Cybrarians Journal
E-ISSN 1687-2215
دورية إلكترونية محكمة تعنى بمجال المكتبات والمعلومات
العدد 39، سبتمبر 2015

Problems of Encoding Digital Text Documents in Arabic Using the International Standard TEI

Rachid ZGHIBI

Assistant professor

Higher Institute of Documentation,

University of Manouba, Tunisia

rachid_zghibi@yahoo.fr

Abstract

This research aims to study the most important technical problems that we can be exposed to when encoding text documents in Arabic using the TEI technique - one of the most important international standards and the most commonly used by many libraries, information centers and archives around the world in the digitization process of its records.

In the first part of the research, we will discuss the origins and evolution of this international standard and its most important characteristics which distinguishes it from other coding systems. We will also discuss the structure and the components of the TEI file. The second part will discuss these problems that particularly relate to the inclusion of some Arab graphemes and the relation between letters in Arabic, as well as issues related to printing and the display of bidirectional texts. In this part, we also suggest some solutions to overcome these problems such as using some of additional characters and codes that can be integrated directly with the tag standards or can be written in external files. It should be noted that these solutions are related to letters in Arabic. They are intended to show Arabic letters in their correct form or as in the original texts.

Cybrarians Journal
E-ISSN 1687-2215
دورية إلكترونية محكمة تعنى بمجال المكتبات والمعلومات
العدد 39، سبتمبر 2015

إشكاليات ترميز الوثائق النصيّة الرقميّة باللّغة العربيّة باستعمال المعيار الدّولي TEI

د. رشيد الزغيبي

أستاذ مساعد، المعهد العالي للتوثيق، جامعة منوبة

مدير قسم نظم المعلومات، المعهد العالي للتوثيق

تونس

rachid_zghibi@yahoo.fr

المستخلص

يهدف هذا البحث الى دراسة أهم المشاكل الفنيّة التي يمكن أن نتعرّض لها عند ترميز وثائق نصية باللّغة العربيّة باستعمال تقنية الTEI الذي يعتبر من أهمّ المعايير الدوليّة وأكثرها استخداما من قبل العديد من المكتبات ومراكز التوثيق والارشيف في العالم في رقمنة أرصدتها وتبادلها وإتاحتها على الخطّ.

سنتناول في الجزء الأوّل من البحث نشأة وتطور هذا المعيار الدّولي وأهمّ الخصائص التي يتميّز بها مقارنة بنظم الترميز الأخرى وسنتناول أيضا هيكلة وتركيبة ملفّ الTEI وسنستعرض في الجزء الثّاني للبحث هذه المشاكل والتي تتعلّق خاصة بإدراج بعض الرموز العربيّة واتصّال الحروف وطباعة وعرض النصوص ثنائيّة الاتّجاه مع اقتراح بعض الحلول المناسبة التي تتمثّل في استخدام بعض المحارف والشفرات الإضافيّة التّي يمكن دمجها مباشرة مع وسوم وصفات المعيار أوكتابتها في ملفّات خارجيّة. وتجدر الإشارة إلى أنّ هذه الحلول تتعلّق بإظهار وطباعة الحروف العربيّة في شكلها الصحيح أومثلما وردت في النصوص الأصليّة الورقيّة.

الاستشهاد المرجعي

Cybrarians Journal
E-ISSN 1687-2215
دورية إلكترونية محكمة تعنى بمجال المكتبات والمعلومات
العدد 39، سبتمبر 2015

زغيبي، رشيد. إشكاليات ترميز الوثائق النصيّة الرقميّة باللّغة العربيّة باستعمال المعيار الدّولي TEI .- Cybrarians Journal .- العدد 39، سبتمبر 2015 .- <سجل تاريخ الاطلاع على البحث> .- متاح في: <سجل رابط الصفحة الحالية>

مقدّمة

يتميز مجتمع المعلومات اليوم بالاستخدام المكثّف لتكنولوجيا المعلومات والاتّصال الحديثة في إنشاء المعلومات الرقميّة ومعالجتها ونشرها وتخزينها. وتتميزّ هذه التكنولوجيا بتطورّها السّريع جدا وباندماجها وانصهارها في شكل أقطاب تكنولوجيّة مركبّة يستحيل فصل مكوّناتها منهيّة بذلك الحدود التي كانت قائمة منذ زمن ليس ببعيد بين المهن والاختصاصات من ناحيّة وبين منظمّات وهيئات التقييس من ناحيّة أخرى.

كما تتّسم هذه التكنولوجيا أيضا بتعدّد المعايير والمواصفات مختلفة المصادر (أنظمة تقييس دوليّة ومحليّة، ائتلافات صناعيّة أوأكاديميّة...) التي يمكن تصنيفها حسب كاترين ليبوفتشي (Catherine Lupovici)[1] إلى نوعين رئيسيين وهما التشفير والتبادل. ويشمل النوع الأوّل معايير ومواصفات تشفير المحتوى ومعايير ومواصفات ترميز بنية الوثيقة (المنطقيّة والشكليّة)، ويضمّ النوع الثّاني المعايير والمواصفات التّي تهتمّ بشبكات الإرسال وبروتوكولات التحكّم في الإرسال نذكر منها المعيار الدولي ISO 9579 : Open Systems Interconnection ومواصفات TCP/IP وHTTP و MIME...

ينتمي TEI : Text Encoding Initiative[2] إلى النوع الأوّل من المعايير وبالتحديد إلى فئة معايير ترميز البنية المنطقيّة للوثائق الرقميّة ويستخدم خاصّة في ميادين العلوم الإنسانية والاجتماعيّة واللسانيّات مثل الكتب والمقالات والمخطوطات والمعاجم والموسوعات والتسجيلات الصوتيّة والقصائد الشعريّة والروايات المسرحيّة والصور والرّسوم، الخ.

ويتميّز هذا المعيار باعتماده على لغة الXML لترميز البنية المنطقية للوثيقة مهما كانت درجة تشعّبها وتعقيدها ممّا يسمح بإنشاء قواعد بيانات وكشّافات بطريقة آليّة ومن تركيز البحث على بيانات دقيقة في الوثيقة أينما وجدت. كما يمكّن أيضا من فهرسة الوثائق وتكشيفها بواسطة نظام ميتاداتا خاص به لتيسير عمليات البحث والاسترجاع في قواعد البيانات أوعلى الخطّ.

ونظرا لأهميّته التكنولوجيّة فإن العديد من المؤسسّات التوثيقيّة والأرشيفيّة في العالم تستعمل تقنيّة الTEI لرقمنة أرصدتها وإتاحتها على الخطّ نذكر منها على سبيل المثال المكتبة الرقميّة للمخطوطات بسويسرا والمكتبة الرقميّة للعلوم الإنسانية بفرنسا وجامعة ميشغان بالولايات المتّحدة الأمريكيّة[3] ويوفر موقع ويب المعيار معلومات هامة وروابط تحيل مباشرة لأهم المشاريع في العالم.

سنتناول في الجزء الأوّل من البحث نشأة وتطور هذا المعيار الدولي وأهمّ خصائصه ومميزاته وسنتطرّق أيضا بالوصف والتحليل إلى بنيته المنطقيّة وفي الجزء الثّاني سنستعرض أهمّ المشاكل التي يمكن أن نتعرّض لها عند رقمنة وترميز وثائق باللغة العربيّة وسنقترح بعض الحلول المناسبة لها.

إشكاليّة البحث وتساؤلاته

على الرغم من تعدّد مشاريع الرقمنة التي تعتمد على تقنية الTEI لترميز أنواع مختلفة من الوثائق النصيّة بلغات غير لاتينيّة مثل الصينيّة واليابانيّة واليونانيّة، فإن استخدامها لترميز نصوص عربيّة يطرح في بعض الاحيان إشكاليّات فنيّة تستوجب استخدام بعض المحارف والشفرات الإضافيّة.

وبناء على ذلك، يسعى هذا البحث إلى الإجابة على مجموعة الأسئلة التّالية:

1. ما هوالمعيار الدّولي TEI وما هي أهمّ خصائصه ومميّزاته؟
2. ما هي بنيته المنطقيّة؟
3. ما هي المشاكل التي تطرحها النصوص العربيّة عند ترميزها باستخدام المعيارTEI وما هي الحلول المناسبة؟

منهجيّة الدراسة

اعتمدنا في هذا البحث المنهج الوصفي لدراسة المعيار الدّولي TEI مسلّطين الضوء بالخصوص على ظروف نشأته والتطوّرات التي عرفها وأهمّ الخصائص التي تميّزه عن نظم ترميز النّصوص الأخرى وإلى بنيته بشطريها الميتاداتا والمحتوى.

Cybrarians Journal
E-ISSN 1687-2215
دورية إلكترونية محكمة تعنى بمجال المكتبات والمعلومات
العدد 39، سبتمبر 2015

كما اعتمدنا أيضا المنهج التجريبي لدراسة المشاكل التي يمكن أن نتعرّض لها عند رقمنة وترميز وثائق باللّغة العربيّة والحلول المناسبة لها وذلك من خلال ترميز أنواع مختلفة من الوثائق مثل القصائد الشعريّة والنصوص الأدبيّة واختبار الحلول المقترحة.

الدّراسات السّابقة

1. تناول الباحثان محمد صوالح ومحمد حسّون في دراستهما بعنوان[4] A TEI P5 Manuscript Description : Adaptation for Cataloguing Digitized Arabic Manuscripts المشاكل التّي تتعلّق بفهرسة وتكشيف المخطوطات العربيّة القديمة الرقميّة باستخدام المعيار الدولي TEI في نسخته P5. ويقترح الباحثان مجموعة من الوسوم والصّفات الجديدة لإثراء وحدة Manuscript Description التي من شأنها أن تمكّن من وصف أكثر دقّة للمخطوطات العربيّة القديمة.
2. تناول الباحثون هنري هدريزي ورشيد الزغيبي وسهام الزغيدي ومختار بن هندة في دراستهم بعنوان[5] Promoting the linguistic diversity of TEI in the Maghreb and the Arab region موضوع الخصوصيّات الثقافيّة واللّغوية التيّ يتميّز بها التّراث الثّقافي العربي ومدى قدرة المعيار TEI على معالجتها. ويقترح المؤلفون منهجيّة عمل تتضمّن ثلاثة عناصر أساسية :
   - تحليل شامل للخصوصيّات اللّغوية والاجتماعيّة والثقافيّة للتراث الثّقافي العربي،
   - ترجمة المعيار إلى العربيّة وإلى لغات محليّة مثل البربريّة بالنسبة لدول المغرب العربي واثراؤه بوسوم وصفات جديدة،
   - تكوين مختصيّن في ميدان ترميز وفهرسة الوثائق الرقمية باستخدام المعيار TEI.

أهمية البحث

على المستوى النّظري، تنبع أهميّة البحث من ندرة الأبحاث والدّراسات العربيّة التي اهتمت بدراسة المعيار الدّولي TEI وخاصّة فيما يتعلّق بإشكاليّات استخدامه في ترميز وفهرسة الوثائق العربيّة الرقميّة في ميادين العلوم الإنسانيّة والاجتماعيّة واللسانيّات. ومن هذا المنطلق نأمل ان يساهم هذا البحث في إثراء المكتبة العربيّة وأن يكون منطلقا لأبحاث ودراسات جديدة.

أما على المستوى العملي، فيتناول البحث دراسة وتحليل أهم المشاكل التي يمكن أن نتعرّض لها عند ترميز وثائق نصية باللّغة العربيّة باستعمال تقنية الTEI مع اقتراح الحلول الفنيّة المناسبة التيّ يمكن الاسترشاد بها وتطبيقها من طرف المؤسسات التوثيقيّة التي ترغب في رقمنة وإتاحة أرصدتها باستعمال هذه التقنية.

1- المعيار الدّولي TEI

1.1- لمحة تاريخية

بدأ التفكير في إنشاء هذا المعيار خلال ملتقى علمي دولي انعقد في شهر نوفمبر من سنة 1987 بمعهد فاسار (Vassar College) بالولايات المتّحدة الأمريكيّة حول موضوع إشكاليّات إنشاء وتبادل الوثائق الرقميّة على الخط حيث وقع الاتفاق على المبادئ الأساسيّة لهذا المعيار التّي أطلق عليها اسم (Poughkeepsie Principles) نسبة لمدينة بوغكيبسي مكان انعقاد الملتقى والتي نذكر منها على وجه الخصوص :

- تحديد مختلف الخصائص النصيّة للوثيقة بكل دقّة،
- سهولة الاستعمال بدون الحاجة إلى برمجيّات خاصّة،
- قابلية التوسّع والإثراء حسب حاجيّات المستعملين الآنيّة والمستقبليّة،
- التطابق مع المعايير والمواصفات الدّوليّة المستعملة حاليا[6].

انطلق العمل في البداية من قبل ثلاث مؤسسات بحثية بريطانية وهي Association for Computers and the Humanities و Association for Computational Linguistics و Association for Literary and Linguistic Computing ومع موفّى سنة 1989 انخرط في هذا المشروع أكثر من خمسين باحثا ينتمون إلى ميادين علميّة ومهنيّة مختلفة مثل المكتبات والأرشيف واللّسانيات والمعلوماتيّة وتكنولوجيّات الاتصال، الخ.

صدرت النسخة الأولى لهذا المعيار في شهر جوان لسنة 1990 تحت اسم TEI P1 ثم صدرت النسخة الثانية في شهر ماي لسنة 1994 (TEI P3) محتوية على العديد من التعديلات والتنقيحات والإضافات

وتجدر الإشارة إلى أنّ النسختين تعتمدان على المعيار الدولي SGML (Standard Generalized Markup Language) لترميز الوثائق الرقميّة.

وباقتراح من جامعة Virginia (الولايات المتّحدة الأمريكيّة) وجامعة Bergen (النرويج)، وقع سنة 1999 إنشاء منظمّة عالميّة تحمل نفس اسم المعيار أوكلت إليها مهام التطوير والتنسيق على الصعيد العالمي وهي منظّمة غير حكوميّة وغير ربحيّة مفتوحة لجميع الأشخاص والهيئات والمنظمّات الحكوميّة وغير الحكوميّة التي تهتمّ بميدان النشر الآلي للوثائق والرقمنة.

وفي شهر جوان من سنة 2002 صدرت نسخة جديدة للمعيار تعتمد على لغة البرمجة XML تحمل اسم TEI P4 ثمّ وفي شهر نوفمبر من سنة 2007 صدرت النسخة TEI P5 متضمنّة تنقيحات جوهرية وإضافات هامة مقارنة بالنسخ السابقة تتعلّق خاصّة بترميز الصور والمخطوطات، وفي سنة 2011 صدرت النسخة الأخيرة تحت اسم TEI P5 V2.

1.2 – الخصائص

مقارنة بأشكال الوثائق الرقميّة الأخرى يتميّز الTEI بتركيزه على توصيف محتوى الوثيقة بدون التطرّق إلى شكلها المادّي ممّا يمكّن من معالجتها وتبادلها وإعادة استعمالها بدون ضياع البيانات كما يتميّز أيضا باعتماده على المعيار الدّولي يونيكود كنظام أساسي لتشفير المحارف ممّا يسمح بإنشاء وثائق رقميّة متعدّدة اللّغات والكتابات.

1.2.1– البنية المنطقية والبنية الشكلية

تحتوي كلّ وثيقة رقميّة على بنية شكليّة وبنية منطقيّة :

تتمثّل البنية الشكليّة في مجموعة الخصائص المطبعيّة للوثيقة التي يمكن التعرّف عليها وتمييزها بالعين المجرّدة مثل رقم الصفحة ونوع ولون الخط والمسافة بين الخطوط والفقرات وحجم الصفحة، الخ. فعلى سبيل المثال يمكن توصيف البنية الشكلية لوثيقة رقميّة تمثّل رواية كما يلي :

- تتكون الرواية من صفحات (حجم 10.5 سم * 17.5 سم)،
- تحتوي كل صفحة على أجزاء ورقم الصفحة،
- عرض كلّ جزء 8 صم ويحتوي على عنوان أوأسطر،

- يحتوي كلّ سطر على أحرف (اتّجاه الكتابة من اليمين إلى اليسار، نوع الخطّ Arial وحجمه 11 نقطة)،
- يتكوّن رقم الصفحة من أرقام عربيّة (أسفل الصفحة، نوع الخطّ Times وحجمه 8 نقاط)...

وتجدر الإشارة إلى أن هذه الخصائص ليس لها تأثير مباشر على محتوى الوثيقة إذ يمكن استعمالها لتوصيف أنواع أخرى من الوثائق.

والمقصود بالبنية المنطقية هوتبويب وهيكلة محتوى الوثيقة في شكل عناصر منطقية تحددّ بكلّ دقّة ووضوح مختلف أجزاء محتوى الوثيقة ودلالاتها وأيضا دورها في التنظيم العّام للوثيقة (الروابط المنطقية بينها) مثل العناوين والأجزاء والفقرات والملاحظات وشواهد الدعم، الخ. ولذلك فهي تتوجّه إلى إدراك القارئ وليس لبصره كما هوالشأن بالنسبة للبنيّة الشكليّة[7].

أمّا في ميدان الرقمنة وإتاحة الوثائق على الخط تمكّن البنية المنطقيّة خاصة من التحويل الآلي من شكل (Format) إلى آخر حسب الحاجة وبدون ضياع البيانات ومن إنشاء قواعد بيانات وكشّافات بطريقة آلية وتمكّن أيضا من تركيز البحث على بيانات دقيقة في الوثيقة مثل البحث عن اسم مؤلف ورد اسمه في بيبلوغرافيا الجزء الثالث من كتاب يضمّ خمسة أجزاء.

يهتمّ المعيار الدولي TEI فقط بترميز البنية المنطقيّة لأنواع عديدة من الوثائق الرقميّة النصيّة وذلك بتوفير المئات من الوسوم (Tags) والصّفات (Attributes) ومنذ إصدار النسخة P4 سنة 2002 يعتمد المعيار على لغة الـXML (eXtensible Markup Language) لتحرير الملفّات ووصف محتوياتها وبنيتها المنطقيّة وعلى غرار ملفات الـXML العاديّة يتكوّن كلّ ملف TEI من بنية شجريّة تنازليّة تبدأ بالجذر الذّي يعبّر عنه بواسطة الوسم <TEI> ثم تتجزّأ إلى عناصر فرعيّة محدّدة ومعرّفة مسبقا من طرف المعيار حسب نوع البيانات وحسب نوع الوثيقة.

لتحديد البنية الشكليّة للوثيقة يجب استخدام لغة تنسيق الصّفحات CSS (Stylesheet Cascading) أولغة الـXSL (eXtensible Stylesheet Language) اللتين تمكنّان من إنشاء ملفّات جديدة ذات خصائص شكليّة متنوّعة للملف الأصلي. كما تمكّن لغة الـXSL من تحويل ملف الـTEI إلى ملف HTML ممّا يسمح بإتاحته وتبادله على شبكة الانترنت.

Cybrarians Journal
E-ISSN 1687-2215
دورية إلكترونية محكمة تعنى بمجال المكتبات والمعلومات
العدد 39، سبتمبر 2015

1.2.2- نظام تشفير المحارف

يعتمد الTEI على اليونيكود (Unicode) كنظام تشفير للمحارف ممّا يمكّن من إنشاء ومعالجة وثائق متعدّدة اللّغات والكتابات وقد وقع تطوير هذا المعيار الدّولي من قبل منظّمة عالميّة تسمّى The Unicode consortium تضمّ أهمّ مصنّعي الحواسيب والبرمجيات في العالم على غرار أبل (Apple) وهيولت باكرد (HP) وآي.بي.إم (IBM) وجست سستمز (JustSystem) وميكروسوفت (Microsoft) وأوراكـــل (Oracle) وهومتطابق مع المعيار الدّولي إيزو10646 (ISO/CEI 10646).

مقارنة بأنظمة التشفير الأخرى يستخدم اليونيكود 16 بتّ (Bits) لترميز كلّ محرف من المحارف التّي يستخدمها الحاسوب كما يعتمد على تقنية ترميز تتمثّل في أنّ كلّ محرف معرّف باسم وقيمة عدديّة فريدين من نوعهما بغضّ النّظر عن منصّة التشغيل والبرنامج التطبيقي واللّغة المستعملة مما يضمن عدم تلف البيانات عند تبادلها. كما يوفر معلومات إضافيّة عن كلّ محرف واستخداماته.

ومنذ صدور النّسخة الأولى لهذا المعيار سنة 1991 تطوّر عدد المحارف ليتجاوز 109 ألف محرف في نسخته الأخيرة رقم 6.2.0 الصّادرة في 26 سبتمبر 2012 ممّا يمكّن من تشفير جلّ الكّتابات المستعملة حاليا والعديد من الكتابات القديمة والميّتة على غرار الفارسيّة القديمة والرونية والتيفيناغ. وتشتمل هذه النسخة على 732 محرفا جديدا مقارنة بالنسخة السابقة رقم 6.1.0.

نظرا للأهمية التكنولوجيّة لهذا المعيار، فإنّ جلّ لغات البرمجة والمواصفات القياسيّة الحديثة تستعمل اليونيكود كنظام تشفير أساسي مثل XML وHTML وJava وJavaScript وLDAP وWML وكذلك الشأن بالنسبة لأغلب أنظمة التشغيل ومتصفّحات الويب ومنتجات أخرى كثيرة ومتنوّعة.

1.3- بنية الملف TEI

يبدأ كل ملف TEI بوسم <TEI> الذي يعتبر الجذر الذي تتفرع عنه بقية العناصر وينقسم إلى جزأين أساسيين وهما الترويسة ويعبّر عنها بواسطة الوسم <teiHeader> والمحتوى الذّي يعبّر عنه بواسطة الوسم <text> :

Cybrarians Journal
E-ISSN 1687-2215
دورية إلكترونية محكمة تعنى بمجال المكتبات والمعلومات
العدد 39، سبتمبر 2015

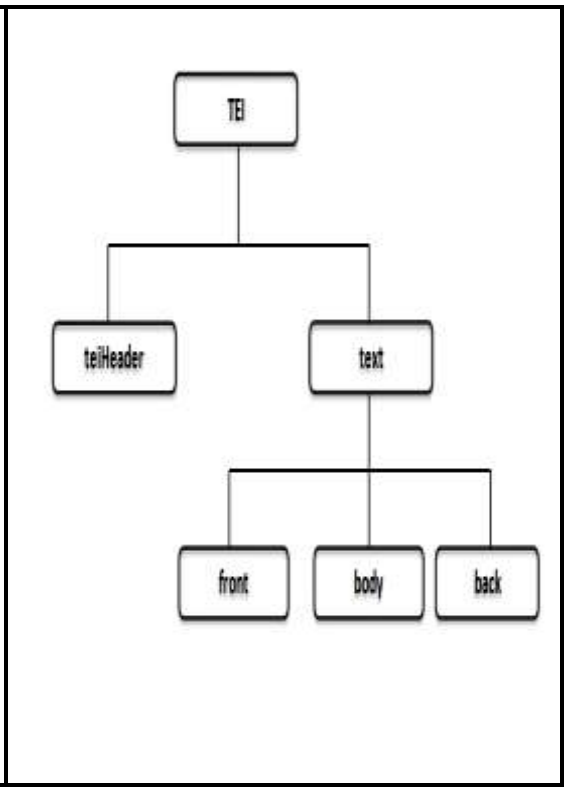
```
<?xml version="1.0"  encoding="utf-8" ?>
<TEI>
        <teiHeader>
           Same tags here
        </teiHeader>

        <text>
                <front>
                   Same tags here
                </front>

                <body>
                   Same tags here
                </body>

                <back>
                   Same tags here
                </back>

        </text>
</TEI>
```



صورة رقم 1 : بنية الملف

1.3.1- الترويسة أوالميتاداتا <teiHeader>

تعتبر الترويسة صفحة العنوان الرّقمي للملف وتستخدم للفهرسة والتكشيف بهدف تيسير عمليّات البحث والاسترجاع على الخطّ أوفي قواعد البيانات وتنقسم إلى أربعة عناصر فرعيّة مركبّة تتفرّع بدورها إلى عناصر فرعيةّ أخرى :

1. **<fileDesc>** : عنصر إجباري يمكّن من وصف ببليوغرافي دقيق للملف الرّقمي على غرار معايير الوصف الببليوغرافي المعتمدة في المكتبات ومراكز المعلومات.
2. **<encodingDesc>** : عنصر غير إجباري يستعمل لتحديد العلاقة بين الملف الرّقمي ومصدره أومصادره الورقيّة.
3. **<profileDesc>** : عنصر غير إجباري يستخدم لوصف الخصائص غير الببليوغرافيّة للملف مثل اللغة أواللّغات المستعملة في الملف وظروف أوأسباب التأليف وهويّة

المساهمين في إعداد المحتوى مع تحديد طبيعة المساهمة. كما يمكّن أيضا من تكشيف الملف بواسطة مستخلصات وكلمات المفاتيح (حرّة أومقيّدة).

4. **<revisionDesc> :** عنصر غير إجباري يستعمل لوصف مختلف التغييرات التي طرأت على الملف الرّقمي منذ نشأته.

يجب أن يتضمّن العنصر <fileDesc> العناصر الفرعيّة التاليّة حتّى يعتبر الملفّ صحيحا :

- <titleStmt> : يوفّر معلومات تتعلّق بعنوان الملف والجهة المسؤولة على نشأته (الأشخاص الماديّون أوالمعنويون). يجب على الأقلّ تحديد عنوان الملف.
- <publicationStmt> : يوفّر معلومات تتعلّق بنشر وتوزيع الملف مثل هويّة الناشر أوالموزّع ومكان وتاريخ النشر.
- <sourceDesc> : يمكّن من وصف ببليوغرافي لمصدر أومصادر الملف الرّقمي بطريقة بسيطة أومهيكلة.

فيما يلي مثال لجذاذة ببليوغرافية حسب ترويسة teiHeader وقد استعملنا في ذلك العنصرين <fileDesc> و<profileDesc> :

المكتبات المتخصصة / تأليف ألن كنت ؛ ترجمة علي الغامدي . – ط1 . – جدة : دار الشروق ؛ 1990 . – 511ص .

المكتبات المتخصصة – العالم العربي

Cybrarians Journal
E-ISSN 1687-2215
دورية إلكترونية محكمة تعنى بمجال المكتبات والمعلومات


```
<teiHeader>
	<fileDesc>
		<titleStmt>
			<title>المكتبات المتخصصة </title>
			<author role="مؤلف">
				<surname> كنت</surname>
				<forename>ألن</forename>
			</author>
			<author role="مترجم">
				<surname> الغامدي</surname>
				<forename>علي</forename>
			</author>
		</titleStmt>
		<editionStmt>
			<edition>ط1</edition>
		</editionStmt>
		<publicationStmt>
			<publisher>دار الشروق </publisher>
			<pubPlace>جدة</pubPlace>
			<date when="1990">1990</date>
		</publicationStmt>
		<sourceDesc>
			<biblStruct xml:lang="ar">
				<monogr>
					<title>المكتبات المتخصصة</title>
					<edition>ط1</edition>
					<author role="مؤلف">" ألن كنت </author>
					<author role="مترجم">علي الغامدي </author>
					<imprint>
						<publisher>دار الشروق </publisher>
						<pubPlace>جدة</pubPlace>
						<date when="1990">1990</date>
					</imprint>
					<extent>ص 511 </extent>
				</monogr>
			</biblStruct>
		</sourceDesc>
	</fileDesc>
	<profileDesc>
		<langUsage>
			<language ident="ar">عربية</language>
		</langUsage>
		<textClass>
			<keywords scheme="مكنز الالكسو">
				<term>المكتبات المتخصصة </term>
				<term>العالم العربي  </term>
			</keywords>
```

Cybrarians Journal
E-ISSN 1687-2215
دورية إلكترونية محكمة تعنى بمجال المكتبات والمعلومات
العدد 39، سبتمبر 2015

```
            </textClass>
        </profileDesc>
</teiHeader>
```

صورة رقم 2 : ترويسة teiHeader

1.3.2- محتوى الملف <text>

يستخدم هذا الجزء لترميز مختلف مكوّنات الوثيقة من صفحة العنوان إلى آخر صفحة وهوينقسم إلى ثلاثة أجزاء :

- **<front>** : هذا الجزء غير إجباري ويستعمل لترميز جميع المعلومات التي تسبق المحتوى الفعلي للملف مثل صفحة العنوان والإهداء والشكر والتمهيد والمستخلصات وكشّاف الموضوعات.
- **<body>** : يعتبر من أهمّ الأجزاء فهوإجباري ويستعمل لترميز المحتوى الفعلي للوثيقة.
- **<back>** : هذا الجزء غير إجباري ويستعمل لترميز مختلف المعلومات التّي توجد عادة في نهاية الوثيقة مثل الملاحق والببليوغرافيا والفهارس، الخ.

يوفّر المعيار الدولي TEI أكثر من 500 وسم والمئات من الصّفات لترميز محتويات أنواع مختلفة من الوثائق الرقميّة النصيّة في مجالات العلوم الإنسانية والاجتماعيّة وهي مبوّبة في شكل وحدات (Modules) حسب نوع الوثيقة ونوع البيانات لتيسير التعرّف عليها واستخدامها بطريقة صحيحة، نذكر منها على سبيل المثال المعاجم (Dictionaries) والمخطوطات (Description Manuscript) والقصائد الشعرية (Verse)، الخ. علما وأنّه بالإمكان دمج وسوم تنتمي إلى وحدات مختلفة لترميز الوثائق ذات المحتوى المركّب. فعلى سبيل المثال يمكن ترميز هذا المقتطف من كتاب مقامات بديع الزّمان الهمذاني باستعمال وسوم تنتمي إلى خمس وحدات مختلفة :

- Elements Available in All TEI Documents : لترميز العنوان (الجزء رقم 1)،
- Performance Texts : لترميز النصّ السردي (الجزء رقم 2)،
- Verse : لترميز القصيدة الشعريّة (الجزء رقم 3)،
- Tables, Formula, Graphics and Notated Music : لترميز الصورة (الجزء رقم 4)،

Cybrarians Journal
E-ISSN 1687-2215
دورية إلكترونية محكمة تعنى بمجال المكتبات والمعلومات
العدد 39، سبتمبر 2015

- Linking, Segmentation, and Alignment : لترميز الروابط التشعبيّة التّي تربط بين بعض المفردات في الجزء رقم 2 وشرحها في الجزء رقم 5.

المقامة الثانية عشر ــ المقامة البغدادية ــ 1

2 حدثنا عِيسى بْنُ هِشامٍ قال :إشتهيتُ **الأُزاذ**، وأنا ببغداد، وليس معي عَقْدٌ، على نَقْدٍ، فَخَرَجْتُ أنتهزُ مَحالَّهُ حَتَّى **الكَرْخِ**، فإذا أنا بِسَوَّادِيٍّ يَسوقُ بالجَهْدِ حِمارهُ، ويُطَرِّفُ بالعَقْدِ إزارَهُ، فَقُلْتُ : ظَفِرْنا واللَّهِ بصَيْدٍ، وَحَياكَ اللَّهُ أبا زَيْدٍ، مِنْ أيْنَ أقْبَلْتَ؟ وأيْنَ نزَلْتَ ؟ ومتى وافَيْتَ ؟ وَهَلُمَّ إلى البَيْتِ، فَقالَ السَّوَاديُّ :لَسْتُ بأبي زَيْدٍ، ولكنِّي أبُوعُبَيْدٍ، فقُلتُ : نَعَمْ ، لَعَنَ اللَّهُ الشيطانَ، وأبْعَدَ النِّسْيانَ، أنْسانيكَ طُولُ العَهْدِ، واتِّصالُ البُعْدِ، فكَيْفَ حالُ أبيكَ ؟ أشابٌّ كَعَهْدي، أَمْ شابَ بَعْدي ؟ فقال : لقدْ نَبَتَ الرَّبيعُ على **دِمْنَتِه**، وأرجوأنْ يُصَيِّرَه اللَّهُ إلى جَنَّتِهِ، فَقُلتُ : إنَّا لله وإنَّا إليهِ راجِعون، ولا حَوْلَ ولا قُوَةَ إلا بالله العلي العظيم، ومَدَدْتُ يَدَ البِدارِ، إليَّ الصِّدار، أريدُ تَمزيقَهُ، فَقَبَضَ السَّواديُّ على خَصْري بِجُمْعِهِ، وقال : نَشَدْتُكَ اللَّه لا مَزَّقْتَهُ، فَقُلْتُ : هِلُمَّ إلى البيتِ نُصِب غَذاءً، أوإلى السُّوق نَشْتَرِ شِواءً، والسُوق أقْرَبُ، وطعامُهُ أطْيَبُ، فاستَفَزَّتْهُ حُمَّةُ القَرَمِ، وعَطَفَتْهُ عاطِفَةُ اللَّقَمِ، وطَمِعَ، ولم يَعْلَمْ أنَّهُ وَقَعَ، ثُمَّ أتَيْنا شَوَّاءً يَتَقاطَرُ شِواؤهُ عَرَقاً، وَتَتَسايَلُ جَوذاباتُهُ مَرَقاً، فَقُلْتُ : افْرِز لأبي زَيْدٍ مِنْ هذا الشِّواءِ، ثُمَّ زِنْ لَهُ مِنْ تِلكَ الحَلواءِ، واخْتَرْ لَهُ مِنْ تِلكَ الأَطْباقِ، وانْضِد عليها أَوراقَ الرُّقاقِ، ورُشَّ عليه شَيئاً مِنْ ماءِ السُّمَّاقِ، لِيأْكُلَهُ أَبوزَيْدٍ هَنيَّا، فأَنْخى الشِّواءُ بِساطورِهِ، على زُبْدَةِ تَنُّورِهِ، فَجَعَلَها كالكُحْلِ سَحْقاً وَكالطِّحْنِ دَقّاً، ثُمَّ جَلَسَ وَجَلَسْتُ، ولا يَئِسَ ولا يَئِسْتُ، حتَّى اِستَوْفَيْنا، وقُلْتُ لِصاحِبِ الحَلْوى، زِنْ لأبي زَيْدٍ مِنَ اللوزينجِ رِطْلَيْنِ فَهوأَجْرى في الحُلوقِ، وأمْضَى في العُروقِ، ولْيَكُن لَيْلِيَّ العُمْرِ، يَوْمِيَّ النَشْرِ، رَقيقَ القِشْرِ، كَثيفِ الحَشْوِ، لُؤلُؤِيَّ الدُّهْنِ، كَوكَبِيَّ اللَّونِ، يَذوبُ كالصَّمْغِ، قَبْلَ المَضْغِ، لِيأْكُلَهُ أَبوزَيْدٍ هَنيّاً، قالَ : فَوَزَنَهُ ثُمَّ قَعَدَ وقَعَدْتُ، وجَرَّدَ وجَرَّدْتُ، حتَّى اُستَوْفَيْناهُ، ثُمَّ قُلتُ : يا أَ با زَيْدٍ ما أَحْوَجَنا إلى ماءٍ يُشَعْشِعُ بالثَّلْجِ، لِيَقْمَعَ هذهِ الصَّارَّةَ، ويُفْثَأَ هذه اللُّقَمَ الحارَّةَ، ، اِجْلِسْ يا أبا زَيْدٍ حتَّى نَأتيكَ بِسِقاءٍ، يأتيكَ بِشُرْبَةِ ماءٍ، ثُمَّ خَرَجْتُ وجَلَسْتُ بِحَيْثُ أَراهُ ولا يَراني أَنْظُرُ ما يَصْنَعُ، فَلَمَّا أَبْطأتُ عليه قامَ السَّواديُّ إلى حِمارِهِ، فاعْتَلَقَ الشَوَّاءُ بإذارِهِ، وقالَ : أَيْنَ ثَمَنُ ما أَكَلْتَ ؟ فَقالَ أَبوزَيْدٍ : أَكَلْتُهُ

Cybrarians Journal
E-ISSN 1687-2215
دورية إلكترونية محكمة تعنى بمجال المكتبات والمعلومات
العدد 39، سبتمبر 2015

ضَيْقاً، فَلَكَمَهُ لكْمَةً، وثَنَّى عليه بلَطْمَةٍ، ثُمَّ قال الشَوَّاءُ : هاكَ، ومَتى دَعَوْناكَ ؟ زِنْ يا أخا القِحَةِ عِشرينَ، فَجَعَلَ السَّواديُّ يَبْكي ويَحُلُّ عُقَدَهُ بأسْنانِهِ ويقولُ : كَمْ قُلْتُ لِذاكَ القُرَيْدِ، أنا أبوعُبَيْدٍ، وهويقول : أَنْتَ أَبوزَيْدٍ، فَأَنْشَدَتْ

3

أَعْمِلْ لِرِزْقِكَ كُلَّ آلَه — لا تَقْعُدَنَّ بِكُلِّ حالِهْ
وانْهَضْ بِكُلِّ عَظيمَةٍ — فالمَرْءُ يَعْجِزُ لا مَحالَهْ

4

5

الأزاذ : نوعٌ من التمر الجيد
الكَرْخ : مكان في بغداد
الدِمْنَة : القبر

Cybrarians Journal
E-ISSN 1687-2215
دورية إلكترونية محكمة تعنى بمجال المكتبات والمعلومات
العدد 39، سبتمبر 2015

1

```
<text><body>
<div type="fable" xml:lang="ara">
<head> ـ المقامة الثانية عشر ـ المقامة البغدادية </head>
```

2

```
<div>
<p> حدثنا عِيسى بْن هِشامٍ قال <lb/> اِشتهيتُ <ref target="#A1">الأزاذ</ref>، على نَقْدٍ، معي عَقْدٌ
، وأنا ببغداد
، فإذا أنا بسَوَادِيٍّ يَسوقُ بالجَهْدِ حِمارهُ، ويُطَرِّفُ بالعَقْدِ <ref target="#A2">الكَرْخ</ref> فَخَرَجْتُ أنتهزُ مَحالَّهُ حَتَّى أحَلَّني
ظَفِرْنا واللّهِ بصَيْدٍ، وَحَياكَ اللّهُ أبا زَيْدٍ، مِنْ أيْنَ أقْبَلْتَ؟ وأيْنَ نَزَلْتَ ؟ ومتى <q who="Issa Banou Hichem"> : إزارَهُ، فَقُلْتُ
، فقال السوادي<q>واقيْتَ ؟ وَهَلُمَّ إلى البَيْتِ
نَعَمْ ، لَعَنَ <q who="Issa Banou Hichem">: ، فقلت <q>لَسْتُ بأبي زَيْدٍ، ولكنّي أبُوعُبَيْدٍ<q who="Bedouin">
<q>اللّهُ الشيطان، وأبْعَدَ النِّسْيانَ، أنْسانيكَ طُولُ العَهْدِ، واِتِّصالُ البُعْدِ، فكَيْفَ حالُ أبيكَ ؟ أشابٌّ كَعَهْدي، أمْ شابَ بَعْدي ؟
، وأرجو أنْ يُصَيِّرَه اللّهُ إلى جَنَّتِهِ، <ref target="#A3">لِمِنْتَهِ</ref> لقدْ نَبَتَ الرَّبيعُ على <q who="Bedouin"> : فقال
</q> العلي العظيم إنَّا لله وإنَّا إليهِ راجِعون، ولا حَوْلَ ولا قُوَّةَ إلا بالله <q who="Issa Banou Hichem"> : ، فقلت<q>
نَشَدْتُكَ اللّه لا <q who="Bedouin"> ومَدَدْتُ يَدَ البدار، إليَّ الصِّدار أريدُ تَمزيقَهُ، فَقَبَضَ السَّواديُّ على خَصْري بِجُمْعِهِ، وقال
: ، فقلت<q> مَزَّقْتَه
هِلُمَّ إلى البيتِ نُصِب غَذاءً، أو إلى السُّوق نَشْتَرِ شِواءً، والسُوق أقْرَبُ، وطعامُهُ أطْيَبُ<q who="Issa Banou Hichem">
فاسْتَفَزَّتْهُ حُمَّةُ القَرَم ، وعَطَفَتْهُ عاطِفَةُ اللَّقَم ، وطَمِعَ، ولم يَعْلَمْ أنَّهُ وَقَعَ، ثُمَّ أتَيْنا شَوَّاءً يَتَقاطَرُ شِواؤهُ عَرَقاً، وَتَتَسايَلُ جَوذابَاتُهُ <q>
مَرَقاً،
اقْرِز لأبي زَيْدٍ مِنْ هذا الشِّواء، ثُمَّ زِنْ لَهُ مِنْ تِلْكَ الحَلواءِ، واخْتَرْ لَهُ مِنْ تِلْكَ <q who="Issa Banou Hichem"> : فَقُلْتُ
، فأنْخى الشِّوَّاءُ بِساطورِهِ، على <q> هَنِيًّا ماءِ السُّمَّاقِ، لِيأكُلَهُ أبوزَيْدٍ الأطْباقِ، وانْضِد عليها أوراقَ الزُّقاقِ، ورُشَّ عليه شَيئاً مِنْ
q> زُبْدَةٍ تَنُّورِهِ، فَجَعَلَها كالكُحْلِ سَحْقاً وكالطِّحْنِ دَقّاً، ثُمَّ جَلَسَ وَجَلَسْتُ، ولا يَئِسَ ولا يَئِسْتُ، حتَّى اِسْتَوْفَيْنا، وقُلْتُ لِصاحِبِ الحَلْوى،
who="Issa Banou Hichem"> زِنْ لأبي زَيْدٍ مِنَ اللوزينج رطْلَيْنِ فَهوأجْرى في الحُلوقِ، وأمْضَى في العُروقِ، ولْيَكُن لِيّاً كالصَّمْغ، قَبْلَ المَضْغ، لِيأكُلَهُ أبوزَيْدٍ هَنِيّا
<q> العُمْر، يَوْميَّ النَشْرِ رَقيقَ القِشْرِ، كَثيفِ الحَشْوِ،لَوْنٌ، يَذوبُ كالصَّمْغ، قَبْلَ المَضْغ، لِيأكُلَهُ أبوزَيْدٍ هَنيّا
فَوَزَنَهُ ثُمَّ قَعَدَ وقَعَدْتُ، وجَرَّدَ وجَرَّدْتُ، حتَّى اُستَوْفَيْناه، ثُمَّ قُلتُ
يا أ با زَيْدٍ ما أحْوَجَنا إلى ماءٍ يُشَعْشِعُ بالثَّلْجِ، لِيَقْمَعَ هذه الصَّارَّةَ، ويَقْتُأ هذه اللُّقَمَ الحارَّةَ، <q who="Issa Banou Hichem">
<q> اِجْلِسْ يا أبا زَيْدٍ حتَّى نَأتيكَ بِسِقاءٍ، يأتيكَ بِشُرْبَةِ ماءٍ
: ، ثُمَّ خَرَجْتُ وجَلَسْتُ بِحَيْثُ أراهُ ولا يَراني أنْظُرُ ما يَصْنَعُ، فَلَمَّا أبْطأتُ عليه قامَ السَّواديُّ إلى حِمارِهِ فاعْتَلَقَ الشَوَّاءُ بإذارِهِ، وقالَ
<q who="Restaurateur">أيْنَ ثَمَنُ ما أكَلْتَ ؟</q>
: ، فَلَكَمَهُ لَكْمَةً، وثَنَّى عليه بلَطْمَةٍ، ثُمَّ قال الشَوَّاءُ <q>أكَلْتُهُ ضَيْفاً<q who="Bedouin"> فَقالَ أبوزَيْدٍ
<q who="Restaurateur"><q> هاكَ، ومَتى دَعَوْناكَ ؟ زِنْ يا أخا القِحَةِ عِشرينَ
كَمْ قُلْتُ لِذاكَ القُرَيْدِ، أنا أبوعُبَيْدٍ، وهويقول : أنْتَ <q who="Bedouin"> : ، فَجَعَلَ السَّواديُّ يَبْكي ويَحُلُّ عُقْدَهُ بأسْنانِهِ ويقولُ
: فَأنْشَدَتْ <q>أبوزَيْدٍ
</p>
```

3

```
<lg type="verse">
    <l>أعْمِلْ لِرِزْقِكَ كُلَّ آلَه <caesura/> لا تَقْعُدَنَّ بِكُلِّ حالِهْ </l>
    <l>فالمَرْءُ يَعْجِزُ لا مَحالَهْ <caesura/> وانْهَضْ بِكُلِّ عَظيمَةٍ</l>
</lg>
```

4

```
<figure>
    <graphic url="maqama.jpg"/>
</figure>
```

5

```
</div>
<div type="glossary">
<list type="gloss">
    <item xml:id="A1">الأزاذ : نوعٌ من التمر الجيد</item>
```

Cybrarians Journal
E-ISSN 1687-2215
دورية إلكترونية محكمة تعنى بمجال المكتبات والمعلومات
العدد 39، سبتمبر 2015

```
        <item xml:id="A2">الكَرْخ : مكان في بغداد</item>
        <item  xml:id="A3">الدِمْنَة : القبر</item>
</list>
</div>
</div> <body> </text>
```

2- ترميز النصوص العربيّة: المشاكل والحلول

تصنّف اللّغة العربية من اللّغات السّامية وبالتحديد من اللّغات الساميّة الوسطى ويرجع المؤرّخون والباحثون نشأة وتطورّ الكتابة العربيّة إلى الكتابة النبطيّة مستدلّين في ذلك بالعديد من النّقوش التّي وقع اكتشافها على غرار نقش أمّ الجمال الذّي عثر عليه جنوب حوران وشرق الأردن ويعود تقريبا للعام 250 ميلادي ونقش حرّان الذي وقع اكتشافه جنوب دمشق ويعود للعام 568 ميلادي[8].

والأنباط هم أقوام عربية قديمة استقرّت في منطقة جغرافيّة تمتدّ من سيناء والجزء الشّمالي من الجزيرة العربية إلى جنوب الشّام ولقد طوّروا كتابتهم انطلاقا من الأبجديّة الآراميّة القديمة التّي اشتُقّت بدورها من الأبجديّة الفينيقيّة. ولقد أخذت الكتابة العربيّة عن الكتابة النبطيّة ارتباط بعض الحروف ببعض وتعدّد أشكال كلّ حرف حسب موضعه من الكلمة (ابتداء وتوسّط وانتهاء وانفراد) وكانت كتابة غير منقوطة ولا مشكولة.

بعد ظهور الإسلام عرفت الكتابة العربيّة العديد من التحويرات تتعلقّ بضبط الحروف عن طريق الشّكل والتنقيط ويعود الفضل في ذلك إلي ثلّة من العلماء النحاة على غرار أبي الأسود الدؤلي الذي ابتكر نظام تشكيل يعتمد على التنقيط باستعمال اللّون الأحمر. فنقطة فوق الحرف للدلالة على الفتحة ونقطة أسفله للدلالة على الكسرة ونقطة من شماله للدلالة على الضمّة ونقطتين بدلا من نقطة للدلالة على التنوين في كلّ موضع. ولحلّ مشكلة الحروف المتشابهة قام نصر بن عاصم الليثي ويحيى بن يعمر العدواني وبتكليف من الحجّاج بن يوسف الثّقفي بابتكار نظام التنقيط باستعمال اللّون الأسود.

لتفادي التباس نقاط الإعجام، ونقاط الشّكل، واختلاطهما على القارئ قام الخليل بن أحمد الفراهيدي بإبدال نقاط الشّكل التّي وضعها أبوالأسود الدؤلي بجرّات علويّة وسفليّة للدلالة على الفتح والكسر وواوًا صغيرة فوق الحرف للدلالة على الضمّ وكرّر هذه الحركات مرّتين إذا كان الحرف منونا وأضاف أشكالا أخرى لضبط القراءة مثل السكون الخفيف والسكون الشّديد واستعار رأس العين للهمزة

ورأس صاد صغير لألف الوصل وغيرها من الإصلاحات فأصبح ممكنا كتابة نصّ بنقاطه وشكله بلون واحد من المداد دونما لبس واستمرّ الشّكل بالطريقة نفسها حتّى يومنا هذا.

مع ظهور تكنولوجيا المعلوماتيّة وتقنيّات الطّباعة الآليّة في أواسط القرن الماضي مثّلت هذه الخصائص الشكليّة عوائق حالت دون معالجتها بطريقة سليمة تراعى فيها جماليّة الخط العربي وتحافظ على أبعاده الفنيّة لحقبة طويلة من الزمن. ومع إصدار المعيار الدّولي يونيكود وتبنّيه من قبل كبار الشركات المعلوماتية في العالم واعتماده في جلّ المواصفات القياسيّة الحديثة أصبح ممكنا إنشاء ومعالجة ونشر وتخزين المعلومات الرقميّة باللّغة العربيّة وبكلّ لغات العالم.

ومع أنّ المعيار الدّولي TEI يعتمد على اليونيكود كنظام تشفير أساسي، هنالك بعض المشاكل التّي يمكن أن نتعرّض لها عند ترميز النصوص العربيّة تتعلّق خاصة بإدراج بعض الرموز العربيّة، واتصّال الحروف، وطباعة وعرض النصوص ثنائيّة الاتّجاه والتّي تستوجب استخدام بعض المحارف والشفرات الإضافيّة التّي يمكن دمجها مباشرة مع وسوم وخاصيّات المعيار أوكتابتها في ملفّات خارجيّة.

نستعرض في هذا الجزء من البحث هذه المشاكل وطرق معالجتها.

2.1- إدراج الحروف

مثلما ذكرنا في الفقرات السّابقة، يعتمد الTEI على اليونيكود كنظام تشفير للمحارف ممّا يمكّن من إنشاء ملفّات متعدّدة الكتابات واللّغات ويخصّص هذا المعيار الدّولي في نسخته الأخيرة 1236 محرفا لتشفير الحروف والرموز المستعملة في الكتابة العربيّة وكذلك اللّغات المكتوبة بالأحرف العربيّة مثل الفارسيّة والأرديّة والبشتونيّة والكرديّة وهي تتوزّع على ستّ خرائط :

- خارطة رقم 06 : 254 محرف (Arabic Range: 0600–06FF)
- خارطة رقم 08 : 39 محرف (Arabic Extended-A : Range: 08A0–08FF)
- خارطة رقم FB وFC وFD : 691 محرف (Arabic Presentation Forms-A Range: FB50–FDFF)
- خارطة رقم FE : 252 محرف (Arabic Presentation Forms-B Range: FE70–FEFF)

Cybrarians Journal
E-ISSN 1687-2215
دورية إلكترونية محكمة تعنى بمجال المكتبات والمعلومات
العدد 39، سبتمبر 2015

تحتوي النسخة الأخيرة لليونيكود على 183 محرفا جديدا لتشفير الكتابة العربية تتمثّل في 143 محرفا لكتابة الريّاضيات بالعربيّة و40محرفا لتشفير علامات قرآنية إضافيّة وبعض الأحرف المستعملة في بعض اللّغات الإفريقية التي تكتب بالأحرف العربيّة.

عند عدم توفّر لوحة المفاتيح العربيّة أوعند استعمال محرّر TEI لاتيني أوإذا كانت ثمّة ضرورة لكتابة حرف أورمز غير متوفّر على لوحة المفاتيح يمكن حينئذ استعمال القيمة العشريّة أوالقيمة ستّ عشريّة للمحارف العربيّة لإدراجها في الملفّات.

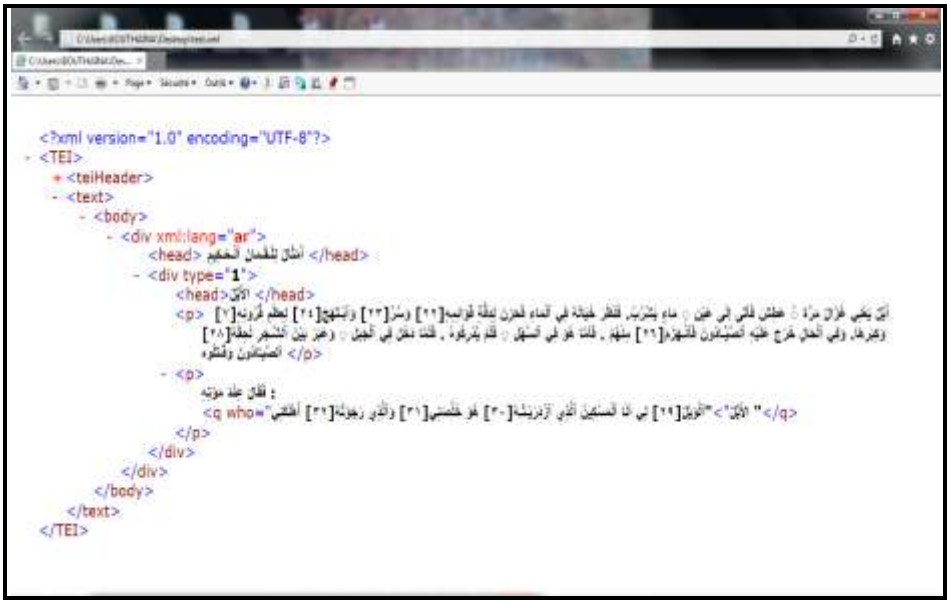
وتتمثّل القيمة العشريّة في الرقم التسلسلي للمحرف من 0 إلى 535 65 أمّا القيمة ستّ عشريّة فتتكوّن من رقمين وهما رقم الخارطة (Row) ورقم الخليّة (Cell) التي تتركّب بدورها من رقم العمود (Column) ورقم الصفّ (Line) فعلى سبيل المثال القيمة العشريّة لحرف "**ب**" هي 1576 وقيمته ستّ عشريّة هي 628 (6 رقم الخارطة و2 رقم العمود و8 رقم الصفّ).

عند استعمال الطريقة الأولى يجب أن تسبق القيمة العشريّة بسلسلة &# وتنتهي برمز الفاصلة المنقوطة في اللغة اللاتينيّة ";" وعند استعمال الطريقة الثّانية يجب أن تسبق القيمة ست عشريّة بسلسلة &#x وتنتهي برمز الفاصلة المنقوطة في اللغة اللاتينيّة ";" وتجدر الإشارة إلى إنّه يمكن استعمال الطريقتين لكتابة جميع محارف اليونيكود بدون استثناء.

وتبيّن الصّورة رقم 3 طريقة إدراج الحروف العربيّة باستعمال القيم ستّ عشريّة وتمثّل الصّورة رقم 4 الشكل النهائي لملفّ الTEI عند عرضه باستعمال متصفّح ويب.

Cybrarians Journal
E-ISSN 1687-2215
دورية إلكترونية محكمة تعنى بمجال المكتبات والمعلومات
العدد 39، سبتمبر 2015

صورة رقم 3 : إدراج الحروف العربيّة باستعمال القيم ستّ عشريّة



صورة رقم 4 : عرض الملف بواسطة متصفّح ويب (Internet Explorer)

Cybrarians Journal
E-ISSN 1687-2215
دورية إلكترونية محكمة تعنى بمجال المكتبات والمعلومات
العدد 39، سبتمبر 2015

2.2- اتصال الحروف

على غرار الكتابات الساميّة الأخرى تتميّز الكتابة العربيّة بارتباط بعض الحروف ببعض ممّا ينجرّ عنه تعدّد أشكال كلّ حرف حسب موضعه من الكلمة (ابتداء وتوسّط وانتهاء وانفراد) وحسب شكل الحرف الذّي يسبقه والذّي يليه.

ومع أنّ هذه الخاصية تعدّ من القواعد الأساسية للكتابة العربيّة، فإنّ العديد من شعراء العصر الجاهلي والعصر الإسلامي مثل عامر بن الطفيل والأعشى وأوس بن حجر وابن الرّومي وأيضا بعض المعاصرين استعملوا لضرورة لغويّة أووزنيّة التدوير في بعض قصائدهم الشعريّة الذي يتمثّل في التقاء الصدر والعجز من البيت في كلمة واحدة ويسمّى حينها البيت مدوّرا أومداخلا كقول الشّاعر عامر بن الطّفيل في قصيدته " قُلْ لزَيْدٍ قد كنتَ تُؤثَرُ بالحِلْـــ" :

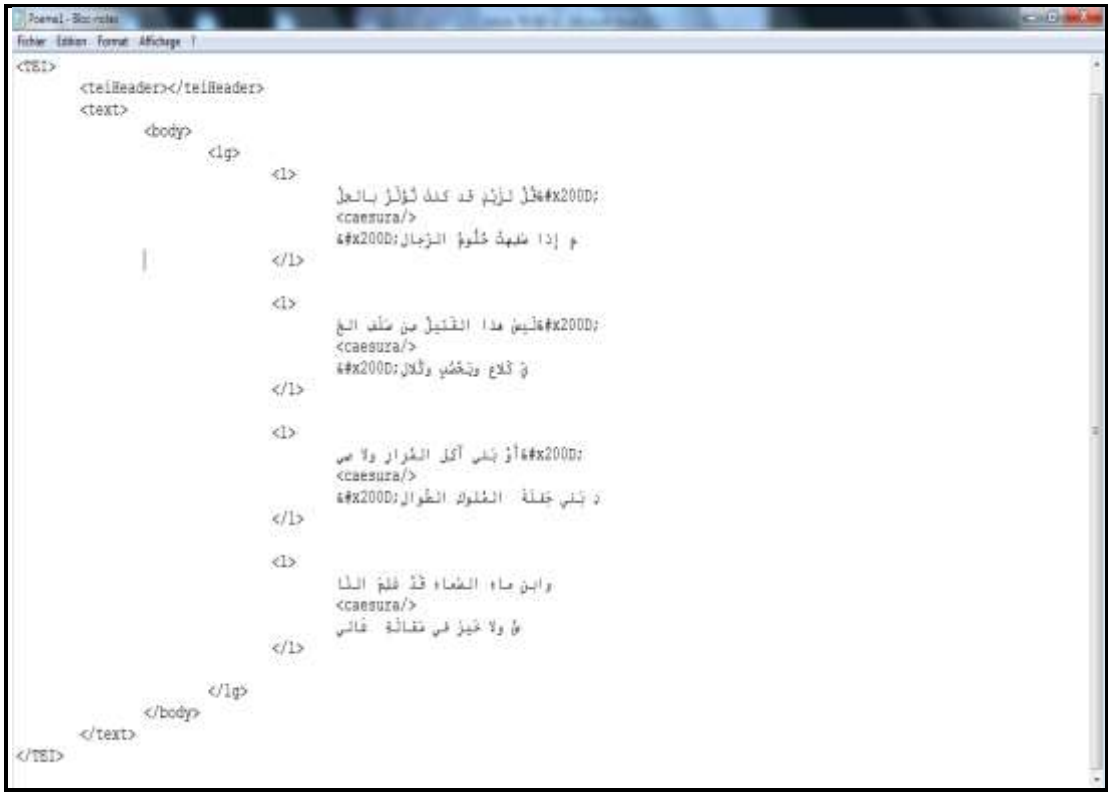
قُلْ لزَيْدٍ قد كنتَ تُؤثَرُ بالحِلْـــ ـــمِ إذا سَفِهتْ حُلُومُ الرّجالِ
لَيسَ هذا القَتيلُ مِن سَلَفِ الحَـــ ـــيّ كَلاعٍ ويَحْصُبٍ وكُلالِ
أوبَني آكِلِ المُرارِ ولا صيـــ ـــدِ بَني جَفنَةَ المُلوكِ الطّوالِ
غَالي وابنِ ماءِ السّماءِ قَدْ عَلِمَ النّا سُ ولا خَيرَ في مَقالَةِ

أفرد المعيار الدولي TEI وحدة خاصة لترميز القصائد الشعرية تحمل اسم **Verse** وهي تحتوي على العديد من الوسوم والصّفات التي تمكّن من توصيف جزئيّات دقيقة لأنواع مختلفة من القصائد الشعريّة إلّا أنّه لم يتعرّض لهذه الخاصيّة باعتبارها من الخصائص الشكليّة للوثيقة ولذلك يجب عند ترميز هذا النوع من الأبيات استعمال محرف يونيكود الواصل بعرض صفر (ZWJ: zero width joiner) لإظهار وطباعة الحرف الوارد في نهاية صدر البيت في شكله الموصول من الجهة اليسرى والحرف الوارد في مطلع عجز البيت في شكله الموصول من الجهة اليمنى.

وتجدر الإشارة إلى أنّ هذا المحرف يعدّ من المحارف غير المطبوعة وعند وضعه بين حرفين يفترض ألّا يتّصلا أوإثر حرف غير موصول من الجهة اليسرى فهويسببّ طباعتها في شكلها المتّصل ويمكن إدراجه في ملفّ الTEI باستعمال قيمته الستّ عشريّة (200D) أوقيمته العشريّة (8205).

Cybrarians Journal
E-ISSN 1687-2215
دورية إلكترونية محكمة تعنى بمجال المكتبات والمعلومات
العدد 39، سبتمبر 2015

تبيّن الصّورة رقم 5 طريقة ترميز هذا المقتطف من القصيدة باستخدام محرف الواصل بعرض صفر بالنسبة للأبيات الثلاثة الأولى (آخر حرف في الصدر وأوّل حرف في العجز) وتمثّل الصورة رقم 6 الشكل النهائي للملف بعد تحويله باستعمال لغة الـXSL وعرضه على متصفّح ويب.



صورة رقم 5 : استخدام محرف الواصل بعرض صفر

Cybrarians Journal
E-ISSN 1687-2215
دورية إلكترونية محكمة تعنى بمجال المكتبات والمعلومات
العدد 39، سبتمبر 2015

صورة رقم 6 : الشكل النهائي للقصيدة بعد عمليّة التحويل بلغة الXSL

Cybrarians Journal
E-ISSN 1687-2215
دورية إلكترونية محكمة تعنى بمجال المكتبات والمعلومات
العدد 39، سبتمبر 2015

2.3- ازدواجيّة الاتّجاه في النصوص

يسمح المعيار الدولي TEI بإنشاء ملفّات متعدّدة اللّغات والكتابات ولكنّه لا يتعرّض مطلقا إلى تحديد اتّجاه الحروف عند عرضها أوطباعتها إذ يكتفي فقط بالتعريف باللّغة المستعملة في كامل الملف إذا كان بنفس اللّغة أوفي جزء من أجزائه وذلك بتوفير الصّفة **@xml:lang** التي يجب إضافتها إلى الوسوم التّي تحتوي على نصوص :

<head xml:lang="ar"> مقدمة **</head>**

عندما يكون النصّ مكتوبا بنفس اللّغة أوبلغات تكتب في نفس الاتّجاه مثل اللّغات الأوربيّة فإنّ تحديد اتّجاه الكتابة عند عرضها أوطباعتها لا تطرح التباسا في معظم الحالات ولكن عندما يكون مزيجا من كلمات أونصوص بعضها يكتب من اليسار إلى اليمين والبعض الآخر يكتب من اليمين إلى اليسار يصبح من الضروري استخدام محارف التحكّم بالاتّجاه حتّى يتسنّى عرض النصّ وطباعته بشكله الصّحيح.

وقبل أن نستعرض محارف التحكّم بالاتّجاه تجدر الإشارة إلى أنّ المعيار يونيكود يصنّف بطريقة مقنّنة كلّ محرف من المحارف إلى ثلاثة أنواع من الاتّجاهات :

- **القوية** : تشمل على سبيل المثال جلّ المحارف الألفبائية العالميّة والأرقام باستثناء العربيّة والأوربيّة التي تكتب من اليسار إلى اليمين وتضمّ أيضا المحارف العربيّة والثانا (Thana) والسيرياك ومعظم محارف التشكيل والترقيم لهذه اللّغات التي تكتب من اليمين إلى اليسار إضافة إلى المحارف العبرية.
- **الضعيفة** : تشمل على سبيل المثال الأرقام العربيّة والأوربيّة ومحارف الترقيم مثل الفاصلة والفاصلة المنقوطة والنقطة وأيضا المحارف الحسابيّة.
- **الحياديّة** : تشمل على سبيل المثال فاصل الفقرات وفاصل الأجزاء وفاصل الأسطر.

تنتمي محارف التحكّم بالاتّجاه إلى النوع الأوّل ويمكن تصنيفها إلى أربعة أنواع[9] :

1. محرفي تضمين النص وتحديد الاتّجاه بصراحة : يستخدمان للإشارة بأنّ هناك نصّا موجودا ضمن نص آخر وأنّ اتّجاه كتابة النصّ المتضمّن مختلف عن اتّجاه كتابة النصّ الأصيل وهما محرف

تضمين النص من اليسار إلى اليمين (LRE: Left-to-Right Embeding) ومحرف تضمين النص من اليمين إلى اليسار (RLE: Right-to-Left Embedding)

2. محرفي إلغاء الاتّجاه بصراحة : يستخدمان لإلغاء اتّجاه النص المتضمّن وهما محرف إجبار الاتّجاه من اليسار إلى اليمين (LRO: Left-to-Right Override) ومحرف إجبار الاتّجاه من اليمين إلى اليسار (RLO: Right-to-Left Override) فعلى سبيل المثال يمكن في بعض النصوص العربيّة استخدام المحرف RLO لتضمين بعض الأرقام أوبعض الحروف اللّاتينية بحيث تكتب من اليمين إلى اليسار.
3. محرف إنهاء الاتّجاه بصراحة : ينهي تأثير محارف الاتّجاه السّابقة (LRO RLO LRE RLE) ويعيد اتّجاه النص إلى ما كان عليه وهومحرف (PDF: Pop Directional Format)
4. علامات التحكّم بالاتّجاه ضمنا : هذه المحارف شبيهة بمحارف تضمين النص وتحديد الاتّجاه بصراحة (RLM وLRM) إلّا أنّ تأثيرها محلّي أكثر من السّابقة وعرضها صفر عند عمليّة الترئية. وهذه المحارف هي محرف الاتّجاه من اليمين إلى اليسار وعرضه صفر(RLM: Right-to-Left Mark) ومحرف الاتّجاه من اليسار إلى اليمين وعرضه صفر (LRM: Left-to-Right Mark) ومحرف الاتّجاه العربي من اليمين إلى اليسار وعرضه صفر (ALM: Arabic Letter Mark).

حسب توصيّات المنظّمة العالميّة لتقييس تقنيّات الويب (World Wide Web Consortium) الصّادرة في 24 جانفي (يناير) 2013 بعنوان Unicode in XML and other Markup Languages[10] لا يجوز استخدام هذه المحارف مباشرة في نصوص الTEI لتفادى كلّ التباس عند معالجتها بل يجب استخدام نظائرها التي وقع تقنينها بالنسبة للغات الترميز (Markup Languages) ولغات تنسيق الصّفحات (Stylesheet Languages).

فعند استعمال لغة الCSS لتحديد البنية الشكليّة لملفّ TEI يجب ترجمة محارف التحكّم في الاتجاه إلى الصّفات التّالية[11] :

- {direction: ltr; unicode-bidi: embed}: لتعويض محرف تضمين النص من اليسار إلى اليمين (LRE)

Cybrarians Journal
E-ISSN 1687-2215
دورية إلكترونية محكمة تعنى بمجال المكتبات والمعلومات
العدد 39، سبتمبر 2015

- {direction: rtl; unicode-bidi: embed}: لتعويض محرف تضمين النص من اليمين إلى اليسار (RLE)
- {direction: ltr; unicode-bidi: bidi-override}: لتعويض محرف إجبار الاتّجاه من اليسار إلى اليمين (LRO)
- {direction: rtl; unicode-bidi: bidi-override}: لتعويض محرف إجبار الاتّجاه من اليمين إلى اليسار (RLO)

في سنة 2014 قام فريق العمل (Text Directionality Workgroup) التابع لمنظمة الTEI بتقديم مجموعة من المقترحات في إطار مشروع أوّلي[12] تتعلّق بمعالجة إشكاليّات عرض وطباعة النصوص مزدوجة الاتّجاه معتمدا في ذلك على خوارزميّات تحديد الاتّجاه للمعيار الدولي يونيكود ومواصفة CSS Writing Modes module[13] ومواصفة CSS Transform modules[14].

وتخصّ هذه المقترحات الكتابات الأفقيّة التي تكتب من اليسار إلى اليمين ومن اليمين إلى اليسار مثل العربيّة والعبريّة والكتابات اللّاتينيّة والكتابات العموديّة التي تكتب من اليمين إلى اليسار مثل اليابانيّة والكوريّة والصينيّة القديمة وأيضا الكتابات العموديّة التي تكتب من اليسار إلى اليمين مثل المنغوليّة القديمة. وتجدر الإشارة هنا إلى أنّ خوارزميّات تحديد الاتّجاه للمعيار الدولي يونيكود لا تعالج أي نوع من الكتابات العموديّة.

ويتمثّل مقترح فريق العمل في استحداث صفة شاملة تسمّى **@style** يمكن استخدامها كمتمّم للصّفة @xml:lang لتحديد اتّجاه ونمط الكتابة بالنسبة لوسوم الTEI ذات المحتوى النصّي. وتحتوي هذه الصّفة على القيم التاليّة :

direction : ltr | rtl
writing-mode : horizontal-tb | vertical-rl | vertical-lr
text-orientation: mixed | upright | sideways-right | sideways-left | sideways | use-glyph-orientation
unicode-bidi: normal | embed | isolate | bidi-override | isolate-override | plaintext

Cybrarians Journal
E-ISSN 1687-2215
دورية إلكترونية محكمة تعنى بمجال المكتبات والمعلومات
العدد 39، سبتمبر 2015

تستخدم القيمة الأولى لتحديد اتّجاه النصّ بالنسبة للكتابات الأفقيّة التي تكتب من اليسار إلى اليمين (direction:ltr) ومن اليمين إلى اليسار (direction:rlt) وتستخدم القيمة الثّانية لتحديد اتّجاه النصّ بالنسبة للكتابات الأفقيّة التّي تكتب من الأعلى إلى الأسفل (writing-mode:horizontal-tb) والكتابات العمّوديّة التي تكتب رموزها من اليمين إلى اليسار (writing-mode:vertical-rl) ومن اليسار إلى اليمين (writing-mode:vertical-lr). وأمّا القيمة الثالثة فتستخدم لتحديد اتّجاه النصّ على نفس السطر بالنسبة للكتابات العمّوديّة فقط وتمثّل القيمة الأخيرة محارف التحكّم بالاتّجاه لنظام التشفير اليونيكود.

فيما يلي مثال لاستخدام الصّفة @style لتحديد الاتّجاه بالنسبة لنصّ باللّغة الانكليزيّة يتضمّن نصّا باللّغة العربيّة :

```
<s xml:lang="en" style="direction: ltr">
 The Arabic term
<term xml:lang="ar" style="direction: rtl; unicode-bidi: embed">قلم رصاص</term> means "pencil".
</s>
```

الخاتمة

يعتبر الTEI من أهمّ المعايير الدّوليّة وأكثرها استعمالا لترميز وفهرسة وتكشيف الوثائق الرقمية وغير الرقمية في ميادين العلوم الإنسانيّة والاجتماعية واللسانيّات. ومنذ صدور نسخته الأولي في سنة 1990، مافتئ هذا المعيار يتطوّر ويتوسّع ليستجيب لحاجيات المستعملين المتزايدة وليواكب آخر تطوّرات تكنولوجيا المعلومات والاتّصال.

ولعلّ أهمّ ما يتميز به مقارنة بنظم الترميز الأخرى هواهتمامه فقط بترميز البنية المنطقية لأنواع مختلفة من الوثائق بواسطة طقم من الوسوم والخاصيات يتجاوز عددها المئات تستعمل لتوصيف مختلف مكونّات محتوى الوثيقة بطريقة مقننّة وعلى مستوى عال من الدقّة ممّا يمكّن من إنشاء قواعد بيانات وكشّافات بطريقة آليّة ومن تيسير عمليّات البحث والاسترجاع. كما يتميّز أيضا باعتماده على العديد من المواصفات القياسيّة الحديثة التّي تتعلّق بتشفير المحارف وتبادل المعلومات على الخطّ.

فيما يتعلّق بترميز النصوص العربيّة باستعمال الTEI، قمنا من خلال هذا البحث بدراسة أهمّ المشاكل التي يمكن أن نتعرّض لها مع اقتراح الحلول المناسبة التي تتمثّل في استخدام بعض المحارف والشفرات الإضافيّة التّي يمكن دمجها مباشرة مع وسوم وصفات المعيار أوكتابتها في ملفّات خارجيّة. وتجدر الإشارة إلى أنّ هذه الحلول تتعلّق بإظهار وطباعة الحروف العربيّة في شكلها الصحيح أومثلما وردت في النصوص الأصليّة الورقيّة.

الاستشهادات المرجعية

[1] Lupovici, Catherine (1993). Révolution électronique et normalisation. *Bulletin des Bibliothèques de France (BBF)*, T.38, n°5, 22-31.

[2] TEI: Text Encoding Initiative. Accessed July 20, 2015. Available at: http://www.tei-c.org

[3] http://www.tei-c.org/Activities/Projects/ لمزيد المعلومات حول أهم المشاريع في العالم

[4] Ourabah Soualah, Mohammed & Hassoun, Mohamed. A TEI P5 Manuscript Description Adaptation for Cataloguing Digitized Arabic Manuscripts. *Journal of the Text Encoding Initiative*, Issue 2, February 2012. Accessed July 20, 2015. Available at: http://jtei.revues.org/398.

[5] Hudrisier, Henri & Zghibi, Rachid & Zghidi, Sihem & Ben Henda, Mokhtar (2013). Promoting the linguistic diversity of TEI in the Maghreb and the Arab region. *The Linked TEI: Text Encoding in the Web. TEI Conference and Members Meeting 2013: October 2-5, Rome (Italy)*. Accessed July 20, 2015. Available at: http://digilab2.let.uniroma1.it/teiconf2013/program/papers/abstracts-paper#C174

[6] Burnard, Lou & Serberg-McQueen (1996). La TEI simplifié : une introduction au codage des textes électroniques en vue de leur échange. *Cahiers GUTenberg*, n°24, juin 1996, 23-151.

[7] André, Jacques & Quint, Vincent (1991). Structures et modèles de documents. *Le document électronique*, 3-60.

[8] ابو الحب، سعد الدين. جذور الكتابة العربية الحديثة : من المسند الى الجزم. كلية بروك، جامعة مدينة نيويورك

Accessed July 20, 2015. Available at: http://www.academia.edu/1958611/_

[9] Davis, Mark & Lanin, Ahoran & Glass Andrew (2014). Unicode bidirectional algorithm. Accessed July 20, 2015. Available at: http://www.unicode.org/reports/tr9/

[10] W3C (2013). Unicode in XML and other Markup Languages. Accessed July 20, 2015. Available at: http://www.w3.org/TR/unicode-xml/

[11] Ishida, Richard (2007). CSS vs. markup for bidi support. Accessed July 20, 2015. Available at: http://www.w3.org/International/questions/qa-bidi-css-markup

[12] Text Directionality Workgroup (2014). Text directionality draft. Accessed July 20, 2015. Available at: http://wiki.tei-c.org/index.php/Text_Directionality_Draft

[13] W3C (2013). CSS Writing Modes Level 1. Accessed July 20, 2015. Available at: http://dev.w3.org/csswg/css-writing-modes/

[14] W3C (2014). CSS Writing Modes Level 3. Accessed July 20, 2015. Available at: http://www.w3.org/TR/css3-transforms