

نظم استرجاع المعلومات الكيانية الشمولية :

دراسة تطبيقية للمعالجة المعلوماتية الجامعة والاسترجاع الموحد عبر الوسائط

د. سيد ربيع سيد إبراهيم

أستاذ المكتبات والمعلومات المساعد

قسم علوم المعلومات-كلية الآداب-جامعة بني سويف (مصر)

قسم المعلومات ومصادر التعلم-كلية الآداب والعلوم الإنسانية-جامعة طيبة (السعودية)

sayedrs1m@hotmail.com

تاريخ القبول: 5 أكتوبر 2024

تاريخ الاستلام: 22 سبتمبر 2024

المستخلص:

اعتمدت نظم استرجاع المعلومات على الوسيط النصي في معالجة وبحث المحتوى المعلوماتي الرقمي بوسائطه المتعددة، ذلك لما لخصائص النص من توافق وطبيعة مستخدمي النظم في التعامل بالنص عنه بالوسائط الأخرى، لكن نسبة التحقق في استرجاع الوسائط النوعية باستخدام النص تكون بدرجة منخفضة لا تحقق الكفاءة المطلوبة، وقد عملت محركات بحث الويب على اعتماد أسلوب المعالجة والاسترجاع المبني على محتوى مادة المعلومات، باستخدام خصائص وسائط الصوت والصورة في الكشف والاسترجاع، وقد أتى ذلك بشماره في نتائج الاسترجاع للوسائط المتعددة، وقد تطورت قدرات نظم استرجاع المعلومات باعتماد تقنيات التحليل الدلالي وشبكات الكيانات المعرفية وقدرات معالجة اللغة الطبيعية، بما زاد من كفاءة النظم في تطوير نتائجها إلى النتائج السياقية الدلالية، وهو ما تمثل في الاسترجاع عبر مولدات المحادثات Chat-GPT.

وتستهدف هذه الدراسة التوجه نحو جيل أكثر تقدما من نظم استرجاع المعلومات يتبنى المنهج الشمولي الجامع لكل مواد المعلومات بوسائطها النوعية، ويتم داخله معالجة وتنظيم واسترجاع وسائط النص والصوت والصورة بعملية معالجة موحدة تشق خصائص كل وسيط على حدة، ثم تجمعها في قاعدة معرفية يمكنها تحقيق الربط الدلالي والمفاهيمي بين محتوى مادة الصوت والصورة والنص، بما يجعل النظام قادرا على تلقي استفسار البحث بوسيط واحد، واسترجاع النتائج بكل الوسائط. وعليه، تدوب الخصائص الشكلية الحاكمة لوسيط المحتوى ليتبقى المحتوى المعرفي الممثل في قاعدة المعرفة، ويمكن لنظام الاسترجاع استدعاء النتائج بوسائط متعددة متصلة بذات الموضوع لعملية البحث الواحدة، كأن يكون البحث بالصوت والنتائج صوتية ومصورة ونصية.

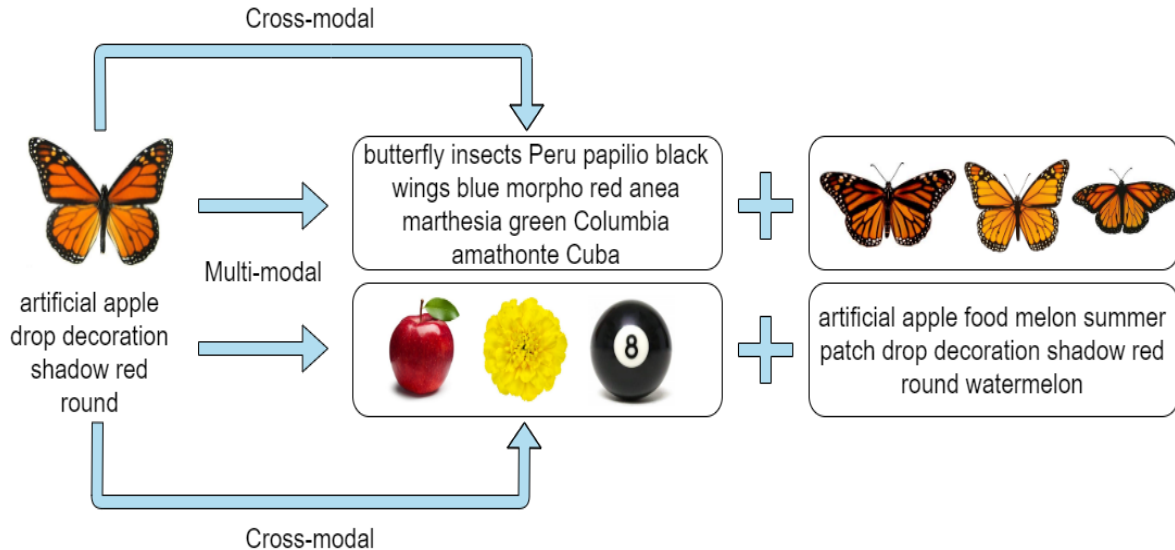
وتستخدم الدراسة منهج البحث الوصفي والتجريبي باعتماد عينة من نظم استرجاع المعلومات على الويب، مستخدمة أدوات جمع البيانات الوثائقية لجمع أدبيات الموضوع، والملاحظة للدراسة التطبيقية، وتمثل أهم نتائج الدراسة في: إن نظم الاسترجاع الكيانية عبر الوسائط تعد أول مستويات نظم إنتاج المعرفة الشمولية التفاعلية في البيئة الرقمية، وهو أمر ليس بعيدا عن الوقت الحالي، وسيدعم تطور تقنيات الذكاء الاصطناعي وتطبيقاتها في مجال معالجة المعلومات، وإدراك المحتوى الرقمي ظهور وتطور نظم معرفية قادرة على تجزئة وبناء الكيانات الموضوعية والدلالية لإتاحتها بما يتوافق والاحتياجات المعرفية لمستخدمي النظم، وقد قدمت الدراسة تصورا لبناء وهيكل نظم الاسترجاع الشمولية متعددة الوسائط.

الكلمات المفتاحية: نظم استرجاع المعلومات؛ إنتاج وتوليد المعرفة؛ قواعد المعرفة؛ الوسائط المتعددة؛ محركات البحث؛ الويب؛ معالجة اللغة الطبيعية؛ الكشف الدلالي؛ الويب الدلالي؛ شبكات الربط المفاهيمي.

أولاً: مقدمة الدراسة:

1 / 1 التمهيد:

اكتسب استرجاع الوسائط المتعددة (CMR) Cross Modal Retrieval الكثير من الاهتمام مؤخرًا بسبب الزيادة السريعة في المحتوى الرقمي متعدد الوسائط مثل: الصور، والنص، والفيديو، والصوت، ويمثل مصطلح الوسيط شكلاً محدداً لحمل المحتوى المعلوماتي، ويرتبط أيضًا بالإدراك الحسي مثل: الرؤية والسمع اللذين يعدان مصدرين رئيسيين للاتصال والاستجابة لدى البشر، وتساعد الأشكال المختلفة من الوسائط المتعددة الأشخاص على فهم الأشياء بشكل أفضل، فغالبًا ما يرغب مستخدمو المعلومات في الحصول على نتائج دقيقة بأشكال مختلفة، مما يخلق حاجة إلى منصة فعالة لاسترجاع المعلومات المتعددة الوسائط، ويعتبر استرجاع المعلومات أحادية الوسائط النهج الكلاسيكي لنظم البحث؛ واستدعاء المعلومات المشتقة من قناة واحدة فقط، من الصور فقط أو من النص فقط (وليس كليهما)، فيتم استخدام استدعاء النص فقط للبحث واسترجاع المعلومات النصية بشكلٍ كفاء، بينما يكون أقل كفاءةً عند التعامل مع المحتوى متعدد الوسائط، ومن ثم فإن نظم الاسترجاع متعددة الوسائط القادرة على ربط أكثر من وسيلة مثل: الصورة والنص والصوت والفيديو في الوسائط المتعددة، يكون وضع الاستعلام النتائج مختلفين بتحقيق الكفاءة الأفضل في الاسترجاع، ومثالاً لذلك، فإن كلاً من النص والصور يمكن استرجاعهما باستخدام استدعاء مصور فقط أو استدعاء نصي فقط، وهو ما يعكس معالجة النظام الموحدة الشاملة لكل الوسائط المتعددة في مجموعاته. ويوضح شكل رقم (1) إمكانية استخدام صورة الفراشة كاستدعاء مصور، فيتم استدعاء النصوص والصور ذات الصلة بموضوع الصورة، أيضاً يمكن استخدام نص واصف لتفاحة كاستدعاء نصي فيتم استدعاء النتائج المتصلة بالموضوع في شكل صور ونصوص (Kaur, P., 2021)



شكل رقم (1) الاستعلامات والنتائج متعددة الوسائط (Tian, J., et al, 2022, July)

وتكمن الفكرة الأساسية للبحث عبر الوسائط المتعددة في دمج العديد من طرائق البحث للحصول على نتائج أفضل من استخدام قناة واحدة فقط، وهو ما يمثل تحدي عملي، لأن التباين الجوهرى بين الوسائط المختلفة يجعل من الصعب قياس أوجه التشابه بين الوسائط المختلفة بشكل مباشر، وتعمل نظم الاسترجاع على تضمين الملامح المشتركة والارتباطات المتقاطعة الجوهرية للوسائط المختلفة بتمثيلات رقمية متسقة، وتوصلت العديد من الدراسات الحديثة إلى اعتماد تقنيات الذكاء الاصطناعي مثل: الشبكات العصبية لتضمين العلاقات والارتباطات بين الوسائط المتعددة (مثل: دلالات الكلمات أو العلاقات الهرمية)، وتستخدم هذه النظم النمذجة الرقمية لملامح المحتوى لكل وسيط، وعليه، فإن نظم استرجاع الوسائط المتعددة

ستعتمد على بنية معالجة من ثلاث وحدات رئيسية؛ أولاً: مستخلصات ومشفرات الملامح والخصائص لكل وسيط يعمل على تحويل تلك الخصائص لتمثيلات رقمية موحدة، وثانيها: وحدة الاندماج والتبادل التي يمكنها فصل بيانات الإدخال إلى ملامح متشابهة ومختلفة تجمع بين محتوى الوسائط ذات الدلالات الواحدة وتفضلها عن الوسائط الأخرى ذات المحتوى المغاير، ثم ثالثها: وحدة اقتراح مخططات الربط والتشابك الدلالي والربط المفاهيمي الموسع بين مختلف التمثيلات الرقمية على مستوى قاعدة بيانات النظام، وذلك لتشكيل هياكل معرفية مسبقة يمكن تبادل الاستدعاء وإعادة تركيبها وفقاً لاحتياجات مستخدم النظام (Hu, P., et al, 2023)

لقد وفر ظهور تقنيات معالجة اللغة الطبيعية والتعلم الآلي تقنيات قابلة للتطبيق لاسترجاع المعلومات متعدد الوسائط؛ ويمكن استخدام تقنيات الشبكات العصبية العميقة لنمذجة البيانات من مصادر متعددة أو وسائل أخرى والتعلم منها، ثم التمكن من فهم ومعالجة الوسائط بمفردها ومجموعة وتوفير إجراءات استرجاعها بآليات موحدة، وتتمثل أهم الإيجابيات المترتبة على هذه التقنيات أنها: تسمح للنظام بدمج المزيد من المحتوى الرقمي متعدد الوسائط وتحقيق نتائج استرجاع أكثر دقة وملاءمة، ويمكن لنظام الاسترجاع بتقنية التعلم العميق متعدد الوسائط استخدام كل من النص والصور المرتبطة به، لفهم أفضل لمحتوى مادة المعلومات، وإجراء تنبؤات أكثر دقة حول ملاءمته لاستعلام بحث معين، أيضاً فإن تقنيات التعلم العميق متعددة الوسائط تمكن نظام الاسترجاع من التعامل بشكل أفضل مع الوسائط غير المتجانسة التي تحمل دلالات متباينة (Hu, P., et al, 2021) ولا تزال نظم استرجاع الوسائط المتعددة تواجه العديد من التحديات غير البسيطة؛ تتمثل في أولاً: تتطلب نظم استرجاع الوسائط المتعددة معالجة المدخلات في وسائل مختلفة متعدد الوسائط أيضاً إنشاء مخرجات وبيانات نتائج في وسائل متعددة، على عكس نظم الاسترجاع أحادية الوسيط التي تركز فقط على معالجة المدخلات وتوفير أسلوب محدد لاسترجاعها، ثانياً: من الأهمية بمكان بالنسبة لنظم الاسترجاع متعددة الوسائط تبني قدرات المعالجة والتنظيم على مستوى الدلالات بين الوسائط المختلفة، مع دمج هذه الدلالات المتباينة والتكاملية بفعالية مع تصفية الدلالات غير المترابطة لمحتوى المعلومات، ثالثاً: تتلقى نظم الاسترجاع متعددة الوسائط أشكالاً متعددة من المدخلات من بيانات المحتوى الرقمي الواسعة، ومن ثم فإن كيفية تطوير نظام استرجاع معلومات متعدد الوسائط لا تزال مواجهة صعوبة تحتاج لمزيد من التمكين، رابعاً: يجب مراعاة بعض الزوايا الأخرى لاسترجاع المعلومات متعدد الوسائط مثل: تمكين أنظمة الاسترجاع من دعم أشكال مختلفة من طريقة التفاعل، والمحادثة التبادلية بين النظام ومستخدميه، أيضاً، استخراج البيانات الوصفية متعددة الوسائط المنظمة من المدخلات ثم تطبيقها على عملية الاسترجاع اللاحقة لتحقيق نتائج أفضل (Fei, H., et al, 2021)

وانطلاقاً من هذا التمهيد، ستعمل الدراسة في إطارها النظري على استقراء قدرات العمل وآلياته في نظم استرجاع الوسائط المتعددة، بما في ذلك معالجة الوسائط الأولية، واستخراج الميزات، وتقنيات التحليل والنمذجة، ومنهجيات التقييم، ومقومات العمل في نظام استرجاع الوسائط المتعددة، ثم آليات وقدرات البحث والاستدعاء متعدد الوسائط، ثم تتناول الدراسة في إطارها التطبيقي عيناً من نظم استرجاع الوسائط المتعددة، ثم مناقشة الرؤى المستقاة من الاستقراء والملاحظات، وهو ما يمكن أن يقدم صورة أفضل لنظم استرجاع المعلومات متعدد الوسائط، وتعزيز فرص التفعيل والبناء الأمثل خاصة في البيئة العربية الإقليمية، (Lerner, P., et al, 2024)

1 / 2 مشكلة الدراسة:

شهدت مختلف أنواع بيانات الوسائط - مثل: الصوتيات والنصوص والصور ومقاطع الفيديو - نمواً هائلاً على الإنترنت، وعادة ما تستخدم أنواع مختلفة من البيانات لوصف الحدث نفسه أو الموضوع، فقد يتكون منشور على إحدى المنصات الرقمية يتكون من محتوى نصي ومصور ومسموع مقترن، وعليه، فإن استرجاع الوسائط المتعددة لمادة المعلومات

يمثل طريقة فعالة للبحث عن نتائج ذات صلة موضوعية ودلالية للأشكال والوسائط المتعددة التي تعبر عن هذا الحدث أو الموضوع، ويتيح ذلك للمستخدمين أنفسهم الحصول على جملة المعلومات النوعية حول الحدث/الموضوع المعني، وأمام هذه الإيجابية والكفاءة لنظام استرجاع الوسائط في تحقيق فائدة نوعية مستحدثة لمستخدميه، إلا أن أحد التحديات الرئيسية لتحقيقها تتمثل في تباين وتشتت دلالات وتمثيلات وتوزيعات الوسائط المتعددة، فضلاً عن تنوع خصائصها الرقمية المادية، مما يجعل التوفيق والاتساق الدلالي بين تلك الوسائط أمراً معقداً، وقد سعت دراسات نظم استرجاع المعلومات إلى محاولات مبكرة لاستكشاف الارتباطات وشبكات العلاقات بين الوسائط غير المتجانسة في المحتوى، وهو ما يعني أن نظم استرجاع الوسائط المتعددة قد انتقلت من مستوى الاسترجاع أحادي الوسائط إلى مستوى الاسترجاع متعدد الوسائط، وقد وجه الباحثون والمتخصصون انتباههم نحو تطوير أنظمة استرجاع المعلومات متعددة الوسائط بتقديم المفهوم الأساسي لهذه النظم، ثم تحديد مفاهيم الاستعلام والهدف والمعالجة والأداء معايير التقييم، وآليات البحث والاسترجاع متعدد الوسائط.

يكتسب استرجاع المعلومات متعدد الوسائط أهمية متزايدة في المشهد الرقمي الحالي، بما له من مردود نفعي في استرجاع نوعي جديد، وما يواجهه من تحديات كبيرة مرتبطة باسترجاع معلومات الوسائط المتنوعة ودمجها بفعالية، وتتمثل مشكلة الدراسة هنا في الحاجة إلى تصور لآليات عمل محسنة يمكنها دمج تقنيات استخراج البيانات والتعلم العميق بسلاسة لتحسين دقة وأهمية وفعالية أنظمة استرجاع المعلومات متعددة الوسائط، وهذه المشكلة تمثل أمراً حيوياً لمعالجة هذه التحديات وتطوير مجال استرجاع المعلومات متعدد الوسائط، وتزويد المستخدمين بإمكانية الوصول الأكثر دقة وأهمية وفعالية إلى كنز المحتوى الرقمي متعدد الوسائط المتاح في البيئة الرقمية، وتقوم مشكلة الدراسة على فرضية (إن نظم استرجاع الوسائط المتعددة الحالية تعمل بآليات الاسترجاع أحادي الوسائط، بينما يحتاج الاسترجاع متعدد الوسائط إلى بناء آليات عمل نوعية تلائم البحث عبر الوسائط).

وقد تمت صياغة هذه المشكلة البحثية في مجموعة من التساؤلات هي:

1. ما التصور المقترح لبناء نظم استرجاع المعلومات عبر الوسائط؟
2. كيف يمكن الاستفادة من تقنيات التعلم الآلي ومعالجة اللغة العربية في معالجة الوسائط المتعددة في قواعد المعرفة؟
3. ما آليات المعالجة والتنظيم المتوافقة والوسائط الرقمية المتجانسة وغير المتجانسة دلاليا وموضوعيا؟
4. ما آليات الاسترجاع والبحث التي تمكن من توحيد إجراءات استدعاء الصورة والصوت والنص مجتمعين في نتائج واحدة؟

1 / 3 أهمية الدراسة: تستمد الدراسة أهميتها من الجوانب التالية:

1. يحتاج المحتوى الرقمي متعدد الوسائط إلى توفير قدرات وآليات البحث الموحد، بما يمكن معه استدعاء نتائج نوعية تختلف عن تلك الآنية في نظم استرجاع المعلومات أحادية البحث، لأن البيئة الرقمية قد وحدت المحتوى الرقمي بوسائطه المختلفة للغة قابلة للإدراك بواسطة الآلة؛ بل والتعامل الدلالي والمفاهيمي لهذا المحتوى متعدد الوسائط.
2. إن نظم استرجاع المعلومات تعايش مجموعة مستحدثة من تقنيات الذكاء الاصطناعي جعل الأجيال السابقة من نظم استرجاع، تبدو وكأنها غير مكافئة لمتطلبات واحتياجات المستفيدين في استدعاء المحتوى الرقمي النوعي متعدد العلاقات متعدد الوسائط.
3. إن تقنيات الذكاء الاصطناعي كتعلم الآلة وقواعد المعرفة والتحليل السياقي للمحادثات الاسترجاعية ومعالجة اللغة الطبيعية وشبكات الربط المفاهيمي، قد أوجد قدرات إضافية يمكن لنظم استرجاع المعلومات استثمارها في إنتاج جيل جديد من نظم إنتاج وتوليد المعرفة متعددة الوسائط.

1 / 4 أهداف الدراسة: تتمثل أهم أهداف الدراسة في:

1. وضع تصور مقترح لما يمكن أن يكون عليه نظام استرجاع المعلومات عبر الوسائط، سواء في تدفق الإجراءات أو آليات معالجة وبحث الوسائط المتعددة.
2. استكشاف آليات عمل نوعية تستوعب تقنيات الذكاء الاصطناعي مثل: استخراج البيانات والتعلم الآلي لتحسين أداء وقابلية استخدام أنظمة استرجاع المعلومات عبر الوسائط.
3. استخلاص معايير الأداء والتقييم لعمل نظم استرجاع المعلومات عبر الوسائط اعتمادا على استقراء الإنتاج الفكري للمجال، وما تخرج به الدراسة التطبيقية الحالية.
4. تحديد جوانب التحديات التي تواجه نظم استرجاع المعلومات عبر الوسائط واقتراح ما يعالجها من قدرات وآليات عمل.

1 / 5 حدود الدراسة:

- الحدود الموضوعية:** تركز الدراسة على مجال نظم استرجاع المعلومات المتعددة الوسائط، لاستخلاص آليات المعالجة الموحدة والاسترجاع عبر الوسائط.
- الحدود الجغرافية:** تهتم الدراسة بنظم استرجاع المعلومات متعددة الوسائط في البيئة الرقمية.
- الحدود النوعية:** تركز الدراسة على نظم استرجاع معلومات الوسائط (النص والصوت والصورة والفيديو).

1 / 6 منهجية الدراسة:

اعتمدت الدراسة على استخدام المنهج الوصفي التحليلي بأسلوب العينة القصدية من نظم استرجاع الوسائط المتعددة؛ وتركز الدراسة في جانبها التطبيقي على استخدام المنهج التجريبي لدراسة واختبار واقع عمل نظم استرجاع الوسائط المتعددة وآليات المعالجة والبحث المتبعة بها، وحتى يتم استيفاء الإطار النظري للدراسة تم استخدام أداة البحث الوثائقي في جمع الإنتاج الفكري المرتبط بظاهرة البحث، واستخدام أداة الإبحار والمعايشة مع الويب إلى جانب استخدام أداة الملاحظة في الجانب التطبيقي للبحث، وحتى تتسم الدراسة التطبيقية بالمنهجية، فقد استخدمت معايير عمل وإدارة نظم استرجاع الوسائط المتعددة لقياس مدى تطبيقها في عينة الدراسة، مضافا إليها مجموعة المعايير المعنية بتطبيق تقنيات الذكاء الاصطناعي المقترح إدخالها لبناء نظم استرجاع المعلومات عبر الوسائط.

1 / 7 مصطلحات الدراسة: تتضمن الدراسة مجموعة من المصطلحات أهمها:

- نظم الاسترجاع الكُلِّيَّانِيَّة الشمولية: جيل من نظم استرجاع المعلومات يمكنه معالجة كل الوسائط المتعددة بشكلٍ جامع شامل موحد، ثم اعتماد أساليب بحث واسترجاع موحدة تلائم البحث في مختلف الوسائط المتعددة واسترجاعها بشكل شامل.
- الاسترجاع عبر الوسائط: آلية البحث والاسترجاع التي تتطوي على قدرة استرجاع مختلف أشكال الوسائط المتعددة بالبحث بأحدها فقط.
- قواعد المعرفة: قاعدة النظام التي تتكون من مجموعات المواد متعددة الوسائط والمفاهيم والمصطلحات التي تستخدم لمعالجة وتنظيم المعلومات.
- معالجة اللغة الطبيعية: تقنية إدارة المصطلحات والكلمات دلاليا والإدراك الآلي لمفاهيم ومعاني تلك المصطلحات والعلاقات بينها.

- شبكات المفاهيم: العلاقات والروابط بين المفاهيم والمواد المعلوماتية التي تتضمنها قاعدة بيانات النظام.
- الكيانات المعرفية: مجموعات الدلالات والتمثيلات المعرفية التي تكون قاعدة المعرفة داخل نظام استرجاع الوسائط.

1 / 8 الدراسات السابقة:

تم البحث في قواعد البيانات العالمية مثل: springer, science direct, IEEE, Dissertation proquest وغيرها، حتى يمكن البحث عن الدراسات السابقة التي عنيت بدراسة ظاهرة الاسترجاع عبر الوسائط المتعددة، وقد نتج عن هذا البحث العديد من الدراسات التي تناولت موضوع نظم استرجاع الوسائط المتعددة والاسترجاع عبر الوسائط، أما الإنتاج الفكري العربي فقد خلا من الدراسات التي تتناول ظاهرة الدراسة، وقد كانت أهم الدراسات السابقة ما يلي:

الدراسة الأولى:

Li, L., Hu, C., Zhang, H., & Maradapu Vera Venkata sai, A. (2023, December). Cross-modal Image-Recipe Retrieval via Multimodal Fusion. In Proceedings of the 5th ACM International Conference on Multimedia in Asia (pp. 1-7).

اهتمت هذه الدراسة بتقديم إطار عمل استرجاع الاندماج متعدد الوسائط لمعالجة وسيط الصورة بشكل مخصص، يستخدم الإطار المقترح تقنيات التعلم العميق لمعالجة بيانات الوصف والصورة بشكل فعال، ثم دمج مجموعة من الواصفات الدلالية عبر وسيط النص والصورة، لتحقيق قدرات استرجاع الصورة بالنص والنص بالصورة، وتختلف الدراسة الحالية عن هذه الدراسة السابقة في التركيز على مختلف الوسائط المتعددة بشكل شمولي، مما يعني التركيز على قدرات معالجة مواد الصوت والفيديو واشتقاق خصائصها الفنية ودلالاتها على المحتوى، وقدرات استرجاعها بشكل شامل من خلال أساليب استرجاع محددة.

الدراسة الثانية:

Beltrán, L. V. B., Caicedo, J. C., Journet, N., Coustaty, M., Lecellier, F., & Doucet, A. (2021). Deep multimodal learning for cross-modal retrieval: One model for all tasks. Pattern Recognition Letters, 146, 38-45.

ركزت هذه الدراسة على مواجهة تحديات استرجاع عبر الوسائط المتعددة وإمكانية قبول الصور أو النصوص كاستعلامات، من أجل استرجاع البيانات ذات الصلة من مجموعة مواد متعددة الوسائط. ولتحقيق هذا الهدف، اعتمدت الدراسة على تقنيات التعلم الآلي متعدد الوسائط لدمج التمثيلات البصرية والنصية لبحث المعلومات، وسمي هذا النموذج: "التمثيلات متعددة الوسائط العميقة (DME)"; ويمكن للنظام تصنيف الموضوعات / المفاهيم الدلالية التي تم تحديدها مسبقاً في مجموعة مواد المعلومات ذات الاهتمام، مما يؤدي إلى تحسين الأداء بشكل كبير في الاسترجاع، وبخاصة في النتائج العليا للقائمة المصنفة، والتي تعد الأهم بالنسبة للمستخدمين، وتهتم الدراسة الحالية عن هذه الدراسة السابقة بتقنيات أخرى داعمة لنظم استرجاع الوسائط المتعددة كاستخدام الشبكات المفاهيمية والربط الدلالي ومعالجة اللغة الطبيعية وقواعد المعرفة، مما يرسم صورة شاملة متكاملة لنظام استرجاع يحقق قدرات البحث عبر الوسائط بشكل مفهوم للآلة وشامل لمستخدمي النظام.

الدراسة الثالثة:

Fei, H., Yu, T., & Li, P. (2021, June). Cross-lingual cross-modal pretraining for multimodal retrieval. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (pp. 3644-3650).

ركزت هذه الدراسة على تقنيات نماذج الرؤية البصرية واللغة المدربة مسبقاً لتضمينها في قدرات نظام الاسترجاع عبر الوسائط باللغة الإنجليزية، وقد اقترحت هذه الدراسة نهجاً جديداً لتعلم التمثيلات عبر اللغات وعبر الوسائط لمطابقة

الصور وتعليقاتها ذات الصلة بلغات متعددة، وذلك من خلال توفير مجموعات بيانات التعليق التوضيحي للصور المسمى مسبقاً للتدريب، بحيث لا تكون النصوص بالضرورة باللغة الإنجليزية، لذا فقد درست أهداف التدريب المسبق عبر اللغات وأهداف التدريب المسبق عبر الوسائط بسلاسة في إطار عمل موحد لتعلم الصورة والنص في فضاء تضمين مشترك من بيانات التعليق التوضيحي للصورة الإنجليزية المتاحة، والجمهور الأحادي اللغوي والمتوازي، وتختلف الدراسة الحالية عن هذه الدراسة السابقة في محاولة تقديم نموذج شامل يراعي جوانب معالجة اللغة الطبيعية بشكل عام وخصائص معالجة الوسائط المتعددة من الناحية المعلوماتية لإمكانية دمجها في نظام واحد يغلب عليه العمل بالذكاء الاصطناعي.

ثانياً: الإطار النظري:

تسعى الدراسة من خلال الإطار النظري لاستقراء جوانب وخصائص الاسترجاع عبر الوسائط، بالنظر لما يمكن أن تكون عليه طبيعة ونوعية المعالجة الفنية لمجموعات نوعية من الوسائط متباينة المحتوى والبنية الرقمية، كما يتم تناول معمارية وأداء نظم الاسترجاع عبر الوسائط، وآليات البحث بها، وذلك سيكون بهدف استقراء جوانب المعالجة ومعايير العمل المطلوبة في هذا الجيل من نظم استرجاع المعلومات.

المبحث الأول: المعالجة الفنية الرقمية الكليانية عبر الوسائط:

يتناول المبحث الأول المعالجة الفنية الرقمية لوسائط المحتوى الرقمي بأشكاله المختلفة من النصوص والمواد المصورة والصوتية، وذلك لما تمثله عملية المعالجة من تهيئة الوسائط المتعددة وجعلها قابلة للتنظيم والاسترجاع داخل النظام، وتتطلب معالجة الوسائط المختلفة قدرات معلوماتية وبرمجية لتحليل خصائص عرض المحتوى والتعبير عنه في كل وسيط على حدة، ثم تحويله إلى تمثيلات دلالية تعبر عن المحتوى والوسيط؛ بل وربطهما مع باقي الوسائط والمحتويات الأخرى، وسيكون تناول المعالجة والتنظيم للوسائط المتعددة بالتعرف على آلية وقدرة إدراك النظام للوسيط ومحتواه، ثم آليات بناء التمثيلات الدلالية المعبرة عن المحتوى الموضوعي والشكلي للوسيط، ومقومات عملية المعالجة الناجعة القادرة على تحقيق ذلك.

2 / 1 / 1 الإدراك الرقمي الشامل للوسائط المتعددة:

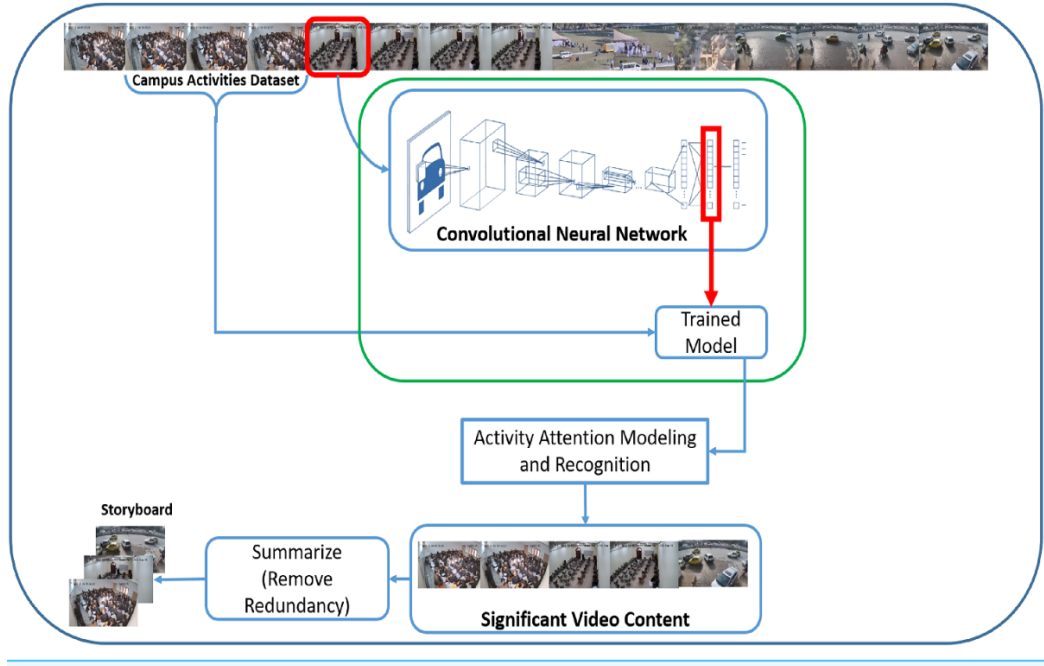
إن نظم استرجاع المعلومات عبر الوسائط لا يمكن لها التعامل مع المحتوى المعرفي المعلوماتي الكامن في الوسيط دون إدراك ملامح هذا الوسيط الموضوعية والشكلية، وقد وفرت تقنيات الإدراك الرقمي قدرات تعامل النظام مع مختلف الوسائط المصورة والصوتية، وتنظيمها وإخضاعها للوصف الفني المبني على المحتوى CBMM والمبني على النص TBMM، ولتوفير قدرات إدراك شمولية وكليانية للوسائط المتعددة، وسيتم دراسة طبيعة الإدراك الرقمي لكل وسيط معلومات على حدة، حتى إذا ما كان النظام قادراً على أدائها، توفرت به صفة الإدراك الشمولي لكل وسائط حمل المعلومات.

2 / 1 / 1 الإدراك الرقمي للوسيط المصور المتحرك:

إدراك الصور المتحركة الرقمية (الفيديو):

تعتمد الصور المتحركة على العناصر البصرية للتعبير عن المعنى، لكنها تنفرد بإضافة عنصر الحركة والتسلسل الزمني، وهذا التميز يجعلها تحمل قدرًا أكبر من المعلومات؛ ويمكن تشبيه هذه العملية بتحليل قصة يتم التركيز على الجمل الرئيسية لفهم المعنى العام للقصة، وتعتمد تقنيات الإدراك للصور المتحركة على تحليل الملامح البصرية الأساسية لكل لقطة فردية تشمل الأشكال والألوان والبنية ثم ربطها معًا لتكوين فهم شامل للمشهد، ومن خلال مقارنة هذه الملامح بين اللقطات المختلفة، يمكن لنظم الإدراك أن تبني تمثيلاً داخلياً للمحتوى المرئي الذي يمكن استخدامه للعديد من التطبيقات، مثل: البحث

في قواعد البيانات المرئية أو توليد ملخصات للمقاطع المرئية، ويوضح شكل (2) البنية العامة والإجراءات التي تمر بها عملية إدراك الفيديو، (Patel, B., & Singh, 2023)



شكل (2) بنية نظام استرجاع الفيديو والصور المتحركة المبني على الإدراك.

وتفرض طبيعة التخصص الموضوعي لنظام إدراك الصور المتحركة آليات ومعاملات تحليل الصور المكونة لملفات الفيديو، وينعكس ذلك إجمالاً على تقنيات وبرمجيات الإدراك المستخدمة لاقتباس واشتقاق الملامح الشكلية أو العناصر الرسومية المكونة لصور الفيديو، ولتوضح ذلك فإن نظم لتحليل ملفات الفيديو العامة تحتاج إلى اعتماد برمجيات وآليات تحليل لون الجلد وتحديد الوجوه face detection، وتحتوي هذه الملفات أكثر ما تحوي صوراً لأشخاص ومشاهد تمثيلية لأحداث معينة (Wasim, M., Ahmed, 2022). أيضاً فإن نظاماً لإدراك ملفات الفيديو الوثائقية يحتاج بطبيعته إلى تقنيات لتحليل خصائص الأشكال والرسومات shapes، كما يحتاج نظام آخر إلى تقنيات تحليل بنية الصورة وخصائص اللون color detection الموزعة على مشاهد الفيديو، ومن ثم تستخدم هذه التقنيات جميعاً بشكل مستقل أو جماعي لإدراك وتحليل محتوى ملفات الصور المتحركة الرقمية على الويب، هذا فضلاً عن استخدام آليات قياس درجات الكثافة للملامح الشكلية المكونة لمشاهد وأطر الفيديو المراد إدراكه (Bommisetty, R.M., 2022)

وتمر عملية إدراك الصور المتحركة بمرحلتين رئيسيتين تتضمن كل منهما إجراءات محددة تتمثل في التالي: (Phan, T., Phan, 2022)

المرحلة الأولى: استخلاص الأطر المفتاحية:

1. تحليل ملف الفيديو: يتم في هذه المرحلة فصل الجزء المرئي عن الجزء الصوتي في الفيديو، وذلك للتركيز على تحليل الصور الثابتة التي تتكون منها المشاهد المتحركة.
2. تصنيف الأطر: يتم تصنيف كل إطار من إطارات الفيديو بناءً على أهميته في نقل المعنى العام للفيديو.
3. ترشيح الأطر المفتاحية: يتم اختيار الأطر التي تحمل أكبر قدر من المعلومات والمعنى وتسمى بالأطر المفتاحية.
4. إدراك الأطر المفتاحية: يتم استخدام تقنيات معالجة الصور لإدراك الأشكال والألوان والبنية في الأطر المفتاحية لاستخراج الملامح البصرية الأساسية.

المرحلة الثانية: إدراك الأشكال الداخلية (Nallappan, M., 2024) :

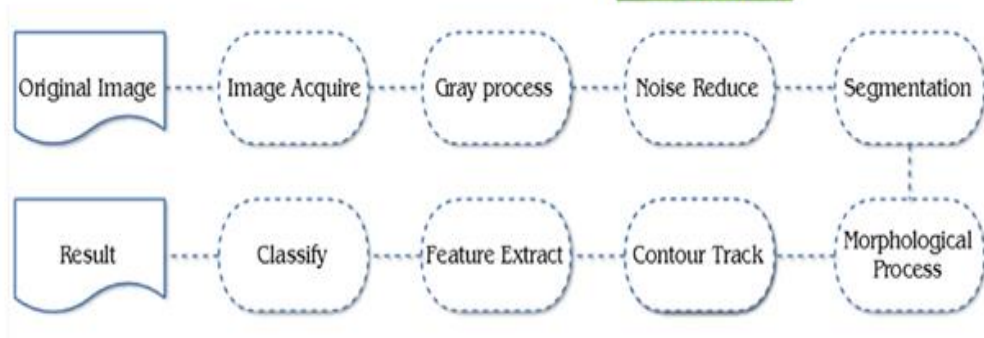
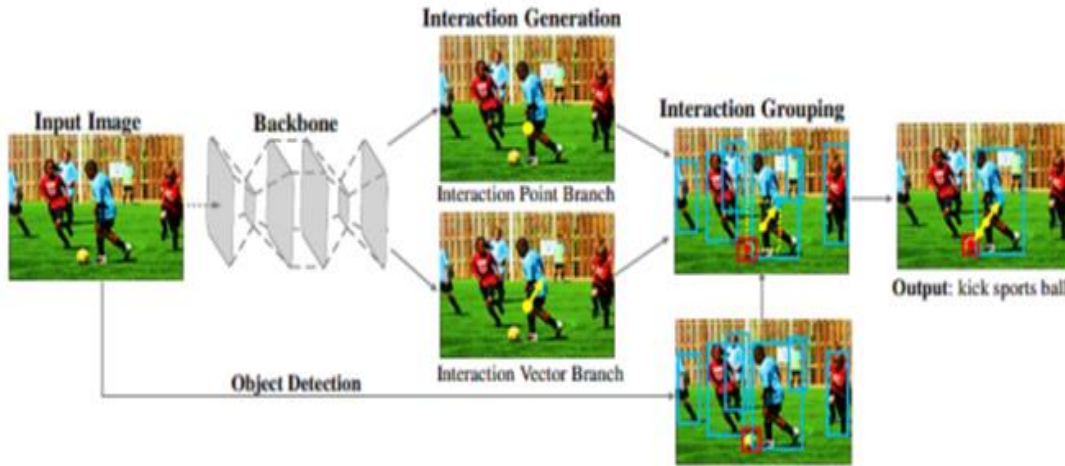
5. التركيز على الأطر المفتاحية : يتم التركيز على تحليل الأشكال والنماذج الموجودة داخل الأطر المفتاحية فقط.

6. إدراك الأشكال الداخلية : يتم تحليل كل شكل أو نموذج بصري داخل الإطار لفهم معناه ودوره في الصورة الكلية.

2 / 1 / 1 / 2 الإدراك الرقمي لوسيط الصورة الثابتة:

إن جميع الصور الرقمية الصور الفنية، والإخبارية، والعلمية تحمل في طياتها معلومات بصرية تتضمن عناصر رسومية وأشكال هندسية، وتهدف تقنيات رؤية الحاسوب إلى تفسير المحتوى المصور من خلال تحليل هذه العناصر واستخلاص الدلالات منها، تعتمد هذه التقنيات بشكل كبير على تقنية التعرف على الأشكال (Shape-Based Image Recognition) التي تسعى لاستخراج الملامح الشكلية للصور وتحليلها. ومع ذلك، فإن تعقيد بنية الصور الرقمية وتنوع الأشكال الموجودة فيها يجعل عملية الفهم أكثر تعقيداً مقارنةً بالنصوص، وقد تنوعت تقنيات إدراك الصور باشتقاق أشكالها الداخلية بين تحليل الخصائص الرسومية shape extraction وإدراك الصور باشتقاق لون الجلد skin detection. (Xing, Y., Meyer, 2023)

وتعتمد تقنيات إدراك الأشكال المصورة على مجموعة من الميزات البصرية الأساسية، وهي: اللون، والبنية، والشكل من خلال تحليل هذه الميزات، ويمكن لنظم الاسترجاع أن تفهم المحتوى الداخلي للصورة وتصنفه، فيمكن استخدام اللون لتمييز الأجسام المختلفة، كما تدل طبيعة البنية الرقمية للصورة على طبيعة ما تحمله من موضوعات، إلى جانب تمييز الأشكال وهيئاتها التي تدل بشكل مؤكد لنظام الإدراك على محتوى موضوعي يختلف من شكل لآخر، ومن ثم فإن هذه الميزات مجتمعة تساهم في بناء تمثيل رقمي دقيق للصورة، مما يسهل عمليات البحث والاسترجاع (Veni, S., Anand, R., 2021).



شكل (3) تدفق عمليات إدراك الصور الرقمية (Jagtap, J., 2021)

يمكن تلخيص عملية إدراك الصور الرقمية من خلال الخطوات التالية (Adjetey, C., 2021) :

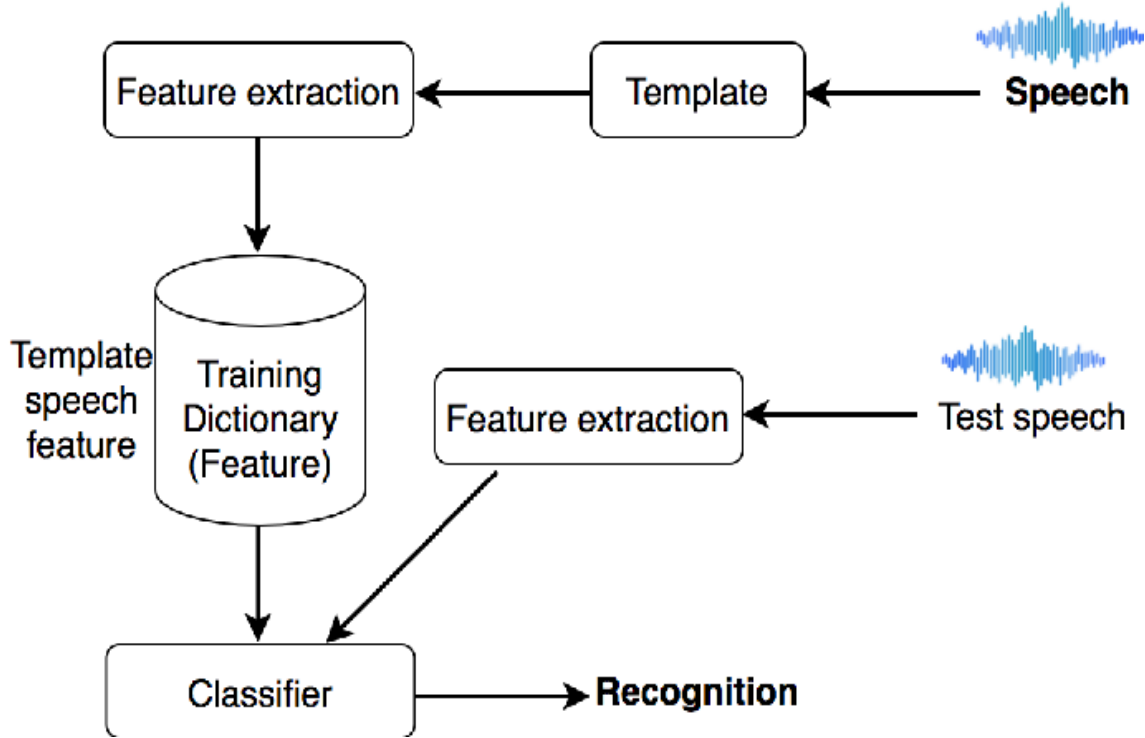
1. التحليل الأولي للصورة : يتم تقسيم الصورة إلى وحدات أصغر (بكسل)، وتحليل قيم اللون لكل بكسل، ثم استخدام هذه المعلومات لحساب الميزات البصرية الأساسية مثل: الحواف والأشكال.
2. استخلاص الملامح: يتم تحديد الملامح البارزة في الصورة مثل: الأوجه، والأجسام، والخطوط، هذه الملامح تمثل التمثيل المجرد للمحتوى المرئي.
3. التصنيف : يتم تصنيف الصورة بناءً على الملامح المستخلصة، وتقسيم الصور التي يتم إدراكها إلى فئات موضوعية أو متخصصة وفقاً للعناصر المشتقة منها.
4. الإدراك : يتم تحويل التمثيل الرقمي للصورة إلى تمثيل دلالي أو لفظي، وهو وصف لمحتوى الصورة يمكن لنظم الاسترجاع التعامل معه، وتعد الخطوة هي الأكثر تعقيداً وتتطلب استخدام تقنيات معالجة اللغات الطبيعية.

2 / 1 / 3 الإدراك الرقمي للوسيط الصوتي:

تتنوع المعلومات الصوتية بين المحادثات، والمحاضرات والموسيقى وغيرها، ورغم هذا التنوع، فإنها جميعاً تحمل معلومات مشفرة في أطوال موجية صوتية، تتميز المعلومات الصوتية بعدم وجود شكل خارجي محدد لحمل المعلومات، على عكس الصور والنصوص، وبالتالي، فإن عملية إدراك المعلومات الصوتية تعتمد على تحليل الخصائص الفيزيائية للأصوات المسجلة، وكما تتنوع ألوان الأشكال والخطوط في المعلومات المصورة والنصية، فإن المعلومات الصوتية تتنوع بحسب المادة المعلوماتية المتضمنة داخل التسجيل، وتتميز المعلومات الصوتية الرقمية عن النصية والمصورة كونها لا تحمل شكلاً خارجياً لحمل المعلومات؛ بل تنصب عملية التحليل والإدراك للصوت على المحتوى المسجل في ملف الصوت، وتواجه عملية إدراك المعلومات الصوتية تحديات عدة، فالأصوات المسجلة قد تحتوي على عدة أصوات متداخلة، أو قد تكون مشوبة بضوضاء خارجية (Park, P.G., et al, 2023)، بالإضافة إلى ذلك، قد تختلف جودة التسجيل بشكل كبير من تسجيل لآخر، ويترتب على ذلك كله أن يمر إدراك الصوت بمجموعة من عمليات التجهيز التقنية السابقة على عملية الإدراك ذاتها؛ كإجراء الفصل التام بين الأصوات المتداخلة لتمييز إدراك صوت بعينه، أيضاً تقسيم لحوار بشكل طولي عبر مدة التسجيل الكاملة، بالإضافة إلى إزالة عمليات التشويش والأصوات الداخلة على الأصوات المحددة المراد إدراكها، وقد وضعت نظم الإدراك بعض قواعد العمل كأن يتم تنقية وتصفية الأصوات المراد إدراكها؛ لأن معدل الإدراك يزيد بانخفاض عدد الأصوات المتداخلة أو عدد المتحدثين في وقت واحد، والعكس من ذلك في حالة ازدياد أعداد الأصوات المسجلة، وتعتمد عملية إدراك المعلومات الصوتية على خصائص وطبيعة نفس هذه المواد المعلوماتية المكونة من أطوال موجية صوتية تحمل إشارات مسموعة للمحتوى المتضمن في مادة المعلومات (Wang, W., 2023).

تحتوي الأصوات على طبقات متعددة من الدلالات، فبالإضافة إلى الكلمات التي تحمل الأفكار والرسائل، تحمل الأصوات أيضاً دلالاتٍ حول حالة المتحدث والبيئة المحيطة، كأن تعكس نبرة الصوت وشدة المشاعر التي يحملها المتحدث، واللهجة والسرعة في الكلام يمكن أن تكشف عن خلفيته الاجتماعية، وبالتالي، فإن المعلومات الصوتية تحمل بُعداً عاطفياً واجتماعياً إلى جانب البعد اللغوي، وإذا كانت الصور قادرة على نقل المشاهد بصرياً من خلال الأبعاد الرسومية، فإن الأصوات قادرة على نقل المشاعر والأوضاع من خلال الأبعاد الصوتية، ولكن هذا البعد الإضافي يمثل تحدياً كبيراً لنظم التعرف على الصوت؛ ويتطلب قياساً دقيقاً لخصائص الصوت كالشدة والتردد، وتحليلاً معقداً للإيقاع والنمط اللغوي (Wen, 2023) ويمثل الإدراك للإشارات الصوتية الواردة في السياق المسموع تحدياً أمام نظم إدراك الصوت، وذلك لما له من متطلبات فنية وتقنية تتعلق بقياس شدة الصوت وانخفاضه والأطوال الموجية للصوت، بحيث يمكن الحكم على الحالة النفسية والعصبية والمشاعر للمتحدث التي قد تحمل دلالات ما للمتلقي في المادة الصوتية، وثمة جوانب أخرى مثل: اللهجة والجنس

والعمر واللباقة لا يمكن لنظم إدراك الصوت الوصول إليها والتعامل معها دون توافر تقنيات وبرمجيات متقدمة تساعد في تحديد هذه الملامح، وهو ما يمكن أن يترتب عليه بفقدانه دلالات سلبية في التعبير عن محتوى التسجيلات الصوتية، والعكس من ذلك صحيح بأن يترتب على اكتسابه دعم لتكشيف وتحليل المحتوى الصوتي والكشف الدلالي عن مختلف المعاني والأفكار التي تضمنتها المادة الصوتية، وقد تركزت عمليات إدراك الصوت فقط في اشتقاق الكلمات اللفظية ودلالاتها اللغوية دون الدلالات الحسية والشكلية الأخرى كلهجة وَجَدَة الصوت لما لها من صعوبة (Fanaras, K., 2022).



شكل (4) إدراك ملفات الصوت بتحليل الإشارات الصوتية (Zheng, W., Ling, 2023).

يوضح الشكل رقم (4) أن نظم التعرف على الصوت تتبع مجموعة من الخطوات الأساسية في عملية إدراك المعلومات الصوتية، وهي كالتالي (Zhu, M., 2022) :

أولاً: تقسيم الإشارة الصوتية : تقوم نظم التعرف على الصوت بتقسيم الإشارة الصوتية إلى مقاطع صوتية أصغر حجماً، قد تكون بضع ثوانٍ، يتم تحديد بداية ونهاية كل مقطع، ويتم ترميز كل فاصل صوتي بأنه كلمة منفصلة، ثم يقارن النظام كل صوت في المقطع بأصوات مشابهة مخزنة مسبقاً في قاعدة البيانات لتحديد مدى تطابقهما.

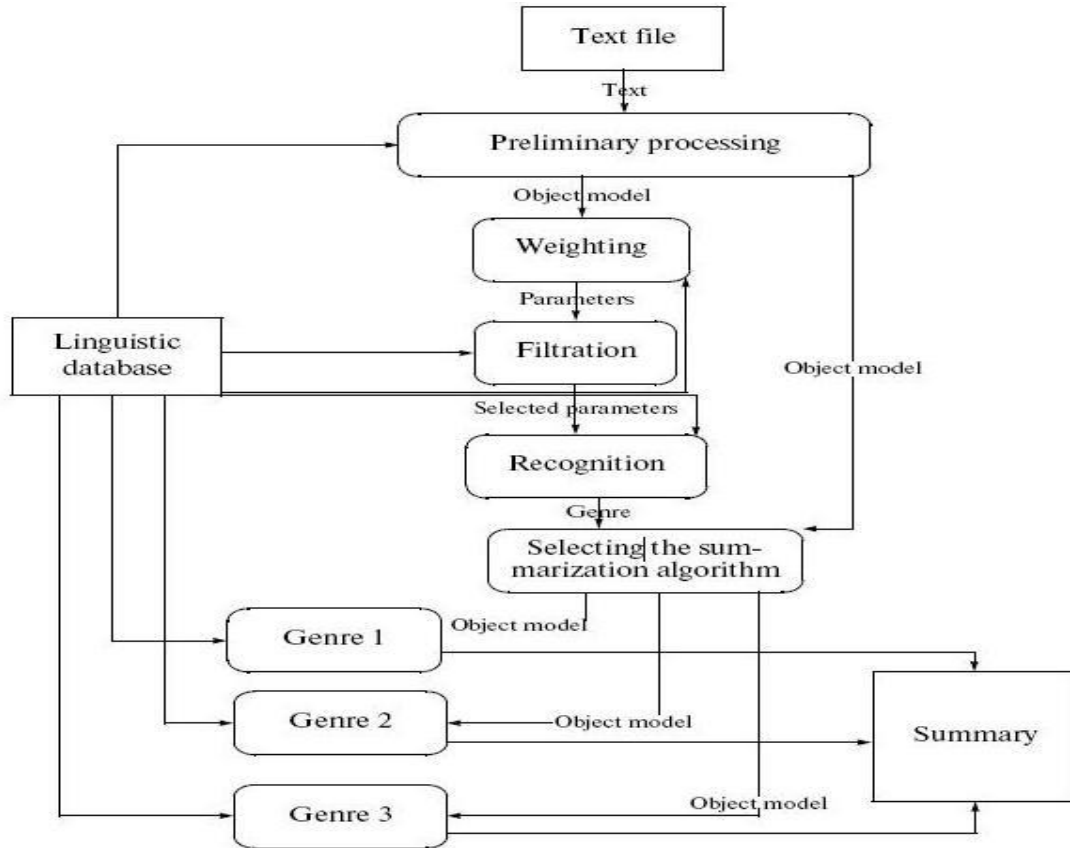
ثانياً: تحديد العقد الصوتية : يتم تحديد نهاية كل مقطع صوتي لكلمة كأنها "عقدة"، هذه العقد تمثل نقاطاً مهمة في الإشارة الصوتية، ويتم نطق الكلمة بشكل كامل، بعد تحديد العقد، يبدأ النظام في البحث عن الكلمات المتوقعة التي قد تأتي بعد كل عقدة، وذلك لتحديد سياق الجملة المحتمل.

ثالثاً: بناء الجمل : بناءً على تسلسل الكلمات والعقد، يقوم النظام ببناء الجمل، فبعد التعرف على كلمة ما بعد عقدة معينة، يمكن للنظام توقع الكلمات التالية في الجملة، كلما زاد عدد الكلمات المتعرف عليها، زادت دقة التنبؤات حول الكلمات اللاحقة.

رابعاً: تحديد بداية الجملة الجديدة : يتم تحديد الكلمة الأخيرة في الجملة كعلامة لبدء جملة جديدة، هذا يساعد النظام على فهم القواعد اللغوية، وتوقع الكلمات التي من المحتمل أن تظهر في بداية الجملة التالية.

2 / 1 / 1 / 4 الإدراك الرقمي للنصي:

لقد ركزت أنظمة إدراك النصوص الرقمية الآلية على المستوى الشكلي والظاهري لمفردات النصوص، ثم قامت بنقلها إلى بيئة رقمية يمكنها التعامل معها كشكل رقمي مفهوم للتعبير عن المعلومات النصي، و التعامل وفقًا لهذا المستوى توقف عند حدود الاسترجاع عن طريق المقارنة والمعالجة عن طريق مقابلة الكلمات المفتاحية والكلمات النصية دون تناول معنى هذه الكلمات وروابطها بشكل موضوعي ومعلوماتي، لذلك، فإن الأهداف الأكثر أهمية لهذه الدراسة هي معالجة الاستخدام المعلوماتي للإدراك بالإضافة إلى الوظيفة التقنية، والتي يمكن أن تقدم جوانب من التحليل الموضوعي وعمليات معالجة المعلومات للنصوص التي يتم إدراكها تلقائيًا (Jeya Christy, A., 2022) ولتحقيق القدرات التحليلية والمعلوماتية لأنظمة الإدراك الآلي للنصوص، يجب أن تتضمن مكوناتها وحدة معالجة تتولى تنفيذ التحليل البنيوي والنحوي، والتحليل الصرفي، والشرح من خلال تصنيف أجزاء الكلام أو من خلال التصنيفات الدلالية وتحليل البنية النحوية للجملة بناءً على مهمة محددة، ويوضح الشكل (5) تدفق العمليات التي تحدث داخل نظام الإدراك والتحليل الموضوعي للنصوص الرقمية التي تخضع للإدراك والتحليل؛ تبدأ هذه العمليات بالوصول إلى المواد الإعلامية النصية أو إضافتها، ثم التطبيق التقني لعمليات الإدراك الآلي، وتنتهي بإحدى عمليات التحليل الموضوعي للمواد الإعلامية الممثلة باستخراج وعرض ملخص للأفكار والمواضيع الواردة في المادة الإعلامية النصية (Kass, D., 2022)



شكل (5) تدفق وتتابع العمليات داخل نظام الإدراك والاستخلاص الآلي للنصوص الرقمية (Li, H, 2020).

يتضح من الشكل السابق أن الإدراك الآلي للنص يتم بناءً على معايير محددة يتم تمثيلها في شكل خصائص لغوية لنص له معاملات وأوزان ثابتة، وتدل هذه القيم والأوزان على أهمية الخصائص والمعاملات اللغوية الواردة بالنص موضع الإدراك، وعلى ذلك فإن النجاح في إدراك النصوص الرقمية باستخدام هذا النظام يمكن من زيادة قدرة المعالجة المعلوماتية للنصوص الرقمية على مستويين: أولهما: تمييز الخصائص اللغوية والتركيبية للنصوص الرقمية، وثانيهما: ضبط المعاملات

والخوارزميات الرياضية اللازمة لتمييز الأوزان النسبية لأهمية ودلالة الكلمات في النصوص (Prieto, J. R., 2020) ، وتواجه عمليات إدراك النصوص مشكلات نوعية تختلف عن تلك المتعلقة بوزن وإدراك الكلمات، ويختلف الغرض من كتابة النصوص بين النصوص العلمية المباشرة والنصوص الفنية التي تحمل رمزية أو استعارة في محتواها النصي، مما يمنع تحديد معايير ثابتة لمعالجة إدراك النصوص الرقمية بخطوات ومعايير محددة، وتصف النصوص الفنية العالم الخيالي، بينما تصف النصوص غير الفنية العالم الحقيقي والمادي، ويلجأ مؤلف النصوص غير الفنية غالباً إلى بناءات شخصية غير محددة لتأكيد الشكل الموضوعي لأحكامه التي تعتبر مكتوبة، وينعكس فيها الطابع الشخصي بما يمثل نتاج خيال المؤلف، وتحمل النصوص الفنية ميلاً فردياً وشخصياً قد يختلف إدراك وتفسير العمل الأدبي نفسه من قبل عدد مختلف من القراء بشكل واضح ويعتمد على الخصائص الفردية للشخصية، وإذا كانت هذه الخصائص البنيوية واللغوية للنصوص الرقمية تمثل تحديات لأنظمة معالجة وتنظيم النصوص الرقمية على الويب، فإنها تمثل تحدياً أكبر لأنظمة إدراك النصوص الرقمية إذا انتقلت من مستوى التعرف الآلي على الكلمات فقط إلى مستوى التحليل الموضوعي والتحليل والمعالجة والتنظيم للنصوص الرقمية التي تم إدراكها (Prieto, J. R., 2021)

لقد تعقد الأمر بسبب وجود تحديات تقنية معلوماتية لفهم النصوص، وهي كالتالي (Chandio, A. A., 2022) :

1. صعوبة التقنية : تقنية النص الرقمي في عملية المعالجة يكون غالباً بصيغة txt ، وعلى العكس من ذلك ، لا يمكن استخدام علامات HTML أو ميزات تقنية أخرى تدعم أنظمة الويب.
2. صعوبة معلوماتية : تحتاج أنظمة المعالجة والتنظيم إلى تطوير قدرات تلخيص قائمة على الإدراك المعرفي؛ وتمثل النصوص الرقمية تصنيفات موضوعية وفكرية مختلفة تؤدي إلى قدرات تحليلية مختلفة وتحديد معايير ثابتة لها. بالإضافة إلى الاختلاف في طرق التعبير والبنى اللغوية بين النصوص، هناك اختلافات أخرى تواجه آليات الإدراك الآلي للنصوص الرقمية، تمثلت في الأدوات اللغوية والنصية التي تساعد في بنية النصوص، مثل: أدوات ربط الكلمات والوصلات وحروف الجر، والتي تمثل في حد ذاتها مكوناً نصياً ضرورياً في بنية النصوص بلغات مختلفة، على الرغم أن هذه الأدوات تكتسب أهمية كبيرة من الناحية اللغوية، إلا أنها تمثل عائقاً وتحدياً أكبر من الناحية المعلوماتية أثناء تنفيذ معالجة وتنظيم المواد المعلوماتية النصية داخل أنظمة الإدراك الآلية، تعيق هذه الأدوات التعبير الدلالي للمواضيع أثناء عمليات التنظيم في عناوين الموضوعات، وأثناء عمليات البحث في بناء وصياغة إستراتيجية بحث عن معلومات نصية رقمية (Zhang, D., Li, 2022)

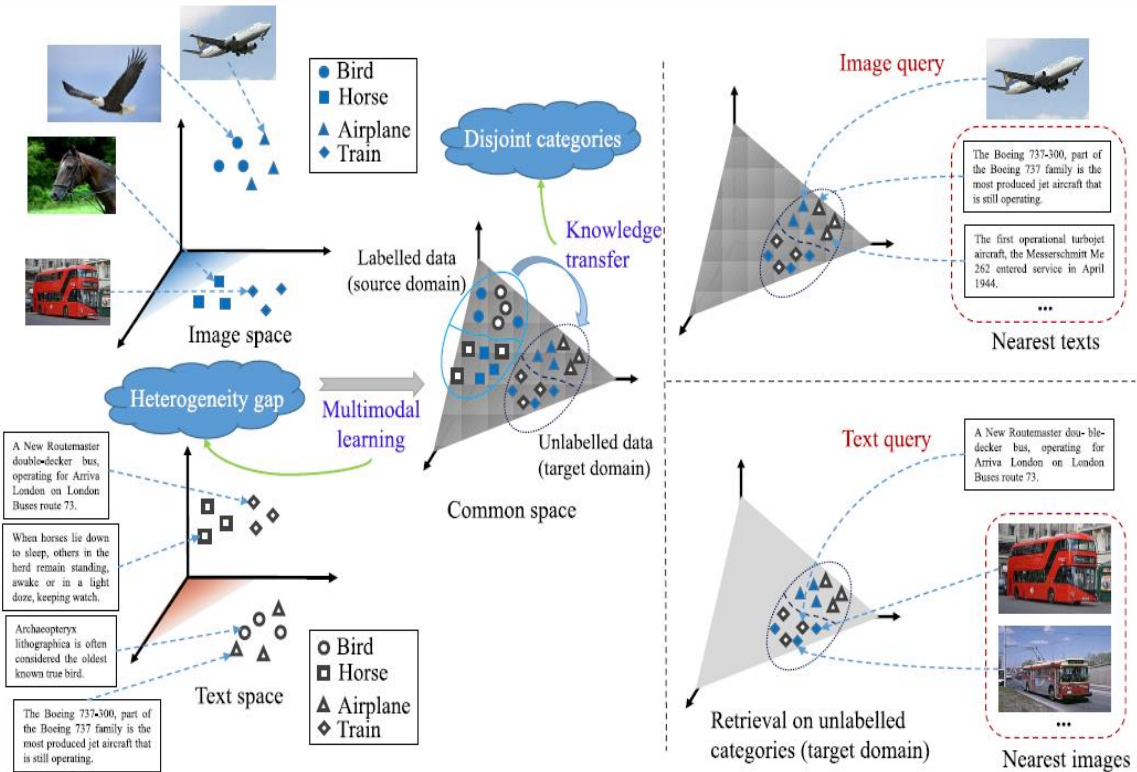
2 / 1 / 2 بناء التمثيلات الدلالية الكَلَيَانِيَّة الموحدة للوسائط المتعددة:

تتم المعالجة الفنية والتنظيم للوسائط المعلومات باستخلاص الملامح الموضوعية والشكلية للوسيط، ثم التعبير عنها في شكل عناصر واصفة كالواصفات للموضوع والحقول للشكل، ويتم تحويل مختلف الملامح الواصفة للمحتوى والشكل إلى ما يطلق عليه التمثيلات الدلالية للوسيط، أي: تلك الكيانات الرمزية الرقمية كالأكواد أو الشفرات التي يتم تخصيصها في ثبوت جامع وكلية داخل النظام، يمكن للنظام من خلالها تمثيل الموضوعات والمجالات المعرفية حتى تكون موحدة لكل الوسائط، ومن ثم تربط بينها لتمييز المتشابه والمختلف (Li, S., et al, 2023) ، وتشابه فكرة هذه التمثيلات الدلالية مع فكرة الرمز في خطط التصنيف، لكن الرمز في خطة التصنيف يكون موجهاً مفهوماً للإنسان، إلا أن التمثيلات الدلالية تكون مستخدمة ومُدركة داخل النظام الرقمي للاسترجاع عبر الوسائط، وليس من الضروري فهمها البشري، نظراً لكون المعالجة بالإدراك والتحليل تتم رقمياً آلياً دون تدخل العنصر البشري، ويمكن للنظام ووفقاً لقدراته التشغيلية أن ينتج هذه الأكواد والشفرات وهذه التمثيلات بشكل ذاتي اعتماداً على مولد رقمي لهذه التمثيلات الدلالية، أو من خلال بنية معدة مسبقاً، ويعمل بها النظام كخطط التصنيف (Zeng, S., Liu, et al, 2022)

لقد منحت تقنيات التحليل الآلي لوسائط المعلومات إمكانات المعالجة الذاتية لنظم الاسترجاع عبر الوسائط من خلال إنتاج تمثيلات ثنائية إرشادية للتجزئة، وتمييز المحتوى الرقمي عبر الوسائط، ويمكن استخدام التمثيلات الدلالية الثنائية التي يتم نمذجتها بواسطة توزيعات متعددة المتغيرات بسهولة وفعالية لتحليل، ورصد علاقات المحتوى بين الوسائط وبعضها البعض، سواء المتشابه منه والمختلف، ولكي تقوم هذه التمثيلات الدلالية بدورٍ فاعل في استرجاع عبر الوسائط، فإن ثمة متطلبات تتمثل في: (Zhen, L., Hu, et al, 2020)

أولاً: يجب أن تلتقط التمثيلات بالضرورة المعلومات من الإدخال، أي: الحفاظ على هوية الوسيط الواحد ومحتواه المعلوماتي. ثانياً: يجب أن تتسم التمثيلات بقدرات المعالجة الرابطة مع الوسائط الأخرى، واكتشاف الروابط المعرفية بين كل وسيط وآخر، حتى يتم رسم خريطة من توزيعات التشابه والتباعد بين وسائط النظام.

ثالثاً: يجب تقليل الفجوة بين الوسائط (أي: الفجوة غير المتجانسة) بين تمثيلات الوسائط المختلفة إلى الحد الأدنى. ويمكن لإجراءات التحليل باستخدام التوزيعات والمصفوفات الدلالية داخل النظام تجنب التشتت بين محتويات الوسائط، وإجراء تحليل أحادي للوسيط، أو جماعي عبر الوسائط بالاشتراك في فضاء دلالي متكامل أنشأه النظام ذاتياً، وتعد تقنيات عمل مثل شبكات التجزئة التوليدية للتمثيلات وسيلة مهمة لدعم قدرات التحليل والمعالجة الدلالية داخل نظام الاسترجاع عبر الوسائط (Lu, H., et al, 2023)، ويوضح شكل رقم (6) أن النظام يستخدم قدرات التحليل الرقمي لوسيطي الصورة والنص في إجراء المعالجة الدلالية لطبيعة كل وسيط، ومن ثم استخراج مجموعات التمثيلات المعبرة عن محتوى كل وسيط، ودمج تلك الوسائط في مجموعات النظام، ويتم استخدام العلاقات والتشابه الدلالي بين التمثيلات للجمع بينها في تصنيفات وتوزيعات على أساس التشابه والتنافر بين تلك التمثيلات، وهذا التوزيع سيكون أساس الاسترجاع عبر الوسائط لاحقاً، من خلال البحث بأحد الوسائط وتحليله، ثم البحث عن التمثيلات الدلالية والرمزية المتوافقة وتمثيلات وسيط الاستفسار، ثم استدعاء مختلف أنواع الوسائط المتعددة ذات الصلة والتي تربطها العلاقات الموضوعية المترابطة (Liu, X., et al, 2022)



شكل رقم (6) إنتاج التمثيلات الدلالية داخل نظم الاسترجاع عبر الوسائط.

إن فاعلية التمثيلات الدلالية لا تكمن فقط في التعبير الدلالي عن الوسائط المتعددة والربط بينها؛ بل تعمل بشكل أساسي على معالجة التحديات والصعوبات التي واجهتها نظم استرجاع الوسيط الواحد، فقد عانت نظم استرجاع الوسيط المصور المبني على تحليل رسوم وأشكال الصورة من صعوبة استرجاع الصور المتوافقة والصورة استفسار البحث، وذلك لما قد تكون عليه الصور من اختلاف رسومي وشكلي تام، في حين تحمل محتوى دلالي وموضعي واحد، وهنا تبرز أهمية التمثيلات الدلالية في تحقيق التوافق والأساس المفاهيمي الدلالي الواحد الذي يحقق للنظام استرجاع كل الوحدات المعلوماتية ذات الصلة لنفس الوسيط، بل ومن الوسائط المتعددة (Pan, Z.K., et al, 2023) ويوضح شكل رقم (7) كيف يمكن أن يكون اختلاف الصور الرقمية في محتواها الشكلي الرسومي، رغم تطابها المعرفي؛ ويتم البحث بصورة تمثال وينستون تشرشل كاستفسار لاستدعاء الصور المتطابقة للمحتوى، وبسبب التباين البصري الواضح بين الصورة الرقمية للتمثال والصور الأخرى للشخصية، يجعل الاسترجاع أحادي الوسائط بالصورة للصورة أمراً صعباً، إلا أن اعتماد النظام للتمثيلات الدلالية وعمل قاعدة المعرفة للنظام يمكن تخطي تلك الصعوبة بالاسترجاع عبر الرموز والتمثيلات الدلالية التي تبني التوافق بين وحدات الوسيط الواحد ووحدات الوسائط المتعددة، بأن يتم العمل والمعالجة للكيان المعرفي الدلالي، وهو هنا شخصية وينستون تشرشل (Jiang, D., et al, 2023).

Visual Question (input)	Relevant Visual Passage in the Knowledge Base
 "In which English palace was this man born?"	 Churchill was born on 30 November 1874 at his family's ancestral home, Blenheim Palace in Oxfordshire.
 "How many avenues radiate from this building?"	 The Arc de Triomphe is located on the right bank of the Seine at the centre of a dodecagonal configuration of twelve radiating avenues.

شكل رقم (7) دور التمثيلات الدلالية في استرجاع وحدات الوسيط الواحد أو عبر الوسائط.

إن إنتاج التمثيلات الدلالية في قاعدة معرفة نظم لاسترجاع يحتاج إلى العمل على مستويين اثنين بشكل متوازي لا يجب الاهتمام بأحدهما أكثر عن الآخر، يتمثلان في المستوى المعرفي المعلوماتي والمستوى المفاهيمي الدلالي؛ فتقنيات رؤية الحاسب يمكنها إدراك الأشكال والرسوم في الصورة الثابتة، وتحليل المشاهد واللقطات في الصورة المتحركة أو التعرف على الأطوال الموجية للأصوات، والتعبير عن هذه الكيانات الرسومية بالكيانات المعرفية المعلوماتية التي تعبر عن محتوى الوسيط (Bin, Y., et al, 2023)، لكن الاهتمام بهذا التحليل المرئي أو المسموع قد يأتي على حساب التحليل الدلالي والمفاهيمي لطبيعة التمثيلات الدلالية وقدرات الربط بينها، لذلك فإن العلاقات الدلالية والتوزيعات المفاهيمية التي تكون قاعدة المعرفة لنظام الاسترجاع، لا يجب أن تقل أهمية عن تقنيات الإدراك، وهو ما يمكن النظام لاحقاً في دمج كل وحدات المعلومات

على اختلاف وسائط المحتوى في نظام واحد شمولي باستفسار واحد واسترجاع كُلي، ولتحقيق المعالجة الفنية والدلالية المتكافئة داخل النظام، فإن ثمة متطلبات أدائية تتمثل في التالي: (Li, B., et al, 2024)

- تضمين السياقات المساعدة في شكل فئات دقيقة يحددها خبراء المجالات الموضوعية المتخصص بتكلفة عمل بشرية إضافية .
- الحصول على المضامين السياقية الدلالية من مجموعة جُمْل لغوية عبر الإنترنت كالموسوعات والمعاجم التي قد تكون متضمنة لتمثيل المعاني الدلالية للفئات الموضوعية النوعية.
- تطبيق مبدأ الحيادية الدلالية في التعامل مع الوسائط المتعددة، دون الاهتمام بمجال أو وسيط محدد.

2 / 1 / 3 مقومات المعالجة الفنية الكَلَيْتَانِيَّة (المكانز الرقمية النوعية):

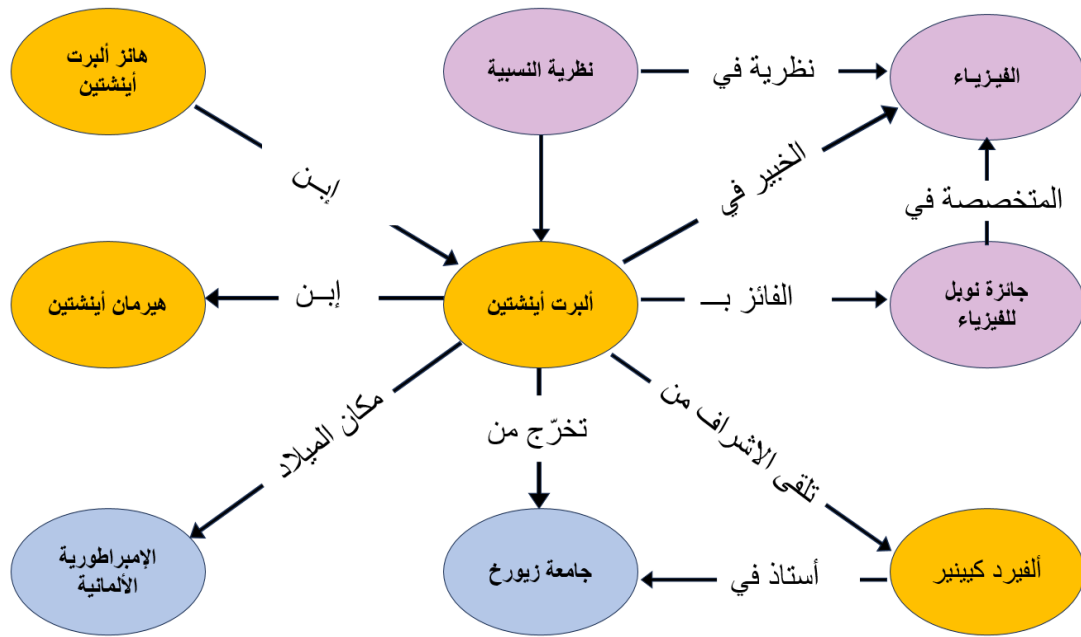
تحتاج نظم الاسترجاع عبر الوسائط في المعالجة الفنية الرقمية للوسائط المتعددة إلى مقومات عمل تتمثل في تقنيات الأداء الرقمي المعتمد على رؤية الحاسب والإدراك لوسائط المحتوى النوعية، إلى جانب مقومات التحليل الدلالي والعلاقات المفاهيمية، وتعد معالجة اللغة الطبيعية والمكانز الموضوعية الرقمية أهم أدوات التحليل المعلوماتي لمحتوى الوسائط المتعددة، نظرا لدورها المهم في تحقيق بنية التوافق والربط لمحتويات الوسائط، لإمكانية البحث بأحدها، واسترجاعها العام الشمولي.

2 / 1 / 3 1 معالجة اللغة الطبيعية (NLP):

تمثل اللغة الطبيعية المحرك الداخلي للعمليات في نظم الاسترجاع لتحليل المحتوى النصي خاصة، والوسائط المتعددة بشكل عام؛ وتدير آليات معالجة اللغة الطبيعية (NLP) natural language processing إجراءات الربط الدلالي والمفاهيم في مجموعات المصطلحات التمثيلات الموضوعية التي تبحث عن استفسارات المستفيدين في المجموعات الداخلية لاستدعاء المتطابق منها وما له علاقة موضوعية بالمعلومات المستدعاة، ورغم كونها تؤدي دورا معلوماتيا إلا أن معالجة اللغة الطبيعية هو الشق اللغوي الممثل للغات الحية بما يفهمه مستخدم نظم الاسترجاع؛ بل الانطلاق في معالجة اللغة الطبيعية حتى لتقريب فهم الاحتياجات الإنسانية للآلة والرد عليها بما يلائمها من جوانبها المختلفة الموضوعية والحسية والدلالية. (Nabankema, H., 2024) وتتطور معالجة اللغة الطبيعية بتطور منظمة الاسترجاع وآليات عملها، وقد عنيت آليات معالجة اللغة الطبيعية برسم شبكات المفاهيم والخرائط المعرفية حتى كانت الأساس في عمل مولدات النصوص الاسترجاعية GPT، وستؤدي اللغة الطبيعية في الدورة الأساسية في بناء الأرضية المشتركة لمعالجة محتوى الوسائط المتعددة؛ وستؤول مختلف أشكال الوثائق إلى المردود اللغوي أو التمثيل الدلالي، وهو كان المستقبل الذي سيفك شفرة الاستفسار أحادي الوسيط، ويعمل في الوقت نفسه على مضاهاتها مع التمثيلات الدلالية الموحدة لكل وسائط ووحدات المعلومات داخل قاعدة نظام الاسترجاع عبر الوسائط، واستدعاء المتوافق منها واحتياجات مستخدمي النظام، وتعتمد اللغة الطبيعية في معالجتها على ما قدم علم الأنطولوجيا في مجالات نظم استرجاع المعلومات؛ ويمكن تتبع دلالات ومفاهيم المجالات المختلفة من خلال أطر وصف المعاني RDF التي تعمل على ربط المفاهيم ببعضها البعض بعلاقات لغوية ودلالية، مثال: إن دلالات السفن والشرع والأسماك تعني الربط بمفاهيم البحار والأنهار والماء، مما يدعم استرجاع المعلومات بامتدادات إضافية (Yu, L., 2024)

لقد شهدت التقنيات الدلالية تقدماً سريعاً، ولعل الويب الدلالي أحد هذه التقنيات المهمة بتحويل المحتوى إلى تنسيق يمكن فهمه بسهولة من قبل الآلة، وقد اكتسبت نماذج البيانات الدلالية مثل: الأنطولوجيات ورسوم المعرفة أهمية كبيرة؛ وتعتبر الأنطولوجيا العمود الفقري لأجيال نظم استرجاع المعلومات الداعمة للتحليل اللغوي والدلالي لمجموعاتها، وهو ما يرجع إلى آليات عمل توفر سياقاً ومعنى للمعلومات، فيمكن استخراج المعرفة وإعادة استخدامها بكفاءة؛ بل إنها توفر علاجاً لمشكلات

التوافقية البنوية والدلالية، التي تعيق تبادل المعلومات وتوافقها بكفاءة عبر الوسائط غير المتجانسة، ونظراً لاستخدام التمثيلات والتصنيفات والأنطولوجيات المختلفة في مجال المعالجة والاسترجاع لتمثيل المعرفة والمعلومات للوسائط غير المتجانسة والمعقدة في كثير من الأحيان، فإنه يلزم إجراء مقارنة واختيار يدويين للمفاهيم والدلالات، وهو ما يمكن أن يقوم به النظام ذاتياً إذا ما توفرت تقنيات معالجة اللغة الطبيعية الداعمة لذلك، ولقد تطورت أدوات معالجة اللغات فقط للغة الإنجليزية أولاً، ثم نالت لغات أخرى اهتمام مطوري معالجة اللغة الطبيعية، وبدأت الأدوات الخاصة بتلك اللغات في التطور أيضاً في مجال استرجاع المعلومات ومعالجة اللغات الطبيعية، بالتركيز على أدوات معالجة النص للتحليل على المستويات اللغوي والنحوي والصرفي والدلالي، واقترح الباحثون العديد من خوارزميات الجذع للغات الوطنية والدولية المختلفة مثل: الإنجليزية والعربية والأردية والصينية، والألمانية، والفرنسية، والإيطالية (Wang, S., 2024).



شكل رقم (8) اشتقاق العلاقات الدلالية والموضوعية.

ولبناء المصفوفات والتوزيعات الدلالية كما يتضح من شكل رقم (8)، فإن ثمة آليات يمكن استخدامها في معالجة اللغة الطبيعية داخل نظم الاسترجاع عبر الوسائط وبناء الشبكات والتوزيعات الدلالية والمفاهيمية، وهي تتمثل في الآتي: (Jabbar, A. et al, 2023)

الآلية الأولى: تعد طريقة بيانية تعتمد على الأهمية الذاتية لكل عبارات ومصطلحات الوصف، حتى يتم بناء دلالي رسومي تكون مركزه عبارة أو كلمة، وتنفيذ مصفوفة اتصال تعمل على مبدأ التشابه الدلالي لإتمام بناء باقي المصفوفات المتجاورة، ويتم الاستمرار بهذا النهج لبناء مصفوفة المصفوفات الدلالية لكل الوسائط داخل قاعدة النظام، اعتماداً على تحديد التشابه الدلالي في عدد "ن" (عدد التخصصات الممكنة) من الاتجاهات الموضوعية، والوسائط المختلفة؛ بل واللغات الحية المتعددة. الآلية الثانية: باستخدام التحليل الدلالي الكامن، والمعروفة أيضاً باسم المعالجة الدلالية الكامنة لمحاكاة ونمذجة جوهر الكلمات الواردة في النص وتحليله على مستوى القيم المفردة، وتتنبأ آلية التحليل الدلالي الكامن بمعنى مصطلح معين من خلال تحليل التزامن للمصطلح السياق النصي، انطلاقاً بأن المعنى يتحدد بموضع الكلمة ومكانها في بنية النص.

الآلية الثالثة: تعتمد على النهج التجريبي باختيار الكلمات ذات الأهمية الأعلى بناءً على توزيع التردد الخاص بها، وتعيين أوزان أعلى للكلمات المذكورة في بداية النص ونهايته، ويتم حساب وزن الأهمية وفقاً لمعادلات ثابتة داخل النظام (Mishra, A., 2022)

2 / 3 / 1 / 2 المعاجم والبنى الدلالية الأنطولوجية:

تتمثل مهمة نظم الاسترجاع في تحديد المعلومات الأكثر ارتباطاً باستعلام المستخدم، سواء أكانت هذه المعلومات مستنداً أم صورة، أم فيديو، أم نصاً، ويتشكل الاستعلام من مجموعة من الكلمات تمثل حاجة المستخدم، مما يجعل لهذه الكلمات أهمية كبرى في نجاح النظام أو فشله لاستدعاء النتائج الصحيحة المرجوة من مستخدميه، وعليه، فإنه لتحسين استرجاع المعلومات لا بد للنظم أن تعمل على تحسين استرجاع المعلومات الدلالية باستخراج المفاهيم، ثم تصفية النتائج باستخدام المعلومات الدلالية التي يوفرها البناء الدلالي داخل النظام (Chebil, W., et al, 2023). وتعمل آلية الاستدلال اعتماداً على شبكة من العلاقات والروابط المفاهيمية التي تميز بين المصطلحات، واستخراج المفاهيم، ومطابقة النتائج، والاستعلامات، ويمكن أن يطلق على هذه البنية المفاهيمية داخل النظام مصطلح المعجم المعرفي؛ ويعد نظاماً لتمثيل إطار المعرفة الموضوعية في مجالات موضوعية متعددة، وتمثل المصطلحات به الكيانات المعرفية التي تحدد رسم الارتباطات الموجهة في نظام المعرفة، ولذلك فإن هذه المعاجم تعد أداة دلالية مهمة في مجال استرجاع المعلومات على وجه التحديد، لما تحمله من بنية تشابكية معرفية وقدرة الاسترجاع الدلالي واستعراض مجالات المعرفة واكتشاف علاقاتها، وتتكون عملية بناء المعاجم المعرفية من ثلاث وحدات (Wang, X. et al, 2021):

أولاً: المعجم الأساسي: حيث يتم وضع علامات على الكلمات الأساسية من جانب النظام أو من قبل خبراء في المجال، ويتم إنشاء نظام التصنيف والتوزيع لهذه الكلمات والدلالات.

ثانياً: البناء الاحترافي: يتم الحصول على الكلمات المهمة من مجموعات المعلومات ووسائطها، ثم دمج الدلالات غير المتجانسة متعددة المصدر وتنظيفها ومعالجتها مسبقاً لتوليد مجموعة شبكة علاقات جديدة.

ثالثاً: المعجم المعرفي: يبدأ باكتشاف الدلالات الموضوعية الجديدة وتصنيفها تلقائياً وفقاً لحساب التشابه الدلالي، ثم إجراء التوسع الدلالي للمصطلحات وفقاً لخوارزمية التصنيف متعدد العلاقات والاستنتاج التلقائي لموقعها في الشبكة المعرفية للمعجم.

إن البنى الدلالية لمعالجة المعلومات تجعل مهام استرجاع المعلومات تنقسم إلى نوعين هما: مهام الوصول والاستدعاء لتلبية احتياج مستخدم النظام والرد على استفساره، ومهام ذات طبيعة تحليلية معلوماتية، تبرز أهميتها أثناء المعالجة لوسائط المعلومات داخل النظام، وتتضمن مهام النوع الأول، العثور على حل يمكن دائماً تمثيله بعملية استدعاء ذات نتائج محددة، أي: تسلسل موجه للأحداث والإجراءات على الكائنات الدلالية المعرفية للنظام، وهذا يجعل من الضروري تقديم نتائج البحث في شكل يعكس التركيز (من نقطة البداية إلى "الاستجابة") لاستجابة النظام، وفي الحالة المثالية، تتولى التمثيلات الدلالية وعلاقاتها إجراءات المعالجة والبحث والوصول إلى النتائج الصحيحة، وتتضمن تسلسل الدلالات المقابلة للأشياء والأحداث والإجراءات، والمفاهيم المعبرة عن كل وسيط معلوماتي، بداية من المفاهيم الأساسية إلى المفاهيم في سياق حل محتمل لمشكلة النشاط الرئيسي (Meng, X., et al, 2021) أما بالنسبة للنوع الثاني المختص بالمهام التحليلية المعلوماتية، فتتضمن إجراءاته تصور سياق المفاهيم المرجعية للمحتوى الرقمي، والمفاهيم الأخرى المتجانسة معه وغير المتجانسة، مع التعمق في كشف المحتوى عبر الوسائط. وعليه، فإن نظم الاسترجاع عبر الوسائط المبنية على البنى الدلالية والمعرفية تتحرك في معالجتها واسترجاعها وفقاً لمخططات الأنطولوجيا، وهو ما يتم على مستويات عمل متباينة: (Cheliger, C, et al, 2022)

- مستوى البحث عن مجموعة من الدلالات والكيانات المعرفية المقابلة لأجزاء المحتوى المتضمن في الوسيط، وتكون النتائج هنا مبنية على الشبكات المعرفية جميعها وتكاملها داخل النظام.
- مستوى البحث عن المفهوم الواحد وعلاقاته الأولية دون التوجه إلى العلاقات المشتتة الأخرى مع باقي شبكات المفاهيم أو الكيانات المعرفية الموزعة؛ ويتم استخدام المفهوم المرجعي كنقطة بداية ونهاية.
- مستوى البحث بمزيج من المخططتين الأولين.

إن البنى والتمثيلات الدلالية المنشأة بواسطة نظام من المفاهيم والعلاقات، ستمثل الرسوم الأنطولوجية التي توفر المساحة التقنية الكافية لنظام الاسترجاع بتحديد "نقاط الدخول" في مصفوفة المعلومات، مما يوفر إمكانية الانتقال المباشر من رؤوس الرسم البياني إلى أجزاء المحتوى الرقمي المتضمن في الوسائط، وستلعب المخططات الأنطولوجية دوراً مهماً في تمكين مستخدمي النظام من تصفح البنى المعرفية المُخلقة من قبل النظام، بتحري الاحتياجات المعرفية والتطلع إلى رؤى استرجاعية جديدة، بجانب ما توفره هذه المخططات من قدرات عمل في استدعاء معلومات مشتتة غير متجانسة ما كان مستخدمو النظام ليعبروا عنها أو التطلع إليها (Sadirmekova, Z., et al, 2023).

المبحث الثاني: بناء نظام الاسترجاع الشمولي عبر الوسائط:

يتناول هذا الجزء من الدراسة بنية ومعمارية نظم الاسترجاع عبر الوسائط، بالتعرف على آليات وإجراءات ومراحل العمل بها، فضلاً عن مكونات النظام ودعائم عمله، وذلك حتى يمكن تقديم الصورة الكاملة لإجراءات المعالجة والاسترجاع والتعامل مع الوسائط المتعددة، وسيتم البداية من المعمارية والهيكلية التي تكون عليها نظم الاسترجاع عبر الوسائط، ثم الدعائم الواجب توافرها حتى يمكن للنظام البدء في عمله.

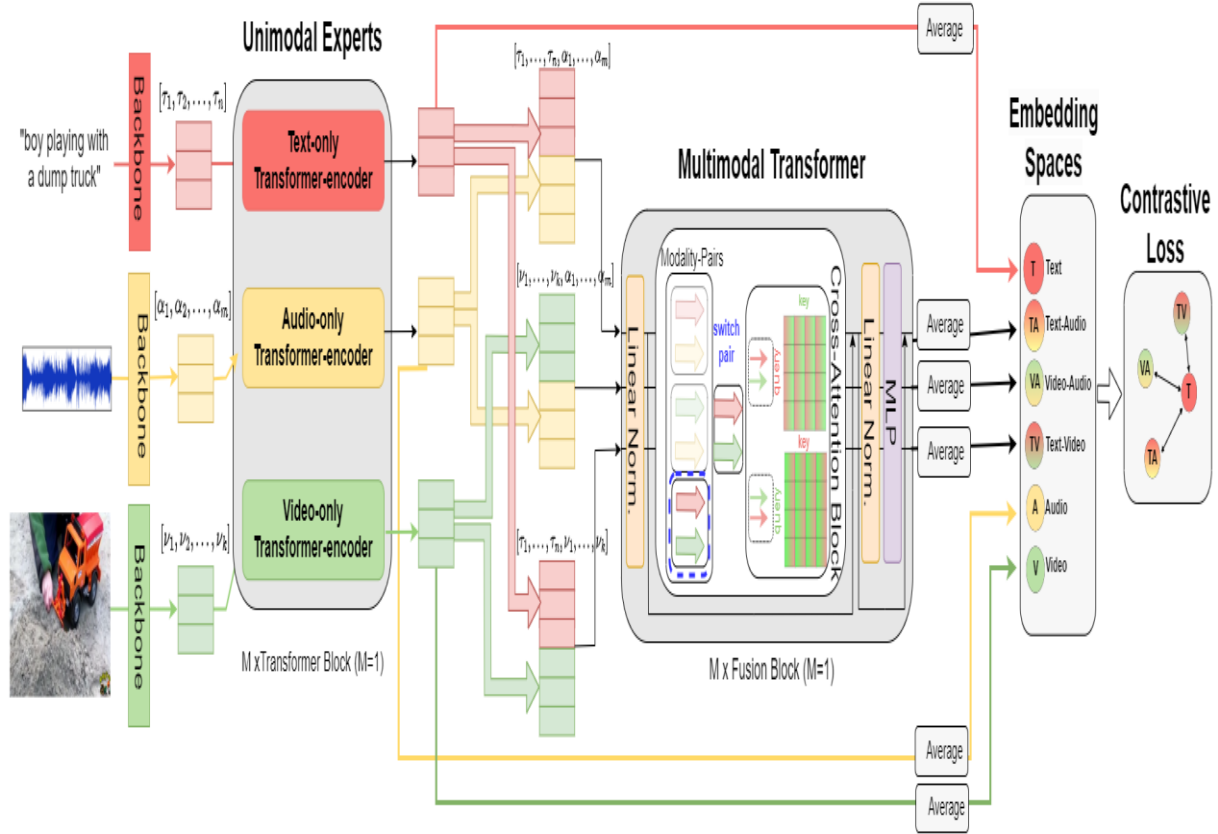
2 / 2 / 1 معمارية نظم الاسترجاع عبر الوسائط:

تستلهم بنية نظم الاسترجاع عبر الوسائط هيكليتها من طبيعة عمل المحولات الاندماجية متعددة الوسائط؛ وتبدأ رحلة المعالجة الفنية للوسائط المتعددة بإجراء بعمليات توليد الرموز الخاصة بملاحج كل وسيط وفقاً لطبيعته المصورة أو النصية أو المسموعة، ثم يتمثل ذلك في البنية الكلية في دمج المحولات المستقلة لكل وسيط، حتى يكون النظام إجمالاً عبارة عن بنية متكاملة لتحويل النص والصوت والفيديو جميعها إلى تمثيلات ورموز دلالية، وينتقل النظام ببنيتها إلى استكمال الحساب التبادلي للتمثيلات الدلالية عبر الوسائط، بما يجعلها شفرات موحدة تتكون من وسائط متعددة بطريقة مرنة وقابلة للتوسعة، وإن ثمة مكونات تقوم عليها بنية نظم الاسترجاع عبر الوسائط تتمثل في المكون الأول المعني بإدراك واشتقاق الملاحج الشكلية والمعبرة عن المحتوى وفقاً لكل وسيط، ثم المكون الثاني المنوط بتضمين التمثيلات الدلالية والرموز المودة الدالة على مجموعات الوسائط داخل قاعدة النظام وتنويعها جميعاً على اختلاف طبيعتها في حالة مادة معرفية واحدة ذات هيئة دلالية ومعلوماتية موحدة، أما المكون الثالث الأخير، فيعني بطبيعة النتائج المسترجعة وهيئاتها النهائية أمام مستخدم النظام وقدرته على الجمع بين الوسائط المتعددة في عرض وموحد لنتائج البحث رغم اختلاف خصائصها الشكلية (Zhang, X., Zhen, T., et al, 2023).

2 / 2 / 1 مكون إدراك ملاحج الوسائط:

يتم معالجة مدخلات جميع وسائط حمل المعلومات النصية والصوتية والفيديو من خلال محولات لاستخراج الملاحج الشكلية، ثم تخضع الميزات المستخرجة لتحليلات نمطية خاصة بالوسائط من خلال طبقات إسقاط الرموز والتمثيلات المعبرة عن كل جزء من أجزاء الوسيط، وتتوسع عملية المعالجة النمطية لتشمل مختلف الإصدارات الرقمية لكل صورة أو جودة مختلف ملفات الفيديو أو البناءات الموجية لكل ملفات الصوت، مما يمكن من خلاله ضم مختلف الوسائط باختلاف حالاتها الرقمية، وينقسم مكون إدراك الملاحج للوسائط إلى جزئين رئيسيين: الجزء الأول: من النموذج من ثلاث محولات أحادية الوسائط عادية للنص والفيديو والصوت كما هو موضح في الشكل (9) (Verschuuren, P.J., et al, 2023)، إن الهدف الرئيسي وراء استخدام المحولات أحادية الوسائط هو إنشاء تمثيل أفضل وتكييف التضمين لكل وسيط بشكل مستقل، وبشكل فعال، وهو ما يتيح أداءً أفضل عبر المهام أحادية الوسائط، كما يعمل أيضاً كمدخل لمحول الاندماج لكل الوسائط، بما يمكن

معه استغلال هذه المدخلات عالية الأبعاد بشكل جيد في نمذجة العلاقات الدلالية متعددة الوسائط، ويتم معالجة كل من الرموز للنص والفيديو والصوت، لتحقيق قدرة الدمج والإقران بالتمثيلات والرموز الدلالية المحولة وتنظيمها من خلال توجيهها لمحول الاندماج، وستكون المخرجات المدمجة مدخلات للمكون الثاني القائم بعمليات الربط الدلالي وبناء شبكات وتوزيعات معرفية بين تلك التمثيلات ورموز الوسائط (Xu, B., Xiong, Y., et al, 2022)



شكل رقم (9) إجراءات الإدراك للوسائط المتعددة في المكون الأول لنظام الاسترجاع عبر الوسائط (Yu, Q., et al, 2023)

أما الجزء الثاني المتعلق بالمحولات متعددة الوسائط، فيعد الأكثر أهمية في إدراك الملامح البنيوية لكل وسيط مزدوج الخصائص مثل: ملفات معلومات الفيديو أو الرسوم الحاسوبية؛ وتستطيع هذه المحولات اشتقاق وتمييز كتلة من الملامح المتقاطعة يمكن أن تكون ثنائية الاتجاه أو مشتقة الخصائص، ملفات الفيديو هي بالأساس معلومات مختلطة بين صوت وصورة ونصوص الترجمة في أحيان كثيرة، مما يعني أن اشتقاق الخصائص الشكلية سيكون متعدد الملامح ومتفق المحتوى الموضوعي والمعرفي، مما يستلزم قدرات تختلف عنها المحولات أحادية الوسائط، وتعطي المحولات أحادية الوسائط المخصصة لكل تمثيلات رمزية لكل وسيط شكلا ثابتا معبرا؛ كأن ترمز بتمثيلة t للنص، و v للفيديو، و a للصوت، أما محول الاندماج متعدد الوسائط، فتعطي ترميزا مزدوجا للوسائط المتعددة، كتمثيلات ta : للنص والصوت، و tv للنص والفيديو، و va للفيديو والصوت، وتودع مختلف الوسائط وتمثيلات الرمز في مساحة التخزين لنظام الاسترجاع، ويتم تطبيق هذه الوسائط وتحديد ما بينها من ارتباطات موضوعية وعلاقات دلالية بسياقات النظام وقاعدة المعرفة، وتبرز في هذا المكون أهمية الاعتماد على تقنيات مثل: رؤية الحاسب وتعلم الآلة والإدراك الرقمي لمختلف المميزات واللامح المرتبطة بوسائط حمل المعلومات (Gong, Y., et al, 2023)

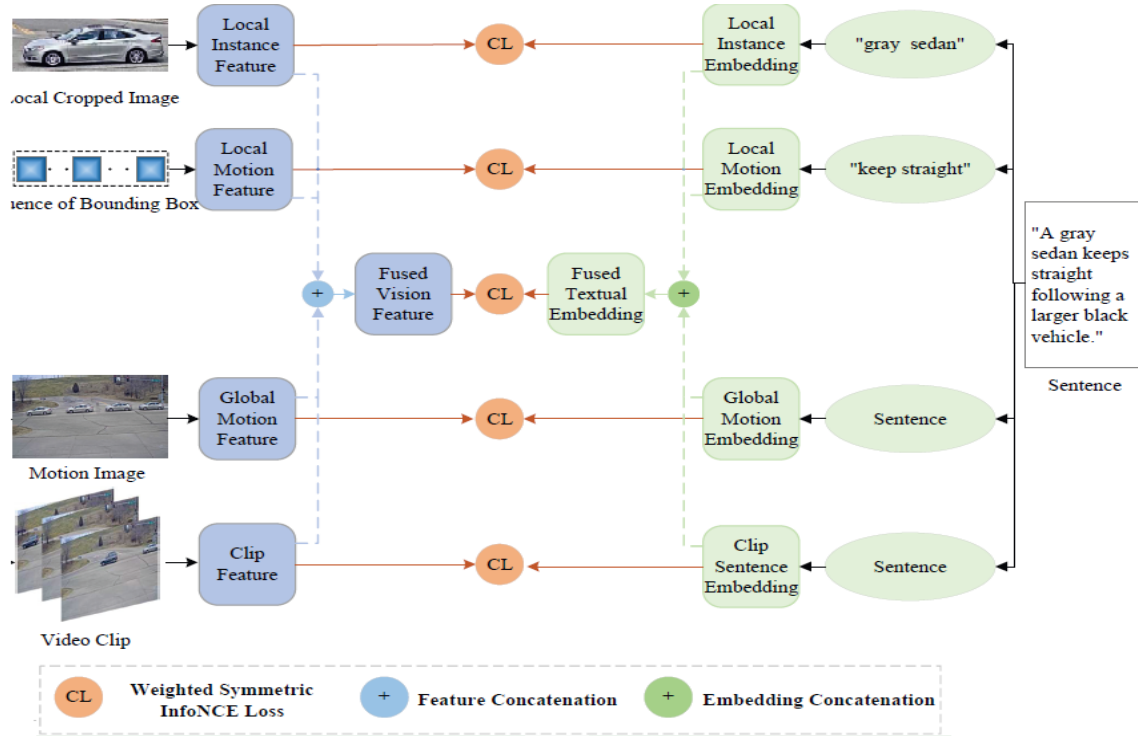
2 / 2 / 1 / 2 مكون المعالجة عبر الوسائط:

يمثل هذا المكون واسطة العقد لنظام الاسترجاع عبر الوسائط، ذلك لما له من دور فاعل في تحويل الخصائص والملامح الشكلية لمختلف الوسائط المتعددة إلى مجموعات من الكيانات المعرفية والدلالية التي تعبر عن المحتوى المعلوماتي لهذه الوسائط متفرقة ومجمعة، كما يعمل على انتقال معالجة الوسائط من المعالجة الأحادية المستقلة لكل نوع أو وحدة من وسائط حمل المعلومات، إلى مجموعة الوحدات المعلوماتية إجمالاً من خلال أولاً: توحيد بنية المعالجة في التمثيلات الدلالية، وثانياً: في شبكات العلاقات والارتباطات بين وحدات المعلومات، بما يدعم تعامل النظام ورؤيته للمحتوى العام وإمكانية الاستجابة لمختلف الاحتياجات التي يطرحها مستخدموه، فضلاً عن المضاهاة والمقابلة للاحتياجات بالنتائج على أسس الصلة المعرفية الموضوعية، وليست على أساس التصنيف النوعي لوسائط المعلومات، ثم ينعكس ذلك على النتائج المستدعاة بما يجمع مختلف الوسائط في واجهة استرجاع واحدة دون تمييز أو تجاهل لوسيط محدد، ولتقييم الأهمية التي يكتسبها مكون المعالجة، فيمكن تخيل كم الاحتمالات التي يمكن أن يطرحها مستخدمو النظام حول جوانب معرفية مرتبطة، بل وكـم الصبغ التي يمكن أن تكون عليها استفساراتهم، وأيضاً قدرتهم في التعبير عنها بدرجات من الوضوح إلى الغموض، أو تلك اللغات الحية التي يتحدثون بها، ويعني ذلك كله أن المعالجة الدلالية والسياقية التي يقوم بها مكون المعالجة ستكون عاملاً فاصلاً في نجاح الاسترجاع عبر الوسائط أو عدمه (Zhang, B., et al, 2021)

ولتوضيح أهمية التحليل الدلالي ومدى تأثيره على النظام، فيمكن النظر في التفاصيل والاتجاهات الدلالية التي يمكن أن تكون عليها استفسارات مستخدم لاسترجاع الوسائط المتعددة عن (سيارة تسير على طريق)، فنظراً لوجود أنواع عديدة من المركبات على الطريق في واقع الأمر، يمكننا وصف الهدف بشكل أساسي من ثلاثة جوانب دلالية ومعرفية مختلفة تتمثل في خصائص المظهر، ومسارات الحركة، والبيئة المحيطة، بالنسبة للجانب الأول، يستخدم الأشخاص عادةً ألوان وأنواع المركبات لوصفها، مثل سيارة سيدان رمادية، أما بالنسبة للجانب الثاني، فعند وصف حركة المركبة، تنجح الاستفسارات إلى استخدام مصطلحات مثل "استدارة إلى اليسار أو السير بخط مستقيم، أما بالنسبة للجانب الأخير، فإن تقديم وصف مفصل بمساعدة البيئة المحيطة يعد طريقة فعالة لتحديد واستدعاء النتائج المستهدفة بفاعلية، فيمكننا تحديد موقع الهدف بسهولة من تدفق حركة المرور باستخدام الجملة "سيارة سيدان رمادية تسير بشكل مستقيم تتبع مركبة سوداء أكبر، ووفقاً للجوانب الثلاثة المذكورة، يتم بناء إستراتيجيات الربط الدلالي والعلاقات المفاهيمية بين هذه الكيانات المعرفية الموصوفة في استفسار مستخدم النظام، ثم مقابلتها بالتمثيلات الدلالية المعبرة عن مختلف الوسائط من نصوص، وصور، وصوت، وفيديو (Banar, N., et al, 2021)

يؤكد المثال السابق على أهمية التحكم الدلالي في مرحلة المعالجة وتأثيره على قراءة الاتجاهات المعرفية للنتائج المسترجعة، ونظراً لهذا لمستوى الأداء الدلالي والمعرفي المطلوب لمكون المعالجة وإنتاج التمثيلات الدلالية، فإن ثمة إجراءات يجب وضعها في الاعتبار عند بناء الشبكات المعرفية والعلاقات الدلالية بين التمثيلات المعبرة عن المحتوى متعدد الوسائط متعدد المجالات متعدد اللغات: (Li, C., Xu, H., et al, 2022)

الإجراء الأول: البناء المعرفي والدلالي الصحيح على مستوى التمثيلة الواحدة، بحيث يكون لكل تمثيلة مفهوم محدد وواضح. الإجراء الثاني: مراعاة العلاقات الدلالية بين التمثيلات وبعضها البعض لكل التوجهات المعرفية الاجتماعية والإنسانية والعلمية... وغيرها، بما يسمح ببناء توزيعات وشبكات مفاهيمية ومعرفية تستطيع استرجاع النتائج على مستويات عدة. الإجراء الثالث: منح مكون المعالجة المرونة الكافية لإعادة الربط وتشكيل العلاقات المعرفية بين التمثيلات الدلالية وبعضها البعض، وفقاً لخبرة الربط السابقة واتساقاً مع توجهات إعادة الاستفسارات، بما يمنح النظام القدرة على التماهي والتوافق مع الاحتياجات المستحدثة والتطورات المعرفية.



شكل رقم (10) إجراءات إنتاج التمثيلات الدلالية في مكون المعالجة بنظام الاسترجاع عبر الوسائط (Pang, N., Guo, et al, 2023) يوضح الشكل رقم (10) بداية عمل مكون المعالجة الفنية للاسترجاع عبر الوسائط من خلال التعبير عن الملامح الوصفية المشتقة من طبيعة الوسيط، وتحويلها إلى تمثيلات دلالية في شكل ألفاظ يمكن بها بناء أساس موحد لكل مجموعات ووحدات المعلومات متعددة الوسائط في النظام واسترجاعها بشكل شمولي بناءً عليه، ولعل مفهوم تعزيز البيانات هنا يحمل أهمية كبيرة لعمل مكون المعالجة؛ ويتم تعزيز التمثيلات الدلالية الواصفة للوسائط بالبناء وفقاً لمعالجة اللغة الطبيعية التي توفر للنظام إدراك السياقات المختلفة للتركيبات اللغوية، فضلاً عن تعزيز البيانات بجعلها تشمل الوصف المكاني والزمني والحركي للوسائط المتعددة، حتى تكون هذه المتغيرات ممثلة دلالياً ودورها المكمل في التعبير عن الاحتياجات المعرفية وفقاً لاشتراطات محددة، وعليه، فإن تمثيل خلفيات الصور ولقطات الفيديو والأصوات المتعددة في الوسائط المتعددة يعد بعداً دلالياً مهماً في إكمال العلاقات المعرفية التي تربط محتوى وسائط حمل المعلومات (Arora, P., et al, 2024)

2 / 2 / 1 / 3 مكون النتائج عبر الوسائط:

يستمد مستخدم نظام الاسترجاع عبر الوسائط حكمه النهائي عن النتائج من واقع الواجهة وطبيعة النتائج المسترجعة، فإذا ما كانت النتائج متسقة شكلاً وموضوعاً مع احتياجات المستخدم، كانت النتيجة الرضا التام عن أداء النظام، أو العكس من ذلك، وتختلف نظم الاسترجاع عبر الوسائط في شكل وطبيعة النتائج المسترجعة، كونها تتعامل مع وسائط متعددة متباينة في شكل وطبيعة التعبير عن المحتوى الموضوعي، بما يجعل الاهتمام بتنسيق وعرض هذه النتائج أمراً مهماً لمكون النتائج في النظام، وستتغير طبيعة واجهات بحث واستعراض النتائج على المتبع الآني لنظم الاسترجاع، بما يوفر للمستخدم مرونة البحث بوسيط محدد والنتائج بوسائط أخرى متعددة، وستقاس كفاءة النظام من الجانب الشكلي بمقدرته على تحقيق العرض الأفضل لمجموعات من النتائج بوسائط متعددة يحتاج كل منها إلى إظهاره وعرضه بتمييز مختلف (Iscen, A., et al, 2023)



شكل رقم (11) واجهة النتائج في نظام الاسترجاع عبر الوسائط (Devillers, B., et al, 2023)

يوضح الشكل رقم (11) بنية الواجهة الأمامية لعرض نتائج الاسترجاع عبر الوسائط بحيث تكون هذه الواجهة جامعة لتقسيمات داخلية تشمل ثلاثة أقسام: واجهة التنقل، وواجهة الاسترجاع، وواجهة المحادثة، ولكل من هذه الأقسام دوره في التفاعل بين النظام ومستخدمه؛ بحيث تُستخدم واجهة التنقل لإدخال استعلام نصي في مربع النص المقدم أو تحميل استعلام بصري باستخدام زر التحميل، وبالنقر على زر البحث، يتم عرض نتائج الاستفسار على واجهة الاسترجاع (Sain, A., et al, 2024). ويمكن لمستخدم النظام الانتقال إلى واجهة المحادثة بحيث يستطيع إدخال استفسارات نصية في مربع النص كمحادثات مع النظام، أو تحميل ملفات الصور والصوت باستخدام زر التحميل، وهو ما يجعل المحادثات متقدمة لدرجة استيعاب السياق النصي أو السياق البصري والمسموع، ويمكن للجانب التقني في نظام الاسترجاع عبر الوسائط لعب دور مهم في التحرك على واجهة النتائج بمرونة كأن يتم الاستعلام النصي لطلب استرجاع نص إلى صورة عند الوقوف على إطار صورة محددة أو على الوصف المعروض عنها، أو أن تصبح الصورة وملفات الفيديو والصوت المحددة بمثابة الاستعلام البصري أو المسموع لطلب استرجاع صورة إلى نص، كذلك يمكن لمستخدم النظام عند سحب صورة مسترجعة، نقلها إلى واجهة المحادثة كاستعلام في شكل صورة أو محادثة في شكل ملف مسموع (Sain, A., et al, 2024).

وتحدد نظم الاسترجاع عبر الوسائط مقاييس للكفاءة من خلال واجهة عرض النتائج؛ وتتبنى العديد من النظم المقاييس الكمية للنتائج المستدعاة، من حيث عدد كل وسيط على حدة، وعدد الوحدات ذات الوسائط المتعددة ذات العلاقات المعرفية أو المعلوماتية الواحدة، أيضا يمكن أن تكون المقاييس الكمية مهمة في عدد المحاولات المسجلة للنظام على استفسار محدد أو مرات التنقيح والتهديب والبحث داخل النتائج المسترجعة، وغالبا ما تحظى النتائج الكمية باهتمام مستخدمي نظم الاسترجاع خاصة في حالة البحث العام أو بداية مرحلة التعرف على الموضوع بجوانبه المختلفة، وتتبنى نظم الاسترجاع عبر الوسائط أيضا المقاييس الكيفية التي تركز على نوعية وسائط حمل المعلومات المسترجعة أو تلك التي تم استخدامها في البحث، وخرجت بها النتائج أو قدرة النظام على استدعاء وسائط متعددة أخرى مرتبطة بالمدى المعرفي وذات صلة قوية باحتياجات مستخدميه، وإن كانت مختلفة في النوع. (Bhattacharyya, A., et al, 2022).

2 / 2 / 2 دعائم بناء النظام:

يحتاج نظام الاسترجاع عبر الوسائط إلى مجموعة إضافية من مكونات العمل لتحقيق غايته بالاسترجاع الشمولي، وتدعم هذه المكونات الإضافية إجراءات العمل بأبعاد تقنية تحقق للنظام الانتقال من اشتقاق الملامح الوصفية للوسائط إلى التمثيلات الدلالية ومعالجتها وربطها معرفياً، ثم استخدامها في عمليات الاسترجاع، وتتمثل أهم دعائم عمل الاسترجاع عبر الوسائط في وجود قاعدة معرفية مرنة تستوعب العلاقات والبنى المفاهيمية بين التمثيلات الدلالية، بجانب تقنيات التفاعل السياقي وإعادة التركيب للغوي لمولدات النصوص GPT.

2 / 2 / 2 1 قاعدة المعرفة:

باتت المحادثات النصية المولدة جزء رئيسياً من بناء نظم استرجاع وإنتاج المعارف على الويب؛ ويعد محركا MSN و GOOGLE أول ما تبنى تقنيات المحادثات الاسترجاعية الحوارية Chat GPT من خلال واجهتي Gemini، Bing، وتعد قواعد المعرفة أساس العمل للمحادثات الاسترجاعية والممد لمولدات النصوص بالمقاطع المعرفية والكيانات المعلوماتية الدلالية؛ و تحتوي قاعدة المعرفة للنظام على مجموعات المحادثات والنصوص المجمعة والمولدة، بجانب الهيكلية المعرفية للمصطلحات والتوزيعات المفاهيمية التي تعمل على إدراك دلالة الكلمات ومعالجتها باللغة الطبيعية ثم ربطها بالعلاقات والروابط الموضوعية، وعلى المدى القريب ستكون المحادثات الحوارية جزء من نظم الاسترجاع عبر الوسائط، بما يجعل هناك قدرة للنظام على محاورة مستخدميه بالمقاطع والسياقات الصوتية والمصورة، بجانب المقاطع والسياقات النصية، ولجعل هذه المحادثات في مستوى الأداء الاسترجاعي الدقيق، فلا بد من بناء قاعدة معرفة متعددة الوسائط خلف المولدات الحوارية القادرة على إدراك السياقات المصورة والصوتية؛ بل والرد بما يوافقها من جانب النظام، وستكون قاعدة المعرفة إضافة لا غنى عنها لنظم الاسترجاع عبر الوسائط داعمة لتحليل التمثيلات الدلالية والشبكات المعرفية التي تربط بينها، سواء أتم دعمها بمولدات الوسائط الحوارية أم لا (Lerner, P., et al, 2024)

2 / 2 / 2 2 مولدات السياق متعدد الوسائط:

تسعى نظم الاسترجاع عبر الوسائط إلى تحقيق تفاعل مع مستخدميها يتوافق وطبيعة تفاعل البشر مع بعضهم البعض، باستخدام أشكال متعددة للتحوار تتمثل في الكلمات والإشارات والحركات والأصوات، لذا فيمكن إضافة جانب المحادثات الاسترجاعية لتملك النظم قدرة الرد بسياق محدد ومحاورة المستخدم، ولكن بوسائط متعددة وليست بشكل نصي فقط كما هو في مولدات السياق النصي Chat GPT، إن مولد السياق متعدد الوسائط Multimodal-GPT التابع لنظام الاسترجاع عبر الوسائط سيكون قادراً على إدراك المعلومات المصورة والصوتية بكفاءة بتقنيات رؤية الحاسبة، وباستخدام طبقات الإدراك متعددة المستويات والرابطة بين الوسائط المتعددة، ولتعزيز قدرة مولدات السياق متعدد الوسائط على الدردشة مع الأشخاص بشكل أكبر، يتم جمع بيانات اللغة وتحديد قالب تعليمات موحد لتدريب MultiModal-GPT بشكل يحسن التدريب المشترك للتعليمات اللغوية والمصورة والصوتية؛ وتستخدم التعليمات التي تغطي مجموعة واسعة من التطبيقات والمجالات، مما يسهل التطور الشامل لأداء مولدات الحوار متعدد الوسائط (Gong, T., et al, 2023)

المبحث الثالث: آليات البحث والاسترجاع الموحد عبر الوسائط:

تعتمد نظم الاسترجاع آليات وأساليب للبحث تختلف وفقاً لطبيعة المحتوى المعلوماتي موضع الاستدعاء، وقد تبنت نظم الاسترجاع جميعها أسلوب البحث الحر بالنص، كونه الأقرب لطبيعة مستخدم نظم الاسترجاع والأيسر تعبيراً عن احتياجاتهم المعلوماتية في شكل كلمات أو عبارات موجهة للنظام، وقد استمر استخدام البحث بالنص مستمراً إلى المراحل الأولى من ظهور محركات بحث الوسائط المتعددة، وتطور قدرات نظم الاسترجاع في ضم الوسائط المتعددة للمحتوى الرقمي

المسترجع، تم إضافة آليات بحث الوسائط المتعددة كالاستعلام بالصورة عن الصورة والبحث المنطوق للمعلومات الصوتية، اعتماداً على تقنيات الإدراك الرقمي لهذه الوسائط، وستبرز آليات بحث نوعية مع نظم الاسترجاع عبر الوسائط التي تتعدد اتجاهاتها بالبحث عن الوسائط المتعددة بأحدها، فيمكن استرجاع الصورة وملفات الفيديو باستعلام صوتي؛ بل وإمكانية المحاورة والدرشة مع النظام باستخدام الوسائط المتعددة المختلفة، وسيتناول هذا الجزء من الدراسة أساليب وآليات البحث ومعاملاتها التي تتوافق والاسترجاع عبر الوسائط (Gadre, S.Y., et al, 2023)

2 / 3 / 1 خصائص واجهة البحث:

تتفرد نظم الاسترجاع عبر الوسائط بخصائص عمل تجعلها في حاجة إلى واجهة بحث واسترجاع تتميز عن باقي واجهات نظم الاسترجاع أحادية الوسيط؛ وستحتاج هذه الواجهة إلى سمات عرض وآليات بحث وقدرات تفاعل مع مستخدم النظم تتوافق وتعدد الوسائط المعروضة عليها، وتتنوع هذه السمات بين الشكلية المدعومة بتقنيات العرض والإتاحة للصور والصوت والرسوم الحاسوبية، وبين الموضوعية الداعمة لاستدعاء احتياجات مستخدم النظام (Lin, X., et al, 2021)، ويمكن حصر خصائص بناء واجهة البحث في الجوانب التالية:

- تقديم مختلف أساليب التفاعل للمستخدم بما يمكنه من استخدام قدرات النظام في الاسترجاع عبر الوسائط، أو استخدام قدرات الاستكشاف لاختبار أولي لمحتويات النظام.
- تقديم مجموعة من مفاتيح العمل وقدرات الاسترجاع والتحكم بالبحث والنتائج المترتبة عليه، من تحديد الوسيط المسترجع أو الأسلوب المستخدم أو الجوانب الكمية للاسترجاع عبر الوسائط أو خصائص العرض للنتائج.
- القدرة على توفير خصائص العرض أحادي الوسيط بجانب العرض الشمولي إذا ما ناسب مستخدم النظام، وتحديد آليات العرض الشمولي وترتيب النتائج المعروضة وفقاً لترتيب مفضل لوسائط المعلومات.
- تحديد اتجاهات وحدود النتائج المسترجعة وعلاقاتها النوعية من حيث الجانب الدلالي أو الجانب الشكلي بين الوسائط المرتبطة المتجانسة وغير المتجانسة.
- إعادة ترتيب أفضل المقاطع التي تم الحصول عليها من البحث الدلالي بناءً على صلتها بوسيط الاستعلام، وإمكانية دمجها في مقطع أطول قبل عرضها بشكل متعدد الوسائط.

2 / 3 / 2 آليات البحث بالخصائص المادية للوسائط:

سيعتمد نظام الاسترجاع عبر الوسائط على بحث واستدعاء النتائج على خصائص المحتوى المصور والمتحرك والصوتي المادية، ويحقق ذلك كفاءة أعلى في استرجاع الوسائط المتعددة؛ ولم يكن البحث النصي عن الوسائط المتعددة من خلال بيانات الوصف الفني (الميتاداتا) يحقق أفضل مضاهاة بين احتياجات المستخدم والنتائج المسترجعة، وأضحت الخصائص المادية أكثر كفاءة في معالجة واسترجاع الوسائط المادية، نظراً للقدرات التقنية الداعمة لعمل نظم الاسترجاع، مع زيادة الوعي بآليات بحثها من جانب مستخدم نظم الاسترجاع، وتتمثل أهم الخصائص المستخدمة للبحث عن الوسائط المتعددة في التوزيع الرقمي الرسومي واللون والبنية للصورة الثابتة، وتداعي اللقطات وهيئة الحركة ومسار الصوت لمقاطع الفيديو، وتحليل الأطوال الموجية وخامة الصوت للمواد المعلوماتية الصوتية، وزادت قدرات النظام الداعمة للبحث بالخصائص المادية في إمكانية البحث بها بشكل منفرد، فيما يطلق عليه البحث أحادي الوسيط كاللون فقط أو الشكل الرسومي فقط مع الصور الثابتة والمتحركة، أو استخدام خصائص مدمجة أو مزوجة كالجمع بين اللون والحركة أو الشكل والصوت للبحث عن الفيديو، أو خامات الصوت والشكل الرسومي في الرسوم الحاسوبية، وذلك لاستدعاء أحد الوسائط فقط أو الاسترجاع الشمولي عبر الوسائط. (Yi, Z., et al, 2021)

2 / 3 / 2 / 1 البحث المادي أحادي الوسيط:

يتيح نظام الاسترجاع عبر الوسائط في هذا المستوى من البحث إمكانية البحث باستخدام خصائص مادية لوسيط واحد؛ كأن يتم استخدام اللون أو الشكل أو البنية لبحث واسترجاع الصور CBIR، أو استخدام خصائص المادة الصوتية للبحث عن المحتوى المعلوماتي المسموع، إن استرجاع الصور يهدف إلى إيجاد أكبر عدد ممكن من الصور المطابقة للمحتوى الرسومي ذات الدلالة الواحدة، خلال تحليل المعلومات المرئية للصورة، مما له قيمة كبيرة على درجة تحقيق النتائج المسترجعة، ومع ظهور تقنيات التعلم العميق، فقد تم تطبيق الميزات التي تم تعلمها من الشبكات العصبية على نطاق واسع على مختلف المهام المرئية؛ وتمتلك معظم الشبكات العصبية كمية كبيرة من الحسابات والمحارف الرقمية للتعامل مع سطح الصورة وتحليل جوانبه، ويحتاج الاسترجاع بالصور المبني على المحتوى إلى جانبين مهمين هما: بناء إطار عمل للاسترجاع يمكن استخدامه لتنفيذ الاسترجاع بالاعتماد على التمثيلات والرموز المحددة لجوانب المحتوى المصور، وأيضا تقوية نموذج التعلم العميق للحصول على ميزات الصورة، بحيث يتم إكمال استرجاع الصور عن طريق تحويل نموذج التعلم العميق إلى نموذج طرفي، وضمه إلى إطار عمل الاسترجاع (Hu, Z., et al, 2023) أما البحث المادي أحادي الوسيط بالفيديو فيعد أكثر تعقيداً مقارنة ببحث الصورة الواحدة؛ ويجمع محتوى الفيديو بين الحركة والصوت والنص، مستهدفاً استرجاع مقاطع الفيديو الأكثر ارتباطاً من قاعدة النظام، معتمداً فقط على محتوى الفيديو دون أية بيانات وصفية إضافية، وتتكون عملية استرجاع الفيديو القائم على المحتوى CBVR من ثلاث خطوات: الخطوة الأولى: هي أخذ عينات من الإطارات المصورة لمعالجة الفيديو بكفاءة وتعزيز أداء الاسترجاع، وخاصة إطارات الفيديو المحورية التي تحتوي على العناصر الأساسية للمقطع المصور، فقد تكون هناك إطارات متشابهة جداً في نفس سياق مدة معينة، وتكون مخرجات الخطوة الأولى هي مجموعة الإطارات الرئيسية ذات الدلالة على المحتوى، وتتمثل الخطوة الثانية في: استخراج الميزات والتمثيلات الرمزية؛ باستخراج مخطط بنية اللون للإطارات الرئيسية باستخدام تقنيات رؤية الحاسب التقليدية أو التعلم العميق، أما الخطوة الأخيرة: فهي حساب المسافة الرقمية التي تقارن الملامح المستخرجة من كل من مقاطع الفيديو في قاعدة النظام والفيديو موضع الاستعلام، من حيث طبيعة الخصائص المصورة أو مقياس تشابه التمثيلات الدلالية (Sowmyayani, S., et al, 2023).

وأما بالنسبة للاسترجاع المادي أحادي الوسيط بالصوت CBAR، فهو يمثل مفهوماً واسعاً لمعالجة المحتوى الصوتي، بما في ذلك معالجة الإشارات الصوتية الرقمية، والسمع، واللغويات، وتكنولوجيا معالجة الإشارات الصوتية والموسيقى، وتقنيات رؤية الحاسب، وتقنيات نظم الاسترجاع عبر الوسائط، وتعد تقنية تصنيف معلومات الصوت النقطة الأساسية لمعالجة معلومات الصوت، التي تملك مجموعة واسعة من التطبيقات في أنظمة معالجة الصوت والفيديو التفاعلية بالإضافة إلى تطبيقات الوسائط المتعددة الأخرى، وتعتبر طريقة مطابقة أنماط الصوت معقدة نسبياً ذلك لتعدد إجراءاتها بثلاث خطوات: تدريب نموذج النظام، ومطابقة الأنماط، واتخاذ قرار التطابق والاسترجاع للمواد الصوتية، ويتم استخدام تقنيات الشبكة العصبية التي تحاكي أساس النشاط العصبي البشري، مع القابلية للتكيف وتقادي الأخطاء وخصائص التعلم، ولديها قدرة تصنيف وترجمة قوية للأصوات التعرف على الكلام (Fan, L., 2020).

2 / 3 / 2 / 2 البحث المادي متعدد الوسائط:

تفوقت قدرات استرجاع المعلومات متعددة الوسائط المتداخلة على تحديات كل من نماذج الاسترجاع أحادية الوسائط ومزدوجة الوسائط، بما في ذلك الصور والنصوص، والنصوص والصوت، والصور والصوت، أو النصوص والصور والصوت، وقد تضمنت هذه القدرات بناء نموذج موحد للبحث يعتمد على الارتباطات الدلالية بين الوسائط المختلفة؛ ويتم تمثيل كل وحدة معلومات من الوسائط المتعددة (النص والصور والصوت) كتمثيل دلالي أو ترميز، بناءً على طبيعة المحتوى المتضمن في كل وحدة ومدى التوافق بينها وبين الوحدات الأخرى المتجانسة وغير المتجانسة في المحتوى، ويتم استخدام خوارزمية

لتحليل عناصر المحتوى تعتمد حساب أوجه التشابه بين الوسائط المختلفة في فضاء المكونات الرقمي، وهو ما يعكس اعتماد نظام البحث للوسائط المتعددة المتداخلة على الارتباطات التحليلية الإحصائية لمكونات الصوت والصورة والمقاطع المصورة، ذلك مما يجعله قادراً لاستعلام نصي لاسترجاع معلومات مصورة ذات صلة بموضوع النص، واستعلام مصور لاسترجاع مواد نصية وصوتية تتوافق وطبيعة الاستعلام المصور لمستخدم النظام، وتدل استطاعة نظام الاسترجاع عبر الوسائط بالبحث متداخل الخصائص على تحقق مفهوم التوحد الرقمي بين مواد المعلومات متعددة الوسائط وما تحمله من محتوى رقمي، أي: إذابة فوارق وشتات التنوع المادي للوسائط المتعددة لصالح تحرير المحتوى المعرفي الرقمي داخلها، والتوجه إلى جيل نظم استرجاع وإنتاج المعارف (Yu, J., & Qin, Z, 2015)

2 / 3 / 3 آليات البحث بالخصائص النصية الدلالية:

تتتمي آليات البحث الدلالي بشكل خاص إلى الوسيط النصي لحمل المعلومات، بطبيعة ما يتضمنه من مفردات لغوية حُصصت في بنائها وارتباطاتها لحمل الدلالات والمعاني، ويمثل البحث بالكلمات والنصوص أسلوب البحث الرئيسي المباشر لدى مستخدم نظم الاسترجاع النصية أو متعددة الوسائط، كون النص الأقرب لطبيعة المستخدم وقدرته على التعبير المباشر عن الاحتياجات المعلوماتية، وبرغم أن المستهدف هنا هو الاسترجاع عبر الوسائط إلا أن النص يظل أهم وسائط المحتوى المعرفي ووسيلته للمعالجة والاسترجاع، إلى جانب أن النص يمثل الأساس المشترك الذي تلقي عنده معالجة وسائط الصوت والصورة الثابتة والمتحركة، باستخدام تمثيلات دلالية بنيت على معاجم لغوية، أو تمثيلات دلالية معرفية نتجت عن شبكات العلاقات والارتباطات المفاهيمية داخل قاعدة المعرفة لنظام الاسترجاع عبر الوسائط، ولعل تقنيات معالجة اللغة الطبيعية هي الأكثر أهمية واستخداماً مع البحث الدلالي، ذلك لاهتمامها بأمرين أساسيين في البنية الدلالية وهما: كيفية استخدام المعلومات الموجودة في قاعدة المعرفة بشكل فعال بإدراك طبيعة النصوص واستيعاب ما بينها من جوانب تشابه النصوص واختلاف مفاهيمية، وكيفية نمذجة العلاقات الدلالية والمفاهيمية بين بنية المعرفة وبنية النص والتعبير عن ذلك بشبكات ربط بين التمثيلات الدلالية، ويتم بناء مجموعة من الرسوم البيانية الدلالية المتعددة تمثل الهيكل المعرفي العام لنظام الاسترجاع، ويحتوي كل منها على معلومات دلالية كُفِّدَ، ترتبط بالسّمات العلائقية بين كل عقدة وأخرى داخل الرسم البياني (Wang, B., et al, 2022)

يتم بناء مصفوفة التشابه الدلالي للكيانات، والمعرفي من الكلمات والمصطلحات باستخدام قدرات النظام على إعادة معالجة المعلومات في قاعدة المعرفة والمعارف الخارجية التي يتفاعل معها بشكل مستمر، والقدرة على استقطاب هذه المعارف ودمجها في الشبكة الدلالية الداخلية للنظام، ثم يعمل النظام على نمذجة العلاقة بين بنية المعرفة وبنية النص للمصفوفة الدلالية دائمة التوسع، من خلال خطوات ثلاث هي: تشفير وترميز المعارف المعالجة والمعارف الخارجية، ثم تحويلها إلى كيانات معرفية وتحديد علاقاتها بالكيانات الأخرى وفقاً للتشابه والاختلاف، ثم تضمين هذه الكيانات وتسكينها في الشبكة الدلالية للنظام، وتقسم نظم الاسترجاع عبر الوسائط المعالجة الدلالية للنصوص اعتماداً على تشابه الجمل القائم على الكلمات، وتشابه الجمل القائم على البنية، وطرق تشابه الجمل القائمة على المعاني، وفي التشابه الدلالي القائم على الكلمات، يتم التعامل مع الجملة كقائمة من الكلمات، ويتم مقارنة الارتباط بين الكلمات في الجملتين، أما في التشابه القائم على البنية، فيتم اقتراح العديد من الطرق التي تستخدم قواعد اللغة وعلامات تجزئة النص وترتيب الكلمات، وتعتمد طرق تشابه الجمل القائمة على المعاني على حساب المفاهيم الضمنية للجمل التي تصف توجه الجملة إلى تخصص أو مجال أو حدث محدد (Chen, Q., et al, 2023)

2 / 3 / 4 الدردشة عبر الوسائط:

لقد تمكنت تقنيات الذكاء الاصطناعي من إنشاء مساعد متعدد الاستخدامات قادراً على اتباع تعليمات متعددة الوسائط للرؤية واللغة تتوافق مع الرؤى البشرية، وقد انتقلت تقنيات الحوار الذكي لنظم استرجاع المعلومات من مستوى الحوار السياقي النصي GPT-Chat إلى مستوى الحوار السياقي متعدد الوسائط Multimodal-GPT، بحيث يمكن لنظم الاسترجاع التحوّل مع مستخدميها بمقاطع صوتية أو برسوم وصور تعبيرية أو بقطاعات حية حول موضوع ما، وأكثر من ذلك مرونة

النظم على دمج الوسائط المتعددة في سياق واحد كأن يرد على الصورة بالصوت أو الرد على النص بالفيديو، ويمكن لنظام الاسترجاع باستخدام مولد الوسائط الحوارية دمج مثل هذه الوسائط موفقا لرؤية الحاسب وتخصيص جوانب المحتوى الدال على الحدود المكانية والزمنية والموضوعية، من خلال توحيد الترميز والتمثيلات الدلالية في قاعدة المعرفة للنظام، ويمكن تصور Multimodal-GPT كنظام استرجاع مستقل، قادر على إنتاج وتوليد المعارف متعددة الوسائط، وأما دوره مع نظم الاسترجاع عبر الوسائط، فيعد واجهة أخرى بالنمط التفاعلي، ليست على غرار الاستفسار والنتائج وإنما على شاكلة الحوار، ويمكن لنظم الحوار عبر الوسائط المتعددة أن تكون أكثر نجاعة في إنتاج المعارف واسترجاعها بتوافر مجموعة من مقومات المعالجة الفنية لمحتوى الوسائط المتعددة، تتمثل في: (Gong, T., et al, 2023).

- دمج المعلومات الدلالية مع ميزات الصورة وملاحها المشتقة لإنتاج شبكة معرفية بين وحدات الصورة والصوت والفيديو مجتمعة، حتى تكون التمثيلات المرئية الدلالية داخل هذه الشبكة موزعة في اتجاهات ومجالات موضوعية متعددة.
- دمج نموذج إضافي لمولد الحوار عبر الوسائط يختص بتحليل المشاعر المتضمنة في محتوى الوسائط Emotional Generative Pretrained Transformer، بحيث يتيح لنا هذا التقاط المحتوى الشعوري الدقيق ومعالجة تمثيلاته باللغة الطبيعية، ويمكن لهكذا إضافة في معالجة وتحليل محتوى الوسائط تقديم أفضل النتائج المرجو استرجاعها من جانب مستخدمي النظام، بما يلائم احتياجاتهم وانفعالاتهم (Jin, Q., et al, 2024).

Please give a detailed description of this image



The image shows a beautiful lake surrounded by lush green grass and trees. The lake is filled with water lilies, creating a tranquil and serene atmosphere.

The lake is situated in a mountainous landscape, with a mountain peak visible in the background.

There are several birds flying around the lake, adding to the tranquility and natural beauty of the scene.

What is the season in the picture

The season in the image is summer, as evidenced by the lush green grass and the presence of water lilies in the lake.

شكل رقم (12) نموذج لمولدات الحوار عبر الوسائط MultiModal-GPT

يوضح شكل رقم (12) أن تقنيات رؤية الحاسبة قد دعمت مولدات الحوار متعدد الوسائط بوجود نموذجين للعمل هما: أحادي التدفق وثنائي التدفق، و تجمع النماذج أحادية التدفق بين الوسائط المرئية واللغوية والتحاور بوسيط أمام آخر، أما الأنظمة ذات التدفق المزدوج، فتعمل على تحليل داخلي للوسائط المتعددة، وتقديمها بشكل مدمج ومتعدد يمكن به استخدام أكثر من وسيط في آن واحد للرد على السياق، ومن ثم تستطيع نظم الاسترجاع عبر الوسائط الاستفادة من أسلوب الحوار متعدد الوسائط بجانب آليات البحث المستخدمة في النظام (Liao, H., et al, 2023)

ثالثاً: الإطار التطبيقي:

يختص الإطار التطبيقي للدراسة باستقراء واقع نظم استرجاع الوسائط المتعددة، لرصد درجة تحقق خصائص الاسترجاع عبر الوسائط، من خلال مجموعة من المعايير الدالة على توفر قدرات الاسترجاع عبر الوسائط في النظام، وسيكون الهدف من الدراسة تحليل النتائج الإيجابية بتوافر عناصر الاسترجاع عبر الوسائط، واستنباط ما يمكن تقديمه لمعالجة الجوانب السلبية والأسباب الكامنة وراءها، وستبنى الدراسة التطبيقية على اختيار عينة قصدية من محركات بحث الوسائط المتعددة وفقاً لأعلى ترتيب صادر عن موقعي Forbes Advisor and SEO PowerSuite للعام 2024م، وقد روعي في اختيار محركات عينة الدراسة كنظم استرجاع للمعلومات تتوفر بها قدرات استرجاع الوسائط المتعددة، وتقنيات البحث الذكية باعتماد معالجة اللغة الطبيعية، فضلاً عن التزود بآليات التحليل الدلالي للعناصر المادية للوسائط المصورة والصوتية، كما تم التركيز على النظم الأكثر اقتراناً بمولدات السياقات الحوارية النصية أو بالوسائط المتعددة، وسيبدأ استيفاء معايير الاسترجاع عبر الوسائط على عينة الدراسة ابتداءً من معايير الاختيار والاضافة لنوعية ملفات الوسائط المتعددة، مروراً بمعايير المعالجة ومقوماتها، ثم دعائم بناء نظام الاسترجاع عبر الوسائط وهيكلته، إضافة إلى آليات الاسترجاع عبر الوسائط ونمط البحث بالحوار متعدد الوسائط، ثم خصائص عرض وتقديم نتائج الوسائط المتعددة.

جدول رقم (1) عينة الدراسة التطبيقية من محركات بحث استرجاع الوسائط المتعددة

م	نظام الاسترجاع	جانب التميز
1	Google	: Best Overall
2	Bing	: Best for Multimedia Searches
3	DuckDuckGo	: Best for Privacy
4	Ecosia	: Best for Environmental Impact
5	Yahoo	: Best for Diverse Content
6	Ask.com	: Best for Unique Results

سيتم قياس واقع الاسترجاع عبر الوسائط من خلال المعايير التالية:

3 / 1 معايير المعالجة الكُليّانية لدى عينة الدراسة:

ستبدأ معايير تقييم أداء نظم استرجاع عينة الدراسة بجانب المعالجة وآليات العمل التي يترتب عليها آليات البحث والاسترجاع لاحقاً، وذلك من حيث طبيعة المحتوى الرقمي للوسائط المتعددة في قاعدة النظام ونوعية الملفات التي يتعامل معها، ثم آليات المعالجة الفنية للوسائط وتحليل المحتوى.

3 / 1 / 1 المحتوى متعدد الوسائط للنظام:

يتنوع المحتوى الرقمي متعدد الوسائط بين المواد الصوتية والصورة والنصية، بحيث يحمل معه كل نوع من خصائص وطبيعة تستلزم قدرات عمل محددة، وسيقيم هذا المعيار أشكال الوسائط المتعددة، ومستوى معالجة النظام لهذه الأنواع إذا كان فردياً أو مزدوجاً، ثم نمط المعالجة المبني على تحليل المحتوى أو الخصائص المادية لوسيط حمل المعلومات، ويسهم ذلك في قياس مدى قابلية عينة الدراسة للتحويل لمستوى نظم الاسترجاع عبر الوسائط.

جدول رقم (2) تقييم معايير المحتوى عابر الوسائط للنظام

نظم الاسترجاع عبر الوسائط						المحتوى عابر الوسائط للنظام	
متعددة الوسائط			الذكية				
Ask.com	Ecosia	Yahoo	DuckDuckGo	Bing	Google		
√	√	√	√	√	√	النص	أنواع المحتوى متعدد الوسائط
√	√	√	√	√	√	الصوت	
√	√	√	√	√	√	الصورة	
√	√	√	√	√	√	الفيديو	
√	√	√	√	√	√	بالملف	درجة التعامل مع المحتوى متعدد الوسائط
---	---	---	√	√	√	بالخصائص المادية	
---	---	---	---	√	√	بالإدراك	
√	√	√	√	√	√	فردى	ازدواجية المحتوى
---	---	---	---	---	---	مزدوج	

يتضح من الجدول السابق الملاحظات التالية:

- اعتمدت الدراسة على عينة من محركات بحث الوسائط المتعددة التي تحتوي بالفعل على مختلف وسائط حمل المعلومات، حتى تكون متوافقة وأهداف الدراسة التطبيقية، ورغم هذا الاتفاق في نوعية الوسائط إلا أنها تباينت تماماً في نمط المعالجة والاسترجاع وآلياتهما مع الوسائط المتعددة، وهو ما يجعل مفردات العينة تنتمي لأجيال مختلفة من استرجاع الوسائط.
- تعاملت كل محركات عينة الدراسة وفقاً لملف الوسيط الذي يشمل مادة المعلومات والبيانات الوصفية المصاحبة لها، ما عزى إلى استخدام آليات الاسترجاع النصية المتعارف عليها بالبحث النصي المباشر، ولم يكن لقدرات الإدراك والتحليل الرقمي لخصائص مادة الوسيط المصور أو المسموع أو المتحرك نصيباً إلا مع نظامي Google, Bing، وذلك لما لهما من إمكانات ذكية في تحليل المحتوى الرقمي متعدد الوسائط.
- لم تستطع كل محركات بحث العينة التعامل مع المحتوى المزدوج لوسائط المعلومات الذي يجمع بين وسيطين في عملية بحث واحدة، رغم وجود نظامي Google, Bing اللذين يتمتعان بقدرات ذكية، لكنها تتعامل مع الوسائط فردياً لكل وسيط على حدة.

3 / 1 / 2 المعالجة الفنية والتنظيم عبر الوسائط:

تعكس طبيعة المعالجة الفنية والتنظيم داخل نظام استرجاع الوسائط درجة وقدرة النظام على التعامل مع المحتوى الرقمي أو التعامل مع الخصائص المادية لوسيط المعلومات، ومدى استيعاب النظام لتحليل جوانب المحتوى المتضمن في

وسيط المعلومات والتعبير عنه، أو اعتماد نظام الاسترجاع على البيانات الوصفة المرافقة لوسائط المعلومات المتعددة، وما يتطلبه ذلك بشكل عام من دعم تقني وقدرات وآليات عمل نوعية.

جدول رقم (3) تقييم معايير المعالجة الفنية والتنظيم عبر الوسائط

نظم الاسترجاع عبر الوسائط						المعالجة الفنية والتنظيم عبر الوسائط	
متعددة الوسائط			الذكية				
Ask.com	Ecosia	Yahoo	DuckDuckGo	Bing	Google		
√	√	√	√	√	√	لنص	توجه المعالجة
---	---	---	---	√	√	للخصائص المادية	
---	---	---	---	√	√	للخصائص الدلالية	
√	√	√	√	√	√	أحادية الوسيط	مستوى المعالجة للوسائط
---	---	---	---			مزودة الوسائط	
√	√	√	√	√	√	البيانات الواصفة	الآليات الداعمة
---	---	---	---	√	√	رؤية الحاسب	

يتضح من الجدول السابق المشاهدات التالية:

- حددت تقنيات الإدراك والمعالجة الذكية قدرات عينة الدراسة في التعامل مع وسائط المعلومات معالجةً وتحليلاً؛ واتبعت أغلب محركات العينة الوسيط النصي كوسيط معالجة الوسائط المتعددة والتعبير عنها في عمليات التنظيم، ومن ثم البحث والاستدعاء، ولم يتمكن من إثبات قدرات التحليل الآلي وإدراك مادة المعلومات سوى نظامي Google, Bing؛ وتبين من واقع الاسترجاع للمواد المصورة قدرة النظامين على قراءة السطح المصور لمادة المعلومات، وإدراك كل جوانب بنائها المصور واستخدامه للبحث مرة أخرى.
- تمكن المعالجة الفردية من تحليل مادة الصورة تماماً إلا أنه يظل وسيطاً مفرداً، أما المعالجة الفنية مزدوجة المستوى، فهي قدرة النظام على التعامل مع وسيطين معا بواقع حالهما كالصوت المصاحب لملف الفيديو، أو بتعمد دمجهما كالنص مع الصورة أو الصوت مع الصورة، وهو المستوى الأعلى من المعالجة المنشود في هذه الدراسة، بما يدعم فكرة وجود نظم الاسترجاع عبر الوسائط.
- تميز نظاما Google, Bing بما يملكان من آلية تحليل رقمية ذكية كتقنية رؤية الحاسب معالجة وتنظيم مختلف وسائط المعلومات، اعتماداً على تحليل مادة المحتوى وخصائص الوسيط، وهو ما لم تصل إليه باقي محركات بحث العينة.

3 / 2 مقومات نظام الاسترجاع الشمولي عبر الوسائط:

تحتاج نظم استرجاع الوسائط المتعددة إلى مقومات عمل نوعية لتمكين نظام الاسترجاع من إدراك المحتوى الرقمي متعدد الوسائط، وإدراكه والتعبير عنه بتمثيلات أو رموز دلالية تعكس المحتوى المعلوماتي أو الخصائص المادية للوسيط،

وتعد هذه الأدوات الميزات النسبية التي ستملكها نظم الاسترجاع عبر الوسائط، فإذا ما توفرت كان النظام على قدرة الاسترجاع الشمولي للوسائط المتعددة.

جدول رقم (4) تقييم مقومات نظام الاسترجاع الشمولي عبر الوسائط

نظم الاسترجاع عبر الوسائط						مقومات الاسترجاع الشمولي عبر الوسائط	
متعددة الوسائط			الذكية				
Ask.com	Ecosia	Yahoo	DuckDuckGo	Bing	Google		
---	---	---	√	√	√	نصية	قاعدة المعرفة
---	---	---	√	√	√	وسائط	
---	---	---	√	√	√	المعالجة	الشبكات الدلالية
---	---	---	√	√	√	البحث	
---	---	---	√	√	√	نصية	مولدات السياق
---	---	---	---	---	---	وسائط	
---	---	---	---	√	√	المعالجة	معالجة اللغة الطبيعية
---	---	---	---	√	√	البحث	

يتضح من الجدول السابق ما يلي:

- إن الوصول إلى مستوى الاسترجاع عبر الوسائط لا يتم فقط باتباع قواعد وتوجهات نوعية في معالجة المعلومات؛ بل بالتمكن من وجود قدرات ذكية ورؤية حاسوبية تدعم تحليل المحتوى والتعبير عنه بنمط الربط المفاهيمي.
- إن قاعدة المعرفة والشبكات الدلالية والعلاقات المفاهيمية هي قدرات لا يمكن تحقيق واقع الاسترجاع عبر الوسائط بدونها، وسيكون نمط المعالجة والاسترجاع نوعياً يفوق المتبع الآن في نظم استرجاع المعلومات التي تتعامل مع الألفاظ والموضوعات بما يربطها من إحصاءات وليس على نطاق المفاهيم والدلالات.
- إن استخدام معالجة اللغة الطبيعية هو لب عمل نظم الاسترجاع عبر الوسائط، ذلك لكون المحتوى الرقمي لكل المواد متعددة الوسائط سيتخلى عن معالجة مادة المعلومات لصالح المحتوى المتضمن بها، أي: إذابة كل مواد المعلومات لاستخلاص المحتوى المعرفي المعلوماتي المتمثل في شكل دلالات أو مفاهيم نصية لقاعدة المعرفة، أو رمزية بقراءة النظام.

3 / 3 آليات الاسترجاع الموحد عبر الوسائط:

تكتسب آليات البحث أهمية كبيرة لمستخدمي نظم الاسترجاع لما لها من تأثير على مستوى النتائج المسترجعة وتوافقها مع احتياجات المستخدم، وستختلف آليات البحث الشمولي عن تلك المستخدم لبحث الوسائط الفردية، ولا ينظر لمدى تأثير هذه الآليات فقط داخل النظام، ولكن لقدرة المستخدم على استخدامها وتطويعها لاسترجاع النتائج المرجوة الشمولي عبر الوسائط.

3 / 3 / 1 آليات بحث الوسائط المتعددة:

تمثل آليات بحث الوسائط المتعددة الجانب الشكلي من التعامل مع الوسائط المتعددة، ويسهل على مستخدم النظام تصفح مجموعات الصور، أو استخدام النصوص لوصفها، أو البحث بصورة مقارنة لما يرغبه من نتائج، وتحتفظ هذه الآليات بأسلوب التفاعل البحث والاسترجاع المستقل لكل عملية بحث، وي طرح المستخدم استفساره ثم يتلقى النتائج فيقبلها أو يغير إستراتيجية البحث ليحتل على نتائج أخرى في عملية بحث مستقلة عن العملية الأولى.

جدول رقم (5) تقييم معايير آليات بحث الوسائط المتعددة

نظم الاسترجاع عبر الوسائط						آليات بحث الوسائط المتعددة	
متعددة الوسائط			الذكية				
Ask.com	Ecosia	Yahoo	DuckDuckGo	Bing	Google		
√	√	√	√	√	√	مستقل	البحث بالنص
---	---	---	---	√	√	مصاحب	
√	√	√	√	√	√	أحادي الوسيط	البحث بالوسائط
---	---	---	---	---	---	مزدوج الوسيط	
---	---	---	---	√	√	اجمالي المحتوى	مستوى البحث
---	---	---	---	√	√	أجزاء المحتوى	

يتضح من الجدول السابق الملاحظات التالية:

- مازالت محركات بحث الوسائط المتعددة تعتمد على النص كوسيط لمعالجة هذه الوسائط، رغم ما بها من تقدم في معالجة مادة الوسائط وخصائصها، وذلك لما للنص من قابلية مع عموم مستخدمي النظام، ويتمثل الوسيط النصي في البحث هنا في الكلمات الكشفية أو الواصفات الموضوعية والبيانات التي تصاحب مادة المعلومات متعددة الوسائط.
- تتماثل قدرات بحث واسترجاع الوسائط المتعددة مع قدرات المعالجة والتحليل في النظام؛ وقد أثبتت الدراسة التطبيقية لمحركات عينة الدراسة اعتماد عينة الدراسة على التعامل مع الوسائط المفردة المستقلة دون الوسائط المزدوجة المندمجة أو المتكاملة، وقد انعكس ذلك على واقع آليات بحث الوسائط لعينة الدراسة في استرجاع وسائط مفردة وليست مزدوجة.
- مكنت القدرات الذكية لنظامي Google, Bing من التعامل مع مواد الوسائط المتعددة على مستوى إجمالي المادة المعلوماتية أو جزء منها، وقد بدا ذلك جليا في استخدام صورة ثابتة للبحث عن الصور داخل النظام، فقد كان يمكن البحث بالصورة كاملة أو بجزء منها واسترجاع ما يناسب هذا الجزء بشكل محدد من إجمالي المواد المصورة داخل النظام.

3 / 2 آليات البحث بالحوار عبر الوسائط:

تمثل آليات الحوار واستدعاء المعلومات عبر التفاعل المباشر في اتجاهين بين النظام ومستخدمه سيكون له أثره على آليات البحث والاسترجاع عبر الوسائط، خاصة في حال تم استخدام الحوار متعدد الوسائط، كالتفاعل بنماذج مصورة أو التحوار بمقاطع صوتية، وقد تم اختبار هذه القدرات في عينة الدراسة، لتوضيح موقف نظم استرجاع الوسائط من استخدام هذه الآليات.

جدول رقم (6) تقييم معايير آليات البحث بالحوار عبر الوسائط

نظم الاسترجاع عبر الوسائط						آليات البحث بالحوار عبر الوسائط	
متعددة الوسائط			الذكية				
Ask.com	Ecosia	Yahoo	DuckDuckGo	Bing	Google		
---	---	---	---	---	---	جزء من النظام	درجة التكامل
---	---	---	---	---	---	نظام فرعي	
---	---	---	√	√	√	نظام مستقل	
---	---	---	√	√	√	نص	أشكال السياقات
---	---	---	---	---	---	صورة	

---	---	---	---	---	---	صوت	
---	---	---	---	---	---	فيديو	
---	---	---	√	√	√	أحادي	مسارات الحوار
---	---	---	---	---	---	مزدوج	
---	---	---	---	---	---	للسياق فقط	حالة الوسائط
---	---	---	---	---	---	مصاحب للوصف	

يتضح من الجدول السابق الملاحظات التالية:

- إن مولدات الحوار قد خرجت للتعامل مع المحتوى الرقمي بشكل مستقل عن نظم الاسترجاع، وذلك كونها تمثل مفهوم إنتاج وتوليد المعارف المشتقة من وحدات المعلومات المدخلة مسبقاً، وهي لا تنتمي لأجيال نظم استرجاع المعلومات، وإنما تمثل نظاماً نوعية تستهدف إدارة وإنتاج المعارف الرقمية، لذلك فإن من الطبيعي أن تكون واقعية مستقلة عن نظم الاسترجاع، حتى مع تلك الشركات التي قامت على أساس بحث واسترجاع المعلومات كما في نظم Google, Bing, DuckDuckGo.
- يؤكد واقع نظم الاسترجاع عبر الوسائط أن فكرة الاسترجاع الشمولي متعدد الوسائط لن يكون في أعلى مستوياته إلا بالاندماج مع آليات عمل مولدات الحوار، ذلك لما لها من آليات نوعية في المعالجة والاستدعاء لوسيط المعلومات تتوافق مع سياق محددة له بداية ونهاية وتطور في الرد والرد المقابل.
- أثبتت نتائج واقع عمل نظم الاسترجاع الذكية في عينة الدراسة أنها مازالت تقف على أدنى مستويات استدعاء واسترجاع المعلومات عبر مولدات الحوار؛ ذلك لأنها أولاً: مازالت تعتمد الوسيط النصي فقط وليس المتعدد للتجاوز بين مستخدم النظام وبين مولدات السياق، ثانياً: تعالج وتحلل المحتوى المعلوماتي المعرفي الرقمي بمعزل عن قاعدة المعرفة المخصصة لمولدات الحوار، وهو ما يجعل إدخال الوسائط المتعددة الفردي أو المزدوج ضمن سياقات الحوار مع مستخدم النظام أمراً بعيداً وأسلوباً غير متبع حتى الآن.

3 / 4 واجهات البحث والاسترجاع عبر الوسائط:

تعكس واجهات نظم الاسترجاع أسلوب وأداء النظم في المعالجة والبحث؛ وتتضمن واجهة النظام التعريف بما يحويه النظام من محتوى معلوماتي متعدد الوسائط، فضلاً عن القدرات المتاحة لمستخدميه في طلب واسترجاع ما يحتاجونه من محتوى، وتتبدل الواجهة في عملية الاسترجاع عنها في إجراءات البحث، فتحتوي واجهة عرض النتائج على الوحدات المتصلة بمحتوى أو وسائط البحث المطلوبة، أما واجهة البحث فتتيح للمستخدم أدوات التحديد والتدقيق في طبيعة النتائج المرغوبة.

3 / 4 / 1 واجهة نظام البحث عبر الوسائط:

تعتبر واجهة البحث عبر الوسائط عن كل ما يملكه نظام الاسترجاع من قدرات الاسترجاع الشمولي الكلي المعبر عن جيل جديد من نظم الاسترجاع عابر للوسائط المتعددة، ويتوقع من هذه الواجهات أن تكون على الشاكلة الذكية في التفاعل والإدراك وسهولة الاستخدام.

جدول رقم (7) قياس معايير واجهة نظام البحث عبر الوسائط

نظم الاسترجاع عبر الوسائط						واجهة نظام البحث عبر الوسائط	
متعددة الوسائط			الذكية				
Ask.com	Ecosia	Yahoo	DuckDuckGo	Bing	Google		
---	---	---	---	---	---	متوافق	قالب العرض للنظم
√	√	√	√	√	√	متفرد	
√	√	√	√	√	√	فردى للوسيط	بناء الواجهة

شمولي للوسائط	---	---	---	---	---	---
شمولي للحوار	---	---	---	---	---	---
متضمنة بالنظام	---	---	---	---	---	---
غير مضمنة	---	---	---	---	---	---
تفاعلية	---	---	---	---	---	---
غير تفاعلية	---	---	---	---	---	---

يتضح من جدول واجهة البحث ما يلي:

- نظم استرجاع الوسائط عينة الدراسة قد اعتمدت على قالب موحد متمثلاً في واجهة نصية يعكس هوية وطبيعة المستخدم أكثر ما يعكس هوية وواقع النظام في استرجاع الوسائط المتعددة، فقد استخدمت نظم عينة الدراسة القالب المبني على البحث الحر بالنص والتعبير عن الوسائط المتضمنة في قاعدة النظام بمسمياتها النصية.
- الدخول على بحث واسترجاع كل وسيط يكون منفرداً مستقلاً عن الوسائط الأخرى، وليس متوافقاً مع ما يمكن أن تكون عليه واجهة الاسترجاع الشمولي للوسائط المتعددة الجامعة بين البحث بالخصائص المادية لوسيط أو مجموعة وسائط معاً، مبنية بالشكل التفاعلي الذكي، ولم تتضمن واجهة الاسترجاع أيضاً قدرات البحث بالحوار المخصص للاسترجاع الشمولي للوسائط.
- التقنيات التفاعلية الداعمة لعرض وتمثيل الوسائط والمعبرة عن هوية الاسترجاع الشمولي للوسائط المتعددة لم تتضمنها واجهات بحث عينة الدراسة، وهو ما يتوافق وطبيعة انتمائها لجيل متأخر من نظم الاسترجاع يحتاج إلى التحديث والتطوير ليصبح قادراً على البحث الشمولي للوسائط المتعددة، وتحليل وإدراك مادتها الشكلية وخصائصها الرقمية، ولم تظهر التقنيات التفاعلية إلا في نظامي Google, Bing لما لهما من صفة استخدام تقنيات الذكاء الاصطناعي.

3 / 4 / 2 واجهة النتائج المسترجعة عبر الوسائط:

تتطلب واجهات عرض نتائج الوسائط المتعددة التحلي بخصائص شكلية تفاعلية تعكس طبيعة وحدات المعلومات الشكلية أو الصوتية المسترجعة، وهو ما لا يعد من الاستيفاء الشكلي لبناء واجهة النظام، وهو من تحقيق كفاءة عرض المعلومات النوعية كالتحكم في أبعاد الصور الثابتة والمتحركة، وفحص جوانبها وإعادة البحث بإحدى النتائج المسترجعة أو بجزء منها لتدقيق الاحتياجات المعلوماتية لمستخدم النظام.

جدول رقم (8) تقييم معايير واجهة النتائج المسترجعة عبر الوسائط

نظم الاسترجاع عبر الوسائط						واجهة النتائج المسترجعة عبر الوسائط	
متعددة الوسائط			الذكية				
Ask.com	Ecosia	Yahoo	DuckDuckGo	Bing	Google		
√	√	√	√	√	√	فردى	عرض النتائج
---	---	---	---	---	---	شمولى	
---	---	---	---	---	---	متكامل	
√	√	√	√	√	√	للمحتوى	ترتيب النتائج
---	---	---	---	---	---	للوسيط	
---	---	---	---	---	---	قابل للتحديد	التحكم بالوسائط المسترجعة
√	√	√	√	√	√	وفقا للبحث	
√	√	√	√	√	√	وسائط مستقلة	مستوى النتائج
---	---	---	---	---	---	وسائط مزدوجة	
						وسائط متعددة	نمط العرض

---	---	---	---	---	---	بانورامي متكامل	التحكم المسبق بالنتائج
√	√	√	√	√	√	الكمي	
√	√	√	√	√	√	النوعي	

يوضح الجدول السابق ما يلي:

- نظم استرجاع الوسائط عينة الدراسة جميعها قد اتسق في خصائص عرض النتائج المسترجعة من الوسائط المتعددة، فلم يتبنَ أحدها العرض الشمولي متعدد الوسائط الجامع لوسائط الصوت والصورة والفيديو معا، أيضا لم يصل أي منها إلى اعتماد الشكل المتكامل لعرض الوسائط بحيث يتم عرض الوسائط التي تعرض جوانب الاحتياجات المعلوماتية وفقا لقيمة التعبير عنها.
- نظم الاسترجاع عينة الدراسة قد اتفقت جميعا في نمط وأسلوب عرض نتائج الوسائط المتعددة وفقا للصلة بالموضوع المتعلق بالمحتوى المعلوماتي، وليس وفقا للتحديد المسبق لتقديم عرض وسيط على آخر، وهذا الأمر يرجع إلى أحادية البحث بالوسيط، وليس البحث مزدوج أو شامل الوسائط.
- طمحت الدراسة التطبيقية في استكشاف نمط العرض البانورامي المستخدم لخصائص الوسائط المتعددة الفنية الشكلية، مما يضيف على استرجاع النتائج قبول الرؤية والتفاعل الإنساني مع مستخدمي النظام، وهو أمر أقرب ما يكون لقدرات نظم الاسترجاع عبر الوسائط المزودة بتقنيات العرض الذكي، إلا أن واقع الاسترجاع للوسائط المتعددة قد اعتمد العرض التقليدي المتداعي لوحدة المعلومات بحسب ملاءمتها للاحتياجات الموضوعية، وذلك على مستوى وحدات الواحد.

رابعا: رؤى ونتائج الدراسة:

- (1) تمثل نظم الاسترجاع الشمولية عبر الوسائط نتائج دمج قدرات الذكاء الاصطناعي مع آليات استرجاع المعلومات، بما يؤثر مباشرة على تحليل وإدراك المحتوى متعدد الوسائط، والتعامل معه في بيئة ديناميكية، وهو ما سيجعل نظم الاسترجاع عبر الوسائط كيانات ذكية قادرة على إدارة وتقديم معارف جديدة من محتوى الوسائط المتعددة.
- (2) إن نظم استرجاع الوسائط المتعددة لم تفارق موقعها الآني بآلياتها المتبعة في معالجة وتحليل واسترجاع الوسائط المتعددة، للتحويل إلى نظم شمولية يمكنها تحقيق المعالجة الفنية الكليانية لوسائط حمل المعلومات جميعها، والعمل بآليات بحث واسترجاع موحدة تستهدف المحتوى متعدد الأشكال.
- (3) تمثل نظم الاسترجاع عبر الوسائط أهم الكيانات الرقمية لإدارة المعلومات وإنتاجها متعدد الوسائط، بما ستملكه من تقنيات الإدراك الدلالي للمحتوى ومعالجته باللغة الطبيعية، ثم بناء الشبكات الدلالية والمفاهيمية، بجانب تقنيات الإدراك الآلي لخصائص مادة الوسائط وتمثيلها رمزيا، والتحكم بها تنظيميا واسترجاعا.
- (4) إن نظم الاسترجاع الكليانية عبر الوسائط تعد أول مستويات نظم إنتاج المعرفة الشمولية التفاعلية في البيئة الرقمية، وهو أمر ليس بعيدا عن الوقت الحالي، وسيدعم تطور تقنيات الذكاء الاصطناعي وتطبيقاتها في مجال معالجة المعلومات، وإدراك المحتوى الرقمي ظهور وتطور نظم معرفية قادرة على تجزئة وبناء الكيانات الموضوعية والدلالية لإتاحتها، بما يتوافق والاحتياجات المعرفية لمستخدمي النظم.
- (5) تحققت رؤية الرقمنة الكلية الشاملة للمحتوى الرقمي متباين ومتعدد الوسائط والأشكال، لتتطلق نظم استرجاع قادرة على توليد المعلومات المعنية بالمحتوى دون التوقف عن شكل ومادة المعلومات المتوافقة، فيما يمكن أن يطلق عليه انصهار مواد المعلومات لصالح المحتوى الواحد.

- (6) استهدفت الدراسة استكشاف آفاق أبعد لعلم المعلومات، تجعله يتحول من علم إدارة المادة المعلوماتية لصالح تمثيل وإنتاج المحتوى المعرفي، بما ستؤطر له الدراسات القادمة عن المعالجة الشمولية للمعلومات، واستخدامها في إنتاج وتوليد المعارف.
- (7) قدمت الدراسة رؤية متكاملة لبنية وآليات عمل نظم الاسترجاع عبر الوسائط كاستشراف لجيل من نظم الاسترجاع، يمثل نقلة نوعية من مستوى معالجة، وتنظيم مواد المعلومات إلى إدراك وتشكيل المحتوى، ومن البحث والاسترجاع بقوالب الوسائط إلى التمازج بالسياقات المعلوماتية المتعددة.
- (8) أثبتت الدراسة التطبيقية أن نظم استرجاع المعلومات تتعامل وفقا للمستوى الأدنى من تلبية احتياجات مستخدميها في بحث واسترجاع مواد الوسائط المتعددة، باعتبار المستوى العام في استرجاع المعلومات بآليات تناسب قدراتهم، في حين تلك النظم يمكنها امتلاك قدرات التحول من استرجاع المعلومات إلى توليد المعارف متعددة الوسائط
- (9) أثبتت الدراسة التطبيقية صحة فرضية الدراسة التي انطلقت من عدم تأهل نظم استرجاع المعلومات الآنية للتحول إلى نظم الاسترجاع عبر الوسائط، ذلك كونها لم تملك قدرات أو آليات العمل المنشودة للتعامل مع محتوى الوسائط المتعددة وإدراك ومعالجته معرفياً.
- (10) خرجت نتائج الملاحظات والمشاهدات المباشرة لأداء نظم استرجاع عينة الدراسة، وتعاملها مع البحث والاسترجاع عبر لصالح نظامي جوجل وبنج؛ ذلك كون هذين النظامين هما الأوفر نصيب في امتلاك دعائم ومقومات الاسترجاع عبر الوسائط، ابتداءً من إدراك المحتوى والوسيط والتمثيل الدلالي للوسائط وبناء قواعد معرفة واستخدام معالجة اللغة الطبيعية، وتقنيات رؤية الحاسب في تمثيل المحتوى المصور والصوتي.
- (11) أثبتت نتائج الدراسة التطبيقية تأخر الوسيط المصور المتحرك (الفديو) عن وسيطي الصورة الثابتة، والصوت في اكتمال عمليات المعالجة الرقمية لمادة المحتوى المصور المتحرك، والتعامل معها إجمالاً في عينة الدراسة وفقاً للمادة والبيانات الواصفة وليس وفقاً للمادة المعلوماتية.
- (12) لقد فرضت البيئة الرقمية التي تقوم على إدارة القدر الأكبر من المحتوى المعلوماتي المعرفي للعالم، مجموعة من القدرات التقنية الداعمة لنظم استرجاع المعلومات، وتمثل القدرات التقنية الميزات النسبية لجيل من نظم استرجاع المعلومات على جيل آخر؛ وستكون نظم الاسترجاع الشمولية عبر الوسائط الأكثر تميزاً والمختلفة نوعياً عن أجيال سابقة بما تملكه من قدرات إدارة المحتوى الموحد على اختلاف مستوياته ووسائطه، وبناء كيان دلالي معرفي كلي.

خامساً: مصادر ومراجع الدراسة:

- Adjetey, C., & Adu-Manu, K.S. (2021). Content-based Image Retrieval using Tesseract OCR Engine and Levenshtein Algorithm. cited at https://www.researchgate.net/profile/Charles-Adjetey/publication/353684209_Content-based_Image_Retrieval_using_Tesseract_OCR_Engine_and_Levenshtein_Algorithm/links/6126069b1e95fe241af60242/Content-based-Image-Retrieval-using-Tesseract-OCR-Engine-and-Levenshtein-Algorithm.pdf
- Arora, P., Pehlivan, S., & Laaksonen, J. (2024). Text-to-Multimodal Retrieval with Bimodal Input Fusion in Shared Cross-Modal Transformer. *International Conference on Language Resources and Evaluation*. cited at <https://aclanthology.org/2024.lrec-main.1374.pdf>
- Banar, N., Daelemans, W., & Kestemont, M. (2021). Multi-modal Label Retrieval for the Visual Arts: The Case of Iconclass. *International Conference on Agents and Artificial Intelligence*. cited at <https://repository.uantwerpen.be/docstore/d:irua:5138>

- Beltrán, L. V. B., Caicedo, J. C., Journet, N., Coustaty, M., Lecellier, F., & Doucet, A. (2021). Deep multimodal learning for cross-modal retrieval: One model for all tasks. *Pattern Recognition Letters*, 146, 38-45. cited at <https://www.sciencedirect.com/science/article/am/pii/S0167865521000908>
- Bhattacharyya, A., Mauceri, C., Palmer, M., & Heckman, C. (2022). Aligning Images and Text with Semantic Role Labels for Fine-Grained Cross-Modal Understanding. *International Conference on Language Resources and Evaluation*. cited at <https://aclanthology.org/2022.lrec-1.528.pdf>
- Bin, Y., Li, H., Xu, Y., Xu, X., Yang, Y., & Shen, H.T. (2023). Unifying Two-Stream Encoders with Transformers for Cross-Modal Retrieval. *Proceedings of the 31st ACM International Conference on Multimedia*. cited at <https://arxiv.org/pdf/2308.04343>
- Bommisetty, R.M., Ponnusamy, P., Sekhar, H.H., & Khare, A. (2022). Content based video retrieval using dynamic textures. *Multimedia Tools and Applications*, 82, 59-90. <https://link.springer.com/article/10.1007/s11042-022-13086-6>
- Chandio, A. A., Asikuzzaman, M. D., Pickering, M. R., & Leghari, M. (2022). Cursive text recognition in natural scene images using deep convolutional recurrent neural network. *IEEE Access*, 10, 10062-10078. cited at <https://arxiv.org/pdf/2103.15348>
- Chebil, W., Wedyan, M.O., Alazab, M., Alturki, R., & Elshaweesh, O.G. (2023). Improving Semantic Information Retrieval Using Multinomial Naive Bayes Classifier and Bayesian Networks. *Inf.*, 14, 272. cited at <https://www.mdpi.com/2078-2489/14/5/272>
- Cheligeer, C., Yang, J., Bayatpour, A., Miklin, A., Dufresne, S., Lin, L.F., Bhuiyan, N., & Zeng, Y. (2022). A Hybrid Semantic Networks Construction Framework for Engineering Design. *Journal of Mechanical Design*. cited at <https://ieeexplore.ieee.org/iel7/4235/4358751/09696004.pdf>
- Chen, Q., Zhao, G., Wu, Y., & Qian, X. (2023). Fine-grained semantic textual similarity measurement via a feature separation network. *Applied Intelligence*, 1-14. cited at <https://link.springer.com/article/10.1007/s10489-022-04448-6>
- Devillers, B., Mayti'e, L., & Rullen, R.V. (2023). Semi-supervised Multimodal Representation Learning through a Global Workspace. *IEEE transactions on neural networks and learning systems*, PP. cited at <https://ieeexplore.ieee.org/iel8/5962385/6104215/10580966.pdf>
- Fan, L. (2020). Audio Example Recognition and Retrieval Based on Geometric Incremental Learning Support Vector Machine System. *IEEE Access*, 8, 78630-78638. cited at <https://ieeexplore.ieee.org/iel7/6287639/6514899/09072167.pdf>
- Fanaras, K., Tragoudaras, A., Antoniadis, C., & Massoud, Y. (2022). Audio-visual Speaker Diarization: Improved Voice Activity Detection with CNN based Feature Extraction. 2022 IEEE 65th International Midwest Symposium on Circuits and Systems (MWSCAS), 1-4. cited at <https://repository.kaust.edu.sa/bitstream/handle/10754/680550/Audio-visual%20Speaker%20Diarization%20Improved%20Voice%20Activity%20Detection%20with%20CNN%20based%20Feature%20Extraction.pdf?sequence=1>
- Fei, H., Yu, T., & Li, P. (2021, June). Cross-lingual cross-modal pretraining for multimodal retrieval. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 3644-3650). cited at <https://aclanthology.org/2021.naacl-main.285.pdf>
- Fei, H., Yu, T., & Li, P. (2021, June). Cross-lingual cross-modal pretraining for multimodal retrieval. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 3644-3650). cited at <https://aclanthology.org/2021.naacl-main.285.pdf>
- Gadre, S.Y., Ilharco, G., Fang, A., Hayase, J., Smyrnis, G., Nguyen, T., Marten, R., Wortsman, M., Ghosh, D., Zhang, J., Orgad, E., Entezari, R., Daras, G., Pratt, S., Ramanujan, V., Bitton, Y., Marathe, K., Mussmann, S., Vencu, R., Cherti, M., Krishna, R., Koh, P., Saukh, O., Ratner, A.J., Song, S., Hajishirzi, H., Farhadi, A., Beaumont, R.,

- Oh, S., Dimakis, A.G., Jitsev, J., Carmon, Y., Shankar, V., & Schmidt, L. (2023). DataComp: In search of the next generation of multimodal datasets. ArXiv, abs/2304.14108. cited at https://proceedings.neurips.cc/paper_files/paper/2023/file/56332d41d55ad7ad8024aac625881be7-Paper-Datasets_and_Benchmarks.pdf
- Gong, T., Lyu, C., Zhang, S., Wang, Y., Zheng, M., Zhao, Q., Liu, K., Zhang, W., Luo, P., & Chen, K. (2023). MultiModal-GPT: A Vision and Language Model for Dialogue with Humans. ArXiv, abs/2305.04790. cited at <https://arxiv.org/pdf/2305.04790>
 - Gong, T., Lyu, C., Zhang, S., Wang, Y., Zheng, M., Zhao, Q., Liu, K., Zhang, W., Luo, P., & Chen, K. (2023). MultiModal-GPT: A Vision and Language Model for Dialogue with Humans. ArXiv, abs/2305.04790. cited at <https://arxiv.org/pdf/2305.04790>
 - Gong, Y., & Cosma, G. (2023). Boon: A neural search engine for cross-modal information retrieval. arXiv preprint arXiv:2307.14240. cited at <https://arxiv.org/pdf/2307.14240>
 - Hu, P., Huang, Z., Peng, D., Wang, X., & Peng, X. (2023). Cross-modal retrieval with partially mismatched pairs. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(8), 9595-9610. cited at http://pengxi.me/wp-content/uploads/2023/03/Cross-Modal_Retrieval_with_Partially_Mismatched_Pairs.pdf
 - Hu, P., Peng, X., Zhu, H., Zhen, L., & Lin, J. (2021). Learning cross-modal retrieval with noisy labels. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 5403-5413). cited at http://openaccess.thecvf.com/content/CVPR2021/papers/Hu_Learning_Cross-Modal_Retrieval_With_Noisy_Labels_CVPR_2021_paper.pdf
 - Hu, Z., Hu, Z., & Bors, A. G. (2023). Co-attention enabled content-based image retrieval. Neural Networks, 164, 245-263. cited at <https://www.sciencedirect.com/science/article/pii/S0893608023001867>
 - Iscen, A., Caron, M., Fathi, A., & Schmid, C. (2023). Retrieval-Enhanced Contrastive Vision-Text Models. ArXiv, abs/2306.07196. cited at <https://arxiv.org/pdf/2306.07196>
 - Jabbar, A., Iqbal, S., Tamimy, M.I., Rehman, A., Bahaj, S.A., & Saba, T. (2023). An Analytical Analysis of Text Stemming Methodologies in Information Retrieval and Natural Language Processing Systems. IEEE Access, 11, 133681-133702. cited at https://gredos.usal.es/bitstream/handle/10366/138828/SoftComputing_DecisionSupportTools_postprint.pdf?sequence=1
 - Jagtap, J., & Bhosle, N.P. (2021). A comprehensive survey on the reduction of the semantic gap in content-based image retrieval. Int. J. Appl. Pattern Recognit., 6, 254-271. cited at <https://www.academia.edu/download/91417673/ijapr.2021.pdf>
 - Jeya Christy, A., & Dhanalakshmi, K. (2022). Content-based image recognition and tagging by deep learning methods. Wireless Personal Communications, 123(1), 813-838. cited at <https://link.springer.com/article/10.1007/s11277-021-09159-8>
 - Jiang, D., & Ye, M. (2023). Cross-Modal Implicit Relation Reasoning and Aligning for Text-to-Image Person Retrieval. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2787-2797. cited at https://openaccess.thecvf.com/content/CVPR2023/papers/Jiang_Cross-Modal_Implicit_Relation_Reasoning_and_Aligning_for_Text-to-Image_Person_Retrieval_CVPR_2023_paper.pdf
 - Jin, Q., Chen, F., Zhou, Y., Xu, Z., Cheung, J.M., Chen, R., Summers, R.M., Rousseau, J.F., Ni, P., Landsman, M.J., Baxter, S.L., Al'Aref, S.J., Li, Y., Chiang, M.F., Peng, Y., & Lu, Z. (2024). Hidden flaws behind expert-level accuracy of multimodal GPT-4 vision in medicine. NPJ Digital Medicine, 7. cited at <https://arxiv.org/pdf/2401.08396>
 - Kass, D., & Vats, E. (2022, May). AttentionHTR: Handwritten text recognition based on attention encoder-decoder networks. In International Workshop on Document Analysis Systems (pp. 507-522). Cham: Springer International Publishing. cited at <https://arxiv.org/pdf/2201.09390>

- Kaur, P., Pannu, H. S., & Malhi, A. K. (2021). Comparative analysis on cross-modal information retrieval: A review. *Computer Science Review*, 39, 100336. cited at http://eprints.bournemouth.ac.uk/36355/7/cross_modal_survey.pdf
- Lerner, P., Ferret, O., & Guinaudeau, C. (2024). Cross-modal Retrieval for Knowledge-based Visual Question Answering. *European Conference on Information Retrieval*. cited at <https://arxiv.org/pdf/2401.05736>
- Lerner, P., Ferret, O., & Guinaudeau, C. (2024, March). Cross-modal retrieval for knowledge-based visual question answering. In *European Conference on Information Retrieval* (pp. 421-438). Cham: Springer Nature Switzerland. cited at <https://arxiv.org/pdf/2401.05736>
- Li, B., Wu, Y., & Li, Z. (2024). Team HUGE: Image-Text Matching via Hierarchical and Unified Graph Enhancing. *Proceedings of the 2024 International Conference on Multimedia Retrieval*. cited at <https://dl.acm.org/doi/abs/10.1145/3652583.3658001>
- Li, C., Xu, H., Tian, J., Wang, W., Yan, M., Bi, B., Ye, J., Chen, H., Xu, G., Cao, Z., Zhang, J., Huang, S., Huang, F., Zhou, J., & Si, L. (2022). mPLUG: Effective and Efficient Vision-Language Learning by Cross-modal Skip-connections. *Conference on Empirical Methods in Natural Language Processing*. cited at <https://arxiv.org/pdf/2205.12005>
- Li, H. (2020). Text recognition and classification of english teaching content based on SVM. *Journal of Intelligent & Fuzzy Systems*, 39(2), 1757-1767. cited at <https://content.iospress.com/articles/journal-of-intelligent-and-fuzzy-systems/ifs189313>
- Li, L., Hu, C., Zhang, H., & Maradapu Vera Venkata sai, A. (2023, December). Cross-modal Image-Recipe Retrieval via Multimodal Fusion. In *Proceedings of the 5th ACM International Conference on Multimedia in Asia* (pp. 1-7). cited at <https://dl.acm.org/doi/abs/10.1145/3595916.3626389>
- Li, S., Xu, X., Yang, Y., Shen, F., Mo, Y., Li, Y., & Shen, H.T. (2023). DCEL: Deep Cross-modal Evidential Learning for Text-Based Person Retrieval. *Proceedings of the 31st ACM International Conference on Multimedia*. cited at <https://dl.acm.org/doi/abs/10.1145/3581783.3612244>
- Liao, H., Shen, H., Li, Z., Wang, C., Li, G., Bie, Y., & Xu, C. (2023). GPT-4 Enhanced Multimodal Grounding for Autonomous Driving: Leveraging Cross-Modal Attention with Large Language Models. *ArXiv*, abs/2312.03543. cited at <https://www.sciencedirect.com/science/article/pii/S2772424723000276>
- Lin, X., Luo, W., Xu, P., Qiao, Y., & Yang, S. (2021). PopDMMO: A general framework of population-based stochastic search algorithms for dynamic multimodal optimization. *Swarm Evol. Comput.*, 68, 101011. cited at <https://dora.dmu.ac.uk/bitstream/handle/2086/21483/SWEVO21.pdf?sequence=1&isAllowed=n>
- Liu, X., Zeng, H., Shi, Y., Zhu, J., & Ma, K. (2022). Deep Rank Cross-Modal Hashing with Semantic Consistent for Image-Text Retrieval. *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 4828-4832. cited at <https://arxiv.org/pdf/2309.03064>
- Lu, H., Huo, Y., Ding, M., Fei, N., & Lu, Z. (2023). Cross-modal Contrastive Learning for Generalizable and Efficient Image-text Retrieval. *Machine Intelligence Research*, 1-14. cited at <https://www.mi-research.net/en/article/pdf/preview/10.1007/s11633-022-1386-4.pdf>
- Meng, X., Li, D., & Han, Q. (2021). Personalized Semantic Retrieval System based on Statistical Language Model. *2021 IEEE/ACIS 20th International Fall Conference on Computer and Information Science (ICIS Fall)*, 253-257. cited at <https://ieeexplore.ieee.org/abstract/document/9472983/>

- Mishra, A. (2022). Information Retrieval in Friction Stir Welding of Aluminum Alloys by using Natural Language Processing based Algorithms. ArXiv, abs/2204.12309. cited at <https://arxiv.org/pdf/2204.12309>
- Nabankema, H. (2024). Evaluation of Natural Language Processing Techniques for Information Retrieval. European Journal of Information and Knowledge Management. cited at https://www.myecole.it/biblio/wp-content/uploads/2020/11/7semt_2018-Natural_Language_Processing.pdf
- Nallappan, M., & Velswamy, R. (2024). Exploring deep learning-based content-based video retrieval with Hierarchical Navigable Small World index and ResNet-50 features for anomaly detection. Expert Syst. Appl., 247, 123197 cited at <https://www.sciencedirect.com/science/article/pii/S0957417424000629>
- Pan, Z.K., Wu, F., & Zhang, B. (2023). Fine-grained Image-text Matching by Cross-modal Hard Aligning Network. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 19275-19284. cited at Fine-grained Image-text Matching by Cross-modal Hard Aligning Network.
- Pang, N., Guo, S., Yan, M., & Chan, C.A. (2023). A Short Video Classification Framework Based on Cross-Modal Fusion. *Sensors (Basel, Switzerland)*, 23. cited at <https://www.mdpi.com/1424-8220/23/20/8425>
- Park, P.G., Jang, S., Cho, Y., & Kim, Y. (2023). SAM: cross-modal semantic alignments module for image-text retrieval. Multimedia Tools and Applications, 83, 12363-12377. cited at <https://link.springer.com/article/10.1007/s11042-023-15798-9>
- Patel, B., & Singh, B. (2023). Content-based Video Retrieval Systems: A Review. 2023 3rd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), 441-449. cited at <https://ieeexplore.ieee.org/abstract/document/10425939/>
- Phan, T., Phan, A., Cao, H., & Trieu, T. (2022). Content-Based Video Big Data Retrieval with Extensive Features and Deep Learning. Applied Sciences. cited at <https://www.mdpi.com/2076-3417/12/13/6753>
- Prieto, J. R., Bosch, V., Vidal, E., Alonso, C., Orcero, M. C., & Marquez, L. (2021, January). Textual-content-based classification of bundles of untranscribed manuscript images. In 2020 25th International Conference on Pattern Recognition (ICPR) (pp. 3162-3169). IEEE. cited at <https://joserprietof.github.io/files/docClasICPR20.pdf>
- Prieto, J. R., Bosch, V., Vidal, E., Stutzmann, D., & Hamel, S. (2020, September). Text content based layout analysis. In 2020 17th International Conference on Frontiers in Handwriting Recognition (ICFHR) (pp. 258-263). IEEE. cited at
- Sadirmekova, Z., Sambetbayeva, M.A., Daiyrbayeva, E., Yerimbetova, A.S., Altynbekova, Z., & Murzakhmetov, A. (2023). Constructing the Terminological Core of NLP Ontology. 2023 8th International Conference on Computer Science and Engineering (UBMK), 81-85. cited at <https://ieeexplore.ieee.org/abstract/document/10286646/>
- Sain, A., Chowdhury, P.N., Koley, S., Bhunia, A.K., & Song, Y. (2024). Freeview Sketching: View-Aware Fine-Grained Sketch-Based Image Retrieval. *ArXiv, abs/2407.01810*. cited at <https://arxiv.org/pdf/2407.01810>
- Sain, A., Chowdhury, P.N., Koley, S., Bhunia, A.K., & Song, Y. (2024). Freeview Sketching: View-Aware Fine-Grained Sketch-Based Image Retrieval. *ArXiv, abs/2407.01810*. cited at <https://arxiv.org/pdf/2407.01810>
- Sowmyayani, S., Arockia, P., & Rani, J.V. (2023). Content based video retrieval system using two stream convolutional neural network. Multimedia Tools and Applications, 1-19. cited at <https://link.springer.com/article/10.1007/s11042-023-14784-5>
- Tian, J., Wang, K., Xu, X., Cao, Z., Shen, F., & Shen, H. T. (2022, July). Multimodal disentanglement variational autoencoders for zero-shot cross-modal retrieval. In Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 960-969). cited at

- <https://scholar.archive.org/work/rfz3222hzvbjdytandkjuuxfq/access/wayback/https://dl.acm.org/doi/pdf/10.1145/3477495.3532028>
- Veni, S., Anand, R., Mohan, D., & P, S. (2021). Leaf Recognition and Disease Detection using Content based Image Retrieval. 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), 1, 243-247. cited at <https://link.springer.com/article/10.1007/s00521-020-05017-z>
 - Verschuuren, P.J., Gao, J., Eeden, A.V., Oikonomou, S., & Bandhakavi, A. (2023). Logically at Factify 2: A Multi-Modal Fact Checking System Based on Evidence Retrieval techniques and Transformer Encoder Architecture. DE-FACTIFY@AAAI. cited at <https://arxiv.org/pdf/2301.03127>
 - Wang, B., Xu, C., Liu, X., Cheng, Y., & Li, B. (2022). SemAttack: Natural Textual Attacks via Different Semantic Spaces. NAACL-HLT. cited at <https://arxiv.org/pdf/2205.01287>
 - Wang, S. (2024). Exploring the fusion of natural language processing and information retrieval. Applied and Computational Engineering. cited at <https://link.springer.com/article/10.1007/s42979-021-00775-6>
 - Wang, W. (2023). Music chord sequence recognition method based on audio feature extraction algorithm. 2023 IEEE International Conference on Control, Electronics and Computer Technology (ICCECT), 1286-1289. cited at <https://ieeexplore.ieee.org/abstract/document/10140586/>
 - Wang, X., Xu, X., Zhang, J., Zhu, Y., Fan, Y., & Xu, P. (2021). Research on Intelligent Construction Algorithm of Subject Knowledge Thesaurus Based on Literature Resources. Journal of Physics: Conference Series, 1955. cited at <https://iopscience.iop.org/article/10.1088/1742-6596/1955/1/012038/pdf>
 - Wasim, M., Ahmed, I., Ahmad, J., Nawaz, M., Alabdulkreem, E.A., & Ghadi, Y.Y. (2022). A video summarization framework based on activity attention modeling using deep features for smart campus surveillance system. PeerJ Computer Science, 8. cited at https://www.itl.gr/~bmezaris/publications/ism2021a_preprint.pdf
 - Wen, W. (2023). Design of Music Audio Rhythm Recognition Algorithm Based on Feature Extraction Algorithm. 2023 International Conference on Telecommunications, Electronics and Informatics (ICTEI), 76-80. cited at <https://ieeexplore.ieee.org/abstract/document/10592158/>
 - Xing, Y., Meyer, B.J., Harandi, M., Drummond, T., & Ge, Z. (2023). Multimorbidity Content-Based Medical Image Retrieval and Disease Recognition Using Multi-Label Proxy Metric Learning. IEEE Access, 11, 50165-50179. cited at <https://ieeexplore.ieee.org/iel7/6287639/6514899/10130284.pdf>
 - Xu, B., Xiong, Y., Zhang, R., Feng, Y., & Wu, H. (2022). Natural Language-Based Vehicle Retrieval with Explicit Cross-Modal Representation Learning. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 3141-3148. cited at https://openaccess.thecvf.com/content/CVPR2022W/AICity/papers/Xu_Natural_Language-Based_Vehicle_Retrieval_With_Explicit_Cross-Modal_Representation_Learning_CVPRW_2022_paper.pdf
 - Yi, Z., Yu, J., Tan, Y., & Wu, Q. (2021). A Multimodal Adversarial Attack Framework Based on Local and Random Search Algorithms. Int. J. Comput. Intell. Syst., 14, 1934-1947. cited at <https://www.atlantis-press.com/article/125958419.pdf>
 - Yu, J., & Qin, Z. (2015). Correlations for Cross-Modal Multimedia Information Retrieval. cited at <https://link.springer.com/article/10.1007/s10044-015-0478-y>
 - Yu, L. (2024). Research on Retrieval Method of Online English Grammar Questions Based on Natural Language Processing. International Journal of Emerging Technologies in Learning (IJET). cited at <https://pureportal.strath.ac.uk/files/131112/strathprints002611.pdf>

- Yu, Q., Liu, Y., Wang, Y., Xu, K., & Liu, J. (2023). Multimodal Federated Learning via Contrastive Representation Ensemble. *ArXiv*, *abs/2302.08888*. cited at <https://arxiv.org/pdf/2302.08888>
- Zeng, S., Liu, C., Zhou, J., Chen, Y., Jiang, A., & Li, H. (2022). Learning Hierarchical Semantic Correspondences for Cross-Modal Image-Text Retrieval. Proceedings of the 2022 International Conference on Multimedia Retrieval. cited at <https://scholar.archive.org/work/6epjmanpanhkjjptlh7ug5e6yq/access/wayback/https://dl.acm.org/doi/pdf/10.1145/3512527.3531358>
- Zhang, B., Wei, C., Jin, Y., & Zhang, W. (2021). ACE-BERT: Adversarial Cross-modal Enhanced BERT for E-commerce Retrieval. *ArXiv*, *abs/2112.07209*. cited at <https://arxiv.org/pdf/2112.07209>
- Zhang, D., Li, M., Tian, D., Song, L., & Shen, Y. (2022). Intelligent text recognition based on multi-feature channels network for construction quality control. *Advanced Engineering Informatics*, 53, 101669. cited at <https://www.sciencedirect.com/science/article/pii/S147403462200132X>
- Zhang, X., Zhen, T., Zhang, J., Wang, Y., & Liu, S. (2023). SRCB at SemEval-2023 Task 1: Prompt Based and Cross-Modal Retrieval Enhanced Visual Word Sense Disambiguation. International Workshop on Semantic Evaluation. cited at <https://aclanthology.org/2023.semeval-1.60.pdf>
- Zhen, L., Hu, P., Peng, X., Goh, R. S. M., & Zhou, J. T. (2020). Deep multimodal transfer learning for cross-modal retrieval. *IEEE Transactions on Neural Networks and Learning Systems*, 33(2), 798-810. cited at https://liangli-zhen.github.io/assets/pdf/TNNLS2022_DMTL_CMR.pdf
- Zheng, W., Ling, X., & Hai, H. (2023). Residual-based feature enhancement for forgery audio detection. Conference on Advanced Algorithms and Signal Image Processing. cited at <https://www.spiedigitallibrary.org/conference-proceedings-of-spie/12799/127993B/Residual-based-feature-enhancement-for-forgery-audio-detection/10.1117/12.3005833.short>
- Zhu, M., & Lee, J.H. (2022). Deep Learning-Based 3D Shape Feature Extraction on Flash Animation Style. *Wireless Communications and Mobile Computing*. cited at <https://onlinelibrary.wiley.com/doi/pdf/10.1155/2022/7999312>

Holistic information retrieval systems: An applied study of overall informative processing and unified cross-modal retrieval

Dr. Sayed Rabie Sayed Ibrahim

Assistant Professor of Libraries and Information

Sayeds1m@hotmail.com

Taibah University (Saudi Arabia)

Faculty of Arts and Humanities

Department of Information & Learning Resources

Beni-Suef University (Egypt)

Faculty of Arts

Department of Information Sciences

Abstract:

Information retrieval systems have relied on the textual medium in processing and searching digital information content through its various media, due to the compatibility of the text characteristics and the nature of system users in dealing with the text over other media, but the verification rate in retrieving qualitative media using the text is at a low level that does not achieve the required efficiency. Web search engines have worked on adopting the processing and retrieval method based on the content of the information material, using the characteristics of audio and video media in indexing and retrieval, and this has borne fruit in the retrieval results of multimedia. The capabilities of information retrieval systems have developed by adopting semantic analysis techniques, knowledge entity networks, and natural language processing capabilities, which has increased the efficiency of systems in developing their results into semantic contextual results, which is represented in retrieval via Chat-GPT conversation generators. This study aims to move towards a more advanced generation of information retrieval systems that adopts a comprehensive approach that includes all information materials with their specific media, where text, audio, and image media are processed, organized, and retrieved through a unified processing process that derives the characteristics of each medium separately, then collects them in a knowledge base that can achieve semantic and conceptual linkage between the content of the audio, image, and text material, making the system capable of receiving a search query through a single medium and retrieving results through all media. Accordingly, the formal characteristics governing the content medium dissolve to leave the cognitive content represented in the knowledge base, and the retrieval system can retrieve results through multiple media related to the same topic for a single search process, such as searching by voice and the results are audio, visual, and textual... etc. The study uses the descriptive and experimental research approach by adopting a sample of information retrieval systems on the web, using the tools of collecting documentary data to collect the literature of the subject and observation for the applied study. The most important results of the study are that holistic retrieval systems across media are the first levels of interactive comprehensive knowledge production systems in the digital environment, which is not far from the present time, and the development of artificial intelligence technologies and their applications in the field of information processing and digital content perception will support the emergence and development of cognitive systems capable of segmenting and building objective and semantic entities to make them available in accordance with the cognitive needs of system users. The study presented a vision for building and structuring comprehensive multi-media retrieval systems.

Keywords: *Information retrieval systems; knowledge production and generation; knowledge bases; multimedia; search engines; the web; natural language processing; semantic indexing; semantic web; conceptual linkage networks.*