أرشفة المحتوى الرقمي السعودي على شبكة الويب: تصور مقترح

د. محمد بن صالح الطيار
قسم إدارة المعلومات، كلية علوم الحاسب
والمعلومات
جامعة الإمام محمد بن سعود الإسلامية، الرياض

المستخلص:

يعد الأرشيف من أهم مكونات التراث الوطني لأي دولة وهو بمثابته ذاكرتها الرسمية التي تحفظ ثقافتها وتاريخها، حيث ينشأ الأرشيف نتيجة لجمع وحفظ وإتاحة المعلومات المدونة على كافة مصادر المعلومات أيًّا كان نوعها وشكلها، بما في ذلك مصادر المعلومات الرقمية. ونظرًا لأهمية الأرشيف ودوره الحيوي في حفظ تراث وثقافة الأمة، عملت الدول والحكومات والمؤسسات على تهيئة السبل اللازمة التي تضمن جمعه وحفظه وإتاحته للأجيال الحالية والمستقبلية. تهدف هذه الدراسة إلى تقديم تصور مفاهيمي لبناء أرشيف رقمي لمحتويات مصادر الويب المتاحة على النطاق السعودي. اعتمدت الدراسة على منهج تحليل الوثائق، وذلك من خلال تحليل المحتويات أو ما يعرف بالنتاجات (المكونات)، واستعراض لما تم في مجال أرشفة محتويات الويب الرقمي من خلال التعرف على أبرز التجارب والممارسات الدولية وعمل بعض المقارنات المرجعية، بالإضافة إلى تحليل خصائص وسمات المحتوى الرقمي السعودي المتاح على شبكة الإنترنت. وأظهرت نتائج الدراسة أنه لا يوجد حتى الآن أرشيف رقمي يجمع ويحفظ ويؤرشف محتويات الويب المتاحة على النطاق السعودي.

وبلغ مجموع النطاقات المتاحة على الويب السعودي ما يقارب من ٩٢٤٥٠ نطاقًا، وتغطي مجالات مختلفة. كما أن محتويات الويب متنوعة بين المحتويات الرسمية وغير الرسمية. وتشير النتائج إلى أن عملية أرشفة محتويات الويب يمكن أن تتم معظمها بشكل آلي من خلال استخدام العديد من الأدوات التقنية مثل أداة وي باك WayBack، وأداة اركيف ات Archive-It ، وأداة هيريتريكس Heritrix، وأداة Wget. كما أظهرت النتائج وجود العديد من التحديات التي تواجه هذا المشروع مثل الحجم الهائل لمحتويات الويب، تنوع الملفات وأشكالها، المحتويات المغلقة، حقوق الملكية الفكرية بالإضافة إلى بعض التحديات الأخرى. وتختم الدراسة بالعديد من التوصيات والتي من أهمها أن هذا المشروع المقترح يمكن أن يساهم في تعزيز الحضور الرقمي للمحتوى السعودي على الإنترنت في المؤشرات الرقمية الدولية، بالإضافة إلى تحقيق بعض من مستهدفات رؤية السعودية ٢٠٣٠ فيما يتعلق بالتحول الرقمي.

Abstract:

The archive is one of the most important components of the national heritage of any country, and it is the official memory that preserves its culture and history. In addition, the archive is created as a result of collecting, preserving and making available information written on all sources of information of any kind and form, including digital sources of information. In view of the importance of the archive and its vital role in preserving the heritage and culture of the nation, states, governments and institutions have worked to prepare the necessary means to ensure the collection, preservation and availability knowledge for current and future generations. This study aims to provide a conceptual framework for building a digital archive of web resources available on the Saudi domain. The study used content and document analysis approach, and a review of what was done in the field of archiving the contents of the digital web by identifying the most prominent international experiences practices and benchmarking. In addition, it analysed the characteristics of the Saudi digital content available on the web. The results shows that there is no digital archive that collects, preserves and archives web contents

available on the Saudi domain. The total domains available are approximately 83,109 domains, covering various domains. The contents of the web vary between official and unofficial contents. Furthermore, the results indicates that the process of archiving web contents can be done mostly automatically through the use of many technical tools such as WayBack tool, Archive-It tool, Heritrix tool, and Wget tool. Moreover, the findings reveal that there are many challenges facing this project including the huge size of web contents, files formats, closed contents, intellectual property rights, as well as other challenges. The study concludes with several recommendations, and it argues that the proposed project can contribute to enhancing the digital presence of Saudi content on the Internet in international digital indicators, as well as achieving some of the objectives of Saudi Vision 2030 in relation to digital transformation.

الكلمات المفتاحية

الويب، الأرشفة الرقمية، الإنترنت، السعودية، أرشفة محتويات الويب، المحتوى العربي على الإنترنت.

Keywords:
The World Wide Web, Digital archiving, Internet, Saudi Arabia, Web content archiving, Arabic content on the Internet.

١. المقدمة:

يعد الأرشيف من أهم مكونات التراث الوطني لأي دولة وهو بمثابته ذاكرتها الرسمية التي تحفظ ثقافتها وتاريخها، حيث ينشأ الأرشيف نتيجة لجمع وحفظ وإتاحة المعلومات المدونة على كافة مصادر المعلومات أيًّا كان نوعها وشكلها. ونظرًا لأهمية الأرشيف ودوره الحيوي في حفظ تراث وثقافة الأمة، عملت الدول والحكومات والمؤسسات على تهيئة السبل اللازمة التي تضمن جمعه وحفظه وإتاحته للأجيال الحالية والمستقبلية، وذلك من خلال الاستخدام الأمثل للموارد والامكانيات والتقنيات والسياسات. ومع الاعتماد المتزايد على استخدام تقنيات الاتصال والمعلومات، أصبحت شبكة الإنترنت

مصدرًا هامًا ورئيسيًا للمعلومات بكافة أنواعها، الإخبارية والترفيهية، ومعلومات البحث العلمي، والتجارة الإلكترونية، والمعلومات الحكومية، والتعليم، والتعلم عن بعد، ومعلومات السفر، ومعلومات البحث عن عمل وأنشطة أخرى على الإنترنت، بالإضافة إلى انتشار استخدام شبكات التواصل الاجتماعي التي تعتمد على المحتوى المنتج بواسطة المستخدمين.

وخلال العشرين سنة الماضية، أصبح الويب عبارة عن مكون أساسي يحتوي على مجموعة متعددة الأوجه من المصادر المعلوماتية التي توثق الحياة الاجتماعية والثقافية والسياسية المعاصرة (مصطفى وقوميد ٢٠٢٢)، وفي مقابل ذلك تزايد الاهتمام من قبل الحكومات والمؤسسات والباحثين في التركيز على ما ينشر على شبكة الويب من محتوى ومعلومات، والعمل على تهيئة السبل اللازمة التي تضمن الاستفادة من هذا الكم الهائل من المحتويات والمعلومات وذلك من خلال جمعها وحفظها وأرشفتها وإتاحتها للأجيال الحالية والمستقبلية. ويعزو الاهتمام بهذا الاتجاه إلى العديد من الأسباب. أولًا، تعد الويب أكبر قاعدة معلومات عرفها الإنسان منذ تاريخه وهي وعاء لحفظ الثقافة (قوميد ومصطفى ٢٠٢١). ثانيًا، ومع مرور الوقت يزيد ارتباط الناس بهذه الشبكة بشكل تصاعدي ويتزايد مع ذلك نمو حجم المحتوى الرقمي، حيث تشير آخر الإحصائيات إلى أن حجم صفحات الويب وصل إلى ما يقارب ٦٠ مليار صفحة (إحصاء الإنترنت العالمي ٢٠٢٣). ثالثًا، هناك كم هائل من المعلومات والمحتويات المتاحة على شبكة الويب لا تقل أهمية عن المصادر المطبوعة، وقد لا يوجد لها نظير مطبوع، وبالتالي فقد أصبحت مصدرًا رئيسيًا وهامًا في الوقت الحاضر للكثير من المعلومات (عكاشة ٢٠١٥) ويستفيد منها الباحثين على كافة اهتماماتهم ومستوياتهم، وتلبي حاجات

معلوماتية مختلفة. وأخيرًا، فإن من أبرز المشاكل والتحديات التي تواجه المحتوى والمعلومات على شبكة الويب هو أنه لا يمكن ضمان أنها ستكون متاحة إلى الأبد، وأنها سريعة الزوال والتغير والتحديث، حيث تشير بعض الإحصاءات إلى أن متوسط عمر الصفحة ما يقارب أربعة وأربعين يومًا (قوميد ومصطفى ٢٠٢١). وهذا يشير إلى أن الكثير من المعلومات القيمة والمواد الهامة قد تختفي بعد نشرها، وتتغير عناوينها وقد لا تكون متوفرة ومتاحة في المستقبل، مما يعني فقدانها.

مما سبق من عرض، تتضح أهمية أرشفة الويب ودوره في حفظ المحتويات والمعلومات وإتاحتها للأجيال الحالية والمستقبلية. تحاول هذه الورقة تقديم تصور مقترح لإنشاء أرشيف رقمي للمحتوى السعودي المتاح على شبكة الإنترنت. وتتكون هذه الدراسة مما يلي. في القسم التالي، يعرض الباحث مشكلة الدراسة وأسئلتها، متبوعة بأهمية الدراسة وأهدافها. بعد ذلك، يتم التعريف بمصطلحات الدراسة وحدودها. ويتبع ذلك استعراض الدراسات والتجارب السابقة في هذا المجال. ثم تعرض الورقة منهجية البحث وإجراءات الدراسة، وبعد ذلك نتائج الدراسة ومناقشتها، ومن ثم تقديم عرض تصوري مفاهيمي للمشروع المقترح، وتختم الدراسة بأبرز النتائج والتوصيات.

١,١ مشكلة الدراسة وأسئلتها:

تنبثق مشكلة الدراسة من طبيعة الموضوع الذي تتناوله، حيث تسلط الضوء على أهمية جمع وحفظ وأرشفة المحتوى الرقمي المنشور على الإنترنت كمصدر معلومات وإتاحته للأجيال الحالية والمستقبلية. وكما تمت الإشارة إليه سابقًا فإن من أبرز المشاكل التي تواجه المعلومات على شبكة الإنترنت هو أنه لا يمكن ضمان أنها ستكون متاحة إلى الأبد، وأنها سريعة الزوال والتغير

والتحديث، والبعض منها قد يختفي للأبد. وهذا يشير إلى أن الكثير من المعلومات القيمة والمواد الهامة قد تختفي من شبكة الإنترنت بعد ما تنشر، وتتغير عناوينها URLs وقد لا تكون متوفرة ومتاحة في المستقبل، وهو ما يعني فقدانها، بالإضافة إلى أن الكثير من هذه المعلومات والمحتويات قد تظهر وتولد أساسًا بشكل رقمي وقد لا يوجد لها نظير مطبوع في المصادر التقليدية. وبالتالي فإن هذه المحتويات لا تقل أهمية عن مصادر المعلومات الأخرى، حيث إن عدم الاهتمام بها وتوثيقها وحفظها قد يكون سببًا في إيجاد فجوة رقمية وتاريخية للأجيال القادمة. إضافة لذلك، وفيما يخص سياق هذه الدراسة، فإنه لا يوجد حتى الآن أي مشروع أو مبادرة وطنية تهدف لجمع وتوثيق وأرشفة المحتوى المنشور على نطاق الإنترنت السعودي، حيث تقدم هذه الدراسة مقترح وتصور في هذا الخصوص. وتسعى هذه الدراسة للإجابة على العديد من التساؤلات وذلك كما يلي:

- ما أهمية جمع وحفظ وأرشفة المحتوى الرقمي المنشور على شبكة الإنترنت كمصدر من مصادر المعلومات؟
- ماهي أبرز التجارب والممارسات الدولية في مجال حفظ محتوى الويب وأرشفته وتوثيقه؟
- ما هي الآليات والتقنيات المستخدمة في جمع وحفظ وأرشفة المحتوى الرقمي المنشور على شبكة الإنترنت؟
- ما هو واقع المحتوى الرقمي السعودي المتاح على شبكة الإنترنت وما أبرز خصائصه ومكوناته؟
- كيف يمكن بناء نموذج آلي يعمل على جمع المحتوى الرقمي السعودي على الإنترنت وحفظه وأرشفته؟

١,٢ أهمية الدراسة وأهدافها:

تسعى هذه الدراسة إلى تحقيق العديد من الأهداف من أهمها:

- التعريف بأهمية جمع وحفظ وأرشفة المحتوى الرقمي المنشور على شبكة الإنترنت كمصدر من مصادر المعلومات.
- استعراض بعض التجارب والممارسات الدولية في مجال حفظ محتوى الويب وأرشفته وتوثيقه.
- عرض لأبرز الآليات والتقنيات المستخدمة في جمع وحفظ وأرشفة المحتوى الرقمي المنشور على شبكة الإنترنت.
- التعريف بواقع المحتوى الرقمي السعودي المتاح على شبكة الإنترنت وأبرز خصائصه ومكوناته.
- بناء نموذج مقترح تصوري لإنشاء نظام يعمل على جمع المحتوى الرقمي السعودي على الإنترنت وحفظه وأرشفته.

١,٣ مصطلحات الدراسة

شبكة الإنترنت Internet

الإنترنت هو نظام شبكي متصل عالميًا يعمل على تسهيل الاتصال العالمي والوصول إلى موارد البيانات من خلال مجموعة واسعة من الشبكات الخاصة والعامة والتجارية والأكاديمية والحكومية، وذلك وفقًا لبروتوكولات عالمية مقننة. غالبًا ما يتم استخدام مصطلحات الإنترنت والشبكة العنكبوتية العالمية (الويب) بشكل تبادلي، لكنهما مختلفان تمامًا، حيث يشير الإنترنت إلى نظام الاتصال العالمي بما في ذلك الأجهزة والبنية التحتية، بينما يعد الويب أحد الخدمات التي تقدمها الإنترنت (Techopedia 2020).

المحتوى الرقمي Digital Content

يقصد بالمحتوى الرقمي كل ما هو متاح على شبكة الإنترنت ويشمل محتويات مواقع الويب وقواعد البيانات والمعلومات، والملفات الرقمية التي تأتي في أشكال متنوعة مثل الملفات السمعية، والبصرية، والمرئية، والنصية، والبرمجيات وغيرها. ومن أمثلة منتجات المحتوى الرقمي مواقع المؤسسات والمنظمات، البرامج والألعاب، والصور، الكتب والصحف والمجلات، البحوث والدراسات العلمية والتقارير، الوثائق الرسمية الحكومية والغير حكومية، المخطوطات والوثائق، المواقع الشخصية مثل المدونات، محتويات مواقع وشبكات التواصل الاجتماعي، المنتديات وساحات النقاش (عبود ٢٠١٣).

أرشفة الويب Web Archiving

هي عملية جمع الصفحات والمواقع الإلكترونية ومحتوياتها المنشورة على شبكة الويب وأخذ نسخ أو صور (لقطات) snapshots منها، بهدف جمعها وحفظها في أرشيف رقمي على المدى الطويل، وذلك عن طريق استخدام برامج آلية أو ما يعرف بزواحف الويب web crawlers والتي تقوم بتتبع المواقع الإلكترونية المراد أرشفتها بصفة دورية (Niu 2012).

شبكات التواصل الاجتماعي Social Media Networks

في أبسط أشكالها، تعد شبكات التواصل الاجتماعي منصات قائمة على تقنيات الإنترنت، وتتضمن العديد من الأدوات والبرامج التي توفر للمستخدمين فرص لإنشاء وتبادل المحتويات والمعلومات، وتعزز مفهوم المحتوى المنتج بواسطة المستخدمين، أو ما يعرف User Generated Content. ومن أمثلة هذه الشبكات موقع يوتيوب، تويتر، فيسبوك، انستغرام، وغيرها من المنصات المعروفة (Khan 2015; Altayar 2016).

زواحف الويب Web Crawlers

تعرف الموسوعة التقنية (Techopedia ٢٠١٧) زواحف الويب على أنها برامج قائمة على الروبوتات، وتقوم بالوصول والدخول لصفحات الويب بشكل مستمر وتعمل على فهرسة وتكشيف الويب ومحتوياتها. كما تجمع هذه البرامج معلومات عن الصفحات والمحتويات مثل العناوين URL لمواقع الويب والبيانات الوصفية Metadata والروابط الموجودة في صفحات الويب والوجهات المؤدية من وإلى هذه الروابط وأي معلومات أخرى ذات صلة. وتقوم عملية جمع الصفحات والمحتويات ضمن سياسة معينة يتم تحديدها في هذه البرامج مثل عدد مرات الزيارة، اختيار المواد والمحتويات، والنطاق أو المجال Domain بالإضافة إلى عناصر أخرى.

١,٤ حدود الدراسة

تتناول هذه الدراسة البحث محل الاستقصاء بالحدود التالية:

الحدود الموضوعية:

تركز هذه الدراسة على أرشفة محتوى الويب الرقمي بكافة أشكاله وأنواعه والمتاح على النطاق العام السعودي على شبكة الإنترنت.

الحدود المكانية:

تركز هذه الدراسة على أرشفة محتوى الويب الرقمي المتاح على شبكة الإنترنت في المملكة العربية السعودية، مع استعراض لتجارب بعض الدول في بناء أراشيف متميزة.

الحدود الزمانية:

تم إعداد هذه الدراسة خلال الفترة من شهر جون ٢٠٢٢ إلى شهر ديسمبر ٢٠٢٢.

الحدود النوعية:

تقتصر هذه الدراسة على المحتويات المنشورة ضمن نطاق الويب السعودي والمتاح للعامة على شبكة الإنترنت، ويشمل ذلك محتويات المؤسسات الحكومية وغير الحكومية، المواقع الشخصية، مواقع الصحف والمجلات، حسابات شبكات التواصل الاجتماعي، المنتديات وساحات النقاش، البحوث والدراسات العلمية، وغيرها من محتويات الويب الأخرى.

٢. الإطار النظري:

الدراسات السابقة:

تناولت العديد من الدراسات السابقة العربية والأجنبية موضوع إنشاء وبناء أرشيف الويب من زوايا متنوعة، منها ما ركز على الجانب الفني والتقني، والبعض الآخر على نوعية المحتويات والمعلومات، كما ركز البعض الآخر على الأدوات والوسائل اللازمة لذلك، بالإضافة إلى الجوانب القانونية المتعلقة بحقوق الملكية الفكرية. كما يوجد نوع آخر من الدراسات ركز على دراسات الحالة ووصف بعض الممارسات والتجارب في بناء وإنشاء أراشيف للويب خاصة ببعض الدول.

٢,١ الدراسات العربية:

تعد دراسة خليفة (٢٠٠٦) من أوائل الدراسات التي سلطت الضوء على أرشفة محتويات الإنترنت باللغة العربية، حيث تناول الباحث من خلال دراسة ميدانية تحليلية نموذج مقترح لأرشفة المواقع المصرية على الإنترنت. واعتمدت الدراسة على المنهج المسحي الميداني، وتحليل عينات من مواقع ومحتويات الإنترنت المصرية. وتوصلت الدراسة إلى العديد من النتائج، من أهمها ضعف الضبط الببليوجرافي الوطني لمواقع الإنترنت في مصر، حيث لا

توجد أداة تعرف بهذه المواقع والمحتويات، وعدم التغطية الشاملة لكافة نطاقات الإنترنت المصري، وقصور في أدلة البحث المتاحة على الإنترنت، بالإضافة إلى أن المواقع المصرية تعاني من قصور في تقديم البيانات التوثيقية للمواقع والمحتويات مثل تاريخ إنشاء المواقع.

وتناولت دراسة (قوميد ومصطفى ٢٠٢١) ممارسات حفظ الثقافة على الويب. واستعرض الباحثان أهمية الويب ودوره في حفظ التراث والثقافة، حيث إنها أصبحت أضخم قاعدة معلومات عرفها الإنسان، وأن نمو الويب يتزايد بشكل مستمر وكبير، وأن كثير من المعلومات والمحتويات المتاحة على الويب قد تتغير أو تختفي بعد نشرها. كما استعرضت الدراسة العديد من التجارب المتعلقة بحفظ التراث الشعبي على الإنترنت مع ذكر لنماذج ومشاريع في مجالات محددة في بعض الدول العربية والأجنبية. كما تناولت الدراسة عرض لمشروع أرشفة الويب Internet Archive في الولايات المتحدة الأمريكية. وعرضت الدراسة أبرز المشاكل التي تواجه حفظ المحتوى الرقمي بما في ذلك تركيب وصيانة البنية التحتية التقنية اللازمة، توفر الخبرات والمهارات، الحفاظ على السياق المرتبط بالمعلومات ذات الدلالة مثل روابط النصوص التشعبية، إيجاد معايير للممارسة الجيدة وسياسات الاقتناء التي تحدد نوعية المجموعات ومصادرها وأصالتها، ومشكلات تغيير صفحات الويب عند نقلها إلى أجهزة جديدة. وختمت الدراسة بالتأكيد على أهمية المحتوى الرقمي المنشور على الويب، والعمل المبكر لحمايته والمحافظة عليه، حيث إن جمعه وتنظيمه وأرشفته سيقدم سجلًا مميزًا وفريدًا من نوعه ولا يوجد له نظير في مصادر أخرى، بحيث يمكن إتاحته للأجيال المستقبلية.

وقدمت دراسة سعيدي وآخرون (٢٠٢١) نموذج مقترح لبناء مشروع الأرشفة الإلكترونية. واستخدمت الدراسة المنهج الوصفي لجمع بيانات الدراسة وتحليلها لبناء إطار مفاهيمي ودلالي للأرشفة الإلكترونية. وتقدم الدراسة تصور للقيام بأي مشروع أرشفة الكتروني من خلال العديد من الخطوات المتتابعة. وتأتي المرحلة الأولى وهي مرحلة التخطيط والإعداد للمشروع ويشمل ذلك تعيين مدير المشروع وفريق عمل، دراسة الجدوى، تحديد الأهداف ورسم الخطة الزمنية، وإيجاد مصادر التمويل، متطلبات البنية التحتية والتقنية، استراتيجية الحفظ الرقمي، بالإضافة إلى توثيق إجراءات مشروع الأرشفة الرقمية. بعد ذلك تأتي المرحلة الثانية، وهي مرحلة تنفيذ مشروع الأرشفة الإلكترونية والتي تتضمن بناء الأرصدة الأرشيفية الرقمية وإدارتها وإتاحتها. والمرحلة الثالثة هي مرحلة الصيانة والمتابعة والتي تهدف لضمان سير عمل المشروع وفق الخطة المعتمدة والخطوات المتفق عليها لتحقيق الأهداف المرجوة.

كما قام الحباطي (٢٠٢١) بعمل دراسة تحليلية لحفظ التراث الثقافي غير المادي بالوادي الجديد في جمهورية مصر العربية مع تصميم موقع ويب لإتاحته. واستخدم الباحث المنهج التاريخي والميداني في جمع البيانات وتحليلها. وقدمت الدراسة تصور مقترح لجمع وحفظ وأرشفة المحتويات ذات العلاقة بموضوع الدراسة بما في ذلك الصور الفوتوغرافية للأحداث التاريخية، الملفات الصوتية والأفلام، المخطوطات، وغيرها من مصادر المعلومات. واقترحت الدراسة أبرز أدوات الوصف الببليوجرافي التي يمكن استخدامها في هذا الخصوص مثل مارك ٢١، معيار وصف وإتاحة المصادر RDA ، ومعيار VRA.

وتناول عبدالراضي (٢٠٢٢) أبرز معايير واصفات البيانات الأرشيفية في بيئة الويب، واستخدم الباحث المنهج المسحي لحصر معايير واصفات البيانات المتاحة ودراسة خصائصها وسماتها وأبعادها. واستعرضت الدراسة تسلسلًا تاريخيًا لنشأة وتطور معايير واصفات البيانات الأرشيفية، مع عرض لما يقارب من عشرين معيارًا. وتوصلت الدراسة إلى أن نمو الويب المتسارع وظهور المحتويات رقمية المنشأ digital born التي لا يوجد نظير مطبوع لها يتطلب معايير خاصة بها لتناسب طبيعتها، مثل محتويات مواقع شبكات التواصل الاجتماعي. وتتكون معايير واصفات البيانات من أربعة مكونات رئيسية، وهي: محتوى البيانات، أشكال الاتصال المعيارية، قوالب البيانات، وتبادل البيانات. وتشير الدراسة إلى أن هذه المعايير مناسبة لفهرسة وأرشفة محتويات الويب بما في ذلك النصوص، الملفات الصوتية، الشروحات، الروابط، التعليقات، ملفات الفيديو، الصور، وغيرها من أشكال المحتويات المتاحة على الويب.

كما قام مصطفى وقوميد (٢٠٢٢) بدراسة دور أرشيف الويب وأهميته كمصدر للبحث العلمي في مجال الإنسانيات الرقمية. واعتمدت الدراسة على المنهج الوصفي القائم على التحليل. وتوصلت الدراسة إلى أن هناك ثلاثة أسباب رئيسية تبرر أرشفة محتويات الويب، وهي الحفاظ على التراث الثقافي الرقمي، تحقيق الاستقرار والحفاظ على مواد الويب ككائنات للبحث، والقدرة على توثيق وتوضيح الادعاءات بناءً على تحليلات لكائنات الويب. وتشير الدراسة إلى أن الويب وبشكل متزايد أصبح مصدرًا مهمًا للدراسة والبحث من قبل الباحثين في مجال العلوم الإنسانية والاجتماعية، وهناك اتجاه لأن تكون مصادر المعلومات رقمية بشكل واسع في المستقبل مع

تضخم في حجم إنتاج مصادر المعلومات. وفي نفس الوقت ونظرًا للطبيعة المتغيرة والهشة للويب، فإن محتوياتها سريعة الزوال والتغيير ولا يمكن ضمان بقائها على المدى البعيد، إلا في حال أرشفتها وحفظها، وتوفير أدوات الوصول لها. ويذكر المؤلفان أن هناك ثلاثة طرق استراتيجية لأرشفة محتويات الويب: الشاملة، الموضوعية، والانتقائية. يقصد بالاستراتيجية الشاملة تلك التي تقوم بحصد النطاقات الوطنية بأكملها أو حتى الويب بأكمله المتاح على النطاق العام. أما الاستراتيجية الموضوعية فهي التي تركز على موضوع معين أو مجال محدد أو حدث معين. أما الاستراتيجية الانتقائية فهي القائمة على أساليب التقييم التقليدية أو الاختيار، بحيث يتم انتقاء مواقع الويب عن قصد لإدراجها في الأرشيف بناءً على قابليتها للتطبيق على مهمة المؤسسة وأهدافها. وتختم الدراسة بأنه ومع انتشار الويب في جميع أنحاء العالم، ستصبح أرشيفات الويب أدوات حاسمة لدعم البحث العلمي في مجال الإنسانيات الرقمية، كما أنها قد تكون من أهم المصادر التي يمكن استخدامها لفهم الظواهر الإنسانية والاجتماعية المختلفة.

٢,٢ الدراسات الأجنبية:

ناقشت دراسة ميراندا (Miranda ٢٠٠٦) بعض التحديات التي تواجه أرشفة محتويات الويب. على سبيل المثال فإن النمو السريع والمستمر والكبير الذي يتصف به الويب يجعل من الصعوبة بمكان تحديد ما يجب جمعه وأرشفته وحفظه. كما أن خصائص الويب الديناميكية المتغيرة تجعل من محتوياتها قابلة للتحديث والتغيير بشكل مستمر، أو أنها قد تختفي دون وجود أي أثر لها، بالإضافة إلى أن عناوين المواقع والنطاقات قد تتغير أو تختفي بعد أن كانت موجودة. ويشير المؤلف إلى أن القضايا القانونية وحقوق التأليف

والملكية والنشر وحماية خصوصيات بعض البيانات والمحتويات هي مشكلة أخرى. وهناك تحدٍ آخر يتمثل في أن مجموعات أرشيف الويب عادة ما تكون كبيرة الحجم، ومتنوعة في أشكالها كالنصوص والصور وعناصر الوسائط المتعددة ومحتويات شبكات التواصل الاجتماعي، وهذا قد يؤثر على ضبطها وإدارتها والتحكم بها. كما أن جمع محتويات الويب وأرشفتها وحفظها وصيانتها تكلف الكثير من المال والجهد والوقت وتحتاج لجهود تنسيقية متكاملة لضمان نجاحها واستمراريتها. وأخيرًا فإن عدم وجود بعض التشريعات والآليات المتعلقة بنظام الإيداع القانوني وذات علاقة بمحتويات الويب هي تحدٍ آخر لا يقل أهمية عن التحديات الأخرى.

واستعرضت دراسة كروك (Crook ٢٠٠٩) في وصف تحليلي أرشفة محتويات الويب في استراليا، أو ما يعرف بأرشيف باندورا PANDORA ، وهو أرشيف ويب يهدف للمحافظة على المحتويات الأسترالية على الإنترنت. تم إنشاء هذا الأرشيف في عام ١٩٩٦م بتعاون العديد من المؤسسات لجمع وأرشفة وحفظ محتويات الويب المنشورة على النطاق الأسترالي. وتضمنت بعض المحتويات مثل المجلات الإلكترونية، منشورات المؤسسات الحكومية، محتويات المواقع البحثية والعلمية، الصور، الفيديو، الملفات الصوتية، بالإضافة إلى المحتويات ذات الأهمية الثقافية. ويبلغ حجم هذا الأرشيف ما يقارب ٥٤ تيرابايت. يتم جمع محتويات الويب من خلال استخدام تقنية أداة الحصد الآلي هيريتريكس Heritrix وأداة الأرشفة Archive-It. ومن أبرز التحديات التي واجهها هذا المشروع تنوع الملفات وأشكالها، بالإضافة إلى محتوى الوسائط المتعددة.

وقام قوميز وآخرون (Gomes et al ٢٠١١) بإجراء دراسة مسحية لمبادرات أرشفة الويب. وتوصلت الدراسة إلى أن المحافظة على التراث الرقمي يعد أمرًا بالغ الأهمية للمجتمعات الحديثة. وأظهرت النتائج وجود ٤٢ مبادرة أرشيفية على شبكة الإنترنت تم إنشاؤها في العديد من الدول حول العالم منذ عام ١٩٩٦، وهذه المبادرات في تزايد مستمر. وبلغ حجم المحتويات ما يقارب ٦,٦ بيتابايت (PB). وأشارت الدراسة إلى أن إتاحة الوصول لبيانات الويب التاريخية وإنشاء آليات بحث موحدة عبر أرشيفات الويب وتطوير تطبيقات جديدة بواسطة جهات خارجية من شأنها أن تسهم في استكشاف إمكانات هذه المصادر القيمة للمعلومات التاريخية. ومن أبرز التحديات التي واجهت هذا النوع من المبادرات وجود بعض العوائق القانونية التي تقيد الوصول إلى محتويات الويب التاريخي. وتوصي الدراسة بأهمية سن قوانين جديدة تتعلق بالحفظ الرقمي وتمديد نطاق ومجال الإيداع القانوني ليشمل وبشكل أوسع محتويات الويب.

وهدفت دراسة نيو (Niu ٢٠١٢) إلى تقديم وصف العمل في أرشفة محتويات الويب والتي تتضمن العديد من العمليات المتتابعة، وتشمل: التقييم والاختيار، الحصول على المحتوى، التنظيم والتخزين، والوصف والبيانات الوصفية. وتركز المرحلة الأولى على تقييم واختيار مصادر الويب المراد أرشفتها وذلك بناءً على العديد من المعايير، ويشمل ذلك إيجاد سياسة للتقييم والاختيار بما في ذلك مجال النطاق، الموضوع، الأحداث، نوع الوسائط وأشكالها. وعملية الاختيار والتقييم قد تكون آلية أو يدوية. والمرحلة الثانية هي الحصول على المحتوى، حيث يتم تغذية أرشيف الويب من مصادر متعددة وباستخدام العديد من الطرق والأدوات مثل لغة اكس ام إل XML ، حصاد

الويب الآلي من خلال زواحف الويب، مثل أداة هيريتريكس Heritrix. وفي بعض الحالات يتضمن الحصول على المحتوى أذن من مالك المحتوى للوصول إليه وجمعه وأرشفته لأسباب عديدة، كما يمكن استخدام نظام الإيداع القانوني كوسيلة للحصول على المحتويات. وتأتي المرحلة الثالثة وهي التنظيم والتخزين، والتي تهدف للمحافظة على المحتوى الفكري لمصادر الويب. ويشير الباحث إلى أن أرشيفات الويب تستخدم ثلاث طرق لتنظيم وتخزين محتويات الويب، وهي: نظم الملفات المحلية، الأرشيفات القائمة على الويب، والأرشيفات غير القائمة على الويب. وتتضمن المرحلة الرابعة الجوانب المتعلقة بالوصف والبيانات الوصفية، حيث يتم تقديم وصف لمحتويات الويب من خلال واصفات البيانات Metadata والتي قد يتم إنشاؤها آليًا أثناء عمليات الجمع والفهرسة الآلية مثل عناوين المحتويات، أحجام الملفات وأنواعها، الروابط، وبعض حقول توصيف البيانات الموجودة في صفحات نصوص الملفات والروابط التشعبية، أو ما يعرف بـ HTML tags. كما يمكن استخدام بعض المعايير الدولية الخاصة بالوصف المعياري مثل دبلن كور، ونظام مارك وغيرها من المعايير في أرشفة محتويات الويب.

واستعرضت دراسة دولي وآخرون (Dooley et al 2017) من خلال مراجعة الأدب المنشور إلى تقديم أبرز واصفات البيانات والأدوات التقنية المناسبة لأرشفة محتويات الويب. وفيما يتعلق بواصفات البيانات، فأظهرت الدراسة أن هناك العديد من واصفات البيانات التي يجب أن تكون متاحة عن مصدر الويب مثل بيانات مالك أو صاحب المحتوى، المؤسسة المستضيفة، الناشر، العنوان، تاريخ النشر، طريقة وصف المصدر، أصل المصدر وارتباطه بالموقع، التقييم، الشكل أو صيغة المحتوى، معرفات الموقع، وهل يتضمن

واصفات بيانات معيارية مثل مارك ٢١. كما قدمت الدراسة أهم الأدوات التقنية التي يمكن استخدامها في عمليات الأرشفة مثل أداة Archive-It ، وأداة HTTrack ، وأداة مستكشف الويب Web Archive Discover ، وأداة مسجل الويب Web Archive Discover.

وقدمت دراسة ميلو وركمباخ (Melo and Rockembach 2021) وصفًا للتجربة البرازيلية في بناء أرشيف لمحتويات الويب الحكومي. وتوصلت نتائج الدراسة إلى أنه وعلى الرغم من وجود اهتمام رسمي وحكومي في بناء أرشيف وطني لمحتويات الويب للمواقع الرسمية الحكومية، إلا أنه وحتى الآن لا توجد سياسة محددة تنظم هذا العمل. كما تم إنشاء البوابة البرازيلية لأرشفة محتويات المواقع الحكومية بما في ذلك الأخبار، الأنظمة والتعليمات، القوانين، التقارير الحكومية، والعديد من المحتويات الأخرى بصيغ وأشكال متنوعة مثل الصور وملفات الفيديو وملفات الوسائط المتعددة.

٢,٣ التجارب والممارسات الدولية:

وفي مجال الممارسة العملية لإنشاء وبناء أراشيف للويب، ونظرًا لأهمية المحتوى الرقمي المنشور على الويب ودوره في حفظ تراث الأمة وثقافتها، وجعل هذا المحتوى محفوظًا ومتاحًا للأجيال الحالية والقادمة، بادرت العديد من الدول مثل الولايات المتحدة الأمريكية وبريطانيا وأستراليا وسنغافورة، ونيوزلندا، وغيرها بالقيام بمشاريع لفهرسة وأرشفة محتويات الويب، والتي يمكن الاستفادة من تجربتها في هذا المجال، وفيما يلي عرض لأبرز تلك التجارب.

٢,٣,١ أرشيف الإنترنت Internet Archive

أرشيف الإنترنت هو مشروع غير ربحي، مقره في مدينة سان فرانسسكو، بولاية كاليفورنيا، بالولايات المتحدة الأمريكية. يعمل هذا

الأرشيف كمكتبة رقمية أمريكية ويهدف لأن يقدم وصولًا شاملًا لجميع المعارف، ويركز على مجموعات المواد الرقمية على الإنترنت، بما في ذلك مواقع الويب، التطبيقات، البرامج والألعاب، والموسيقى والأفلام، مقاطع الفيديو والصور المتحركة، الكتب، الصحف والمجلات، البحوث العلمية، وغير ذلك من المواد المتاحة على الفضاء العام. ويقدم هذا الأرشيف الوصول لهذه المحتويات مجانًا للباحثين والمؤرخين والعلماء والمستخدمين بشكل عام.

بدأ العمل في هذا الأرشيف عام ١٩٩٦م من خلال أرشفة الإنترنت نفسها، حيث كانت هذه الفترة هي فترة نمو الإنترنت، والتي لم يتم التركيز على حفظ محتواها، حيث كان ذلك المحتوى سريع الزوال. وبناء على آخر إحصائية من الأرشيف (يناير ٢٠٢٣م) يوجد ما يقارب ٧٣٥ مليار صفحة مؤرشفة، ٤١ مليون كتاب ونص، ١٤ مليون ملف سمعي، ٨ مليون فيديو، ٤ مليون صورة، ٨٩ ألف برنامج وتطبيق.

ويتم تمويل هذا المشروع من خلال التبرعات والهبات الخيرية بالإضافة إلى المنح. ويشترك في عضوية هذا الأرشيف الكبير مجموعة من المؤسسات والجمعيات مثل جمعية المكتبات الأمريكية، رابطة مجموعات التسجيلات الصوتية، تكتل مكتبات بوسطن، الاتحاد الدولي للحفاظ على الإنترنت، مجلس المكتبات وموارد المعلومات، اتحاد المكتبات الرقمية، المكتبة الرقمية الأمريكية العامة، الاتحاد الدولي لجمعيات ومؤسسات المكتبات والعديد من المؤسسات الأخرى. يتم فهرسة وتكشيف وأرشفة محتوى الإنترنت من خلال استخدام تقنية Wayback Machine وتقنية Archive-It ، وهي من التقنيات الرائدة في أرشفة وجمع وحفظ المحتوى الرقمي على الإنترنت. (Internet Archive 2023).

٢,٣,٢ أرشيف الويب البريطاني The UK Web Archive

قامت المملكة المتحدة ببناء وتطوير أرشيف الويب للمملكة المتحدة (Archive The UK Web)، حيث بدأ العمل على هذا المشروع منذ عام ٢٠٠٥م. يهدف هذا الأرشيف إلى حصر وجمع المعلومات المنشورة على النطاق البريطاني من ملايين المواقع على الإنترنت بما في ذلك الإصدارات القديمة. وانطلق هذا الأرشيف من الفكرة القائمة على أن محتويات الويب تعد شاهدًا مهمًا على العصر، وعادة ما يتم إنشاؤها بسرعة، وتتغير بشكل مستمر، وقد لا توجد في مصادر أخرى، وفي بعض الأحيان تختفي تمامًا دون سابق إنذار. وبالتالي، يحاول هذا الأرشيف تحقيق مبدأ الاستمرارية من خلال جمع هذه المواد وحفظها وإتاحة الوصول إليها للباحثين الحاليين والمستقبليين من جميع الأنواع، والعلماء والمهنيين والمؤرخين والمستفيدين بشكل عام.

واستنادًا لنظام ولوائح الإيداع القانوني لغير المطبوعات، يحق لمكتبات الإيداع القانوني في المملكة المتحدة جمع وحفظ مواقع الويب الموجودة على النطاق الخاص بالدولة. وتشمل التغطية لهذا الأرشيف ملايين من النصوص والصور ومقاطع الفيديو وبعض محتويات مواقع وشبكات التواصل الاجتماعي المتاحة على الفضاء العام، ولا يتضمن المحتوى والمعلومات الخاصة مثل رسائل البريد الإلكتروني أو المحتوى الخاص الذي يتطلب تسجيل الدخول مثل قواعد البيانات الخاصة. ويقدم هذا الأرشيف الوصول لمحتويات متنوعة مصنفة في ما يقارب ١٣٧ مجالًا موضوعيًا. ويبلغ حجمه ما يقارب من ٥٠٠ تيرابايت، بزيادة سنوية في حدود ٧٠ تيرابايت. كما أن هذا الأرشيف متاح بثلاث لغات هي الإنجليزية والويلزية الإسكتلندية. ويمكن الاستفادة من محتويات هذا الأرشيف من خلال التصفح الموضوعي للمحتويات أو استخدام خاصية

البحث في محرك البحث الخاص بالموقع. يتم جمع معظم محتويات الأرشيف من خلال عملية الفهرسة الآلية والتكشيف الآلي بشكل دوري لنطاق إنترنت المملكة المتحدة، حيث يتم أرشفة المحتوى بشكل متكرر استنادًا إلى عوامل محددة مثل معدل التغيير في موقع الويب أو علاقته بمجموعة معينة. على سبيل المثال، يتم جمع العديد من المواقع الإخبارية على أساس يومي بينما تتم زيارة مواقع أخرى بشكل أقل تكرارًا. ويتم جمع وحفظ وتخزين أرشيف الويب في نظام المكتبة الرقمية، وهو مستودع رقمي طورته المكتبة البريطانية وبدعم من مكتبات الإيداع القانوني الأخرى في المملكة المتحدة. ويشترك في هذا المشروع العديد من المكتبات في المملكة المتحدة بقيادة المكتبة الوطنية البريطانية. (The UK Web Archive 2023).

٢,٣,٣ أرشيف الويب الأسترالي Australian Web Archive

يعد أرشيف الويب الأسترالي (Australian Web Archive ٢٠٢٣) قاعدة بيانات متاحة للجمهور على الإنترنت لمواقع الويب الأسترالية المؤرشفة، وتستضيفها المكتبة الوطنية الأسترالية من خلال منصتها الرئيسية (تروف Trove). ويتكون هذا الأرشيف من أرشيف الويب الخاص بالمكتبة الوطنية الأسترالية PANDORA وأرشيف الويب للحكومة الأسترالية ومجموعات نطاقات المكتبة الوطنية الأسترالية.

تم إنشاء هذا الأرشيف عام ٢٠٠٩م، وتقوم فكرة هذا الأرشيف على أن من مهام المكتبة الوطنية جمع التراث الوطني والثقافي للدولة، حيث إن هذا جزء من متطلبات نظام الإيداع القانوني الصادر عام ١٩٦٨م، والذي تم توسيعه في عام ٢٠١٦م ليشمل ما ينشر من محتويات على شبكة الإنترنت. وهذا المشروع الوطني هو عبارة تعاون بين المكتبة الوطنية الأسترالية ومئات من

المنظمات الشريكة في جميع أنحاء أستراليا والتي تساهم بالمجموعات والمحتوى الرقمي.

يتم حصد وجمع المحتوى من الإنترنت بشكل آلي من خلال مجموعة من الروبوتات التي تستخدم عملية الزحف الآلي Auto Crawling وذلك من خلال فهرسة وتكشيف ما ينشر على النطاق العام الأسترالي، ويشمل ذلك أنواع مختلفة من مصادر المعلومات مثل النصوص والصور والمواد السمعية والبصرية وملفات الفيديو، والبحوث والدراسات العلمية، والصحف والمجلات والوثائق الحكومية، والمخطوطات والرسائل العلمية ومحتويات مواقع المنظمات والمؤسسات والمتاحف والأراشيف المحلية بالإضافة إلى مجموعة واسعة من محتويات مواقع وشبكات التواصل الاجتماعي والمواقع الشخصية مثل المدونات المتاحة على الفضاء العام. وبلغت عدد المواد المتاحة في هذا الأرشيف ما يقارب من ٦ مليارات سجل. ويمكن الاستفادة من هذا الأرشيف من خلال التصفح الموضوعي للمحتويات أو استخدام خاصية البحث في محرك البحث الخاص بالموقع.

٢,٣,٤ أرشيف الويب السنغافوري Web Archive Singapore

قامت حكومة سنغافورة ببناء أرشيف الويب السنغافوري في عام ٢٠٠٦م. وهو عبارة عن مجموعة واسعة من مواقع الويب في سنغافورة تتكون من مجموعة من المؤسسات الثقافية والجمعيات والشركات والأفراد والتي توثق الأحداث والأخبار المهمة على نطاق الويب السنغافوري. وتعد هذه المواقع موارد قيمة للباحثين لدراسة مواقع الويب التي لم تعد موجودة أو تلك المواقع التي يطرأ عليها تغييرات معينة بمرور الوقت، حيث يتم التقاط وحفظ هذه المصادر بشكل رقمي. يعكس هذا الأرشيف جوانب مختلفة من الحياة والثقافة

والتراث والتاريخ السنغافوري في القرن الحادي والعشرين، حيث يغطي جوانب مهمة من حياة المجتمع السنغافوري.

يحتوي هذا الأرشيف على أكثر من ٦٠,٠٠٠ موقع ويب مؤرشف فريد وصفحة ويب وأكثر من ١٧٠,٠٠٠ نسخة مؤرشفة (لقطات تم رقمنتها من نفس المواقع بمرور الوقت)، ويتضمن ذلك جميع ما ينشر على مواقع الويب المتاحة للجمهور على النطاق السنغافوري .sg والتي يوجد تفويض لجمعها بموجب قانون مجلس المكتبات الوطنية منذ عام ٢٠١٩م، بالإضافة إلى مجموعة مختارة من مواقع الويب غير المسجلة على النطاق السنغافوري، والتي يملكها السنغافوريون أو المنظمات والمؤسسات السنغافورية من خلال طلب الإذن بفهرستها وأرشفتها وحفظها.

بشكل عام، يتم أرشفة مواقع الويب على الأقل مرة واحدة في السنة، أما مواقع الويب المختارة وذات الأهمية التاريخية أو التراثية فإنه يتم أرشفتها بشكل متكرر. يمكن الاستفادة من هذا الأرشيف من خلال البحث باستخدام الكلمات المفتاحية الرئيسية، أو من خلال إدخال اسم الموقع URL في حال معرفته، أو عن طريق استكشاف مجموعة الأرشيف من خلال التصفح الموضوعي للمحتويات والمصنفة في ١٥ موضوعًا رئيسيًا مثل الاقتصاد، التعليم، الفن، المجتمع والخدمات الاجتماعية، التراث والثقافة ، وغير ذلك من المواضيع، بالإضافة إلى مجموعة واسعة من المجموعات الخاصة Web Archive (Singapore 2023).

٢,٣,٥ أرشيف الويب النيوزيلندي The New Zealand Web Archive

يتضمن أرشيف الويب النيوزيلندي (The New Zealand Web Archive 2023) مجموعة واسعة من مواقع الويب المؤرشفة لنيوزيلندا وتراث

وثقافة المحيط الهادئ، حيث يعد هذا الأرشيف جزءًا من مجموعات مكتبة اليكسندر تيرنبل Alexander Turnbull. ويغطي هذا الأرشيف مواقع الويب منذ عام ١٩٩٩م وتتضمن مواقع الويب التي لم تعد متاحة على الإنترنت وهي مجموعة مختارة من مواقع الويب حول نيوزيلندا والنيوزيلنديين، بالإضافة إلى محتويات مختارة ذات علاقة بتراث وثقافة المحيط الهادئ. ويبلغ محتوى هذا الأرشيف ما يقارب من ٨٥٠٠ عنوان موقع فريد وأكثر من ٣٨٥٠٠ نسخة رقمية. وتتم أرشفة مواقع الويب لأغراض الحفظ والبحث على المدى الطويل. يتم جمع معظم مواقع الويب الموجودة في هذا الأرشيف على فترات منتظمة.

ويتضمن الأرشيف النيوزلندي مجموعة واسعة من المواضيع مثل الثقافة والفنون، الترفيه والرياضة، البيئة، الموسيقى، المجتمع والخدمات الاجتماعية، بالإضافة إلى مجموعة متنوعة من المجموعات الخاصة. ويمكن الاستفادة من هذا الأرشيف من خلال التصفح الموضوعي للمحتويات أو استخدام خاصية البحث في محرك البحث الخاص بالموقع.

يتضح مما سبق أهمية إنشاء أرشيف رقمي لمحتويات الويب، حيث اتضح من خلال عرض الدراسات السابقة بالإضافة إلى استعراض لأبرز الممارسات والتجارب الدولية في هذا المجال اهتمام الدول والمؤسسات بإنشاء أراشيف رقمية لمحتويات الويب بغرض جمعها وحفظها وأرشفتها وجعلها متاحة لجمهور المستفيدين. وتنوعت هذه الأراشيف من حيث المجال والتغطية والنطاق، البعض منها على مستوى الدولة والبعض الآخر في مجال متخصص معين. كما أن ما تم عرضه من مبادرات هو عبارة عن نماذج، حيث إن هناك العديد من المبادرات والمشاريع الكثيرة في هذا الخصوص (Wikipedia 2023).

٣. إجراءات الدراسة

٣,١ منهج الدراسة وجمع البيانات:

من أجل بناء النموذج التصوري وتقديم مقترح لأرشفة المحتوى الرقمي المنشور على النطاق السعودي، استخدم الباحث المنهج التحليلي الوصفي، وقامت الدراسة على منهج تحليل الوثائق، بالإضافة إلى تحليل المحتويات أو ما يعرف بالنتاجات (المكونات) Artifacts. يعرف مايلز وهوبرمان (Miles and Huberman 1994) النموذج التصوري على أنه أي منتج مصور أو مكتوب يشرح في رسم بياني أو في سرد تحليلي العوامل الرئيسية أو المفاهيم أو المتغيرات والعلاقات التي تتكون منها ظاهرة ما. وبهذا فإن النموذج التصوري يساعد على فهم الظاهرة محل البحث بشكل مفصل ودقيق. ويصف زيروينسكي (Czerwinski 2017) النتاجات أو المكونات artifacts بأنها كل ما أنتجه الإنسان لممارسة حياته أو عمله وكجزء من الروتين الحياتي وليس نتيجة لظاهرة طبيعية، ويمكن الإحساس بها والتعامل معها. ومن هذا المفهوم يمكن النظر لشبكة الإنترنت على أنها ظاهرة عالمية ومحلية في نفس الوقت (Robson ٢٠١١) وأصبحت تشكل فضاءً واسعًا يوازي الحياة الحقيقية الواقعية، وأن ما تتضمنه من محتوى ومعلومات يمكن وصفها على أنها نتاجات أو مكونات أوجدها الإنسان كجزء من حياته، بحيث تسجل وتوثق جزءًا كبيرًا من ثقافته وتراثه. وانطلاقًا من هذه المفاهيم، جاءت منهجية هذه الدراسة من خلال مرحلتين رئيسيتين. تضمنت المرحلة الأولى سردًا تحليليًّا واستعراضًا لما تم في مجال أرشفة محتويات الويب الرقمي من خلال مراجعة الأدب المنشور والتعرف على أبرز التجارب والممارسات الدولية وعمل بعض المقارنات المرجعية، وتحليل الوثائق المرتبطة بهذه المشاريع، بالإضافة إلى استعراض لأبرز التقنيات المعلوماتية

التي يمكن توظيفها في هذا المجال. كما هدفت المراجعة السردية التحليلية لتحديد المفاهيم ذات العلاقة بموضوع الدراسة. وجاءت المرحلة الثانية من هذه الدراسة وركزت على جمع وتحليل خصائص وسمات المحتوى الرقمي السعودي المتاح على شبكة الإنترنت، وذلك من خلال تحليل الوثائق المتاحة بالإضافة إلى فحص النتاجات أو المكونات artifacts لمحتويات الويب المتاح على النطاق السعودي وذلك بهدف تحديد خصائصه وملامحه.

٤. نتائج الدراسة

فيما يلي عرض وتحليل لأبرز نتائج الدراسة.

٤,١ الأهمية والأهداف الاستراتيجية:

كما تمت الإشارة إليه سابقًا، فإن الهدف من هذا المقترح والتصور هو إنشاء أرشيف رقمي يهدف لجمع وحصر وأرشفة المحتوى الرقمي السعودي المتاح على شبكة الإنترنت والمحافظة عليه وأرشفته وتوثيق الأحداث الجارية المتصلة بذلك المحتوى والأنشطة والأعمال الفكرية. وتمكن أهمية هذا المشروع من حيث توافقه مع العديد من الرؤى والمبادرات والخطط الوطنية، ومن ذلك على سبيل المثال لا الحصر:

- رؤية المملكة العربية السعودية ٢٠٣٠، والتي نصت في مواضع كثيرة على أهمية التحول الرقمي وحوكمته ضمن إجراءات محددة، بالإضافة إلى توسيع نطاق الخدمات الإلكترونية. وبمواءمة هذا التوجه مع المشروع المقترح، فإن جمع وتوثيق وحفظ المحتوى الرقمي المنشور على الإنترنت الخاص بالمملكة يمكن أن يتوافق مع هذا الهدف، بما يحقق أهداف التحول الرقمي بشكل عام.

- يتوافق هذا المشروع مع توجه وزارة الثقافة في خطتها الاستراتيجية المتضمنة توثيق التراث الثقافي الرقمي وأرشفته في المملكة العربية السعودية (وزارة الثقافة ٢٠٢٢)، حيث إن ما ينشر على الإنترنت يمكن أن يعد جزءًا من المكون الثقافي للمجتمع.
- تعزيز دور المكتبة الوطنية (مكتبة الملك فهد الوطنية) والقائم على جمع وحفظ وصيانة الرصيد المعرفي الوثائقي الوطني، بالإضافة إلى تطبيق نظام الإيداع القانوني، ويشمل ذلك المحتوى المنشور على الصيغ الرقمية والمتاح على نطاق شبكة الإنترنت السعودية.
- التكامل مع الخطط الوطنية للتحول لمجتمع المعلومات والمعرفة.

ويمكن أن يحقق هذا المشروع العديد من الفوائد والتي من أهمها:

- تعزيز الريادة السعودية في مجال تنظيم المحتوى الرقمي على المستوى العربي والإقليمي، والانضمام إلى المبادرات الدولية المتعلقة بأرشفة وحفظ محتويات الويب.
- يساهم المشروع في تعزيز مكانة المملكة دوليًا ضمن مؤشرات الأمم المتحدة الخاص بالحكومة الإلكترونية والتحول الرقمي، وذلك عن طريق توفير منصة رسمية واحدة توفر وتقدم المحتوى والمعلومات الخاصة بما ينشر على الإنترنت السعودية، حيث إن هذا جزء مهم في مجال المشاركة الإلكترونية.
- إتاحة المحتوى السعودي على شبكة الإنترنت لأطول فترة ممكنة وضمان استمراره، والتغلب على مشكلة اختفاء المحتويات والمعلومات على الإنترنت، وإتاحتها للباحثين والمستفيدين والدارسين وصناع

القرار في الوقت المناسب، بالإضافة إلى حفظ هذه المحتويات للأجيال القادمة في المستقبل.

- بناء قاعدة معلومات للمحتوى الرقمي السعودي على شبكة الإنترنت لتكون مخزونًا ومرجعًا معرفيًا وتاريخيًا يستفاد منه في العديد من الجوانب الاجتماعية والاقتصادية والثقافية ودعم خطط التنمية المستدامة.
- رصد نمو وتطور المواقع الإلكترونية والمحتوى الرقمي المنشور على الإنترنت.
- تعزيز جهود الضبط الببليوجرافي الوطني، من خلال حصر وجمع محتويات الويب السعودية المنشورة على شبكة الإنترنت، حيث لا يوجد حتى الآن أداة متكاملة تقوم بهذه المهمة.

٤,٢ التكامل بين المؤسسات والشراكة التعاونية:

عند مراجعة نتائج تحليل الدراسات السابقة والتجارب الدولية في هذا المجال، يتضح أن هذا النوع من المشاريع النوعية يتطلب العمل الكبير والدؤوب بين العديد من الجهات الفاعلة والتي يمكن أن تشارك وتساهم في العديد من الجوانب الفنية والتقنية والإجرائية واللوجستية وغيرها التي تضمن بناء أرشيف رقمي مستدام. ومن هذا المنطلق يجب أن يكون هناك جهة محددة تقوم بقيادة هذا المشروع، على أن يكون ضمن اهتمامها وأولوياتها، وتعمل على تنسيق العمل المشترك بين الجهات ذات العلاقة. ونظرًا لأهمية هذا المشروع الاستراتيجية وضخامته من حيث الحجم، واشتراك العديد من الجهات من حيث المهام والمسؤوليات لتوفير الدعم اللازم، فيمكن تنسيق العمل بين هذه الجهات، ومنها على سبيل المثال لا الحصر:

- مكتبة الملك فهد الوطنية (قيادة المشروع).

- دارة الملك عبدالعزيز.
- مركز الملك عبد العزيز الثقافي العالمي (إثراء).
- هيئة الحكومة الرقمية.
- المركز الوطني للوثائق والمحفوظات.
- مكتبة الملك عبدالعزيز العامة.
- الفهرس العربي الموحد.
- وزارة الاتصالات وتقنية والمعلومات.
- هيئة البيانات والذكاء الاصطناعي.
- وزارة الثقافة.
- المكتبة الرقمية السعودية.
- المكتبات الجامعية.
- وزارة الإعلام.

بالإضافة إلى أي مؤسسات حكومية أخرى وشبة حكومية أو مؤسسات القطاع الخاص والتي لها علاقة مباشرة أو غير مباشرة بهذا المشروع.

٤,٣ محتويات الويب المتاحة على النطاق السعودي:

أظهرت نتائج الدراسة من خلال المسح العام لطبيعة ونوع المحتوى الرقمي أنه يتضمن كمًّا هائلًا وأنواعًا مختلفة من المحتويات والمعلومات والمواقع المتاحة والمنشورة على نطاق الإنترنت السعودي، بالإضافة إلى ما كتب ونشر عن السعودية خارجيًا. وفيما يلي عرض لأبرز الخصائص والسمات.

٤,٣,١ نمو الإنترنت ومحتوياتها في السعودية:

تبلغ نسبة انتشار الإنترنت في المملكة ما يقارب ٩٨% من إجمالي السكان ولكافة الأعمار، حيث تصل شبكة الإنترنت لمعظم كافة مناطق

المملكة بنسبة ما يقارب ٩٧% (هيئة الاتصالات والفضاء والتقنية ٢٠٢١). ويتوازى هذا الانتشار مع الاستخدام والاعتماد المتزايد على شبكة الإنترنت، حيث أصبحت جزءًا مهمًا من حياة المواطنين اليومية، وسهلت على الجميع القيام بالمهام والأعمال الشخصية والرسمية. يوضح الجدول رقم (١) إجمالي عدد النطاقات الرسمية المسجلة في السعودية حسب آخر ثلاث سنوات.

جدول ١: إجمالي عدد النطاقات في السعودية لآخر ثلاث سنوات

السنة	٢٠٢٠	٢٠٢١	٢٠٢٢
عدد النطاقات	٧٢٦٥٢	٨٣١٠٩	٩٢٤٥٠

ويلاحظ من الجدول رقم (١) أن عدد النطاقات السعودية المسجلة وصل بنهاية عام ٢٠٢٢م ما يقارب من ٩٢٤٥٠ نطاقًا. كما يوضح الشكل رقم (١) النمو الزمني لعدد النطاقات في السعودية من عام ٢٠١٠م إلى عام ٢٠٢٢م، حيث يلاحظ أن نمو النطاقات ومحتوياتها بكافة أشكالها وأنواعها تنمو وتزيد بشكل مستمر، حيث سجلت نمو ما يقارب من ١٠ % سنويًا خلال الخمس سنوات الماضية (هيئة الاتصالات والفضاء والتقنية ٢٠٢٢).

شكل رقم ١. معدل النمو السنوي للنطاقات في السعودية

وتتنوع نطاقات الإنترنت من حيث تكوينها والمحتويات التي تقدمها، حيث توجد النطاقات الخاصة بمواقع الشركات والعلامات التجارية المسجلة،

المؤسسات التعليمية والحكومية والصحية، مواقع خدمات الإنترنت، والنطاقات العامة. الجدول ٢ يعرض توزيع النطاقات حسب التفريعات.

جدول ٢: توزيع النطاقات بحسب التفريعات

النطاق	النسبة	الوصف
.com.sa	٤٩,٩%	مواقع الشركات والعلامات التجارية المسجلة.
.sa	٣٨,٥%	مواقع تحمل نطاق السعودية.
.org.sa	٣,٥%	مواقع المنظمات الغير ربحية في السعودية.
السعودية.	٣,٣%	مواقع مسجلة على النطاق السعودي باللغة العربية.
.edu.sa	٢,٣%	ويشمل هذا التقسيم مواقع المؤسسات الأكاديمية والبحثية مثل الجامعات والكليات ومؤسسات التعليم العالي.
.gov.sa	٢,٣%	المواقع الحكومية.
.net.sa	٢,٣%	مواقع خدمات الإنترنت.
.sch.sa	٠,٨%	مواقع المدارس الابتدائية والمتوسطة والثانوية.
.med.sa	٠,٦%	ويشتمل على مجال مواقع القطاع الطبي والصحي.
.pub.sa	٠,٢%	ويتضمن المواقع التي لا تقع ضمن التصنيفات السابقة أعلاه. (مثل مواقع الأسماء الشخصية، القبائل،... الخ).

ويستنتج من الجدول رقم ١ ورقم ٢ تزايد نمو محتويات الويب المنشورة على النطاق السعودي، بالإضافة إلى تنوعها من حيث مجال النطاقات المسجلة. كما أظهرت نتائج المسح والمراجعة لمحتويات الويب أنها تشتمل على أنواع مختلفة، ومن ذلك على سبيل المثال:

- مواقع الويب بشكل عام والتي تدخل ضمن اهتمام الأرشيف.
- مواقع المؤسسات والمنظمات الحكومية وشبه الحكومية والخاصة.
- البرامج والألعاب.
- الملفات السمعية والمرئية والصور (الوسائط المتعددة).
- الكتب والصحف والمجلات.

- البحوث والدراسات العلمية والتقارير.
- الوثائق الرسمية الحكومية وغير الحكومية.
- المخطوطات والوثائق.
- المواقع الشخصية مثل المدونات.
- بعض محتويات مواقع وشبكات التواصل الاجتماعي (الحكومية والخاصة).
- المنتديات وساحات النقاش.
- البيانات المفتوحة.

وغير ذلك من المواد المتاحة على الفضاء العام وتدخل ضمن نطاق الأرشيف.

ويجب أن تخضع هذه المحتويات والمصادر لعملية التقييم والاختيار ضمن سياسة محددة بناء على العديد من المعايير، ويشمل ذلك معايير تقييم واختيار مجال النطاق، الموضوع، الأحداث، نوع الوسائط وأشكالها. كما أن عملية الاختيار والتقييم قد تكون آلية أو يدوية، وذلك بناء على طبيعة ونوعية المحتويات المستهدفة.

٤,٣,٢ أنواع الملفات وامتدادها Files type and Extension

أظهرت نتائج المسح لبعض محتويات الويب المتاحة على النطاق السعودي تنوع الملفات وأشكالها، وهذا يعتمد على نوعية المحتويات مثل النصوص والصور والمواد السمعية وملفات الفيديو وغيرها. يوضح الجدول رقم (٣) أنواع الملفات وامتدادها.

جدول ٣: أنواع الملفات حسب طبيعة محتويات الويب

الملفات النصية	
HTML	Hypertext Markup Language

ODF	Open Document Format
PDF	Portable Document Format
RTF	Rich Text Format
TXT	Plain Text
XML	Extensible Markup Language
الملفات الصوتية	
AIFF	Audio Interchange File Format
BWF	Broadcast Wave Format
FLAC	Free Format Lossless Audio Codec
MIDI	Musical Instrument Digital Interface
MP3	MPEG-1 Audio Layer 3
WAV	Wave File
ملفات الفيديو	
AVI	Audio Video Interleaved
MPEG2	Moving Picture Expert Group
Motion JPEG2000	
Falsh	
MXF	Material Exchange Format
MOV, WMV	
ملفات الصور	
BMP	Bitmap File
GIF	Graphic Interchange Format
JPEG, JPG	Joint Photographic Experts Group Format
PNG	Portable Network Graphics
SVG	Scalable Vector Graphics
TIFF, TIF	Tagged Image Format

ونظرًا لتنوع أشكال الملفات وامتدادها، فإنها تعد من أبرز التحديات التي يمكن التعامل معها. على سبيل المثال فإن ملفات الوسائط المتعددة تتطلب مساحة كبيرة من حيث التخزين، وإمكانيات عالية من حيث الإتاحة والاستخدام. كما أن هذا التنوع يتطلب استخدام بعض المعايير الخاصة بعمليات الوصف الببليوجرافي، والتي تضمن وجود واصفات معرفة مسبقًا تناسب طبيعة كل محتوى وشكل، مثل معيار دبلن كور، ونظام مارك، معيار وصف وإتاحة المصادر RDA ، ومعيار VRA، بالإضافة إلى بعض المعايير الأخرى التي تناسب المحتويات بأشكال وأنواع مختلفة.

٤,٣,٣ أنشطة استخدام الإنترنت في السعودية:

ويرتبط بأنواع المحتويات المذكورة أعلاه نمط استخدام المستفيدين وسلوكهم في الاستفادة من الإنترنت ومحتوياتها، حيث إن الأنشطة التي يقوم بها مستخدمو الإنترنت متنوعة، وذلك كما هو موضح في الجدول رقم (٤).

جدول ٤: أنشطة استخدام الإنترنت في السعودية

النشاط	النسبة
شراء السلع والخدمات	70.1%
تحميل البرامج والتطبيقات	62.9%
تحميل الألعاب والأفلام والصور	55.5%
متابعة الأخبار والمدونات	55.1%
البحث عن المعلومات حول السلع والخدمات	52.7%
استخدام البريد الإلكتروني	35.9%
البحث عن المعلومات الصحية	34.5%
الحصول على معلومات من المؤسسات الحكومية	33.9%
الخدمات المتعلقة بالسفر	24.3%
الدراسة عبر الإنترنت	23.7%

المصدر: هيئة الاتصالات والفضاء والتقنية ٢٠٢١.

يتضح من الجدول رقم (٤) تنوع سلوكيات المستفيدين في استخدام الإنترنت والاستفادة من محتويات الويب المتاحة، ومن هذه محتويات قد تكون مفيدة للباحثين على المدى الطويل، ومنها ما يوثق أحداثًا معينة، والبعض الآخر يتضمن البحث عن معلومات من مواقع المؤسسات الحكومية، الحصول على المعلومات الصحية، ومتابعة الأخبار والأحداث الجارية. وبالتالي فإن جمع وأرشفة هذه المحتويات سيساهم في حفظها وإتاحتها للأجيال الحالية والمستقبلية وسد الفجوة الرقمية التي قد تحدث مستقبلًا في حال وضعها في أرشيف متاح للعامة.

٤,٣,٤ التحديات التي تواجه المشروع وأسلوب مواجهتها:

أشارت العديد من الدراسات السابقة إلى وجود العديد من التحديات التي تواجه هذا النوع من المشاريع، على سبيل المثال (٢٠٠٦ Miranda) (٢٠٠٩ Crook) (٢٠١١ Gomes et al) (قوميد ومصطفى ٢٠٢١). ونظرًا لأنه يتوقع أن ينتج عن هذا المشروع عمل وطني كبير، فإنه يتوقع في نفس الوقت وجود العديد من التحديات التي يمكن أن تظهر أثناء مراحل التخطيط والتنفيذ والتشغيل والصيانة، ويمكن حصر تلك التحديات وسبل مواجهتها في النقاط التالية، كما هو موضح في الجدول رقم (٥).

جدول ٥: أبرز التحديات المتعلقة بأرشفة محتويات الويب وسبل مواجهتها

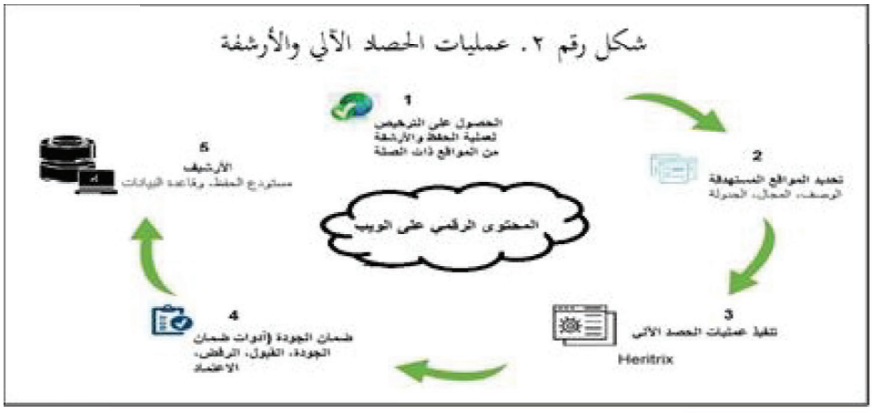
التحدي	سبل التعامل
العدد الكبير من المواقع والصفحات الإلكترونية والمنشورات الرقمية المتاحة على الإنترنت وبلغات متعددة، وصيغ مختلفة مما يجعل من الصعوبة بمكان حصرها بشكل دقيق.	تحديد النطاق الذي سيتضمنه المشروع، وذلك بتنسيق الجهود مع الجهات ذات العلاقة لحصر وتحديد هذه المواقع ومحتوياتها وصفحاتها الفرعية بجميع اللغات المتوفرة.
أنواع مختلفة لصيغ الملفات وأشكالها	الاستعانة ببعض الممارسات والتوجيهات الارشادية المتعلقة بالتعامل مع أنواع مختلفة لصيغ الملفات، على سبيل المثال مصفوفة المسح الضوئي للتعامل مع المحتوى الإلكتروني وتنسيق الملفات المقترحة من إدارة السجلات والمحفوظات الوطنية الأمريكية (٢٠٢١).
آلية الأرشفة الرقمية والحصاد الآلي ومناسبتها للمحتوى العربي	يوجد العديد من التقنيات التي يمكن استخدامها مثل أداة وي باك WayBack ، أداة اركيف ات Archive-It ، أداة هيريتريكس Heritrix، وأداة Wget.

التحدي	سبل التعامل
الدعم المالي	تخصيص ميزانية للمشروع من خلال الدعم الحكومي، بالإضافة إلى دعم قائم على التبرعات والهبات والمنح لضمان استدامة المشروع.
المحتوى المغلق	طبيعة بعض المحتويات تكون مناسبة للإتاحة في الفضاء العام، إلا أنها قد لا تكون متاحة لعدة أسباب، ويمكن التفاوض مع أصحابها لإتاحتها والسماح بالوصول والدخول إلى هذه الصفحات وأرشفتها وحفظها.
حقوق الملكية الفكرية	يتم التعامل معها حسب الأنظمة والتعليمات والقوانين الرسمية والصادرة من الجهات ذات العلاقة.
تحديات أخرى قد تظهر أثناء مراحل التخطيط والتنفيذ والتشغيل والصيانة	يتم التعامل معها في حينه وحسب ما يقتضيه كل تحدي.

ولنجاح هذا المشروع، يجب العمل على ترتيب وتنسيق الأمور ذات العلاقة بالجوانب الفنية والتقنية والإدارية والمالية واللوجستية والقانونية المتعلقة بعمليات الحفظ الرقمي للمحتوى المطلوب، بالإضافة إلى التعامل مع التحديات الآنفة الذكر من خلال تنسيق الجهود والتكامل بين الجهات ذات العلاقة، وذلك لضمان أن يكون ذلك المحتوى محفوظًا ومتاحًا بشكل مستمر وقابلًا للاستخدام على المدى الطويل.

٤,٣,٥ عمليات الحصاد الآلي والأرشفة الرقمية:

تعتمد عملية الحصاد الآلي والأرشفة الرقمية لمحتويات الويب لغرض جمعها وحفظها وأرشفتها على العديد من المراحل، كما هو موضح بالشكل رقم (٢).



وفيما يلي شرح لهذه الخطوات:

١. الحصول على التراخيص لعملية الحفظ والأرشفة من المواقع ذات الصلة:

وهذه هي الخطوة الأولى، وفيها يتم عمل كافة الإجراءات اللازمة للحصول على أذن وترخيص لعمليات جمع وحفظ وأرشفة المحتويات المستهدفة. ويقصد بذلك الحصول على الأذونات اللازمة من أصحاب ومالكي المحتوى في حال تطلب ذلك (المحتويات المغلقة)، وبالتالي يتم إتاحة تلك المحتويات للقيام بعمليات الجمع والفهرسة الآلية. وبشكل عام فإن المحتويات المتاحة على النطاق العام لا تتطلب موافقة من أصحابها لأنها تدخل ضمن المحتويات المنشورة على النطاق السعودي. ويمكن أن يعزز نظام الإيداع القانوني الصادر عن مكتبة الملك فهد الوطنية ليشمل محتويات الويب.

٢. تحديد المواقع والمحتويات المستهدفة:

نظرًا للحجم الكبير للمحتوى المتاح على الإنترنت، فإنه من الصعوبة بمكان جمع وأرشفة كل ما ينشر. ومن هذا المنطلق، يجب أن يكون هناك سياسة للاختيار والتقييم تتضمن تحديد المواقع والمحتويات المستهدفة والتي ستدخل ضمن عمليات الجمع والفهرسة والحفظ. وعادة ما يكون هناك العديد من العناصر التي بموجبها يتم تحديد المواقع والمحتويات مثل النطاق أو

المجال (مثل .gov أو .edu)، والموضوع أو الحدث، ونوع المحتوى، وأشكال مصادر المعلومات. وأظهر المسح الذي تم إجراؤه كجزء من هذه الدراسة أن المحتويات ومواقع الويب المتاحة على نطاق الويب السعودي متنوعة، مثل مواقع المؤسسات والمنظمات الحكومية وشبه الحكومية والخاصة، البرامج والألعاب، الملفات السمعية والمرئية والصور، الكتب والصحف والمجلات، البحوث والدراسات العلمية والتقارير، الوثائق الرسمية الحكومية وغير الحكومية، المخطوطات والوثائق، المواقع الشخصية مثل المدونات، بعض محتويات مواقع وشبكات التواصل الاجتماعي (الحكومية والخاصة)، المنتديات وساحات النقاش، ومواقع المؤسسات بكافة أنواعها عامة أو خاصة. وبشكل عام، فإن جميع ما ينطبق عليه نظام الإيداع الصادر بالمرسوم الملكي الكريم رقم (م/٢٦) في ١٤١٢/٩/٧هـ، والذي تم تطبيقه عام ١٤١٤هـ وينص على تسجيل كل ما ينشر داخل المملكة من أوعية المعلومات (مكتبة الملك فهد الوطنية ٢٠٢٣)، يمكن أن يتضمن محتويات الويب، ولهذا يجب إعادة النظر في نظام الإيداع وتوسيعه ليشمل محتويات الويب المنشورة على النطاق السعودي.

٣. تنفيذ عمليات الحصد الآلي للويب Web Harvesting

وفي هذه المرحلة يتم استخراج المحتوى وجمعه وأرشفته باستخدام العديد من التقنيات والبرامج والتي تستهدف المحتويات وبياناتها من مصادر الويب المتنوعة، وذلك من خلال فهرسة وتكشيف صفحات النصوص التشعبية HTML ومحتوياتها الكاملة (Ferrara et al ٢٠١٤). وهناك العديد من التقنيات التي يمكن استخدامها في هذا المجال مثل أداة وي باك WayBack ، أداة اركيف ات Archive-It ، أداة هيريتريكس Heritrix، وأداة Wget. يتم

برمجة هذه الأدوات والتقنيات على معايير معينة لجمع وأرشفة المحتوى المستهدف، وعادة ما تتم بشكل آلي دون أي تدخل بشري.

٤. ضمان الجودة

عادة ما يتم وضع السجلات الموجودة في أرشيف الويب من خلال عملية ضمان الجودة، والتي تقيس جودة مصدر المعلومات من خلال مقارنته بالمعايير التي وضعت ابتداءً ويجب الالتزام بها. ويعد ضمان الجودة عاملًا مهمًا في هذا الخصوص. وتأتي عملية ضمان الجودة عادة بعد الحصول على مصادر المعلومات من الويب، ويجب أن تتحقق من مجموعة من العناصر التي تضمن جودة المصدر المختار من خلال العديد من الإجراءات مثل مراقبة الجودة والتحليل والمراجعة والتقييم. معظم عمليات ضمان الجودة تتم بشكل يدوي وبالتالي فهي تتطلب التزامًا كبيرًا من حيث الوقت والجهد من قبل مسؤولي الأرشيف بالإضافة إلى ضرورة التدريب المتخصص، والحاجة إلى العديد من المهارات والمعارف في عملية ضمان الجودة (Ayala et al ٢٠١٤).

٥. الأرشيف: مستودع الحفظ وقاعدة البيانات

ينتج عن عملية جمع وأرشفة وحفظ محتويات الويب وجود مستودع رقمي وقاعدة معلومات تضمن عملية تخزين وإتاحة موارد الويب والمحافظة عليها كمعلومات تاريخية، وسجلات تثبت أحداث معينة. يقدم الأرشيف خدمة الوصول للمحتويات التي تم جمعها وأرشفتها وحفظها والتعريف بها من خلال واصفات البيانات مثل عنوان المصدر، الرابط، المؤلف (مالك أو صاحب المحتوى) نوع وشكل المحتوى، بالإضافة إلى بعض البيانات الوصفية الأخرى. وعادة ما يكون هذا الأرشيف متاحًا للمستخدمين للاستفادة منه على المدى الطويل، ويمكن الوصول لمحتوياته من خلال موقع أو بوابة رقمية تفاعلية تقدم

للمستخدمين العديد من الخدمات مثل البحث في الأرشيف، استعراض وتصفح المحتويات حسب الموضوعات، الإحصاءات، وخدمات أخرى تسهل من عملية البحث والاستفادة من محتويات الأرشيف.

٥. النتائج والتوصيات:

بناء على ما تم مناقشته في الأقسام السابقة، توصلت الدراسة إلى العديد من النتائج والتوصيات التي يمكن إجمالها فيما يلي:

٥,١ النتائج:

- يعد الويب وما ينشر عليه من محتويات من أهم مصادر المعلومات في الوقت الحاضر، ولا تقل أهميته عن المصادر المطبوعة، وقد لا يوجد له نظير مطبوع يعادل أهميته كمصدر للمعلومات.
- لا يمكن ضمان إتاحة محتويات الويب إلى الأبد، فهي سريعة الزوال والتغير والتحديث، وبالتالي فإن عدم جمعها وحفظها وأرشفتها قد يعني فقدانها، وقد لا تكون متوفرة ومتاحة في المستقبل.
- نظرًا لأهمية المحتوى الرقمي المنشور على الويب ودوره في حفظ تراث الأمة وثقافتها، وجعل هذا المحتوى محفوظًا ومتاحًا للأجيال الحالية والقادمة، بادرت العديد من الدول مثل الولايات المتحدة الأمريكية وبريطانيا وأستراليا وسنغافورة ونيوزلندا وغيرها من الدول بإنشاء أرشيفات رقمية لحفظ هذه المحتويات.
- يتم فهرسة وتكشيف وأرشفة محتويات الويب من خلال استخدام العديد من التقنيات مثل تقنية Wayback Machine وتقنية Archive-It .
- من أبرز التحديات التي تواجه أرشفه محتويات الويب الحجم الكبير للمحتويات، تنوع صيغ الملفات، آلية الأرشفة الرقمية، الدعم المالي،

المحتويات المغلقة، وحقوق الملكية الفكرية، بالإضافة إلى بعض التحديات الأخرى.

- لا يوجد حتى الوقت الحاضر مشروع وطني يهدف لجمع وحفظ وأرشفة محتويات الويب المتاحة على نطاق الويب السعودي.
- يبلغ عدد نطاقات الإنترنت في السعودية السعودي ٩٢٤٥٠ نطاقًا، وبنسبة نمو سنوية ما يقارب ١٠%.
- تتنوع محتويات الويب المتاحة على النطاق السعودي من حيث الأشكال والأنواع، مثل محتويات مواقع المؤسسات والمنظمات الحكومية وشبه الحكومية والخاصة، البرامج والألعاب، ملفات الوسائط المتعددة (صور وفيديو وسمعية)، الكتب والصحف والمجلات، البحوث والدراسات العلمية والتقارير، المواقع الشخصية مثل المدونات، بعض محتويات مواقع وشبكات التواصل الاجتماعي (الحكومية والخاصة).

٥,٢ التوصيات:

- ضرورة تبني مشروع على نطاق وطني يهدف لجمع وحفظ وأرشفة محتويات الويب المتاحة على نطاق الويب السعودي.
- سن قوانين جديدة تتعلق بالحفظ الرقمي وتمديد نطاق ومجال نظام الإيداع القانوني ليشمل وبشكل أوسع محتويات الويب المتنوعة والتي يمكن أن ينطبق عليها النظام.
- إيجاد سياسة شاملة إرشادية لاختيار وتقييم محتويات الويب التي يمكن أن تدخل ضمن الأرشيف الرقمي، بما يضمن جمع وحفظ وأرشفة تلك المحتويات، ويكون ذلك بناء على المعايير المعتمدة وأفضل الممارسات والمقارنات المرجعية.

- العمل على تنسيق الأمور ذات العلاقة بالجوانب الفنية والتقنية والإدارية والمالية واللوجستية والقانونية المتعلقة بعمليات الحفظ الرقمي للمحتويات المستهدفة، وتنسيق الجهود والتكامل بين الجهات ذات العلاقة، وذلك لضمان أن تكون تلك المحتويات محفوظة ومتاحة بشكل مستمر وقابلة للاستخدام على المدى الطويل.
- ضرورة إجراء العديد من الدراسات مثل دراسة جدوى للمشروع وتقييم آثاره المتوقعة والعوائد الإيجابية التي سينتج عنها المشروع.

٥,٣ الخاتمة

هدفت هذه الدراسة إلى تقديم تصور مفاهيمي مقترح لبناء أرشيف رقمي لمحتويات مصادر الويب المتاحة على النطاق السعودي. وتؤكد الدراسة أنه لا يوجد حتى الآن أرشيف رقمي يجمع ويحفظ ويتيح محتويات الويب المنشورة على النطاق السعودي. واستعرضت الدراسة بعض التجارب والممارسات الدولية في هذا المجال، والتي أظهرت مبادرة العديد من الدول والحكومات والمؤسسات على تهيئة السبل اللازمة لهذا النوع من المشاريع الرقمية والتي تضمن جمع وتوثيق وحفظ وأرشفة ما ينشر على الويب وإتاحته للأجيال الحالية والمستقبلية. كما قدمت الدراسة عرضًا لأبرز خصائص المحتويات المتاحة على نطاق الويب السعودي، والخيارات المتاحة لأرشفة هذه المحتويات، وتقديم خطة عمل مقترحة لإنشاء هذا الأرشيف. كما ناقشت الدراسة العديد من التحديات التي تواجه هذا المشروع وسبل التعامل معها. وانتهت الدراسة إلى تقديم العديد من التوصيات والمقترحات، والتي من أهمها أن هذا المشروع المقترح يمكن أن يساهم في تعزيز الحضور الرقمي للمحتوى

السعودي على الإنترنت في المؤشرات الرقمية الدولية، بالإضافة إلى تحقيق بعض من مستهدفات رؤية السعودية ٢٠٣٠ فيما يتعلق بالتحول الرقمي.

قائمة المراجع:

أولًا: المراجع العربية

- إحصاء الإنترنت العالمي (٢٠٢٣). حجم محتويات الويب والإنترنت. http://www.worldwidewebsize.com.
- إدارة السجلات والمحفوظات الوطنية الأمريكية (٢٠٢١). توجيهات لرقمنة المواد الأرشيفية للوصول الالكتروني. https://www.archives.gov/preservation/technical/guidelines.html
- الحباطي، محمد خميس السيد محمد. (٢٠٢١). التراث الثقافي غير المادي بالوادي الجديد: دراسة تحليلية لحفظه مع تصميم موقع ويب لإتاحته. المجلة الدولية لعلوم المكتبات والمعلومات، مج ٨، ع ٤، ص ٤٤٩ – ٤٩٣ . مسترجع من: http://search.mandumah.com/Record/1224922
- خليفة، محمود عبد الستار (٢٠٠٦). أرشيف المواقع المصرية على الإنترنت: دراسة ميدانية تحليلية / إشراف محمد فتحي عبد الهادي .– القاهرة. (أطروحة ماجستير، جامعة القاهرة. كلية الآداب. قسم المكتبات والوثائق والمعلومات). http://www.aplis.cybrarians.info/index.php/abstracts/372-ia
- رؤية السعودية ٢٠٣٠. (٢٠٢٣). https://www.vision2030.gov.sa/ar
- سعيدي، سليمة، بلال حجاز، وإبتسام السعيدي. (٢٠٢١). مشروع الأرشفة الإلكترونية للوثائق: نموذج مقترح. المجلة العربية للمعلوماتية وأمن المعلومات، ع ٤ ، ص ١٠٩-١٢٩

- عبدالراضي، أحمد إبراهيم. (٢٠٢٢). معايير ميتاداتا الوصف الأرشيفي في بيئة الويب: مدخل نظري. مجلة بحوث في علم المكتبات والمعلومات، ع ٢٨ ، ص ٥٤٣ – ٥٧٤ . مسترجع من:
http://search.mandumah.com/Record/1274976
- عبود، رامي (٢٠١٣). نحو استراتيجية عربية لصناعة المحتوى الرقمي. مركز دراسات الوحدة العربية. بيروت.
- عكاشة، منال جابر محمد. (٢٠١٥). أنظمة إدارة المحتوى الرقمي على الشبكة المعلوماتية العنكبوتية (الويب): دراسة تقييمية للخروج بمواصفات معيارية مناسبة لتطبيقات مرافق المعلومات العربية. المجلة الدولية لعلوم المكتبات والمعلومات، مج. ٢، ع ٢ ، ص ٣٧٥-٣٨٠.
- مصطفى، عمرون، قوميد، فتيحة. (٢٠٢٢). أرشيف الويب مصدر للبحث العلمي في الإنسانيات الرقمية. مجلة العلوم الإنسانية والاجتماعية، مج ٨، ع ٢، ص ٤٥١ – ٤٦٩ . مسترجع من:
http://search.mandumah.com/Record/1286735
- مكتبة الملك فهد الوطنية (٢٠٢٣). الإيداع القانوني.
https://kfnl.gov.sa/Ar/About/Pages/Standards.aspx
- وزارة الثقافة، (٢٠٢٢). دليل توثيق التراث الثقافي وأرشفته الرقمية في المملكة العربية السعودية.
https://www.moc.gov.sa/ar/HeritageDocumentation
- قوميد، فتيحة، عمرون، مصطفى. (٢٠٢١). ممارسات حفظ الثقافة على الويب. المجلة الدولية أبحاث في العلوم التربوية والإنسانية والآداب واللغات، مج ٢، ع ٢، ١١٢ – ١٢٧ . مسترجع من:
http://search.mandumah.com/Record/1131929
- هيئة الاتصالات والفضاء والتقنية. (٢٠٢١). تقرير إنترنت السعودية لعام ٢٠٢١م.

https://www.cst.gov.sa/ar/indicators/Pages/saudi_internet.aspx

- هيئة الاتصالات والفضاء والتقنية. (٢٠٢٢). تقرير أسماء النطاقات السعودية لعام ٢٠٢٢م.

https://www.cst.gov.sa/ar/indicators/Pages/saudi_domain.aspx

ثانيًا: المراجع الأجنبية:

- Altayar, M. (2016). Investigating the use of web 2.0 technologies and their presence in Saudi government agencies' websites. International Journal of Technology Diffusion, 7(2), 63–80.
- Australian Web Archive. (2023). https://trove.nla.gov.au
- Ayala, R., Phillips, B., E, M., Lauren, K. (2014). Current quality assurance practices in web archiving. https://digital.library.unt.edu/ark:/67531/metadc333026/
- Crook, E. (2009). Web archiving in a Web 2.0 world. The Electronic Library, Vol. 27 No. 5, pp. 831-836. https://doi.org/10.1108/02640470910998542
- Czerwinski, A. (2017). Artifact selection. In: The SAGE Encyclopedia of Communication Research Methods. SAGE Publications, Inc, https://dx.doi.org/10.4135/9781483381411
- Dooley, J., Farrell, K., Tammi, K., Venlet, J. (2017). Developing web Archiving metadata best practices to meet user needs. Journal of Western Archives: Vol. 8, (2), DOI: https://digitalcommons.usu.edu/westernarchives/vol8/iss2/5
- Ferrara, E., De Meo, P., Fiumara, G., Baumgartner, R., (2014). Web data extraction, applications and techniques: A survey. Knowledge-Based Systems، 70 pp. 301-323. https://arxiv.org/pdf/1207.0246.pdf
- Gomes, D., Miranda, J., Costa, M. (2011). A survey on web archiving initiatives. International Conference on Theory and Practice of Digital Libraries, Berlin, Germany, pp. 408-420.
- Internet Archive. (2023). https://archive.org
- Khan, G. F. (2015). The Government 2.0 utilization model and implementation scenarios. Information Development, 31(2), 135 149. https://doi.org/10.1177/0266666913502061

- Melo, J., Rockembach, M. (2021). International initiatives and advances in Brazil for government web archiving. In: Bisset Álvarez, E. (eds) Data and Information in Online Environments. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 378. Springer, Cham. https://doi.org/10.1007/978-3-030-77417-2_6
- Miles, M., Huberman, A. (1994). Qualitative Data Analysis: An Expanded Sourcebook; Sage Publications: Thousand Oaks, CA, USA, 1994; ISBN 0803955405.
- Miranda, J. (2006). Web harvesting and archiving. http://web.ist.utl.pt/joaocarvalhomiranda/docs/other/web_harvesting_and_archiving.pdf
- The New Zealand Web Archive. (2023). https://natlib.govt.nz/collections/a-z/new-zealand-web-archive
- Niu, J. (2012). An Overview of Web Archiving. School of Information, Faculty Publications. 308. http://scholarcommons.usf.edu/si_facpub/308
- Robson, C. (2011). Real world research: a resource for users of social research methods in applied settings. 3rd ed. Chichester: John Wiley.
- Techopedia, (2017). Web Crawlers. https://www.techopedia.com/definition/10008/web-crawler
- Techopedia, 2020. The Internet. https://www.techopedia.com/definition/2419/internet
- The UK Web Archive. (2023) https://www.webarchive.org.uk/en/ukwa
- Web Archive Singapore. (2023). https://eresources.nlb.gov.sg/webarchives
- Wikipedia (2023). List of Web archiving initiatives. https://en.wikipedia.org/wiki/List_of_Web_archiving_initiatives