أوزان الطبقة التلايفية أثناء عملية التدريب.
3. نكمل تدريب النموذج باستخدام بيانات تقطيع صفحات الويب لتحسين دقتها في هذه المهمة.

بهذا النهج، يمكن للنموذج الاستفادة من المعرفة المكتسبة من تحليل تحظيط المستندات لتحقيق توازن أفضل وأسرع في تدريب مهام تقطيع صفحات الويب، مما يؤدي إلى نتائج أفضل. الشكل 61 يوضح هذه العملية بالتفصيل.



الشكل 61-تقنية نقل التعلم.

المراجع

- [1] <https://www.statista.com/statistics/273018/number-of-internet-users-worldwide>
- [2] <https://www.w3.org/>
- [3] Michael Cooper, Ben Caldwell, Loretta Guarino Reid, Gregg Vanderheiden, “**Web Content Accessibility Guidelines (WCAG) 2.0**” Available: “<https://travesia.mcu.es/server/api/core/bitstreams/6c7e00c8-cd2d-4066-bba0-2203ba2281ad/content>”
- [4] Padraig Cunningham, Matthieu Cord, and Sarah Jane Delany, “**Supervised Learning**”, Available: “https://www.researchgate.net/profile/Sarah-Delany/publication/225849227_Supervised_Learning/links/0fcfd50d0c1d228420000000/Supervised-Learning.pdf”
- [5] Zoubin Ghahramani, “**Unsupervised Learning**”, September 16, 2004.
- [6] Mohammad Javad Salehpour, Mohammad Javad Hossain “**Leveraging machine learning for efficient EV integration as mobile battery energy storage systems: Exploring strategic frameworks and incentives**” Received in revised form 6 May 2024; Accepted 17 May 2024/. Available: “<https://doi.org/10.1016/j.est.2024.112151>”.
- [7] Shi Na, Liu Xumin, Guan yong “**Research on k-means Clustering Algorithm an Improved k-means Clustering Algorithm**” Available: “DOI 10.1109/IITSI.2010.74”.
- [8] Soni Madhulatha, “**Comparison between K-Means and K-Medoids Clustering Algorithms**” April 2011.
- [9] Fionn Murtagh, Pedro Contreras, “**Algorithms for hierarchical clustering: An overview**” January 2012, Available: “DOI: “[10.1002/widm.53](https://doi.org/10.1002/widm.53)”
- [10] Pinaki MITRA, “**A survey of density based clustering algorithms**” Front. Comput. Sci., 2021, Available: “<https://doi.org/10.1007/s11704-019-9059-3>”.
- [11] Dongkuan Xu, Yingjie Tian “**A Comprehensive Survey of Clustering Algorithms**” Published online: 12 August 2015, Springer-Verlag Berlin Heidelberg 2015. Available: “<https://link.springer.com/article/10.1007/S40745-015-0040-1>”.
- [12] Licensee MDPI, Basel, “**Research on the Real-Time Detection of Red Fruit Based on the You Only Look Once Algorithm**” Published: 20 December 2023. Available: “<https://doi.org/10.3390/pr12010015>”.
- [13] Xuran Pan Tianzhu Ye Dongchen Han Shiji Song Gao Huang, “**Contrastive Language-Image Pre-Training with Knowledge Graphs**” 2022.
- [14] Alexander Kirillov Eric Mintun Nikhila Ravil, Hanzi Mao Chloe Rolland Laura Gustafson Tete Xiao Spencer Whitehead Alexander C. Berg Wan-Yen Lo Piotr Dollar’

Ross Girshick⁴ 1 project lead joint first author equal contribution directional lead
“*Segment Anything*”, Meta AI Research, FAIR, 5 Apr 2023 Available:
[https://arxiv.org/pdf/2304.02643](https://arxiv.org/pdf/2304.02643.pdf).

- [15] Nikhila Ravi, Valentin Gabeur, Yuan-Ting Hu, Ronghang Hu, Chaitanya Ryali, Tengyu Ma, Haitham Khedr, Roman Rädle, Chloe Rolland, Laura Gustafson, Eric Mintun, Junting Pan, Kalyan Vasudev Alwala, Nicolas Carion, Chao-Yuan Wu, Ross Girshick, Piotr Dollár, Christoph Feichtenhofer, “*SAM 2: Segment Anything in Images and Videos*”, Available: <https://ai.meta.com/sam2>.
- [16] Junnan Li, Dongxu Li, Caiming Xiong, Steven Hoi, “*BLIP: Bootstrapping Language-Image Pre-training for Unified Vision-Language Understanding and Generation*”, 2022, Available: <https://proceedings.mlr.press/v162/li22n.html>.
- [17] Mingxi Zhang, Xuemin Li, Shubo Yue, and Liuqian Yang “*An Empirical Study of TextRank for Keyword Extraction*” September 29, 2020.
- [18] Yaru Hao, Li Dong, Furu Wei, Ke Xu, “*Visualizing and Understanding the Effectiveness of BERT*” 15 Aug 2019.
- [19] GOKUL YENDURI, M. RAMALINGAM, G. CHEMMALAR SELVI, Y. SUPRIYA, GAUTAM SRIVASTAVA, (Senior Member, IEEE), PRAVEEN KUMAR REDDY MADDIKUNTA, G. DEEPTI RAJ, RUTVIJ H. JHAVERI, (Senior Member, IEEE), B. PRABADEVI, WEIZHENG WANG, (Graduate Student Member, IEEE), ATHANASIOS V. VASILAKOS, (Senior Member, IEEE), AND THIPPA REDDY GADEKALLU 8,9,10, (Senior Member, IEEE), “*GPT (Generative Pre-Trained Transformer)— A Comprehensive Review on Enabling Technologies, Potential Applications, Emerging Challenges, and Future Directions*” 15 April 2024.
- [20] HAOPENG ZHANG, PHILIP S. YU, JIAWEI ZHANG “*A Systematic Survey of Text Summarization: From Statistical Methods to Large Language Models*”, 17 Jun 2024.
- [21] Yashal Shakti Kanungo, Gyanendra Das, “*COBART: Controlled, Optimized, Bidirectional and Auto-Regressive Transformer for Ad Headline Generation*” 22, August 14–18, 2022, USA, Available: <https://doi.org/10.1145/3534678.3539069>
- [22] Lei Gao, Haolong Hong, and Fan Min “*Stable first-arrival picking through mathematical morphology and edge detection*” publication 2023 October 18, Available: <https://doi.org/10.1093/gji/ggad410>.
- [23] <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>.
- [24] D. Cai, S. Yu, J. R. Wen, and W. Y. Ma, “*VIPS: a vision-based page segmentation algorithm*,” Technical Report MSR-TR-2003-79. Microsoft Research, Nov. 2003. [Online]. Available: <http://research.microsoft.com/pubs/70027/tr-2003-79.pdf>
- [25] C. Kohlschütter and W. Nejdl, “*A densitometric approach to web page segmentation*” in Proceeding of the 17th ACM conference on Information and knowledge mining - CIKM '08, 2008, p. 1173. Available: “<https://dl.acm.org/doi/10.1145/1458082.1458237>”.
- [26] Saeedeh Sadat Sajjadi Ghaemmaghami, James Miller “*Integrated-Block: A New Combination Model to Improve Web Page Segmentation*” Received 19 April 2021 Accepted 31 January 2022; Publication 15 April 2022. Available : “<https://journals.riverpublishers.com/index.php/JWE/article/view/6917>”.

- [27] Kiril Griaev, Simona Ramanauskaitė. “***Web Page Content Block Identification with Extended Block Properties***” Published: 5 May 2023 Available: “<https://doi.org/10.3390/app13095680>”.
- [28] Jocher, Glenn; Chaurasia, Ayush; Stoken, Alex; Borovec, Jirka; NanoCode012; Kwon, Yonghye; Michael, Kalen; TaoXie; Fang, Jiacong; Imyhxy; Lorna; 曾逸夫, "Zeng Yifu"; Wong, Colin; V, Abhiram; Montes, Diego; Wang, Zhiqiang; Fati, Cristi; Nadar, Jebastin; Laughing; UnglvKitDe; Sonck, Victor; Tkianai; YxNONG; Skalski, Piotr; Hogan, Adam; Nair, Dhruv; Strobel, Max; Jain, Mrinal “***ultralytics/yolov5: v7.0 - YOLOv5 SOTA Realtime Instance Segmentation***” Published November 22, 2022 | Version v7.0. Available: “<https://doi.org/10.5281/zenodo.7347926>”.
- [29] Li Dai, Zunwang Ke, Wushour Silamu, “***YOLO-WS: A Novel Method for Webpage Segmentation***” CNIOT ’23, May 26–28, 2023, Xiamen, China, “<https://doi.org/10.1145/3603781.3603862>”.
- [30] L. Hubert and P. Arabie, “***Comparing partitions***” Journal of classification, vol. 2, no. 1, pp. 193–218, 1985.
- [31] Deepak Dharrao, Manasvi Mishra, Aqsa Kazi, Madhuri Pangavhane, Priya Pise, Anupkumar M. Bongale, “***Summarizing Business News: Evaluating BART, T5, and PEGASUS for Effective Information Extraction***”, 21 June 2024 Available “<http://creativecommons.org/licenses/by/4.0/>”
- [32] <https://www.bbc.co.uk/news>
- [33] <https://github.com/rkrzr/dataset-popular>
- [34] <https://github.com/rkrzr/dataset-random>
- [35] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick “***Microsoft COCO: Common Objects in Context***” Publishing Switzerland 2014, Available: <http://mscoco.org/>.
- [36] <https://huggingface.co/datasets/EdinburghNLP/xsum>
- [37] https://huggingface.co/datasets/abisee/cnn_dailymail
- [38] Ylber Januzaj, Edmond Beqiri, Artan Luma “***Determining the Optimal Number of Clusters using Silhouette Score as a Data Mining Technique***”, Final acceptance 2023-01-29. Available: “<https://doi.org/10.3991/ijoe.v19i04.37059>”.
- [39] Omar Tahri, François Chaumette, “***Point-Based and Region-Based Image Moments for Visual Servoing of Planar Objects***” December 2005. Available: “<https://www.researchgate.net/publication/3450251>”.
- [40] <https://huggingface.co/sshleifer/distilbart-cnn-12-6>
- [41] Top Programming Languages 2021 - IEEE Spectrum. “<https://spectrum.ieee.org/top-programming-languages-2021>” (accessed Jun. 15, 2024).
- [42] P. Ramya, V. Sindhura, and P. V. Sagar, “***Testing using selenium web driver***” in 2017 Second International Conference on Electrical, Computer and Communication Technologies (ICECCT), Feb. 2017, pp. 1–7. Available: “<https://www.semanticscholar.org/paper/Testing-using-selenium-web-driver-Ramya-Sindhura/c816af4f20fb9115905daf50e11e51ed72335217>” (accessed Jun. 15, 2024).

- [43] G. L. Hajba, “*Website Scraping with Python*”. Berkeley, CA: Published 2018. Available: “<https://link.springer.com/book/10.1007/978-1-4842-3925-4>”
- [44] Guifen Zhao, Yanjun Liu, Wei Zhang, Yiou Wang. “*TFIDF based Feature Words Extraction and Topic Modeling for Short Text*” 13 January 2018.
- [45] Alabhyaa Farkiya, Prashant Saini, Shubham Sinha, “*Natural Language Processing using NLTK and WordNet*” International Journal of Computer Science and Information Technologies, 2015.
- [46] Travis E. Oliphant, PhD, “*Guide to NumPy*” 2006.
- [47] Wes McKinney, “pandas: a Foundational Python Library for Data Analysis and Statistics”, 2011.
- [48] Ali Hassan Sial, Syed Yahya Shah Rashdi, Dr. Abdul Hafeez Khan, “*Comparative Analysis of Data Visualization Libraries Matplotlib and Seaborn in Python*”, 2021, Available Online at <http://www.warse.org/IJATCSE/static/pdf/file/ijatcse391012021.pdf>
- [49] Fabian Pedregosa, Gael Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, “*Scikit-learn: Machine Learning in Python*”, 2011.
- [50] Polina Lemenkova, “*PROCESSING OCEANOGRAPHIC DATA BY PYTHON LIBRARIES NUMPY, SCIPY AND PANDAS*” Available: “<https://doi.org/10.3153/AR19009>”.
- [51] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierrick Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, Alexander M. Rush, “*Transformers: State-of-the-Art Natural Language Processing*”, 2020.
- [52] Alessandro Fantechi, Stefania Gnesi, Samuele Livi, Laura Semini, “*A spaCy-based tool for extracting variability from NL requirements*”.
- [53] Qibin Hou, Daquan Zhou, and Jiashi Feng. “*Coordinate Attention for Efficient Mobile Network Design*”. 4 Mar 2021 Available: “<http://arxiv.org/abs/2103.02907>”.
- [54] Ningning Ma, Xiangyu Zhang, and Jian Sun. “*Funnel Activation for Visual Recognition*”. 24 Jul 2020, Available : “<https://doi.org/10.48550/arXiv.2007.11824>”.
- [55] Dongxian Wu, Yisen Wang, Shu-Tao Xia, James Bailey, and Xingjun Ma. “*Skip Connections Matter: On the Transferability of Adversarial Examples Generated*” 14 Feb 2020, Available: <http://arxiv.org/abs/2002.05990>
- [56] Roman Solovyev, Weimin Wang, and Tatiana Gabruseva. “*Weighted boxes fusion: Ensembling boxes from different object detection models*”. March 2021. Available: “<https://doi.org/10.1016/j.imavis.2021.104117>”