



# من روائع الأدب العربي القديم

الإطار التكنولوجي للغة العربية من خلال إنجاز و توزيع تطبيق معلوماتية تخص  
رقمنة بعض الأعمال من الأدب العربي القديم

مختار بن هدة

المعهد الأعلى للتوثيق. جامعة منوبة. الجمهورية التونسية

2002

## الملخص

تم إنجاز هذا العمل في إطار برامج دار الكتب الوطنية التونسية لبعث سلسلة من النشريات الإلكترونية تساهم بها في برنامج المكتبة الافتراضية التونسية. من خصائص هذا العمل تغطيته لمجموعة من النصوص الأدبية العربية تمت معالجتها آليا من خلال رقمنة النصوص الكاملة حسب تقنيات التشفير الرقمي الثنائي النصي وليس باعتماد الحلول الصورية و ذلك لتمكين المستفيد من القيام بعمليات البحث المباشر في النصوص واسترجاعها.

للاستجابة إلى هذه المتطلبات العملية، كان من الضروري اختيار الحلول البرمجية الملائمة لطبيعة المحتوى وخصوصية اللغة العربية أثناء عمليات الإدخال للبيانات واسترجاعها. فكانت الحلول مختلفة ومتنوعة حسب اختلاف وتنوع البيانات امتزجت بين الحلول الصورية على مستوى تطبيق Director ونصية بالنسبة إلى تقنية Html المتوخاة في تشفير النصوص الكاملة. كما كان للبيانات متعددة الوسائط حيز من الأهمية في العمل استوجب استخدام تقنيات خاصة لضمان تفاعلها مع المستفيد أثناء عمليات الإبحار في النصوص.

## التقديم

منذ ظهور تقنية التشفير الموحدة للمحارف العالمية التي تعتمد 16 بنة والمسماة بمقياس اليونكودUnicode، أتيح إلى اللغة العربية مجال جديد يتناسب مع ضرورة الدفع بها كلغة عالمية يمكن التعامل معها على الشبكات المعلوماتية و القنوات التكنولوجية للنشر و التوزيع بكل شفافية كمثيلاتها اللاتينية دون الالتجاء إلى تهيئة خاصة للمعدات. و لعل من أهم هذه المكاسب، التخلي عن اعتماد الحلول الصورية (Bitmap) للحرف العربي واستبدال ذلك بالتشفير الرقمي الثنائي النصي (Binary Text Coding) الذي يمكّن من تطويع الحرف والتعامل معه في مختلف التطبيقات ومختلف الأنظمة المحوسبة. فقد صار اليوم ممكنا اعتماد الحرف العربي في كافة الخدمات المتعلقة بالتكشيف و الفهرسة و التعرف الضوئي على المحارف موازنة مع ما هو عليه الحرف اللاتيني منذ بداية عهدنا بالمعلوماتية.

إلا أن هذه التطورات على مستوى تشفير الحرف العربي و التعرف عليه آليا تفتقر إلى حد الآن إلى الدعم اللازم من قبل صانعي الأجهزة و مصممي التطبيقات و مؤلفي المحتوى حتى تكون للغة العربية ركانزها الضرورية التي تمكنها من محاذاة اللغات العالمية الأخرى و توسيع نطاق استخدامها و ترويجها.

في إطار هذا الرؤية المستقبلية لدعم اللغة العربية كأداة اتصال إلكترونية و كلغة نشر و توزيع للثقافة و الأدب و العلوم العربية بتوخي الوسائط الرقمية و المعدات الإلكترونية، قامت دار الكتب الوطنية التونسية بمساهمة منظمة اليونسكو بإصدار قرص ضوئي يشمل النصوص الكاملة لتسعة من أهم الأعمال الأدبية العربية القديمة تم ترقيمها بالتشفير الرقمي الثنائي للحروف تسهيلا لعمليتي تكشيفها و البحث فيها.

وسنحاول من خلال هذه الورقة تحديد أهم الإشكاليات التقنية و اللغوية التي اعترضتنا عند إنجازنا هذا العمل و سعينا إلى إيجاد أوفر الحظوظ لتوزيعه و نشره في مجالات لغوية عربية و غير عربية. كما سنسعى إلى تقديم دراسة جدوى لمختلف الحلول المتاحة في السوق العالمية اليوم لضمان استخدام اللغة العربية كلغة بديلة تروج بواسطتها المعرفة الإنسانية كسائر اللغات "المهيمنة" دون قيود ملزمة أو متطلبات مادية أو برمجية محددة.

سيتمحور هذا العمل إذن حول ثلاثة أقسام أساسية : قسم أول نعرض من خلاله أهداف المشروع وأبعاده الأدبية و التقنية، قسم ثاني نتعرض من خلاله إلى الاختيارات التقنية و البرمجية المعتمدة في معالجة المحتوى و تصميم الإطار العام لهذا المنتج ثم ترويجه في فضاءات أحادية أو متعددة اللغات. و القسم الثالث يسعى إلى تقديم دراسة جدوى للبدائل المتوفرة اليوم لمعالجة اللغة العربية في المجال التكنولوجي و مدى تطابقها مع أهداف عولمة اللغة العربية و شفافتها الألسنية.

## 1. "من روائع الأدب العربي القديم" : دعم للإشعاع الثقافي و التجديد التكنولوجي العربيين

يمثل هذا العمل نقطة انطلاق لسلسلة المكتبة الافتراضية التونسية التي تساهم بها دار الكتب الوطنية التونسية في برنامج المكتبة المتوسطة "مدليب" Medlib الذي تدعمه منظمة اليونسكو من خلال برامجها للحفاظ على التراث الثقافي العالمي. و قد جاء العمل مركزا على ثلاث ميزات أساسية : تنوع المضمون، طبيعة النص و شكل الوعاء، تقنيات البحث و الاسترجاع.

## 1.1 تنوع المضمون الأدبي :

إن الغاية الأساسية من هذا العمل هو التعريف بالأدب العربي القديم ليس بالوصف و التحليل فقط، بل أيضا بتوزيع النص الكامل لهذه الأعمال كحلقة أولى من سلسلة أعمال تسعى دار الكتب الوطنية من خلالها إلى إرساء قنوات إلكترونية موازية لأساليبها التقليدية لنشر المعرفة. فكما أشرنا إلى ذلك، جاء هذا العمل كنواة أولى لسلسلة أعمال رقمية تشمل عددا كبيرا من أهم الأعمال الأدبية والعلمية من مخزون دار الكتب الوطنية ككتاب "إتحاف أهل الزمان"، "ألف ليلة و ليلة"، "مقدمة ابن خلدون" و جملة من المخطوطات القيمة هي الآن على قائمة الأعمال الرقمية المرتبعة.

يخص مشروع "من روائع الأدب العربي القديم" تسعة من الأعمال الأدبية العربية القديمة تمت الموافقة عليها وفقا لجملة من المقاييس التقنية والأدبية أقرتها لجنة مختصة ضمت باحثين و تقنيين من داخل دار الكتب الوطنية و خارجها. فقد سعت هذه اللجنة من خلال المقاييس التي وضعتها إلى تغطية أهم الأعمال الأدبية العربية القديمة مع التأكيد على تنوع أشكالها الأدبية و مصادرها الجغرافية و امتداد فتراتهما التاريخية. لذلك جاءت القائمة النهائية مزيجا من الدواوين الشعرية والأعمال النثرية (قصة و رواية و مقامة) غطت كافة أنحاء العالم العربي القديم من الشام إلى الأندلس مرورا بإفريقية و ذلك على امتداد الحقبة التاريخية من الجاهلية إلى بداية العصر الوسيط. فكانت الأعمال المدرجة كالتالي :

العنوان	الشكل الأدبي	المجال الجغرافي	حياة المؤلف
المعلقات السبع الطوال	شعر	الجزيرة العربية	الجاهلية
ديوان الخنساء	شعر	الجزيرة العربية	575م - 664م
كليلة و دمنة	قصة	الشام	720م - 756م
كتاب البخلاء	رواية	الشام	776م - 869م
رسالة الصداقة و الصديق	رواية	الشام	922م - 1023م
ديوان ابن رشيقي	شعر	إفريقية	1000م - 1064م
مقامات الهذلي	مقامة	الشام	969م - 1007م
طوق الحمامة	شعر	الأندلس	994م - 1064م
قصيدة يا ليل الصب و معارضاتها	شعر	الأندلس	1029م - 1095م

جدول I : قائمة الأعمال المرقمة

و قد صاحبت هذه الأعمال المتنوعة مقدمات تحليلية إلى أبعادها الحضارية و تعريفا بمؤلفيها و بآراء النقاد حولها. كما صاحبتها جملة من الرسوم و التسجيلات الصوتية المدعمة لمحتوياتها.

## 2.1 طبيعة النص و شكل الوعاء :

لعل من أهم أوجه الحداثة في هذا العمل هو اعتماد النص الكامل للأعمال الأدبية المذكورة. فقد اعتدنا من خلال النظم المعلوماتية المعربة التعامل خاصة مع نظم إدارة قواعد البيانات البيبليوغرافية التي تقتصر على الحصر البيبليوغرافي للكتب دون توفير مباشر لمحتوياتها أو بعض الأعمال ذات النص الكامل لكن في صياغة صوتية للمحارف أو في إصدارات مقيدة بنظم تشغيل أو تطبيقات جد ضيقة الرواج و التوزيع. كما ظهرت هنا وهناك نماذج عديدة من المنتوجات الأدبية العربية المحوسبة ذات النص الكامل و الترقيم الثنائي للحروف لكن دون أي دعم إلى الأبعاد التكنولوجية الحالية للنظم المفتوحة و الموزعة و دون أي تركيز على بعدي التصدير والاستيراد للبيانات بين التطبيقات. لذا سعينا في هذا العمل إلى التقيد أكثر ما يمكن بمبادئ الشمولية والانفتاح والشفافية على مستوى التطبيقات و المعدات المستعملة حتى لا

يكون العمل موجها إلى فئة محدودة من المستفيدين دون سواهم و مقتصر على نوعية واحدة من التطبيقات ومن مجالات استخدامها.

أما شكل الوعاء، فرغم توفر الأقراص الليزرية و عدم حداثتها كوسيلة لنشر البيانات و تبادلها على مستويات دولية، فهي لا تزال حديثة العهد إشعاعا و رواجاً في العالم العربي و ذلك لعدة أسباب منها التقنية و الاقتصادية و اللغوية. فعلاوة عن تاريخ ظهورها الحديث نسبياً (1978<sup>1</sup>) مقارنة بتاريخ ظهور الأوعية المغناطيسية في بداية الأربعينيات<sup>2</sup>، بقيت تقنيات صنع الأقراص الليزرية حكراً على الدول المصنعة التي اعتمدها بكثافة لترويج إنتاجها الأدبي و العلمي. و لم تدخل هذه التقنيات العالم العربي إلا بعد رواجها عالمياً إثر تبسيط طرق تصنيعها و هبوط تكاليف تسويقها. إلا أنها اصطدمت إثر ذلك بالحاجز اللغوي الذي قلص كثيراً من مجال رواجها واعتمدها كوعاء لنقل المعلومات و تبادلها خارج الحدود الجغرافية للغة العربية. فقد كانت للغة العربية فعلاً إشكاليات متعددة مع الإعلامية قبل ظهور المواصفة الدولية الموحدة لتشفير المحارف العالمية المسماة باليونيكود. لذا كان مفروضاً على أي إنتاج فكري عربي أن لا يتجاوز المجال اللغوي العربي إذا تم نشره و توزيعه بواسطة الأقراص الليزرية. أما اليوم، و بعد تطور تقنيات التشفير و توحيدها باتباع مبدأي الشفافية اللغوية (Linguistic Transparency) و التعرف الآلي على اللغات (Linguistic negotiation) أصبح من الممكن نشر و ترويج أي عمل فكري عالمي دون التقيد بلغة "مهيمنة" معينة أو بإطار تطبيقي محدد.

مواكبة لهذا التطور الهام في مجال المعالجة الآلية للغات العالمية و خاصة منها العربية، حاولنا التركيز على هذا البعد الدولي للنشر و التوزيع متعدد اللغات لنوفر إلى هذا العمل الأدبي مجال انتشار دولي لا يقتصر على العالم العربي فحسب بل يتعداه ليقرأ على أجهزة الحواسيب العالمية في نصوص كاملة و بتقييم ثنائي.

### 3.1 تقنيات البحث والاسترجاع :

<sup>1</sup> أصدرت شركة فيليبس أول قرص مدمج سنة 1978. ثم صار يسمى بالقرص المدمج ذو الذاكرة المقروءة فقط (CD-ROM) سنة 1983.

<http://frederic.ferre.free.fr/seconde/informatique/seconde2/puin1/>

<sup>2</sup> قدم هوارد أيكن (Howard Aiken) أول نموذج للحاسب الألكترومغناطيسي سنة 1944.

<http://frederic.ferre.free.fr/seconde/informatique/seconde2/puin1/>

من أهم فوائد الاعتماد على النص الكامل في معالجة الوثائق الإلكترونية إمكانية البحث عن المعلومات و استرجاعها أينما وجدت بالوثيقة دون التقييد بفهرسة يدوية مسبقة أو الاقتصار على جزء من النص دون غيره. كما تُمكن نظم الفهرسة و التكتيف الحديثة من البحث والاسترجاع على مستوى مجموعات كاملة من النصوص تتم فهرستها مسبقا لاستخراج محتوياتها النصية وإضافتها إلى فهارس موحدة تكون أداة بحث مركزة على قاعدة معرفية نصية مجمعة.

وتمثل الطريقة التي اعتمدها في البحث والاسترجاع للمعلومات داخل النصوص الأدبية المرقمة حلا وسطا مقارنة بما وصلت إليه التقنيات الحديثة في هذا المجال. وللعوائق التقنية واللغوية التي ذكرناها أنفا مسؤولية أولى في عدم اكتمال هذه التقنية البحثية بتوخي الفهارس المجمع لكافة النصوص. فواجهة البحث التي اعتمدها تخول البحث المفرد في النص الواحد دون إمكانية البحث الشامل في الفهارس المجمع و دون اعتماد المنطق البوليني (Boolean Logic). ويعود هذا القصور إلى محدودية البرنامج الذي اعتمده في التعامل مع اللغة العربية خلافا لما هو عليه الحال بالنسبة للغات أخرى.

على أن لهذه الطريقة إضافة هامة تساعد القارئ على الوصول بسرعة إلى أي جزئ من العمل الذي هو بصدد قراءته و ذلك بالبحث على أي مفردة في النص بغض النظر عن مكان تواجدها. وتبقى إمكانية تطوير هذه التقنية و الوصول بها إلى أعلى درجات البحث المركب و استراتيجيات البحث المعقدة رهينة قرارات أصحاب البرنامج المعتمد وذلك بإضافة العربية إلى اختياراتهم اللغوية حتى تشملها التحيينات الأخيرة و تصبح ضمن اللغات ذات الفهرسة الشاملة للقواعد النصية المجمع.

## 2. الاختيارات التقنية و البرمجية لتصميم القرص و ترويجه

على ضوء هذه الاعتبارات التقنية و اللغوية، حددنا اختيارنا للبرامج و المعدات المستخدمة لتصميم القرص و صياغته حسب تطابقها مع الأهداف التالية :

- الاستجابة لمبدأ الشمولية و استهداف أكبر عدد ممكن من المستفيدين، الشفافية في استخدام اللغة العربية في فضاءات لغوية عالمية،
- الاعتماد على النص المرقم ثنائيا للبحث و القراءة الموجهة،
- إمكانية تصدير البيانات المعروضة واستخدامها بواسطة تطبيقات خارجية،
- التركيبية متعددة المستويات و المداخل (Modularity).

لهذا جاءت تركيبية القرص في شكل ثلاثة مستويات متفاوتة القيمة الإعلامية و الصبغة التقنية من حيث نوعية البرمجيات المستعملة و طريقة المعالجة للبيانات المعربة من خلالها.



الرسم 1 : التصميم العام للقرص

### 1.2 المدخل الرئيسي : بيانات تعريفية بالإطار العام

يمثل المدخل الأول للقرص الجانب الأكثر عمومية بحيث أنه يشتمل على البيانات التعريفية بالإطار العام للمنتج مع تحديد طريقة استعماله و التعريف بفرق العمل المساهمة في إنجازها. ساهمت هذه العوامل في تحديد الخيارات التي وقع اتباعها في التصميم و تحديد المحتوى ومعالجة البعد اللغوي على هذا المستوى.

## المحتوى :

كما هو معمول به في الأوساط التكنولوجية والعلمية المهمة بتطوير أوعية المعلومات الإلكترونية، وكما تنص عليه التوصيات الناجمة عن البحوث في مجال صياغة و تطوير واجهات التطبيق الإلكترونية (Man-Machine Interfaces) التي تنص على مبدأ التدرج من العام إلى الخاص في عرض البيانات إلى المستخدم، فقد سعينا توخي هذا المبدأ بتخصيص محتوى هذا المدخل الرئيسي إلى البيانات العامة المتعلقة بالإطار المؤسسي المشرف على الإنجاز والتكنولوجي من حيث متطلبات الاستعمال و البشري من حيث فرق العمل التي واكبت إنجاز المنتج منذ بدايته. فالإطار المؤسسي ينص على إصدار القرص كعدد أول من سلسلة أعمال ارتأت دار الكتب الوطنية أن تدرجها في إطار برنامج المكتبة الافتراضية التونسية الذي تساهم به في البرنامج المتوسطي "مدليب". أما الإطار التكنولوجي فينص على الخيارات التكنولوجية التي يستوجب على المستخدم اعتمادها لاستخدام أمثل لمحتوى القرص. و تأتي البيانات البشرية لتحديد تركيبة فرق العمل المساهمة على مستويات الإشراف و التهيئة و الإنجاز. كما صاحبت هذه المعطيات إشارة إلى بيانات التسجيل والإصدار وفق القوانين الدولية للوصف البيبليوغرافي والإيداع القانوني.

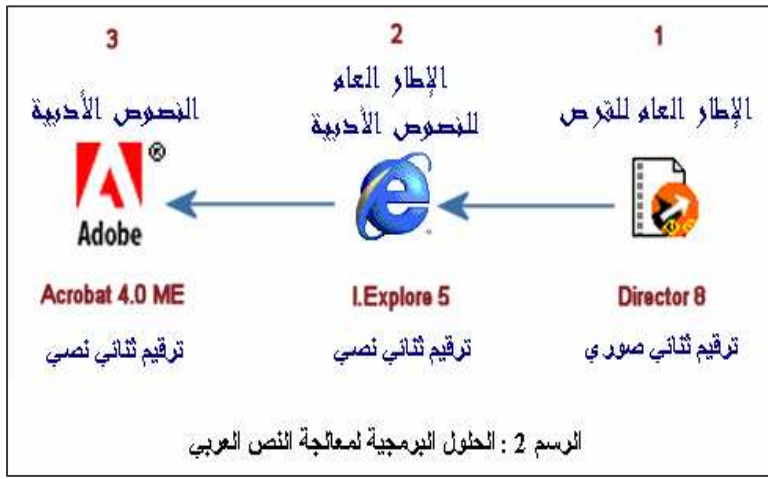
## الخيارات التقنية :

نظرا لأهمية هذه المعطيات في بدئ التعامل مع محتوى القرص وتيسير عملية التفاعل بين المستخدم والواجهة الأولى للتطبيق، كان اختيارنا للحلول التقنية في تصميم هذا المستوى الأول من القرص موجهة إلى عنصرين أساسيين : التحميل الآلي للتطبيق عند إدخال القرص بالفارئ الضوئي (Autorun) ثم التأكد من استيعاب المستخدم لكافة جوانب و متطلبات استخدام القرص والوصول إلى محتواه الأدبي. لذا اعتمدنا النسخة الثامنة من برنامج "Director" الدولي كخيار أثبت جدواه و رواجه عالميا لتصميم أربع صفحات ضمت محتوى هذا المستوى الأولي من تعريفات و إرشادات. هدفنا من ذلك ضمان انطلاقة التطبيق على كافة المعدات الإلكترونية و نظم التشغيل المحتملة إثر ترويج القرص عالميا. كان خيارنا موجهة في بداية الأمر إلى اتباع نفس الخيار التقني بالنسبة إلى كافة المستويات للقرص. لكن العنصر اللغوي فرض علينا التأقلم مع تنوع قيمة المعلومات المدرجة و طبيعتها.

## حلول الدعم اللغوي :

تفاديا للمفاجآت التي قد تطرأ عند بدئ تشغيل القرص بسبب إشكال لغوي غير منتظر، الأمر الذي قد يجبر المستخدم على عدم المضي قدما في التعرف على المضمون الأدبي للنصوص المرقمة، التجأنا إلى اعتماد الحل الصوري للنص العربي في أشكال تتماشى مع الطابع الجمالي و متعدد الوسائط لبرنامج "Director". الغاية من هذه الطريقة هي التأكد من بدئ التفاعل بين المستخدم والمحتوى بلغة عربية تضمن لنا إعطاء التوصيات الضرورية اللازمة حتى يتمكن المستخدم من تهيئة محيطه الآلي والبرمجي لتقبل النص العربي المرقم ثانيا في المستويات المتبقية من القرص. وعلاوة عن هذا الاحتياط المتمثل في توفير التوصيات اللازمة لتهيئة الإطار العام لاستخدام اللغة العربية في أي جهاز محتمل، حملنا على القرص نفسه النسخ المعربة من البرامج الضرورية لقراءة النصوص المرقمة على كافة الأجهزة.





## 2.2 المستوى الثاني : تقديم

### الأعمال الأدبية

عند الوصول إلى هذا المستوى من القرص يكون المستفيد قد أتم تهيئة

إطار عمله اللغوي وانتقل إلى مستوى ثان يشرع من خلاله في التعامل مع الإطار العام للأعمال الأدبية. أهمية هذه المعلومات بالنسبة إلى القارئ من حيث إمكانية استخدامها في مجالات أخرى فرضت ضرورة معالجتها بالطريقة الثنائية حتى تكون قابلة للنشر و التبادل بين التطبيقات. لذا جاء تصميم هذا المستوى متناسقا مع الاعتبارات التالية :

### المحتوى :

تماشيا مع مبدأ التدرج في الوصول إلى نواة العمل الحقيقية والتمثلة في النصوص الأدبية المرقمة، اشتمل هذا الجزء من القرص على المعطيات التمهيديّة التي تعنى بتقديم المؤلفين وبالتعريف بعصورهم و بآراء النقاد حولهم. الهدف من هذه البيانات يتمثل في وضع الأعمال الأدبية في إطارها التاريخي و الحضاري و الأدبي حتى يكتمل الإطار العام للاستفادة القصوى من العمل. ولعل من أهم فوائد هذه المعطيات لدى القارئ هي تثبيت الروابط بين الأعمال الأدبية واسعة الصيت و مؤلفيها الذين لا تقل شهرتهم عن شهرة أعمالهم. و يلعب الوعاء التكنولوجي متعدد الوسائط على هذا المستوى دورا رئيسيا لتثبيت هذه الروابط باعتماد الحس البصري الثلاثي الأبعاد كما تؤكد البحوث في مجال علوم الإدراك (Cognitive Sciences).

### الخيارات التقنيّة :

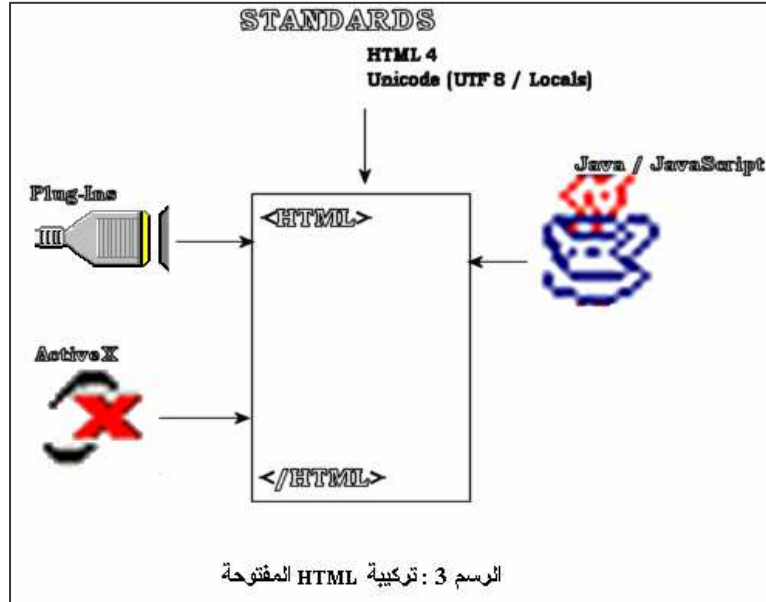
انطلاقا من هذا المستوى، تبدأ ضرورة التعامل مع النص العربي في شفرة عربية موحدة تمكن من التعامل مع الحرف العربي كوحدة معلوماتية ذات ترقيم ثنائي نصي يخول تحميلها تحت تطبيقات خارجية كمعالجي النصوص و نظم إدارة قواعد البيانات. لهذا الغرض وقع اختيارنا على الحل الأكثر رواجاً في عالم النشر الحر من قيود حقوق الملكية التجارية. و نقصد بذلك لغة HTML في إصدارها الرابع بكل الإضافات المدعمة لها في معالجة البيانات متعددة الوسائط واللغات و تبادلها حسب المقاييس الدولية في التشفير و التراسل للمعلومات.

عرفت لغة HTML منذ أن وضعها المهندس Tim Berners Lee بمخابر مركز الدراسات النووية السويسري سنة 1993 تطورات كبيرة في هيكلية الوثائق الإلكترونية وتوزيعها على شبكة الإنترنت العالمية. فكما يشير إلى ذلك الرسم الموالي، تم إثراء هذه اللغة البرمجية على عدة مستويات نذكر منها :

- تطويعها إلى المقاييس الدولية بحيث أن إصدارها الرابع صار يتفاعل مع أغلبية التوصيات والمقاييس الدولية المتعلقة بصياغة الوثائق و تراسلها على شبكة الإنترنت نذكر منها خاصة مقياس اليونكود (Unicode) للتشفير الموحد للمحارف العالمية باستعمال 16 بتة و تفرعاتها للتشفير المختصر بواسطة 7 و 8 بتات (Utf 8-Utf7). كما تمكن هذه اللغة من التعرف آليا على الخصوصيات المحلية

للتقافات و اللغات العالمية (Locals) من ذلك مثلا استعمال الأرقام العربية أو الهندية، اعتماد اتجاهي الكتابة من اليمين أو اليسار حسب النسق، استعمال الرموز النقدية حسب الانتماء الجغرافي.

● إمكانية تشغيل برامج الدعم المدمجة (Plug-Ins). تمكن لغة HTML من الالتجاء أليا إلى برامج



إضافية لدعم تحميل أو تشغيل نوع خاص من البيانات أو الخدمات التي يتعذر القيام بها بواسطة وسوم HTML (TAGS) الأساسية. نذكر من ذلك مثلا تشغيل مقاطع من الفيديو بواسطة برنامج الدعم Media Player. وقد استعملنا في عملنا برنامج الدعم الصوتي لويندوز لإدراج الموسيقى المصاحبة لوثائق HTML التشعبية.

● إمكانية تحميل موارد خارجية بتقنية الربط و الإدماج (ActiveX). عوض الالتجاء إلى برامج الدعم الخارجية، يمكن للغة

HTML أن تحمل داخل إطار عملها بيانات أو موارد تنتسب أصلا إلى تطبيقات خارجية. من ذلك مثلا إدراج جدول حسابات أو قاعدة بيانات تم تصميمها بواسطة أحد برامج Office. الفرق بين برامج الدعم المدمجة و تقنيات الربط و الدمج هي أن النوع الأول يتولى من خلال أوامره وتعليماته الأصلية التحكم في الموارد المدرجة بينما يتولى متصفح الواب إظهار المعلومات المستوردة بالطريقة الثانية والتحكم فيها بواسطة أوامره الخاصة.

### حلول الدعم اللغوي :

أبرز الإشكاليات التي تضعها اللغة العربية أثناء صياغة صفحات الواب و المواقع متعددة اللغات تتمحور حول عنصرين أساسيين : التعرف أليا على الحروف العربية بتنوع أشكالها حسب تواجدها بالكلمة و اتجاه كتابتها من اليمين إلى الشمال.

تعريف الهوية اللغوية للوثيقة (Linguistic Tagging) : أمام تنوع شفرات المحارف العالمية وتنوعها، وردت مختلف إصدارات لغة HTML مدعومة بالوسوم الضرورية لتمكين المتصفحات من التعرف على الهوية اللغوية للصفحات وذلك من خلال شيفرة (Script) مقننة يتم إدراجها كعنصر أساسي في الجزء المخصص لبيانات التعريف (Metadata) في رأس الصفحة على مستوى الوسم <Head>. يتم من خلال هذه الشيفرة تحديد الشفرة المستعملة في صياغة محتوى الوثيقة حتى يتمكن المتصفح من اعتمادها أثناء قراءة رموز الصفحة الداخلية (Coding Sources) و بالتالي ملاءمة الرمز الثنائي النصي (Binary Text Code) مع شكل الحرف الذي يمثله (Glyph). وتتوفر اليوم عدة شفرات للمحارف العربية لا يزال استخدامها واردا على شبكة الإنترنت. و سنتولى التعرض إليها في جزئ لاحق من هذه الوثيقة.

<META HTTP-EQUIV="Content-Type" Content="text-html; charset=utf-8">



اتجاه الكتابة : لئن سمحت مواصفة اليونيكود من إيجاد الحل الملائم للتعرف آليا على أشكال الحروف العربية وإظهارها حسب النسق (في بداية الكلمة، في وسطها أو في نهايتها أو منعزلة) تبقى إشكالية اتجاه الكتابة أمرا يستوجب تدخل المصمم مباشرة لاختيار أنسب الحلول لذلك.

تمتاز النسخة الرابعة من لغة HTML بالإضافات التي تحتويها لمعالجة هذه المعضلة. فقد كانت التجارب الأولى في بداية عهد تصميم المواقع متعددة اللغات تلجئ إلى الوسوم العامة لضبط بنية الصفحات وتحديد محاذاة النصوص (Alignment) يمينا أو شمالا أو وسطا <Div align="right">، وهذا الأمر من شأنه أن يمنع محاذاة النصوص يمينا وشمالا في نفس الوقت <Div align="justify">. إلا أن الإضافات الأخيرة للنسخة الرابعة من لغة HTML مكنت من التحكم بأكثر دقة في اتجاه الكتابة و تسوية النصوص من خلال وسم الاتجاه <Dir> الذي يأخذ القيمتين <RTL> و <LTR> نسبة إلى الاتجاهين "يميني/شمال" و "شمال/يميني". يرد هذا الوسم دائما كخاصية تصاحب وسوما أخرى تحدد اتجاه النص الكامل كوسم <HTML> الذي يرد في بداية موارد الصفحة الرمزية (<html Dir="RTL">)، كما يمكن إضافته إلى وسوم الفقرات و الجداول إن كان في الوثيقة الواحدة تداول بين نصوص عربية و أخرى لاتينية.

```
<html>
<Head>
<META HTTP-EQUIV="Content-Type" Content="text-html;
charset=utf-8">
</Head>
<Body>
<Div align="Center">عنوان الوثيقة</Div>
<p dir="rtl" align="justify">
سيظهر النص العربي تحت المتصفح في اتجاه من اليمين و في محاذاة متوازية من اليمين و الشمال.
<p dir="ltr" align="left">
سيظهر النص اللاتيني في هذه الفقرة في اتجاه يساري و في محاذاة يسارية.
</body>
</html>
```

نموذج من صفحة وab مزدوجة اللغات و الاتجاه

## 3.2 المستوى الثالث : الأعمال الأدبية

يمثل هذا المستوى الأخير من العمل النواة المركزية للقرص. كما يشكل على مستوى واجهة التطبيق الجزء الأكثر عمقا يصله المستفيد بعد تثبيت البرمجيات الضرورية لاستخدامه و التعرف على إطاره العام. ونظرا لخصوصية هذا العنصر الجوهري و أهميته بالنسبة إلى الهدف الرئيسي للعمل تمت إحاطته بعناية فائقة على جميع الأصعدة تخص المحتوى والخيارات التقنية وحلول الدعم اللغوي.

### المحتوى :

كما ورد ذلك في بداية هذا العمل، تم اختيار محتوى هذا العمل الأدبي وفقا لمقاييس أدبية و ثقافية و حضارية تم الاتفاق عليها من قبل لجنة مكونة للعرض. كما تم أثناء تجميع النصوص و ترفيمها التوقف على بعض الإشكاليات التي اعترضت التوصل إلى ترفيم النصوص على شكلها الحالي.

كان العنصر الأول مركزا على اختيار النسخ التي لا تخضع لحقوق التأليف نظرا إلى عدم توفر الاعتمادات الضرورية لتغطية تكلفة حقوق المؤلفين. وقد تم تكليف لجنة قراءة ضمت أخصائين في اللغة و الأداب و التاريخ و الحضارة للثبوت من كل هذه الأبعاد في بنية النصوص النهائية ومطابقتها مع النسخ الورقية التي أخذت منها. أما الإشكال الثاني فهو يتعلق بمدى تطابق الوثيقة الإلكترونية مع نظيراتها الورقية على مستوى الشكل. فبمراعاة الاختلافات بين نوعية الوعاءين من ناحية و طبيعة إطار قراءة محتوييهما من ناحية أخرى، تم الاتفاق على عدم الالتزام بالشكل الورقي و إخضاع النصوص المرقمة إلى متطلبات البرمجيات التي تم الترقيم بواسطتها و التي تتم قراءة النصوص من خلالها. فكان الفرق على مستوى نمطية الحروف و زخرفة الأطر للصفحات و الصور المصاحبة لمضمونها الأدبي.

## الخيارات التقنية :

لإتمام هذا الاتجاه في بناء هيكلية النصوص المرقمة و لإعطائها أوفر الحظوظ للتناسق مع الإطار التكنولوجي الذي سيمكن من استخدامها و مع تطلعات المستفيدين من حيث جمالية الواجهة و سهولة البحث و التنقل داخل الوثائق المرقمة، تمت معاينة جملة من الحلول التقنية التي توفرها السوق العالمية اليوم. و قد اتجه اهتمامنا إلى حلول ثلاثة رئيسية هدفنا من خلالها توفير أكبر قسط ممكن من الانسجام بين الخيارات التقنية للمستويات الثلاثة للقرص.

1. اعتماد برنامج Director واستخدام وحداته و حقله النصية لإدخال النصوص المرقمة في واجهات تتناسب مع المستوى الأول من القرص. لكن تم التخلي عن هذا الحل نظرا لاعتبارين اثنين أولهما عدم توفر النسخة المعربة من نظام Director و ثانيهما عدم تلاؤم النصوص الطويلة المناسبة (Scrolling Text) مع صفحات الملتيميديا أحادية الشاشة ذات الطابع التشعبي (Hypertext).
2. اختيار لغة HTML لترقيم البيانات النصية للأعمال الأدبية المعنية خاصة و أن هذه اللغة قد اكتسبت كما أوردنا ذلك أنفا إضافات تخول لها التعامل بقسط كبير من الشفافية مع اللغة العربية. كان هدفنا من ذلك إيجاد تواصل بين الصيغة التقنية لهذا المستوى النهائي من العمل والمستوى الذي سبقه. لكن نظرا للإشكاليات التي لا تزال تصاحب التعرف الآلي على اللغة العربية على مستوى بعض المتصفحات و التي سنعرض إليها لاحقا، خیرنا البحث عن حل أمثل يتناسب مع مختلف ظروف العمل للمستفيدين.
3. يتمثل الحل الثالث في اختيار الإصدار الرابع من برامج Acrobat لإنشاء نصوص مرقمة ثنائيا بواسطة معالج النصوص MSWord الذي ساعد إثر ذلك على تحويلها إلى ملفات من نوع Post Script ثم إلى ملفات من نوع PDF بواسطة برنامج Acrobat Distiller. من أهم فوائد ملفات PDF أنها شائعة الاستعمال على شبكة الإنترنت لترويج النصوص الطويلة و بالتالي فهي سهلة التحميل على المتصفحات بواسطة برنامج قارئ الأكروبات (Acrobat Reader) الحر من حقوق الملكية. من فوائد هذه الملفات أيضا المحافظة على جودة الطباعة للنصوص (Printing Quality) حسب نوعية مصادرها. لكن الأهم من ذلك كله هو قدرة هذا القارئ على توفير خدمة البحث المباشر في النصوص الكاملة من خلال واجهات بحثية متطورة تستخدم من خلالها كافة تقنيات البحث على الخط كتقنيات البحث البوليني و تقنيات البحث باللغة الطبيعية والبحث بالتواريخ و المرادفات الصوتية (Phonetic search)، وذلك من خلال فهارس متعددة يتولى نظام Acrobat Catalog إنشائها بتكشيف مجموعة من ملفات PDF التي تقدم له من قبل مصمم التطبيق. كما يخول قارئ الأكروبات البحث المبسط بالمفردة الواحدة في النص الواحد دون ضرورة الالتجاء إلى عملية التكشيف المسبقة.

نظرا لهذه الاعتبارات و غيرها من المميزات لهذا النوع من الملفات النصية، وقع الاتفاق على اعتماد نمطية ملفات الأكروبات لتكون الحل المناسب لتوزيع الأعمال الأدبية و ضمان إيصالها إلى المستفيد حتى وإن عجز جهازه على إظهار المستويين الأولين. فالتركيبية متعددة المستويات و المداخل للعمل (Modularity) و كذلك النسخة المحملة على القرص من نظام Acrobat Reader تمكن أي قارئ

من إظهار الأعمال الأدبية في صيغتها الحالية و البحث داخلها دون أي إشكال يذكر إذا توفرت الأرضية اللغوية للإطار العام للجهاز كما سنبين ذلك في الفقرات الموالية.

### حلول الدعم اللغوي :

كان من الممكن أن تستجيب كافة الحلول التقنية المبينة أعلاه إلى صياغة القرص في الشكل الذي أردناه له لو كانت وثائقنا مقتصرة على اللغة اللاتينية فحسب، لكن ضرورة استعمال اللغة العربية جعلتنا نلتجئ إلى البحث عن البدائل اللغوية الملائمة. حتى أن اختيارنا لبرنامج Acrobat لبناء ملفات PDF المعربة كان افتراضيا رغم قدرته على التعرف على الحروف العربية عن طريق شفرة اليونيكود التي صمم بواسطتها وذلك نظرا لعدم توفر القارئ المعرب من نظام Acrobat Reader. فالنسخ البرمجية من هذا النظام الموزع بشكل عام و مجاني على شبكة الإنترنت وعلى الأقراص الليزرية هي إصدارات لاتينية (إنجليزية و فرنسية) لا تتوافق مع النصوص العربية من حيث اتجاه الكتابة في خانات واجهة البحث المباشر. فلئن مكنت هذه الإصدارات فعلا من إظهار النصوص العربية و طباعتها بشكل جيد معتمدة في ذلك على الخصوصيات المتقدمة لتقنية PostScript، فإنها تقتصر إلى دعم اتجاه الكتابة من اليمين إلى اليسار عند القيام بعمليات البحث في النص رغم قبول واجهة البحث للحروف العربية. ولا يمكن بالتالي البحث عن المفردات إلا إذا تم إدخالها بطريقة عكسية ("ثحب" عوضا عن "بحث").

إلا أننا تمكنا في آخر الأمر من الحصول على النسخة المعربة لقارئ الأكروبات الذي توزعه شركة Adobe على موقعها بشبكة الإنترنت كواحدة من الإصدارات متعددة اللغات لمنتجاتها البرمجية. دعمت هذه النسخة "الشرق أوسطية" (Acrobat Reader ME) خدمات البحث المباشر بقدرتها على إدخال مصطلحات البحث من اليمين حسب ما تقتضيه قوانين الكتابة العربية.

### **الخلاصة**

يمكن اعتبار هذا العمل مجددا في طبيعة العمل الرقمي بدار الكتب الوطنية التونسية نظرا لإضافته طابع التعددية اللغوية في المعالجة الآلية للوثائق الرقمية ذات النص الكامل. لكن السؤال يكمن في طبيعة الحلول المتوخاة في التعامل مع اللغة والحرف العربيين من حيث إشكاليات حوسبتهما والاختيارات البرمجية لإتمام ذلك.

كانت الاختيارات النهائية مبنية على دراسات ميدانية مطولة للوقوف على آخر متطورات السوق العالمية للبرمجيات متعددة اللغات وكان الاختيار على الشاكلة التي تمت بها صياغة هذا العمل.

لكننا نقر في الآن ذاته بأن مجال الصناعة اللغوية والمعالجة الآلية للغات في تطور سريع لا يمكن التوقف خلاله انتظارا لحل نهائي ومثالي في معالجة ما عرف باللغات "المعقدة". لذا كانت الحلول المعتمدة في إنجاز هذا العمل ضمن آخر ما جادت به سوق البرمجيات العالمية. لكن نبقى متأكدين من أن المعالجة الآلية للغة العربية ستعرف في الأجل القريبة جدا تطورات كبيرة تدرج في إطار المساعي الدولية لدعم التوجهات الحالية نحو الشمولية (G11n) Globalisation والتدويل (I18n) Internationalisation والمحلية (L10n) Localisation.