المحتوى الرقمي العربي التنقيب المعلوماتي وتقنياته نموذجًا

إعداد

الدكتور حسين محمد علي البسومي

الأستاذ المساعد بقسم اللغة العربية

جامعة المدينة العالمية

ملخص البحث:

تستجمع الحضارات المختلفة في عصر العولمة كل قواها لبناء محتوى رقمي يمثل ثقافتها، ويحفظ عليها خصائصها وسماتها، وينوب عنها في حوار ثقافي متوازن مع الحضارات الأخرى في غير تميع ولا صراع.

وها هو المحتوى الرقمي العربي يتسارع نموه في الآونة الأخيرة عله يلحق بمن سبقه، ويسد بعضًا من الفجوات المعرفية التي خلفت دوامات من التبعية، والانبهار بالآخر.

ويحاول البحث دعم هذه الاستفاقة؛ وذلك بربطها بالواقع، حتى تستثمر ما بُذِل من جهود، وترسم لنفسها أهدافًا متدرجة، مقدرةً الصعاب والتحديات اللغوية والحاسوبية والمعرفية والإدارية، فهو يلقي الضوء على أهم مراحل بناء المحتوى الرقمي العربي، مع التركيز على مرحلة التنقيب المعلوماتي، ويعرض لأهم التقنيات والأدوات اللازمة لتفعيل الوصول السهل والمتنوع إلى المعلومات التي يشتمل عليها، كما يتناول أهم التحديات التي تعترض طريق بنائه، وتطوير أدواته وتقنياته، وقد جاء ذلك في أربعة محاور أساسية، هي:

المحور الأول: إدارة المحتوى الرقمي العربي؛ رؤية عامة.

المحور الثاني: تحديات التنقيب المعلوماتي في المحتوى الرقمي العربي.

المحور الثالث: المعالجة الفنية لوثائق المحتوى الرقمي العربي.

المحور الرابع: تقنيات التنقيب المعلوماتي في المحتوى الرقمي العربي.

Research Summary

At the age of Globalization, civilizations are thriving to build up their digital knowledge which forms their own characterizations, and protects their unique cultural features. The digital Knowledge will be the representative of these civilizations in a well-balanced independent and comprehensive cultural dialogue with each other; without losing identities, or melting in others' ideologies.

The Arabic digital knowledge is increasingly growing lately, in order that, it should catch up with others and bridge gabs which were a result of an inheritance of dependence upon others along with an inferiority complex.

The goal of this thesis is to support this Arabic up rise in digital knowledge through a thorough objective and realistic view, in order to invest what already has been done, and gradually plan a concise map that will never ignore hindrances and other computational, management, informational as wells linguistic challenges.

The thesis highlights the most important stages of the Arabic digital knowledge, focuses on the stage of the stage of exploration and information searching process. The thesis will also keep an eye on the technical tools and ways by which the information will be both attainable and various, beside explaining the difficulties and challenges in every stage.

There are four major aspect of this thesis:

- Management of the Arabic digital knowledge. (general view)
- Challenges that face the process of the collecting and info searching.
- Technical processing of the documents of the Arabic digital knowledge.
- The technicalities of info searching of the Arab digital knowledge.

محاور البحث الرئيسية:

مقدمة.

المحور الأول: المحتوى الرقمي العربي؛ رؤية عامة.

مراحل بناء المحتوى الرقمي العربي:

المرحلة الأولى: تحديد الفجوات المعرفية.

المرحلة الثانية: اختيار المصادر المناسبة.

المرحلة الثالثة: رقمنة المحتوى العربي.

المرحلة الرابعة: المعالجة الفنية للوثائق.

المرحلة الخامسة: تخزين المحتوى الرقمي العربي.

المحور الثاني: تحديات التنقيب المعلوماتي في المحتوى الرقمي العربي.

المحور الثالث: المعالجة الفنية لوثائق المحتوى الرقمي العربي.

أولًا: المعالجة الصوتية.

ثانيًا: المعالجة الصرفية.

ثالثًا: المعالجة الموضوعية.

المحور الرابع: تقنيات التنقيب المعلوماتي في المحتوى الرقمي العربي.

الخاتمة.

مصادر البحث ومراجعه.

المحتوى الرقمي العربي
التنقيب المعلوماتي وتقنياته نموذجًا

مقدمة:

"لم تعد المشكلة الآن في ندرة المعلومات، وإنما في كيفية الوصول إليها في هذا الركام المعرفي الضخم" تتكرر هذه العبارة كثيرًا في مناسبات الحديث عن القفزات الهائلة التي يشهدها النمو المعلوماتي المعاصر، أو في عروض الترويج لأدوات وتقنيات حاسوبية يمكنها تنظيم المحتوى المعرفي، وتقديمه على نحو يسهل التعامل معه، أو في المحافل الإقليمية والدولية التي تناقش مظاهر العولمة ومخاطرها على الثقافات التي لا تقوى على المنافسة.

غير أن هذه العبارة –وإن كشفت عن مكامن الصعوبات والتحديات إلا أنها– أهملت الإشارة إلى جوانب أخرى عديدة لا تزال تمثل عقبة كئودًا أمام الإفادة الكاملة من التراث الثقافي العالمي، والإنتاج المعلوماتي المعاصر، ومن هذه التحديات التي تحتاج إلى تشخيص دقيق لمظاهرها وتقديم حلول شاملة لها:

- تحقيق التوازن الكمي والنوعي للمحتوى المعرفي المعبر عن ثقافة من الثقافات.
- شمولية المعالجات الفنية للمحتوى المعرفي المقدم واللغة الحاملة له بما يمكِّن من تمثيل جيد للمعرفة المبثوثة في تضاعيف هذا المخزون النصي أو السمعي أو المرئي.
- تطوير التقنيات والأدوات الحاسوبية المتقدمة التي تحاكي السلوك البشري في التعامل مع اللغة إبداعًا وتحليلًا.
- إدارة هذه العملية الضخمة، والتأليف بين مهامها المتعاقبة والمتزامنة والمتنوعة بما يضمن تحقيق الأهداف وتوفير الجهود.

هذه المشكلات وغيرها من المشكلات الفرعية تمثل التحدي الأكبر في مواجهة هذا النمو المعلوماتي المعاصر، والابتعاد به عن حلبة صراع الثقافات والحضارات إلى أفق التكامل الرحب؛ وعلى قدر نجاح أبناء ثقافة معينة في تجاوز هذه العقبات فإنهم يتمكنون من نشر

ثقافتهم والذود عنها.

وسأهتم في هذه الورقة البحثية بالحديث حول هذه العقبات، وموقف المحتوى الرقمي العربي منها، مقتصرًا في ذلك على الحديث عن المحتوى الرقمي العربي النصي –دون المحتوى الرقمي المسموع أو المرئي– نظرًا لاختلاف المعالجات والأدوات اللازمة لكل؛ كما أركز في الحديث عن التقنيات والأدوات الحاسوبية على إلقاء الضوء على الجانب اللغوي وقواعد البيانات اللازمة لبنائها، وذلك من خلال المحاور التالية:

المحور الأول: المحتوى الرقمي؛ رؤية عامة:

تسارع نمو المحتوى الرقمي العربي في الآونة الأخيرة محققًا قفزاتٍ تشجع على مواصلة بذل الجهود في محاولة لردم الفجوات المعرفية الواسعة التي يعانيها مقارنة بالمحتوى الرقمي للغات الأساسية الأخرى؛ كالإنجليزية والصينية والفرنسية، غير أنه من الأهمية بمكان: أن تنظم الجهود الرامية إلى تطوير المحتوى الرقمي العربي، وأن تُدار وفق أسس ومخططات عامة تراعي تحقيق التوازن الكمي والنوعي له، فعشوائية الجهود المبذولة –وإن جمعت بينها جميعًا حسن النية– تصيب المحتوى الرقمي العربي بالاضطراب، وتبعد به عن تحقيق الأهداف المعقودة به؛ فزيادة حجم المحتوى الرقمي العربي عن غيره ليس هو الهدف، وإنما الهدف هو أن يغطي المحتوى الرقمي العربي كل فنون الثقافة العربية قديمها وحديثها، وأن يُقدَّم على نحو يلبي حاجات مستخدميه.

ولعل من صور الاضطراب التي يعانيها المحتوى الرقمي العربي الحالي:

1- نصوص ووثائق مُكررة؛ يغني قليلها عن كثيرها.

2- افتقار بعض الفنون والمجالات إلى الوثائق الكافية، بل إن بعض فنون الثقافة العربية قد تغيب كلها عن المحتوى الرقمي.

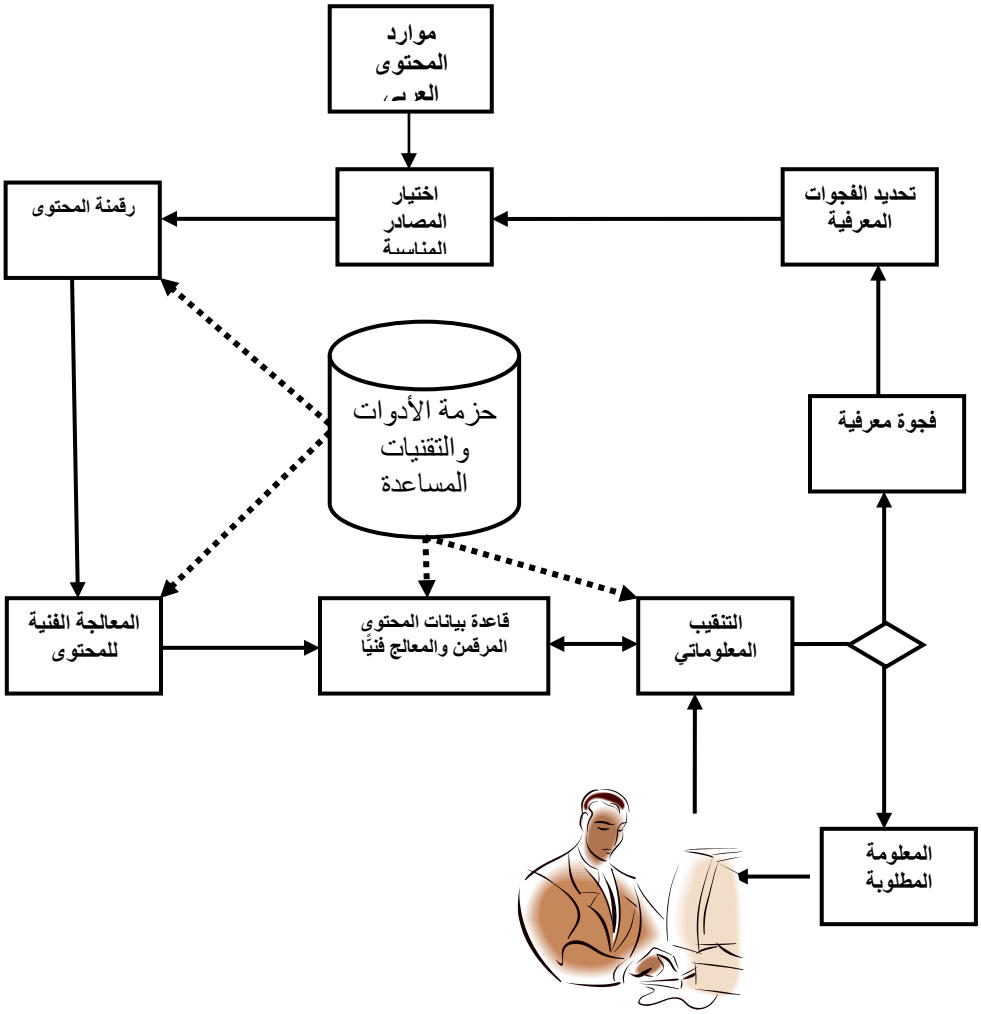
3- ضعف المعالجة الفنية والرقمية للوثائق بما يجعل كثيرًا من المعلومات الحاملة لها حبيسة كلمات الوثائق وجملها.

4- ضعف الأدوات والتقنيات الحاسوبية الملحقة بالمحتوى الرقمي، بما يصعب من

الوصول إلى المعلومة المطلوبة في الوقت المناسب.

ولعل صور الاضطراب السابقة –وغيرها– يعود إلى: غياب الرؤية العامة للأهداف القريبة والبعيدة للمحتوى الرقمي العربي، وآليات تحقيق هذه الأهداف، وحجم وأنواع تحديات بناء المحتوى الرقمي العربي وسبل التغلب عليها. فمثل هذه المشروعات الضخمة يحتاج إلى مجموعة من النظم الإدارية والفنية لتحقيق الأهداف المتدرجة والمتنوعة للمحتوى، وتحقيق التوازن الكمي والنوعي له، واستثمار كل المصادر المتاحة، وابتكار مصادر جديدة داعمة، والإفادة من جهود فرق العمل المتكاملة، وتوظيف الجهود التطوعية، وتحديد الأدوات والتقنيات وإجراء المعالجات الفنية اللازمة لتهيئة المعرفة المباشرة والضمنية للوثائق، وتمكين المستخدم العادي من استرجاعها والحصول عليها.

وسأعرض في الأسطر القادمة -بشيء من الإيجاز- المراحل الأساسية لبناء المحتوى الرقمي العربي وتطويره، وذلك على النحو التالي:



مراحل بناء المحتوى الرقمي العربي:

المرحلة الأولى: تحديد الفجوات المعرفية:

تهدف هذه المرحلة إلى تشخيص علل المحتوى الرقمي العربي، وتحديد المجالات التي تعاني تراكم الوثائق وتحتاج إلى التخفف من بعضها، وتعيين المجالات الأخرى التي تحتاج إلى دعم بمزيد من الوثائق، وإنشاء الفنون الجديدة التي يحتاجها المستخدم المعاصر ويفرضها واقع الحياة المتجدد؛ وذلك لاستثمار الجهود السابقة، وتحويلها إلى أساس يصلح البناء عليه والإضافة إليه، فكما أن الضرورة تقتضي أن نبدأ من حيث انتهى الآخرون، فإنها تقتضي أيضًا فحص الجهود السابقة، والإبقاء على ما يصلح فقط في ضوء الأهداف والمناهج المقررة.

المرحلة الثانية: اختيار المصادر المناسبة:

تعتمد هذه المرحلة على نتائج المرحلة السابقة؛ حيث تعرض على لجان من الخبراء متنوعي التخصصات للنظر بهدف:

1-انتقاء الوثائق المناسبة من بين الوثائق المكررة في المحتوى الرقمي الحالي.

2-اختيار الوثائق المناسبة من الإنتاج الثقافي العربي لردم الفجوات المعرفية التي تم تشخيصها في المرحلة السابقة.

3-تحليل الواقع الثقافي العالمي، وفحص طلبات المستخدمين المتجددة، لتحديد المجالات الجديدة التي يمكن أن تُضاف إلى المحتوى الرقمي العربي، والتشخيص المستمر للفجوات المتجددة.

المرحلة الثالثة: رقمنة محتوى العربي:

تعني رقمنة المحتوى: تحويل الوثائق الورقية إلى صورة رقمية؛ وذلك بإدخالها إلى الحاسوب، وتدقيقها، وتنظيم، محتواها على نحو يمكن التحكم فيه وعرضه على صور متنوعة وفق طلبات مستخدميه، ويُستعان في هذه المرحلة بعدد من الأدوات الحاسوبية التي توفر

الجهد والوقت وتضمن الدقة؛ مثل: الماسح الضوئي- القارئ الآلي- المدقق الإملائي- المدقق النحوي، وتتفاوت القيمة المضافة لهذه الأدوات في رقمنة الوثائق تبعًا لحالة الوثيقة، فكلما كانت الوثيقة واضحة عَظُم دور هذه التقنيات.

المرحلة الرابعة: المعالجة الفنية للوثائق:

تتنوع المعالجات الفنية التي يخضع لها المحتوى الرقمي ما بين معالجات تهدف إلى تنظيم المحتوى وتصنيفه وفق اعتبارات متنوعة، ومعالجات تهدف إلى تمكين الأدوات الحاسوبية الملحقة بالمحتوى من تفاعلها معه، ومعالجات تهدف إلى توفير إحصائيات وتحليلات تفيد الباحثين وشادين المعرفة، ويُستعان في هذه المرحلة بالعديد من الأدوات والتقنيات، منها: المحلل الصرفي، والمحلل النحوي، والمشكل الآلي، ومحرك البحث، والملخص الآلي، والمصنّف الآلي للوثائق، والناطق الآلي TTS.

وسأتناول في هذه الورقة بشيء من التفصيل المعالجات الفنية التي تحتاجها أدوات وتقنيات التنقيب المعلوماتي في المحتوى العربي المرقمن[1].

المرحلة الخامسة: تخزين المحتوى الرقمي العربي:

يحتاج المحتوى العربي المرقمن والمعالج فنيًّا إلى أن يخزن في قواعد بيانات ذات سعة تخزينية كبيرة؛ نظرًا لضخامة حجم المحتوى ونموه المستمر، وذات مرونة تسمح بأداء أكبر عدد من المعالجات في نفس الوقت، وتسمح ببناء التطبيقات المختلفة عليها، ومن أهم هذه التطبيقات هو التنقيب عن المعلومات.

المحور الثاني: تحديات التنقيب المعلوماتي في المحتوى الرقمي العربي:

التنقيب المعلوماتي يعني: البحث عن المعلومات المبثوثة في الوثيقة محل البحث، وهذا الإجراء غاية للعديد من المعالجات المعقدة، وقبل أن نتحدث عن هذه المعالجات يجمل بنا أن نشير إلى أمرين مهمين: الأول هو طبيعة المعرفة وكيفية تعبير العقل البشري عنها في صورة جمل

1-انظر المحور الرابع من البحث.

ونصوص، والآخر يتعلق بخصائص النصوص العربية الحاملة للمعرفة والثقافة العربية.

الأمر الأول: المعرفة البشرية وكيفية تعبير العقل البشري عنها:

ينتظم الواقعَ الذي يعيشه الإنسان شبكةٌ معقدة من المفاهيم المترابطة فيما بينها بروابط منطقية تشكل في مجملها المعرفة التي يحصلها الإنسان في تجاربه، ويمارسها في أقواله وأفعاله، ويختزن العقل البشري شبكة مفهومية توازي تلك الموجودة في الواقع، وتماثل –أو تقارب إلى حد كبير– ما يختزنه بقية أفراد مجتمعه في عقولهم من مفاهيم وخبرات، ويستطيع العقل البشري –بما أودعه الله من قدرات– أن يُؤَلِّف بين المفاهيم التي يختزنها ليعبر عما يريد في عبارة أو حركة أو رسم...إلخ، كما يستطيع أن يحلل ما يرد إليه من عبارة أو حركة أو رسم إلى عدد من المفاهيم وتعيين ما بينها من روابط، فيفهم ما يرد إليه من معرفة، ويتمكن من التفاعل مع الآخرين إفهامًا وفهمًا.

ولقد جعل النظام اللغوي للغة العربية رموزًا صوتية وكتابية وقواعد وقوانين دلالية وتركيبية تمكن متكلمي اللغة العربية من التعبير عن هذه المفاهيم، وما ينشأ بينها من روابط منطقية في صورة عبارات منطوقة أو جمل مكتوبة. وعلى ذلك: فأي نص أو وثيقة مكتوبة أو منطوقة ما هي إلا شبكة من المفاهيم المترابطة نسجها عقل مبدع الوثيقة للتعبير عن معرفة معينة يريد توصيلها للآخرين.

الأمر الثاني: خصائص النظام اللغوي للغة العربية:

يتألف نظام اللغة العربية من أكثر من مستوى؛ فهناك: النظام الصوتي، والنظام الصرفي، والنظام المعجمي، والنظام الدلالي، والنظام التركيبي، ولكل نظام من هذه الأنظمة قوانينه ومبادئه التي تضبط العلاقات الداخلية بين مكوناته وعناصره، وتنظم علاقته ببقية أنظمة اللغة، وتتداخل أنظمة اللغة فيما بينها تداخلًا لا يمكن فصله على مستوى الاستعمال الحقيقي للغة، وإن كان ذلك ضروريًّا على المستوى البحثي. والتمثيل الحاسوبي للغة العربية عليه أن يصف هذا النظام بما يتسم به من تداخلٍ بين مستوياته، واقتصاد في وحداته اللغوية، ووضوحٍ وإفادة في التعبير عن مراد المتكلم، وثراءٍ في البدائل الممكنة للتعبير عن الشيء

الواحد؛ وتوضيح ذلك على النحو التالي:

أولًا: المستوى الصوتي:

اعتمد النظام الصوتي للغة العربية على 28 صامتًا؛ منها 26 صوتًا صحيحًا وصوتان ليّنان هما الواو والياء مسبوقين بفتحة، وقد استطاع توظيف هذا العدد المحدود لإنتاج جميع جذور اللغة العربية، وجميع ما يتولد منها من صور اشتقاقية مع الاستعانة بالصوائت فـ"أول ما يلفت النظر في العربية ... أن عدد الوحدات الصوتية أو الحروف الهجائية ثمانية وعشرون حرفًا، ومع ذلك "كانت كافية لأداء كافة أنواع النشاط اللغوي شعرًا ونثرًا، وتخاطبًا عاديًّا، إلخ" [1]

ولعل حرص اللغة العربية على تحقيق الاقتصاد على مستوى النظام الصوتي، راجع لضرورة ثبات هذه الصوامت وعدم تعرضها للحذف أو الزيادة لأنها تحمل أبرز السمات التي تميز اللغة عن غيرها من اللغات، وتأكيدًا لذلك فإن النظام الصوتي العربي حريص على أن يكون لديه بما يعتمده من صوامت من التتابعات الصوتية الاحتياطية التي يمكنه استخدامها إذا ما دعت الضرورة إلى ذلك دون الحاجة إلى إضافة صامت جديد.

أما الصوائت في اللغة العربية وهي ثلاثة قصار: الفتحة والضمة والكسرة، وثلاثة طوال: الألف مطلقًا، والواو المسبوقة بضمة والياء المسبوقة بكسرة، فإنها تقوم في النظام الصوتي للغة العربية بدور مهم على مستوى المعجم لا على مستوى الجذور، فهي مناط لتقليب صيغ الاشتقاق المختلفة داخل المادة الواحدة؛ فالفرق بين (عَمَلٌ، وعَمِلَ، وعُمِل) إنما هو راجع إلى الصوائت وليس إلى الصوامت، وعلى ذلك: فإن الصوائت وحروف الزيادة والتشديد والمد تتحمل أخطر الوظائف في تركيب الصيغ الاشتقاقية للغة العربية[2].

وهذه الصوائت تتعاون مع الصوامت وتضاعف من الإمكانية الاشتقاقية للغة العربية،

1 – اللغة العربية معناها ومبناها، د. تمام حسان، ص 72.

2 –اللغة العربية معناها ومبناها د. تمام حسان، ص 72.

وهذه الحركات القليلة العدد تؤدي دورها في الكلام دون لبس أو إخلال؛ وذلك بأن الحركات لها دور كبير في تنوع الصيغ في الأسماء والأفعال على حدٍّ سواء، كما أنها تفرق بين المباني المتشابهة؛ كالتفريق بين لام الجر التي تَرِدُ مكسورةً ولام التوكيد التي تَرِدُ مفتوحةً، والحركات مع قلتها تؤدي إلى تكثير الصيغ وتنوعها بشكل بارز سواء أكانت الصيغ أفعالًا أو أسماءً، وسواء أكانت مجردة أم مزيدة، كما تُسهم الصوائت –أيضًا– في بناء النظام النحوي؛ فهي تأتي للدلالة على الوظائف النحوية[1].

ثانيًا: المستوى الصرفي:

النظام الصرفي من أهم دعائم اللغة العربية؛ حيث يكفل لها النمو والتطور، والصمود في مواجهة المطالب المتجددة التي تفرزها الأزمنة المتعاقبة والأمكنة المتنوعة، فهو المولد الرئيسي للبنى اللغوية التي انبنى منها المعجم العربي، ولا يزال لديه مخزون ضخم يمكن الاستعانة به وقت الحاجة، كل ذلك مع حرصه الشديد على تحقيق مبدأ الاقتصاد في وحداته التي يوظفها، ومن أهم ما يتسم به النظام الصرفي للغة العربية:

1- الاشتقاق:

انبني النظام الصرفي للغة العربية على خاصية اشتقاق الصيغ الكثيرة من عدد محدود من الصيغ النواة، والصيغة هي القالب الذي يستوعب الجذر اللغوي مكونين وحدة معجمية، تكون مجردة إذا كانت الصيغة خالية من أحرف الزيادة، أو مزيدة إذا كانت الصيغة تشتمل على حرف أو أكثر منها، والصيغة تكون للأفعال والأسماء المتصرفة، فلا صيغة للحروف ولا للأسماء غير المتصرفة؛ كالضمائر وأسماء الإشارة والأسماء الموصولة[2].

والنظام الصرفي للغة العربية وظف عددًا محدودًا من الصيغ الصرفية المجردة، واستبدل بكثرتها عددًا محدودًا –أيضًا– من الحروف يمكن زيادتها إلى تلك الصيغ المجردة لتحقيق التنوع

1 – الاقتصاد في النحو العربي، د. وليد خيرالله، ص145.

2 – اللغة العربية معناها ومبناها، د. تمام حسان، ص 72.

الكافي في الصيغ للنهوض بالمعاني الوظيفية المعقودة بها، والحروف المزيدة محصورة في كلمة (سألتمونيها)، فالصيغ المجردة للفعل الثلاثي تنشأ من اختلاف حركة عينه في الماضي والمضارع، والاحتمالات الممكنة للصيغ المجردة للثلاثي هي تسع صور؛ حيث يحتمل تشكيل عينه في الماضي بالفتحة أو الضمة أو الكسرة، وكذلك عينه في المضارع، غير أن المستعمل منها، ستة فقط، هي:

-"فعَل يفعُل"؛ مثل: نَصَر ينصُر.

-"فَعَل يفَعَل"؛ مثل: فَتَح يفْتَح.

-"فَعَل يفْعِل"؛ مثل: ضَرَب يضْرِب.

-"فَعُل يفْعُل"، مثل: كَبُر يكبُر.

-"فَعِل يفْعَل"، مثل: فَرِح يفْرَح.

-"فَعِل يفْعِل"، مثل: وَثِق يثِق.

أما المجرد الرباعي فله صورة واحدة، هي: فَعْلَلَ، المضارع منها هو: يُفَعْلِل.

تعد هذه الصيغ السبعة المجردة هي النواة الاشتقاقية للغة العربية، حيث يعتمدها النظام الصرفي لتوليد كل الصيغ الفعلية والأسماء المتصرفة، وذلك عن طريق زيادة أحد أحرف الزيادة المجتمعة في كلمة (سألتمونيها)، ثم تأتي الأسماء المتصرفة وتشتق من هذه الصيغ الفعلية المجردة أو المزيدة الثلاثية أو الرباعية.

2- تعدد المعاني الوظيفية للصيغة الواحدة:

وعلى الرغم من تكثير عدد الصيغ الصرفية عن طريق هذه الزيادات فإنها لم تغطِّ المعاني الوظيفية المطلوب التعبير عنها؛ فلجأ النظام الصرفي إلى نسبة أكثر من معنى وظيفي إلى الصيغة الواحدة اعتمادًا على وسائل تمييزية أخرى يوفرها أحد المستويات اللغوية الأخرى.

فصيغ الأسماء والأفعال محدودة والمعاني التي تعبر عنها كثيرة؛ ولذلك فإن اللغة تلجأ إلى الاشتراك[1]، وكما يكون هذا الاشتراك على مستوى الصيغة فإنه يقع في بنية الحروف والأسماء غير المتصرفة، ومن أمثلة ذلك:

دلالة صيغة "استفعل" على الطلب كاستخرج، والصيرورة كاستحجر، واعتقاد الشيء على صفة ما كاستصغر، والمطاوعة كاستقام، والاتخاذ كاستشعر، وحكاية الشيء كاسترجع، وقوة العيب كاستهتر والاستحقاق كاستحصد[2]، ودلالة "اللام" على الجر والأمر والتعليل، ودلالة (هما، أنتما) على المذكر والمؤنث[3].

ثالثًا: المستوى التركيبي:

يحمل النظام التركيبي للغة العربية العديد من الخصائص التي تتميز بها اللغة العربية، ومن أبرز هذه الخصائص **المرونة التركيبية**؛ فقد أفاد النظام التركيبي من الحركات الصوتية التي وفرها المستوى الصوتي للغة العربية في الدلالة على نوع الوظيفية النحوية التي تؤديها الكلمة؛ فصارت هذه الحركة الصوتية أو العلامة الإعرابية قرينة على وظيفة الكلمة وعنوانًا لدورها في الجملة مهما اختلف مكانها، هذا بالإضافة إلى ما يقدمه المستوى الدلالي من منطق عام يحكم الجملة، يعين أنواع العلاقات التركيبية التي يمكن أن تنشأ بين كلمات الجملة -وإن تغير مكانها في الجملة، واعتمادًا على ذلك؛ فقد أجاز النظام التركيبي لوحدات الجملة الواحدة العدول عن التزام الرتبة إلى حرية التقدم والتأخر، والعدول عن الإظهار إلى الإضمار، والعدول عن الذكر إلى الحذف، وقُيِّد ذلك العدول بأمن اللبس في الفهم البشري لها، هذه المرونة التركيبية في أحد وجهيها تعكس ما يتسم به النظام التركيبي للغة العربية من اقتصاد وثراء وجمال، وفي الوجه الآخر ينبئ عن حجم التحديات التي تقابل التمثيل الحاسوبي

1 مقترحات ضرورية في قواعد اللغة العربية، مصطفى جواد، ص25.

2 مقالات في اللغة والأدب، د. تمام حسان، 293.

3 قد يقع الترادف على مستوى الصيغة، انظر مثلا دلالة ما و لا على النداء، ودلالة صيغة انفعل وتفاعل على المطاوعة

للمستوى التركيبي للغة.

رابعًا: المستوى الدلالي:

لعل أهم الخصائص التي يتصف بها النظام الدلالي هي تعدد المعنى للكلمة الواحدة؛ فعلى الرغم من أن منطق الوضوح والإفادة يقتضي أن تدل الكلمة الواحدة على معنى واحد لئلا يدرك السامع غير ما يقصده المتكلم، فيقع اللبس وسوء الفهم- على الرغم من ذلك: فإن العقل الجمعي يسعى باللغة إلى إحداث التعدد في المعنى عن طريق التغيير في المعنى والإبقاء عليه في الثروة اللفظية المستعملة ما لم يؤدِّ إلى لبس؛ وذلك سعيًّا للاختصار وتوفير الجهد؛ إذ من شأن تخصيص كلمة لكل معنى أن تزيد عدد الكلمات بزيادة عدد المعاني، ولما كانت المعاني غير متناهية فهذا يقتضي أن تكون الكلمات هي الأخرى غير متناهية، وهذا لا يمكن أن تتحقق معه عملية الاتصال اللغوي؛ لأنه سيصعب حينئذ على متكلم اللغة وفقًا لقدراته العقلية أن يستوعب هذا الكم الهائل، فيكتفي منه بما يستطيعه عقله، ومن ثم قد يختلف ما استوعبه عما استوعبه بقية أفراد أبناء اللغة فينقطع الاتصال، إذن لم يكن من بد أن تدل الكلمة الواحدة على أكثر من معنى حتى تقل عدد المفردات بما يمكن استيعابها.

ما يترتب على تعدد المعنى ودور السياق:

"إن من طبيعة المعنى المعجمي أن يكون متعددًا ومحتملًا، وهاتان الصفتان من صفاته تقود كل منهما إلى الأخرى، فإذا تعدد معنى الكلمة المفردة حال انعزالها تعددت احتمالات القصد، وتعدد احتمالات القصد يعتبر تعددًا في المعنى"[1]، والعقل الجمعي للغة العربية -إذ يُقدِم على هذه الخطوة، أي: السماح للكلمة الواحدة أن تدل على أكثر من معنى دون خشية اللبس والوقوع في الغموض، فهو- إنما يعتمد على السياق الذي تَرِد فيه، فالسياق يكسب الكلمة المتعددة المعنى معنى سياقيًّا واحدًا، ومن ثم ترتفع صلاحية بقية المعاني، فذلك

1 - اللغة العربية معناها ومبناها، د. تمام حسان، ص 323.

هو "نفوذ السياق الذي يجعلنا نعطي كلمة ما بضعة معاني مختلفة دون خشية الخلط. ونعتمد على السياق الذي يحدد المعنى المراد ويستبعد المعاني الأخرى من الذهن... فمن أمثلة ذلك في اللغة العربية كلمة "عين" التي تستعمل حتى الآن في أكثر من معنى دون خوف الالتباس اعتمادًا على دلالة السياق. فقولنا: تفجرت عين في الصحراء غير قولنا: دمعت عين فلان"[1]، "وقد أدى الاعتماد على السياق إلى أن تعيش كثير من كلمات المشترك اللفظي جنبًا إلى جنب عدة قرون في اللغة الواحدة دون أن يسبب ذلك غموضًا أو سوء فهم أو حتى صعوبة من نوع ما"[2]، "وإنه لمما ينهض دليلًا قاطعًا على أهمية السياق والمقام في التبادل اللغوي أن الناس يستطيعون في مثل هذه الظروف أن يتفاهموا فيما بينهم تفاهمًا وضاحًا صريحًا لا غموض فيه، إن مقدرة الكلمات على أداء وظيفتها لا تتأثر بحال من الأحوال بعدد المعاني المختلفة التي قدر لها أن تحملها، بدليل أن بعض هذه الكلمات تستطيع بالفعل أن تقوم بعشرات الوظائف في سهولة ويسر"[3].

وعلى الرغم من دور السياق في تحديد المعنى المناسب من بين المعاني التي تشير إليها الكلمة وهي منفردة فإن السياق قد يعجز عن تحديد المعنى المراد فيترتب على ذلك غموض قد يفسد بقية الجملة بدلًا من أن يكتسب منها الوضوح، وهذه الحالة هي المسئول الرئيسي عن توليد الغموض ونموه [4].

المحور الثالث: المعالجة الفنية لوثائق المحتوى الرقمي العربي:

تختلف طبيعة المعالجات الفنية لوثائق المحتوى الرقمي العربي باختلاف المستوى اللغوي محل المعالجة؛ فمعالجة المستوى الصوتي تختلف عن المستوى الصرفي، وهما يختلفان عن المستوى الدلالي والنحوي، فلكلٍّ خصائصها وأدواتها، كما تتنوع هذه المعالجات الفنية تبعًا لمدى

1 – علم الدلالة ص، د. أحمد مختار عمر، 186.

2 – علم الدلالة، د. أحمد مختار عمر، ص 187.

3- دور الكلمة في اللغة، ستيفن أولمان، ترجمة د. كمال بشر، ص 137.

4 علم الدلالة، د. أحمد مختار عمر، ص 187، دور الكلمة في اللغة، ستيفن أولمان، ترجمة د. كمال بشر، ص 136.

التحكم المستهدف في هذه النصوص، والتطبيقات التي تبنى عليها، وفيما يلي عرض لمستويات المعالجة الفنية لمحتوى الوثيقة العربية:

أولًا: المعالجة الصوتية:

تستهدف المعالجة الصوتية تجهيز المحتوى العربي النصي للتقنيات والتطبيقات الصوتية التي يمكن أن تبنى عليه، ولعل أهمها: تقنية تحويل النص المكتوب إلى كلام منطوق TTS، وتقنية التعرف الآلي على الصوت البشري SR، ويمر التحليل الصوتي اللازم لبناء هاتين التقنيتين بأكثر من خطوة، ويحتاج للعديد من الإجراءات التحليلة، توضيحها على النحو التالي:

أولًا: التحليل الصوتي اللازم لبناء تقنية TTS (Text To Speech): التي تحول النص المكتوب إلى كلام منطوق يمر بثلاث خطوات أساسية، هي [1]:

1-ضبط المحتوى اللغوي المكتوب ضبطًا كاملًا.

2-التسجيل الصوتي للمحتوى اللغوي، ويُراعَى فيه دقة التسجيل، وتنوُّع الأصوات التي تُسجَّل؛ من حيث: السن، والجنس، واللهجة، إلى غير ذلك من الاعتبارات التي تُراعَى في ضوء الأهداف المقررة.

3-التحليل الصوتي للمادة المسجلة، ويمر بالخطوات التالية:

3-1-المونتاج؛ لحذف أخطاء التسجيل والملفات غير النقية أو التي زادت بها نسبة الضوضاء عن المعيار المسموح به.

3-2-تقطيع الملف إلى مجموعة من الدفقات الصوتية (Utrans).

3-3-إعادة ضبط الملفات النصية وتشكيلها تشكيلًا صوتيًّا حسب المنطوق.

3-4-توليد الكتابة الصوتية حسب قواعد الأبجدية الدولية (IPA) أو (SAMPA).

3-5-إدخال ملفات المادة الصوتية (Wave.*) وملفات الكتابة الصوتية (trans.*) على

1- فونولوجيا القرآن الكريم، د. أحمد راغب أحمد، ص53-57.

آلية (HTK) (Hidden Markov Model Tool Kit) لاستخراج ملفات التحليل الزمني لقاعدة البيانات (Lab.*).

3-6-مراجعة التحليل الزمني بواسطة متخصصي فونولوجيا العربية (Hand Lab).

3-7-وضع الرموز الخاصة بالنبر الأولي والثانوي ونبر المقطع والنبر الدلالي.

3-8-وضع الرموز الخاصة بالمستوى التنغيمي حسب قواعد الرموز الدولية.

وهذه التحليلات إنما تتم على عينة لغوية فقط وليس على عموم نصوص الوثائق، غير أنه يراعى في هذه العينة أن تشتمل على جميع التنوعات الصوتية التي يتسم بها النظام الصوتي للغة العربية، ثم يبنى من نتائج تحليل هذه العينة أداة حاسوبية قادرة على تحويل أي نص خارج العينة إلى أداء صوتي، تتوقف دقته وجودته على انضباط العينة المحللة، ودقة التحليلات الصوتية السابقة التي أجريت عليها. وقد يقتصر هدف المعالجة الصوتية على تسجيل العناوين والعبارات المهمة في الوثائق بأصوات طبيعية، ليسترجعها المستخدم على حالتها دون تدخل آلي في الأداء الصوتي لها، غير أن هذا القدر من المعالجة يقيد حرية المستخدم في الاستماع للمحتوى اللغوي الذي سبق تسجيله فقط، كما لا يسمح بالتنقيب الصوتي عن معلومات الوثائق.

ثانيًا: التحليل الصوتي اللازم لبناء تقنية التعرف الآلي على الصوت البشري SR (Speech Recognition):

يعتمد هذا النوع من التحليل الصوتي على بناء قاعدة بيانات صوتية، تشتمل على تحليل صوتي لمادة لغوية، تتمثل فيها جميع التنوعات الفونيمية للغة العربية، مسجلة بعدد كبير من الأصوات البشرية تغطي التنوعات الأدائية المختلفة بما يؤثر فيها من سن أو جنس أو ثقافة أو لهجة، ويمر بناء قاعدة البيانات هذه بالمراحل التالية[1]:

1- تحديد المادة اللغوية المراد تسجيلها، وضبطها ضبطًا كاملًا.

1-فونولوجيا القرآن الكريم، د. أحمد راغب أحمد، ص53-57.

2- تسجيل الكلمات المراد تحليلها، ويُرَاعى في التسجيل الضوابط التالية:

2-1- تنوُّع دقة التسجيل الصوتي للمادة اللغوية؛ فلا تقتصر على المواصفات القياسية، بل يتم تسجيل بعضها في مواقف بها نسب ضوضاء متفاوتة ومتنوعة.

2-2- تعدد الأصوات التي تسجل وتنوُّعها؛ بحيث تشمل مئات الرجال، والسيدات، والأطفال.

2-3- تنوع المستويات اللغوية للأصوات المسجلة؛ بحيث تشمل مثقفين لغويًّا وغير مثقفين.

2-4- تنوع اللهجات واللكنات، وقد تشمل تسجيلات لغير الناطقين بالعربية كلغة أولى إذا كانوا في إطار الفئة المستهدفة من المعجم.

2-5- الإفادة من العادات النطقية؛ كالكحة والعطس والتثاؤب واللعثمة والتمتمة.

3-التحليل الصوتي للمادة المسجلة، ويمر بالخطوات التالية:

3-1-المونتاج، لحذف أخطاء التسجيل والملفات غير المطابقة والمكررة.

3-2-تقطيع الملف إلى مجموعة من الدفقات الصوتية (Utrans).

3-3-إعادة ضبط الملفات النصية وتشكيلها تشكيلًا صوتيًّا حسب المنطوق.

3-4-توليد الكتابة الصوتية حسب قواعد الأبجدية الدولية (IPA) أو (SAMPA).

3-5-إدخال ملفات المادة الصوتية (Wave.*) وملفات الكتابة الصوتية (trans.*) على آلية (HTK) (Hidden Markov Model Tool Kit) لاستخراج ملفات التحليل الزمني لقاعدة البيانات (Lab.*).

3-6-مراجعة التحليل الزمني بواسطة متخصصي فونولوجيا العربية (Hand Lab).

3-7-وضع الرموز الخاصة بالنبر الأولي والثانوي ونبر المقطع والنبر الدلالي.

3-8-وضع الرموز الخاصة بالمستوى التنغيمي حسب قواعد الرموز الدولية.

3-9-استنباط شجرة الأخطاء النطقية التي يمكن من خلالها التنبؤ بمجموعة الأخطاء النطقية التي تعرض للمتحدث عند نطقه، كنطق القاف جيمًا أو همزة أو جافًا

فارسية أو كافًا...إلخ.

ثانيًا: المعالجة الصرفيّة:

يهدف هذا النوع من التحليل إلى توفير المعلومات الصرفية اللازمة لبناء محلل صرفي للغة العربية، وبناء تقنيات التنقيب المعلوماتي المعتمد على بنية الكلمة، ويتمثل التحليل الصرفي للكلمة في:

1- تحليل الكلمة إلى عناصرها الصرفية الأساسية، وهي: الجذر والوزن والسابقة واللاحقة.

2- تحديد الصفات الصرفية الأساسية لكل كلمة، كتحديد نوعها من حيث: النوع، التعيين، والجنس، والعدد، والجمود والاشتقاق ... إلخ.

3- فك صور الالتباس الصرفي المتوقعة بين الكلمات، سواء في كانت في الجذور أو في الأوزان، أو في السوابق، أو في اللواحق.

ثالثًا: المعالجة الموضوعية للمحتوى الرقمي:

يهدف التحليل الموضوعي للمحتوى الرقمي إلى تمكين أدوات التنقيب المعلوماتي عن البحث عن أصناف المعلومات المختلفة التي يشتمل عليها النص اللغوي، وتفعيل الإفادة من المعلومات الضمنية المبثوثة بين كلماته، والتي لا تتمكن أدوات التنقيب المعلومات التي تعتمد على بنية الكلمة من الوصول إليها، بالإضافة إلى تمثيل المنطق العام الذي تترابط به الكلمات فيما بينها المتمثل في العلاقات الدلالية والمنطقية التي تعيد تقديم النص اللغوي على هيئة شبكة من المفاهيم المترابطة، ونظرًا لأهمية هذا المستوى من المعالجة وجدته فسأعطيه قدرًا أكبر من التفصيل، وذلك على النحو التالي:

أولًا: مفاهيم أساسية:

1- الدلالة المعجمية والدلالة التركيبية: الكلمة -وهي بمعزل عن السياق- يكون لها دلالة معجمية، محفوظة لها في المعجم، تستخدم بها في سياقاتها المتنوعة مع قليل أو كثير من التغير الدلالي توسيعًا أو تضييقًا، رقيًّا أو انحطاطًا، أو انتقالًا إلى معاني أخرى اعتمادًا على

إحدى طرق المجاز اللغوي، أما الدلالة التركيبية فهي مجموع دلالة الكلمة المعجمية والدلالات الوظيفية الأخرى: الصوتية والصرفية والنحوية.

2- نسبية إدراك دلالة الكلمة: يتفق أبناء الجماعة اللغوية الواحدة في قدر مشترك من الثروةِ اللفظية لتلك اللغة، والنظامِ اللغوي الذي يحكم تفاعل تلك الثروة فيما بينها دلاليًّا وتركيبيًّا، مع احتفاظ كل فرد من أفراد الجماعة اللغوية الواحدة بخصائص تميزه في إبداع اللغة وفهمها تبعًا لقدراته العقلية والنفسية والاجتماعية وتبعًا لمخزونه من رصيد لغته، مما أدى إلى نسبية بين أبناء الجماعة اللغوية الواحدة في إدراك معنى ما، أو في التعبير عنه.

3-تدرج دلالة الكلمة من المركزية إلى الهامشية: تصطحب الكلمة في سياقاتها المتنوعة عددًا من المعاني الجزئية، من هذه المعاني ما يثبت معها في جميع السياقات، ومنها ما يكثر اصطحابها له، ومنها ما تختص به بعض السياقات القليلة، فالمعاني التي تلازم الكلمة في استخداماتها المتنوعة هي المعاني المركزية، وهي التي يتفق المتكلمون باللغة في فهمها والتعبير عنها، أما المعاني التي تتخلف عن الكلمة -كثيرًا أو قليلًا- فإن دورها في الإفصاح عن معنى الكلمة يتدرج من القوة إلى الضعف أو من المركزية إلى الهامشية بمقدار اصطحاب الكلمة لها في سياقاتها أو تخلفها عنها، وهذه المعاني هي التي يتفاوت أبناء اللغة في إدراكها.

4-الالتباس الدلالي: يقصد به -على المستوى المعجمي- إشارة الكلمة الواحدة إلى أكثر من معنى؛ مثل إشارة كلمة [العين] إلى المعاني: جزء من الرأس، بئر الماء، الجاسوس...إلخ، أما على المستوى التركيبي فيقصد به عدم قدرة العقل البشري على ترجيح أحد معاني الكلمة لافتقار السياق إلى القرائن اللغوية وغير اللغوية اللازمة لذلك الترجيح.

أما على مستوى المعالجة الآلية فإن جميع كلمات اللغة ملبسة لافتقاد الحاسوب المعارفَ والمسلماتِ المنطقيةَ التي يختزنها العقل البشري، ويعتمد عليها في الترجيح بين معاني الكلمة الواحدة حال تسييقها، ومن ثم: فإن التحليل الدلالي الحاسوبي للغة العربية يهدف إلى توصيف نظامها الدلالي وربط كل كلمة بمعانيها المختلفة على نحو يمكن الحاسوب من التعامل معها، مع الأخذ في الحسبان خصائص ذلك النظام؛ كتدرج المعاني الجزئية للكلمة

الواحدة من المركزية إلى الهامشية، ونسبية إدراك الأفراد لها، وقابلية نشوء اللبس والغموض في فهم المقصود منها في السياق.

ثانيًا: التحليل الموضوعي الرقمي للوثيقة:

يهدف هذا التحليل إلى تعيين المعلومات الأساسية المتصلة بالنص، والموضوعات المهمة التي يتناولها، ويتمثل في الإجراءات التالية:

1-تحديد ببليوجرافيا الوثيقة، وتشتمل على المعلومات التالية:

1-1عنوان الوثيقة.

1-2-الفن الذي جاءت الوثيقة في قالبه، مثل: قرآن كريم، حديث شريف، شعر، خطابة، قصة، رواية، كتابة صحفية، كتابة علمية..إلخ

1-3-الناشر، وسنة النشر، ومكان النشر، والطبعة المعتمدة.

1-4-مصدر الوثيقة إذا كانت بحثا أو مقالا، مثل: الإنترنت (مع إدراج الرابط الخاص)، مجلة (مع ذكر الببليوجرافيا الكاملة لها)، أعمال المؤتمرات...إلخ

1-5- ما يتصل بمؤلف الوثيقة: **من حيث:**

1-5-1-نسبه ونشأته ووفاته.

1-5-2-مذهبه الديني واتجاهه السياسي، مثل: سنِّي، أمويّ...إلخ

1-5-3-الوظائف والمناصب التي تقلدها بالإضافة إلى اشتغاله بالتأليف، مثل: الإمارة، القضاء، رئاسة ديوان الإنشاء...إلخ

1-5-4-خصائصه النفسية والاجتماعية، مثل: معتد بنفسه، شعوبي، مقرب من الخلفاء.

1-6- ما يتصل بالمجتمع الذي ألفت فيه الوثيقة: **من حيث:**

1-6-1-الحالة السياسية والاقتصادية.

1-6-2-المستوى الاجتماعي والثقافي والحضاري.

1-7- ما يتصل بالمرحلة الزمانية والبيئة الجغرافية التي ألفت الوثيقة خلالهما.

2- شجرة موضوعات الوثيقة:

وهي عبارة عن بناء تنظيمي للموضوعات المهمة المستخلصة من الوثيقة، وهي تتألف من الموضوع العام الذي يترأس الشجرة، ويتفرع عنه موضوعات أقل عمومية إلى أن تنتهي إلى الموضوعات الفرعية التي تتذيل الشجرة.

3- الكلمات المفتاحية للوثيقة:

وهي الكلمات الأساسية في الوثيقة، المرشدة إلى أهم الموضوعات التي تتناولها الوثيقة، ويتم تحديدها من قبل متخصصي الفن الذي تنتمي إليه الوثيقة في ضوء مبادئ مناهج الفهرسة والأرشفة.

4-ملخص الوثيقة:

وهو عبارة عن عرض موجز لمضمون الوثيقة، يساعد متصفح الوثيقة في الإلمام السريع بمضمونها، وتحديد ما إذا كانت الوثيقة مفيدة له أم لا، وينهض بهذه المهمة المتخصصون في الفن الذي تنتمي إليه الوثيقة ذوو القدرة على إعادة التعبير المختصر المفيد.

ثالثًا: فهرسة الأعلام:

تمثل الأعلام -بأنواعها المختلفة- العمود الفقري في أية وثيقة، فهي محور الأحداث، ولذلك فتحديدها، وتصنيفها يفيد كثيرًا في تحديد موضوع الوثيقة، والقضايا التي تتناولها، وأبعادها الثقافية والحضارية. ومن أهم أنواع الأعلام التي تشتمل عليها الوثائق ما يلي:

1-أسماء الأشخاص وألقابهم، مثل: سعد زغلول، رئيس مجلس النواب.

2-أسماء البلدان وأهم مدنها، مثل: مصر، القاهرة، الإسكندرية.

3-أسماء الأمكنة والمزارات، مثل: قارة آسيا، البحر الأحمر، مكة المكرمة.

4-أسماء الأحداث التاريخية: الهجرة النبوية المشرفة، فتح مكة، موقعة عين جالوت، حرب أكتوبر.

5-أسماء الأزمنة والعصور: العصر الحجري، العصر الحديث.

6-أسماء المنتجات والرموز الاجتماعية والثقافية، مثل: سيارة فورد، الأهرامات، دار الأوبرا، حاتم الطائي، إلياس.

7-مصطلحات العلوم والفنون، مثل: الأكسجين، البناء الضوئي، علم الدلالة، الفن الشعبي.

رابعًا: التحليل المفهومي للمحتوى الرقمي:

يهدف هذا التحليل إلى بناء الشبكة الدلالية التي تعبر عن المفاهيم والمعلومات التي تشتمل عليها كل وثيقة، وبيان الروابط المنطقية والدلالية بينها.

وتتألف الشبكة الدلالية من عدد من العقد الدلالية، يرتبط بكل عقدة دلالية مفهوم من المفاهيم، والخيوط البينية التي تنسج منها الشبكة هي العلاقات الدلالية والمنطقية بين تلك المفاهيم، وتنشأ العقدة الدلالية من التقاء علاقتين أو أكثر عند أحد هذه المفاهيم، بحيث تعد كل عقدة بالمفهوم المرتبط بها مركزًا دلاليًّا ينبعث منه عدد من العلاقات تنتهي كل علاقة منها بعقدة دلالية أخرى مرتبط بها مفهوم آخر، وهذه العلاقات هي تمثيل دقيق لمعنى الكلمة أو الكلمات المندرجة تحت المفهوم؛ فالمعنى المقصود من الكلمة كما يمكن توضيحه بالشرح والتفسير؛ فإنه يمكن توضيحه أيضًا عن طريق الشبكة الدلالية من خلال رصد العلاقات الدلالية التي يرتبط بها مفهوم تلك الكلمة بالمفاهيم الأخرى.

وبناءً على ذلك الشكل الشبكي يمر بخطوتين: الأولى: هي تكوين ثنائيات دلالية من المفاهيم، والثنائية الدلالية هي كل مفهومين ارتبطا بإحدى العلاقات الدلالية، أما الخطوة الأخرى: فهي إنشاء سلاسل دلالية عن طريق الربط بين تلك الثنائيات، وهذه السلاسل الدلالية تزداد طولًا باندماج المزيد من الثنائيات، وتتقاطع فيما بينها مكونة الشبكة الدلالية العامة [1]. ومن أهم العلاقات الدلالية والمنطقية التي تترابط بها المفاهيم مكونة الشبكة

(1) هذا وقد سبق تطبيق هذه الأفكار بنجاح على مفردات اللغة الإنجليزية، وهناك محاولات لتطبيقها على اللغة العربية ، انظر المواقع التالية على شبكة الإنترنت: http://www.VisualThesaurus.com و http://www.visuwords.com/ و =

الدلالية المرئية ما يلي:

1- علاقات معيار العموم والخصوص:

يندرج تحت هذا المعيار ثلاث علاقات تتفق في أنها تربط بين مفهومين؛ أحدهما عام والآخر خاص، وتختص كل واحدة بما يلي:

العلاقة الأولى: الاشتمالية النوعية، ومعكوستها (النوعية الاشتمالية):

تكون بين طرفين الخاص منهما نوع للعام، فإذا كان المفهوم العام هو الطرف الأول من الثنائية الدلالية كانت العلاقة اشتمالية نوعية، مثل: الطائر / الهدهد- حسن الخلق / الصدق، وإذا كان المفهوم الخاص هو الطرف الأول من الثنائية كانت العلاقة نوعية اشتمالية، مثل: الهدهد/ الطائر- الصدق / حسن الخلق

العلاقة الثانية: الاشتمالية العضوية ومعكوستها (العضوية الاشتمالية):

تكون بين طرفين الخاص منهما عضو في العام، فإذا كان المفهوم العام هو الطرف الأول من الثنائية كانت العلاقة اشتمالية عضوية، مثل: النساء/ المرأة، وإذا كان المفهوم الخاص هو الطرف الأول من الثنائية كانت العلاقة عضوية اشتمالية، مثل: المرأة/ النساء.

والفرق بين العضوية والنوعية أن الأفراد المندرجة تحت العام في العضوية متماثلة، أما في النوعية فغير ذلك، فالعلاقة بين مجموعة الناس ومجموعة النساء، وكذلك مجموعة الرجال ومجموعة الأطفال علاقة اشتمالية نوعية؛ لأن مجموعة النساء تختلف عن مجموعة الرجال، وهما يختلفان عن مجموعة الأطفال.. أما العلاقة بين مجموعة النساء والمرأة -وكذلك بين مجموعة الرجال والرجل- فهي علاقة اشتمالية عضوية؛ لأن كل الأفراد المندرجة تحت مجموعة النساء متماثلة وتسمى امرأة، وكذلك كل الأفراد المندرجة تحت مجموعة الرجال متماثلة وتسمى رجلًا.

=

http://www.globalwordnet.org/AWN/AWNBrowser.html و http://www.ArabicWordMap.com/ والأخير موقع بحثي خاص بمقدم البحث.

العلاقة الثالثة: الكلية الجزئية ومعكوستها (الجزئية الكلية):

تكون بين طرفين ماديين، الثاني منهما خاص وهو جزء من الأول العام؛ فإن كان المفهوم العام هو الطرف الأول كانت العلاقة كلية جزئية، مثل: الإنسان/ العين-الغراب/ الجناح، وإذا كان المفهوم الخاص هو الطرف الأول من الثنائية كانت العلاقة جزئية كلية، مثل: العين/ الإنسان- الجناح/ الغراب.

2- علاقتا معيار التماثل والتنافر:

يندرج تحت هذا المعيار علاقتا الترادف والتضاد.

العلاقة الأولى: علاقة الترادف:

تكون بين مفهومين متطابقين أو متقاربين في المحتوى الدلالي لكليهما؛ بحيث ينوب أحدهما عن الآخر في الإشارة إلى مفهوم عام واحد، مثل العدل والقسط؛ فكلاهما يشير إلى المفهوم العام: إعطاء كل ذي حق حقه.

العلاقة الثانية: علاقة التضاد:

تكون بين مفهومين متعاكسين في المحتوى الدلالي لكليهما؛ بحيث يشير أحدهما إلى عكس ما يشير إليه الآخر، مثل: العدل والظلم، فالعدل يشير إلى المفهوم العام: إعطاء كل ذي حق حقه، والظلم يشر إلى المفهوم العام: هضم الحقوق.

3- علاقتا معيار السبب والنتيجة:

تربط علاقة السببية بين طرفين غالبًا ما يكُونان من الأحداث، ويكون أحدهما سببًا، والآخر نتيجة، وتحقق النتيجة إما أن يكون حتميًّا، مثل: الإفساد/الخسارة * الجفاف/الضرر* الغباء/الفشل.

أو أن يكون احتماليًّا، مثل: الأكل/ الشبع* الشرب/الارتواء.. فالأكل والشرب هما -وإن كانا سببًا في الشبع والارتواء- إلا أن هذه النتيجة قد تتخلف أحيانًا.

4- معيار الحالية الزمانية والمكانية:

العلاقة الأولى: الحالية (الظرفية) الزمانية ومعكوستها: تكون بين طرفين، أحدهما ظرف

زمان يتم فيه الطرف الآخر، وهو إما أن يكون حدثًا أو اسم ذات، مثال: السهر/ الليل- الشمس/ النهار

العلاقة الثانية: الحالية (الظرفية) المكانية ومعكوستها: تكون بين طرفين؛ أحدهما ظرف مكان يتم فيه الطرف الآخر، وهو إما أن يكون حدثًا، أو اسم الذات، مثل: الماء/الدورق- القمر/السماء- السباحة/البحر.

المحور الرابع: تقنيات التنقيب المعلوماتي في المحتوى الرقمي العربي:

تتنوع التقنيات اللازمة للتنقيب والبحث عن المعلومات في الوثائق المرقمنة بتنوع المعلومات المنقَّب عنها، وكيفية الوصول إليها، وتوضيح ذلك على النحو التالي:

أولًا: باعتبار نوع المعلومات المنقب عنها، تنقسم إلى نوعين أساسيين:

النوع الأول: تقنيات البحث البنيوي المطابق:

هذا النوع من البحث وحدته الأساسية هي الكلمة، وله أنواع منها:

1- البحث المطابق في جميع حروف الكلمة: إذ يستدعي محرك البحث الكلمات المتطابقة في جميع الحروف مع كلمة البحث، مثل: كلمة (رجال) لا يستدعي لها إلا كلمة (رجال).

2- البحث المطابق في بعض حروف الكلمة: سواء كانت المطابقة في بنية الكلمة الأصلية أو في السوابق واللواحق، مثل: كلمة (رجل) يمكن أن يستدعي لها محرك البحث الكلمات: (رجل- رجلي- رجلان- أرجل- الرجل- الرجال..).

3- البحث المطابق في جسم الكلمة مع ثبات شكله وتغير السوابق واللواحق: مثال: كلمة (رجل) يستدعي لها محرك البحث الكلمات (ورجل، الرجل، فالرجل، برجل، رجلان).

4- البحث المطابق في الجذر وإن اختلفت صيغ الكلمات: مثال: الجذر (ر ج ل) يستدعي لها محرك البحث الكلمات (رَجُل، راجل، رِجْل، ترجَّل...).

قد تشترط صور المطابقة السابقة الاتفاق في التشكيل البنيوي والإعرابي، أو في أحدهما فقط أو التعري منه تمامًا، فالبحث بكلمة (رَجُلٌ) في حالة التطابق في التشكيل البنيوي والإعرابي لا يستدعي محرك البحث إلا كلمة (رَجُلٌ)، وفي حالة التشكيل البنيوي

فقط يمكن أن يستدعي (رَجُلاً- رَجُلٍ)، وفي حالة الاتفاق في التشكيل الإعرابي فقط يمكن أن يستدعي (رِجْلٌ)، وفي حالة إهمال التشكيل البنيوي والنحوي يمكن أن يستدعي الكلمات: (رِجْل، رَجَّلَ، رَجِل..إلخ).

النوع الثاني: تقنيات البحث الموضوعي:

هذا النوع من البحث وحدته الأساسية هي المعنى وليس بنية الكلمة، وله صورتان:

أولًا: البحث الموضوعي الخاص:

وهو ما يكون في إطار البحث في المعلومات المستخلصة من كل وثيقة، وهذه المعلومات هي[1]:

1- الترويسة. 2- شجرة الموضوعات.

3- الكلمات المفتاحية. 4- الملخص.

ثانيًا: البحث الموضوعي العام:

ويكون في عموم مفاهيم الوثيقة، ويقصد به البحث عن مفهوم أو أكثر متصل بالمفهوم المبحوث به بعلاقة أو أكثر من العلاقات الدلالية التي تترابط بها المفاهيم في قاعدة البيانات الدلالية، وله صور متعددة، لعل أهمها[2]:

1- أن يذكر الباحث مفهومًا، ويستدعي محرك البحث ما يرتبط بهذا المفهوم من مفاهيم بجميع العلاقات المحددة، مثل: العدل، يستدعي المفاهيم: الفضيلة، القسط، الأمان، الظلم...إلخ

2- أن يحدد الباحث مفهومًا وإحدى العلاقات الدلالية ويستدعي ما يتصل بالمفهوم المحدد من مفاهيم وفقًا لهذه العلاقة، كأن يحدد مفهوم العدل وعلاقة الترادف، فيستدعي

(1) لمزيد من التفصيل حول هذه المعلومات، انظر المحور الثالث: المعالجة الفنية للمحتوى الرقمي: المعالجة الموضوعية من البحث.

(2) هذه الطرق البحثية وغيرها موظفة في قاعدة البيانات الدلالية التي أشرف عليها البرنامج الإنمائي للأمم المتحدة، ووزارة الاتصالات المصرية، والشركة الهندسية لتطوير نظم الحاسبات RDI. انظر الرابط التالي:

http://www.rdi-eg.com/rdi/Downloads/Process_of_building_the_forward_ Arabic_Lexical_Semantic_DB.pdf.

مفهوم القسط، أو علاقة التضاد فيستدعي مفهوم الظلم، أو علاقة العموم فيستدعي مفهوم الفضيلة.

ثانيًا: باعتبار كيفية الوصول إلى المعلومة المنقب عنها، تنقسم إلى نوعين أساسيين:

النوع الأول: تقنيات البحث الكتابي:

تتم عملية البحث في هذا النوع من التقنيات عن طريق كتابة المستخدم الكلمة أو المفهوم الذي يريد أن يبحث عنه وفق إحدى الطرق السابقة، ويكون ناتج البحث مكتوبًا أيضًا، وقد يكون منطوقًا، وذلك يتطلب تطوير تقنية تحويل النص المكتوب إلى منطوق آليًا[1].

النوع الثاني: تقنيات البحث الصوتي:

تتم عملية البحث عن طريق إدخال الكلمة أو المفهوم المراد البحث عنه نطقًا، وذلك يتطلب تطوير تقنية التعرف الآلي على الصوت المنطوق SR، وتستدعي التقنية الكلمة أو المفهوم المقصود مكتوبًا، وتحويله إلى صوت منطوق يتطلب تطوير تقنية التحويل الآلي للنص المكتوب إلى منطوق TTS، وتتطلب بناء التقنيتين تهيئة النص اللغوي، وذلك بإجراء عدد من التحليلات الصوتية عليه، وبناء قواعد بيانات صوتية متكاملة، تشتمل على جميع التنوعات الفونيمة المختلفة لجميع أصوات اللغة العربية، كما تمثل جميع أنواع الأداء الصوتي، بما يعتريه من نبر وتنغيم وظواهر صوتية أخرى يختص بها نطق الرجل عن نطق المرأة أو نطق الكبير عن نطق الصغير، إلى غير ذلك من الخصائص والإجراءات التي تم تفصيلها سابقا[2].

1-انظر المحور الثالث: المعالجة الفنية للمحتوى الرقمي: المعالجة الصوتية من البحث.

2- انظر المحور الثالث: المعالجة الفنية للمحتوى الرقمي: المعالجة الصوتية من البحث.

الخاتمة

خلص البحث إلى إبراز مجموعة من النتائج العامة، هي على النحو التالي:

1- النمو المتسارع للمحتوى المعلوماتي العالمي وضعف مواجهته وتذليل عقبات التحكم الآلي فيه يدفع به إلى أنفاق مغلقة من صراعات ثقافية وحضارية تحرم الحياة الثقافية العالمية من رحابة التعدد وثراء التنوع.

2- إن نجاح بناء المحتوى الرقمي العربي وجودة تقديمه للمستخدمين يتوقف على إدارة جيدة تلملم الجهود العربية المبعثرة، وتعيد توزيعها على نحو يحسن استثمارها وتوظيفها، وترتب أولوياته وضروریاته في ضوء مستجدات العصر، كما تجمع بين مصادره المتنوعة والمتباعدة أحيانًا -زمانًا أو مكانًا، فتمنع بذلك من وقوع الاضطراب في توازنه النوعي والكمي، أو التأخر في تطوير أدواته وتقنياته الحاسوبية.

3- تعدد صور الاضطراب الكمي والنوعي التي يعانيها المحتوى الرقمي العربي.

4- ضرورة تعدد المعالجات الفنية التي ينبغي أن تجرى على النصوص العربية، لتوفير التحليلات المطلوبة لبناء الأدوات والتقنيات الحاسوبية المتنوعة.

5- ضرورة الابتكار والتجديد في إعادة تمثيل المعرفة الضمنية التي يختزنها النص على نحو مفهومي شبكي يمكن للحاسوب أن يتعامل معه.

6- لا تستطيع أدوات التنقيب المعلوماتي الحالية التي تعتمد المطابقة في بنية الكلمة أن تزود المستخدم بكل المعلومات والمعارف التي يختزنها النص بين كلماته وجمله.

مصادر البحث ومراجعه

1- أصول تراثية لنظرية الحقول الدلالية، د. أحمد عزوز.

2- الاقتصاد في النحو العربي، د. وليد خيرالله، رسالة ماجستير، مخطوطة بمكتبة كلية دار العلوم، جامعة القاهرة، 2005.

3- ثورة المعرفة، رأس المال الفكري ومؤسسة القرن الحادي والعشرين، توماس ستيوارت، ترجمة: د. أحمد صلاح، الدار الدولية للاستثمارات الثقافية.

4- دور الكلمة في اللغة، ستيفن أولمان، ترجمة كمال بشر، مكتبة الشباب، القاهرة، ط10، 1987م.

5- العربية نحو توصيف جديد في ضوء اللسانيات الحاسوبية، د. نهاد الموسى، المؤسسة العربية للدراسات والنشر، ط1، 2000م.

6- العلاقات الدلالية والتراث البلاغي العربي؛ دراسة تطبيقية، د. عبد الواحد حسن الشيخ، مكتبة الإشعاع الفنية، الطبعة الأولى، 1419هـ-1999م، الإسكندرية.

7- علم الدلالة، أحمد مختار عمر، عالم الكتب، الطبعة الخامسة، 1998م، القاهرة.

8- الغموض في الدلالة، د. محمد أحمد حماد، رسالة دكتوراه، مخطوطة بمكتبة كلية دار العلوم، القاهرة.

9- الفجوة الرقمية. رؤية عربية لمجتمع المعرفة، د. نبيل علي، د. نادية حجازي، سلسلة عالم المعرفة، العدد 318، أغسطس 2008م.

10- فعالية إستراتيجية الخريطة الدلالية في تنمية مهارات الفهم القرائي لدى تلاميذ المرحلة الإعدادية، د. جمال سليمان عطية، 1999م، رسالة ماجستير، مخطوطة بمكتبة كلية التربية ببنها، جامعة الزقازيق، مصر.

11- فونولوجيا القرآن الكريم. دراسة لأصوات القرآن الكريم في ضوء علم الأصوات الحديث، د. أحمد راغب أحمد، رسالة ماجستير، مخطوطة بمكتبة كلية الآداب، جامعة عين

شمس، 2003م.

12- اللغة العربية في تكنولوجيا المعلومات. تطور واعد وتطوير متواصل، وقائع الندوة الدولية 2002م، المجلس الأعلى للغة العربية، الجزائر.

13- اللغة العربية: معناها ومبناها، د. تمام حسان، الطبعة الثالثة، عالم الكتب، 1998م.

14- اللغة العربية والحاسوب، د. نبيل علي، دار تعريب، 1988م.

15- المعاجم العربية المدرسية، دراسة لغوية في المادة والمنهج، حسين البسومي، رسالة ماجستير، مخطوطة بمكتبة كلية دار العلوم، القاهرة، 2004م.

16- المعجم الحاسوبي للغة العربية وأبعاده التعليمية، حسين البسومي، أعمال مؤتمر التعليم الإلكتروني والتعلم عن بعد، الرياض، 2009م.

17- المعلومات والمعلوماتية (الجزء الأول)، د. صباح صالح الفداغي، مطبوعات جامعة الكويت، 1999م.

18- مقالات في اللغة والأدب، د. تمام حسان، عالم الكتب.

19- البحوث الإنجليزية:

19-1- Attia, M., Rashwan, M., Ragheb, A., Al-Badrashiny, M., Al-Basoumy, H., A Compact Arabic Lexical Semantics Language Resource Based on the Theory of Semantic Fields, LREC2008 conference http://www.lrec-conf.org/lrec2008, Marrakech-Morocco, May 2008.

19-2- Attia, M., Rashwan, M., Ragheb, A., Al-Badrashiny, M., Al-Basoumy, H., Abdou, S., A Compact Arabic Lexical Semantics Language Resource Based on the Theory of Semantic Fields, Lecture Notes on Computer Science (LNCS): Advances in Natural Language Processing, Springer-Verlag Berlin Heidelberg

20-مواقع على الشبكة العنكبوتية (الإنترنت):

20-1- http://www.VisualThesaurus.com

20-2-http://www.visuwords.com

20-3-http://www.globalwordnet.org/AWN/AWNBrowser.html

20-4-http://www.ArabicWordMap.com

20-5-http://www.rdi-eg.com/rdi/Downloads/ Process of building the forward Arabic_Lexical_Semantic_DB.pdf.