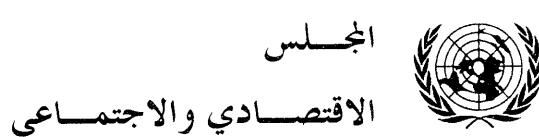


Distr.
LIMITED

E/ESCWA/ICTD/2003/WG.2/14/
2 June 2003
ORIGINAL: ARABIC



اللجنة الاقتصادية والاجتماعية لغربي آسيا
اجتماع خبراء حول تعزيز المحتوى الرقمي العربي
بيروت، ٥-٣ حزيران/يونيو ٢٠٠٣

تقييم استخدام اللغة العربية في الحاسوب

ملاحظة: طبعت هذه الوثيقة بالشكل الذي قدمت به ودون تحرير رسمي. والآراء الواردة فيها هي أراء المؤلف وليس بالضرورة، أراء الإسكوا.

تقنيات استخدام اللغة العربية في الحاسوب

الدكتور أحمد أبو الهيجاء
جامعة العلوم والتكنولوجيا الأردنية
إربد - الأردن

مجموعة المحارف العربية

تمهيد

تستخدم الحروف العربية لكتابة اللغة العربية. وتعتبر الحروف العربية هي الحروف الأساسية لعدد آخر من اللغات، وتستخدم مع حروف أخرى لكتابة لغات أخرى مختلفة منها الفارسية والكردية والأوردو، وغير ذلك. وإن اللغة العربية متصلة بطبيعتها، ولذلك يمكن كتابة نفس الحرف بأشكال مختلفة حسب ظهوره في الكلمة ومعتمد على الحرف الذي يقع على يساره والحرف الذي يقع على يمينه. وي بيان الشكل رقم (١) مثلاً على ذلك لحرف الهاء. كما يمكن استخدام حركات التشكيل وتظهر في الكتابة فوق الحروف المصاحبة لها. ومع أن حركات التشكيل غير شائعة الاستخدام في بعض الدول العربية، إلا أنها ضرورية ومهمة للغاية.

الشكل رقم (١): اتجاه الكتابة وتوصيل الحروف

التخزين في الذاكرة (الاتجاه من اليسار إلى اليمين)	٥ ٥ ٥
تغير الاتجاه (التصبح الحروف من اليمين إلى اليسار)	٥ ٥ ٥
وصل الحروف مع بعضها	هـ هـ هـ

اتجاه الكتابة

تكتب اللغة العربية من اليمين إلى اليسار. وتعطي معظم الموصفات الدولية ترميز المحارف لتخزينها في الذاكرة بما يتناسب مع التنظيم أو الترتيب المنطقي. وعندما يتم ترتيب النصوص المختلفة على شكل أسطر، فإن معظم اللغات تظهر حروفها من اليسار إلى اليمين. وفي عدد من اللغات (ومنها اللغة العربية، واللغة الفارسية، واللغة الأردية)، فإن الحروف تظهر من اليمين إلى اليسار. ولو كان النص المكتوب بكامله له نفس اتجاه الكتابة، لما كان هناك التباس في كيفية تنظيم ترتيبة هذا النص. أما عندما يكون النص ثانوي الاتجاه (أو بعبارة أخرى يكون مزيجاً من نصوص اتجاه بعضها من اليسار إلى اليمين واتجاه البعض الآخر من اليمين إلى اليسار)، قد ينشأ هناك بعض الالتباس في تحديد اتجاه الحروف عند ترتيبها أو طباعتها.

في حالة النصوص مزدوجة الاتجاه، من الضروري تحديد بعض محارف التحكم التي تمكن من ترتيبة وطباعة النصوص وتبادلها عبر شبكة الإنترنت بشكلها الصحيح. وهذا يضمن التحكم بكيفية ترتيب الحروف عند ترتيبها لتبادلها بشكل صحيح، كما يضمن ترتيبة بعض النصوص البسيطة التي تستخدم لتسمية الملفات وغيرها لظهور دائماً بشكلها الصحيح. وتستخدم محارف التحكم بالاتجاه لهذا الغرض فقط لا غير، ويتم استثناؤها في جميع المناحي الأخرى (أي أنها لا تؤثر على مقارنة النصوص، والفصل بين الكلمات، والترتيب، والتحليل العددي، وغير ذلك). وعندما يتم التعامل مع النصوص مزدوجة الاتجاه ومعالجتها، فإنه يتم معالجة المحارف بالترتيب المنطقي لها، حيث لا يتتأثر هذا الترتيب بمحارف التحكم بالاتجاه (التي تؤثر على اتجاه الترتيبة فقط لا غير).

محارف التحكم بالاتجاه

هناك نوعان من محارف التحكم بالاتجاه يمكن استخدامها بصرامة ووضوح لتعديل محارف التحكم بالاتجاه الضمنية المستخدمة في خوارزمية ثنائية الاتجاه المعرفة في المعايير الدولية ISO 10646 . علاوة

على ذلك، هناك محارف ضمنية للترتيب وهي: محرف "من اليمين إلى اليسار" ، ومحرف "من اليسار إلى اليمين". وتؤثر جميع هذه المحارف على الفقرة الحالية، حيث ينتهي تأثيرها عندما يرد "فاصل الفقرات" - paragraph separator . ويجب التمييز بين نوعين من محارف التحكم بالاتجاه. وإن محارف التحكم بالاتجاه "من اليمين إلى اليسار" و "من اليسار إلى اليمين" التي تتحكم باتجاه النصوص، تسمى محارف التحكم بالاتجاه القوية. أما تلك المصاحبة للأرقام، فتسمى محارف التحكم بالاتجاه الضعيفة. عندما يتضمن نص باتجاه معين (من اليسار إلى اليمين مثل) نصا آخر باتجاه معاكس (من اليمين إلى اليسار مثل)، يتم استخدام محارف التحكم بالاتجاه للنص الثاني. وإن النص الذي يتأثر بمحارف التحكم من اليمين إلى اليسار، يعتمد على اتجاه النص الأصيل ولا يكون معزولا عنه. وبعبارة أخرى، فإن ترتيب الحروف داخل النص الثاني (الذي يكتب من اليمين إلى اليسار) يمكن أن يؤثر على ترتيب الحروف داخل النص الأول (الذي يكتب من اليسار إلى اليمين)، والعكس بالعكس.

تضمين النص وتحديد الاتجاه بصرامة

إن محارف التحكم الآتية تشير بأن هناك نصا موجودا ضمن نص آخر. وعلى سبيل المثال، وجود نص باللغة الإنجليزية ضمن جملة باللغة العربية يستلزم وجود محرف تحكم ليدل أن النص الذي اتجاهه "من اليسار إلى اليمين" هو ضمن النص الأصلي الذي اتجاهه "من اليمين إلى اليسار" . كما أن وجود نص باللغة العربية ضمن جملة باللغة الإنجليزية يستلزم وجود محرف تحكم ليدل أن النص الذي اتجاهه "من اليمين إلى اليسار" هو ضمن النص الأصلي الذي اتجاهه من "اليسار إلى اليمين". وإن محارف التحكم هذه تسمح تضمين النصوص بعضها البعض على عدد من المستويات، والمحارف هي:

(أ) **تضمين النص "من اليمين إلى اليسار"** : بحيث يعتبر النص التالي متضمنا من اليمين إلى اليسار ضمن النص الأصيل (Right-to-Left Embedding - RLE) ؛

(ب) **تضمين النص "من اليسار إلى اليمين"** : بحيث يعتبر النص التالي متضمنا من اليسار إلى اليمين ضمن النص الأصيل (Left-to-Right Embedding - LRE) .

إلغاء تحديد الاتجاه بصرامة

إن محارف التحكم الآتية تسمح بإلغاء تحديد الاتجاه للنص المتضمن عندما يكون ذلك مطلوبا في بعض الحالات الخاصة، ومحارف التحكم هي :

(أ) **جعل الاتجاه "من اليمين إلى اليسار"** : التي تجبر المحارف التي تليها ليتم التعامل معها من اليمين إلى اليسار وفق الاتجاه الأقوى (Right-to-Left Override - RLO) ؛

(ب) **جعل الاتجاه "من اليسار إلى اليمين"**: التي تجبر المحارف التي تليها ليتم التعامل معها من اليسار إلى اليمين وفق الاتجاه الأقوى (Left-to-Right Override - LRO) .

وعلى سبيل المثال، فإن محرف التحكم RLO يمكن استخدامه ليجبر سلسلة معينة من المحارف (مثل رقم قطعة في المستودعات) تختلف من بعض أو كل مما يأتي: حروف عربية، وأرقام، وحروف لاتينية، ليتم كتابتها جميعها من اليمين إلى اليسار .

حرف إنتهاء الاتجاه بصراحة

إن حرف التحكم الآتي ينهي تأثير المحارف التي تحدد الاتجاه بصراحة التي تم ذكرها سابقاً (وهي محارف RLO، LRE، RLE)، ويعيد اتجاه النص إلى ما كان عليه قبل ورود حرف تحديد الاتجاه. وإن هذا الحرف هو "حرف استعادة حالة الاتجاه لما كانت عليه قبل آخر ورود لحرف تحديد الاتجاه RL، LRE، أو LR". وهذا الحرف يسمى (Pop Directional Format - PDF).

علامات التحكم بالاتجاه ضمناً

إن تأثير هذه المحارف هو شبيه بتأثير محارف التحكم بالاتجاه "من اليمين إلى اليسار" أو "من اليسار إلى اليمين" ، باستثناء أن هذه المحارف لا يتم ترتيبتها وليس لها معنى آخر. وإن استخدامها عادة يكون ملائماً أكثر من المحارف التي تحدد الاتجاه صراحة حيث أن مداها محلية أكثر من السابقة، وهذه المحارف هي:

- حرف الاتجاه من اليمين إلى اليسار وعرضه صفر (Right-to-Left Mark - RLM).
- حرف الاتجاه من اليسار إلى اليمين وعرضه صفر (Left-to-Right Mark - LRM).

ازدواجية الاتجاه في النصوص

تحدد هذه الورقة طريقة لتمثيل المحارف في الذاكرة، وتعرف هذه الطريقة في الموصفات الدولية (أنظر مواصفة يونيكود) بالترتيب الافتراضي. عندما يتم عرض (أو ترتيب) نص ما في اللغة العربية، يتم العرض بأسطر أفقية وتظهر المحارف مرتبة من اليمين إلى اليسار. وإن معظم اللغات العالمية الأخرى يتم عرضها بأسطر أفقية، ولكن تظهر المحارف مرتبة من اليسار إلى اليمين. ولو كانت جميع النصوص لها نفس اتجاه العرض لما كان هناك أي التباس عند ترتيبية المحارف أو طباعتها. ولكن عندما نريد عرض نصوص في اتجاهات مختلفة (أي أن النص مزيج من عبارات بعضها من اليمين إلى اليسار والبعض الآخر من اليسار إلى اليمين)، فقد يتنشأ بعض الالتباس عند تحديد اتجاه الترتيبة لكل جزء من النص.

يصنف هذا الجزء من الورقة الخوارزمية اللازمة لتحديد الاتجاهات في النصوص مزدوجة الاتجاه. وتراعي هذه الخوارزمية التطبيقات الموجودة حالياً والموصفات الدولية في هذا المجال، وتعرف محارف تحكم لمعالجة بعض الحالات الخاصة. وفي معظم الحالات لا توجد حاجة لوضع معلومات إضافية في النصوص لتحديد اتجاه العرض أو الترتيبة. وفي حالة النصوص مزدوجة الاتجاه، توجد هناك بعض الظروف التي لا يكفي فيها الترتيب الضمني للاتجاه لعرض النصوص بطريقة مقبولة. وللتعامل مع هذه الحالات، تعرف هذه الورقة الحد الأدنى لمحارف التحكم بالاتجاه التي يمكن من ترتيب المحارف لعرضها بالشكل المطلوب. وهذا يمكن من التحكم التام والدقيق بترتيب المحارف لعرضها بأي شكل نريد وكذلك تبادل المعلومات بالاتجاه المطلوب، كما يضمن أن النصوص البسيطة كالتي تستخدم لأسماء الملفات أو العناوين سيتم عرضها دائماً بالشكل الصحيح.

تستخدم شفرات التحكم بالاتجاه للتأثير على ترتيب المحارف من أجل عرضها بطريقة سليمة، ولا تستخدم لأي غرض آخر إطلاقاً. وعلى سبيل المثال، فلا تؤثر هذه الشفرات على المقارنة بين النصوص، أو ترتيب وفرز النصوص، أو الفصل بين الكلمات، أو التحليل العددي، أو أي تطبيقات أخرى. وعند التعامل مع النصوص مزدوجة الاتجاه، يتم التعامل مع المحارف بالترتيب المنطقي لها، وإن اتجاه عرض النصوص هو فقط الذي يتتأثر بمحارف التحكم.

خوارزمية الترئية الأساسية

إن خوارزمية ترئية النصوص ثنائية الاتجاه تأخذ كمدخل لها سلسلة من المحارف وتم عمليتها على ثلاثة مراحل:

- (ا) فصل النص المدخل إلى فقرات. وإن بقية الخوارزمية تعالج النص بين فواصل الفقرات؛
- (ب) تقسيم النص إلى مستويات حسب احتوائه على نصوص باتجاهات مختلفة. وفي هذه المرحلة، تستخدم محارف التحكم بالاتجاه المذكورة سابقاً لتحديد هذه المستويات؛
- (ج) ترتيب النص للطباعة أو الترئية على الشاشة سطراً بعد سطر وذلك باستخدام مستويات الاحتواء الواردة أعلاه، وذلك بعد تقسيم النص إلى أسطر.

وإن الخوارزمية ترتيب النص في الفقرة الواحدة، حيث أن المحارف في فقرة ما لا تأثير لها على المحارف في فقرة أخرى. وتكون الفقرات مفصولة عن بعضها بواسطة فاصل الفقرات (Paragraph Separator) المعروض في المواصفة الدولية إيزو 10646. ولتخزين النصوص العربية وتبادلها وترئيتها بشكل صحيح عبر شبكات المعلومات بحيث يراها الشخص المرسل والشخص المستقبل بنفس الكيفية، يجب استخدام الخوارزمية الموصوفة أدناه، والمتتناسبة مع المواصفات الدولية (إيزو 10646 ومواصفة اليونيكود).

يمكن فصل الفقرات عن بعضها باستخدام وظيفة "سطر جديد" (New Line). ويمكن الاطلاع على المواصفات الدولية الحديثة لمعرفة كيفية استخدام محارف التحكم المتعلقة بهذه الوظيفة، ومنها:

- رجوع العربة (Carriage Return - CR)
- تغذية السطر (Line Feed - LF)
- رجوع العربة وتغذية السطر (Carriage Return Line Feed - CRLF)

ويمكن كذلك تحديد الفقرات بواسطة البروتوكولات ذات المستوى المتقدم. وعلى سبيل المثال، فإن النص في خليتين مختلفتين من جدول ما، سيتم التعامل معهما كفقرتين مختلفتين. نعرف فيما يلي بعض المصطلحات التي ستستخدم في الخوارزمية:

(١) أنواع المحارف ثنائية الاتجاه: هي قيم تخصيص لكل حرف من المحارف.

(٢) مستويات التضمين: هي أرقام تشير إلى مدى عمق التضمين في النصوص، وإلى الاتجاه الأساسي للنص في ذلك المستوى. وإن أدنى مستوى تضمين للنصوص هو صفر، وأعلى مستوى تسمح به مواصفة يونيكود الدولية هو 61. ومن ناحية عملية، فإن مستوى التضمين يكون عادةً رقماً صغيراً، وزيادة مستوى التضمين يؤدي إلى تعقيدات كبيرة في النصوص يصعب التعامل معها. وإن مستويات التضمين عادةً يتم تحديدها بواسطة شفرات الإلغاء (Override codes) أو شفرات التضمين، وكلما زاد مستوى التضمين كلما زاد عمق النص داخل النصوص الأخرى.

(٣) اتجاه التضمين: إن هذا هو الاتجاه الأساسي لمستوى التضمين الحالي (أي لحرف ما في النص). يكون الاتجاه L إذا كان مستوى التضمين زوجي، ويكون الاتجاه R إذا كان مستوى التضمين فردي. وعلى سبيل المثال، في جزء معين من النص، يكون مستوى التضمين صفر عندما يكون النص باللغة

الإنجليزية وحدها، ويكون مستوى التضمين واحد عندما يكون النص باللغة العربية وحدها. ويكون مستوى التضمين ٢ للنص باللغة الانجليزية الذي يكون ضمن نص باللغة العربية له مستوى التضمين ١، وهكذا. عادة يكون مستوى التضمين زوجي للنصوص باللغة الانجليزية والأرقام، ويكون مستوى التضمين فردي للنصوص باللغة العربية (باستثناء الأرقام)، إلا عندما يتم إلغاء الاتجاه باستخدام أحد محارف التحكم الخاصة بذلك (Directional Overrides). وإن المعنى الدقيق لمستوى التضمين سيكون واضحا عند شرح خوارزمية الترئية الأساسية.

(٤) مستوى تضمين الفقرة: هو مستوى التضمين الذي يحدد الاتجاه الأساسي للنص في الفقرة المعنية. وإن اتجاه مستوى التضمين يسمى اتجاه الفقرة أو الاتجاه الأساسي.

(٥) الوضع النهائي للاتجاه (Directional Override Status) : يحدد هذا الوضع فيما إذا كان اتجاه المحارف قد ألغى بواسطة محارف تحكم صريحة. وهذا الوضع له ثلاثة حالات كما يبين الجدول رقم (١).

الجدول رقم (١): الوضع النهائي للاتجاه

الحالات	التفسير
حيادي	لا يوجد حاليا إلغاء للاتجاه
من اليمين إلى اليسار	يعاد اتجاه المحارف ليصبح R (أنظر اتجاه التضمين الوارد أعلاه)
من اليسار إلى اليمين	يعاد اتجاه المحارف ليصبح L (أنظر اتجاه التضمين الوارد أعلاه)

مثال: يوضح المثال الآتي بعض المفاهيم الواردة أعلاه.

blue	car	is	السيارة	in	Arabic	العرض على الشاشة:
blue	car	is	سيارة	in	Arabic	التخزين في الذاكرة:
LLLL - LLL - LL - RRRRRR - RRRRRR - LL - LLLL	0000 0 000 0 00 0 1111111 1 1111111 0 00 0 000000					أنواع المحارف:
						مستوى المحارف:

لاحظ أن المحرف الحيادي (محرف الفراغ) بين "السيارة" و"الزرقاء" يكون له نفس مستوى المحرفين المجاورين، أي المستوى ١. وإذا وضعنا علامات تحديد الاتجاه المناسبة حول المحارف الحيادية ، فإنه يمكن تغيير مستوى هذه المحارف.

أنواع اتجاهات المحارف

يلخص الجدول رقم (٢) أنواع اتجاهات المحارف لكل محرف.

الجدول رقم (٢): أنواع اتجاهات المحارف

الصنف	النوع	الوصف	المجال
L		محرف الاتجاه من اليسار إلى اليمين (LRM) معظم المحارف الألفبائية العالمية، الأرقام باستثناء العربية والأوروبية	من اليسار إلى اليمين
LRE	القوية	محرف التضمين من اليسار إلى اليمين (LRE) محرف جعل الاتجاه من اليسار إلى اليمين (LRO)	محرف التضمين من اليسار إلى اليمين
LRO		محرف جعل الاتجاه من اليسار إلى اليمين (LRO)	محرف جعل الاتجاه من اليسار إلى اليمين
R		محرف الاتجاه من اليمين إلى اليسار (RLM) ()، المحارف العبرية، معظم محارف التشكيل للغة العبرية.	من اليمين إلى اليسار
AL		اللغة العربية من اليمين إلى اليسار محارف التشكيل لهذه اللغات.	المحارف العربية والثانية والسيرياك، معظم محارف التشكيل لهذه اللغات.
RLE		محرف التضمين من اليمين إلى اليسار (RLE)	محرف التضمين من اليمين إلى اليسار
RLO		محرف جعل الاتجاه من اليمين إلى اليسار (RLO)	محرف جعل الاتجاه من اليمين إلى اليسار
PDF		محرف استعادة حالة الاتجاه (PDF) لما كانت عليه قبل آخر ورود لمحرف تحديد الاتجاه.	محرف استعادة حالة الاتجاه
Pop Directional Format	الضعيفة	العدد الأوروبي European Number	الأرقام الأوروبية والعربية
ES		فاصل العدد الأوروبي	الخط المائل (/)
ET		إنهاء العدد الأوروبي European Number	علامة الزائد، علامة الناقص، علامة الدرجة، رمز العملة، وغيرها
AN		العدد العربي Arabic Number	الأرقام العربية، الفواصل العشرية، وفواصل الآلاف ضمن الأرقام العربية، وغيرها
CS		الفاصل العام للعدد Common Number Separator	الفاصلة، الفاصلة المنقوطة، النقطة، محرف الفسحة بدون فراغ (NO-BREAK SPACE)، وغيرها

الجدول رقم (٢) - (تابع)

<p>المحارف المعروفة بـ علامة عدم الفراغ Mn (non spacing mark)</p> <p>وعلامة المحتوى (Me (enclosing mark)) في قاعدة بيانات محارف مواصفة يونيكود العالمية</p> <p>محارف التحكم والتوضيب، باستثناء المحارف التي تم ذكرها صراحة أعلاه، ويتم استثناء هذه المحارف عند معالجة النصوص ثنائية الاتجاه</p> <p>فاصل الفقرات، وظائف السطر الجديد، تحديد الفقرات في البروتوكولات المتقدمة (Tab)</p> <p>حرف فاصل الأجزاء (Tab)</p> <p>حرف الفسحة، شكل الفسحة، فاصل الأسطر، تغذية السطر، والمحارف العامة الأخرى للفسحة</p> <p>جميع المحارف الأخرى التي لم تذكر أعلاه، بما في ذلك حرف OBJECT</p> <p>REPLACEMENT CHARACTER</p> <p>(أنظر مواصفة يونيكود العالمية)</p>	<p>علامة عدم الفراغ NSM</p> <p>Non-Spacing Mark</p> <p>حيادية الحافة BN</p> <p>Boundary Neutral</p> <p>فاصل الفقرات B</p> <p>Paragraph Separator</p> <p>فاصل الأجزاء S</p> <p>Segment Separator</p> <p>الفراغ الأبيض WS</p> <p>White space</p> <p>المحارف الحيادية ON</p> <p>الأخرى</p> <p>Other Neutrals</p>
--	---

تحديد مستويات التضمين

إن الخوارزمية المختصة بمعالجة النصوص ثنائية الاتجاه تستخدم أنواع المحارف والشفرات الصريحة لتعطي قائمة بمستويات النصوص. وإن هذه العملية تتم من خلال خمس خطوات:

- (١) تحديد مستوى الفقرة
- (٢) تحديد مستويات واتجاهات التضمين الصريح (explicit embedding)
- (٣) تحديد الأنواع الضعيفة
- (٤) تحديد الأنواع الحيادية
- (٥) تحديد مستويات التضمين الضمنية (implicit embedding)

تحديد مستوى الفقرة

تقوم خوارزمية ثنائية الاتجاه بمعالجة الفقرات وفق الخطوات الآتية:

- (ف) فصل النص إلى فقرات مستقلة. وإن فاصل الفقرة يبقى ضمن الفقرة السابقة. ويتم تطبيق جميع خطوات وقواعد الخوارزمية على كل فقرة.

(ف٢) في كل فقرة، يتم إيجاد أول حرف من المحارف التي تكون اتجاهاتها من الصنف "القوية"، والتي يكون نوعها L، AL، أو R.

(ف٣) إن تم إيجاد حرف في الخطوة (2) وكان نوعه AL أو R، يتم تحديد مستوى تضمين الفقرة بـ(1)؛ وفيما عدا ذلك يكون هذا المستوى صفرًا.

وتدرك الملاحظة هنا أنه عندما يقوم بروتوكول ذو مستوى متقدم بتحديد مستوى الفقرة، فليس هناك حاجة لإجراء الخطوتين (ف٢) و (ف٣) أعلاه.

تحديد المستويات والاتجاهات الصريحة

يتم تحديد جميع مستويات التضمين الصريحة من خلال شفرات التضمين وإجبار الاتجاه (override)، وذلك بإجراء خطوات المستوى الصريح من (ص١) إلى (ص٩) المبينة لاحقًا. ويتم إجراء هذه الخطوات من خلال معالجة الخوارزمية للنص.

التضمين الصريح

إن شفرة التضمين الصريح تحدد مستوى النص، ولكنها لا تغير نوع اتجاه المحرف للمحارف التي تتأثر بهذه الشفرة. ويتم إجراء الخطوات الآتية لتحديد المستوى في هذه الحالة:

(ص١) يتم أولاً وضع مستوى التضمين الحالي مساوياً لمستوى تضمين الفقرة. كما يتم تحديد وضع إجبار الاتجاه بـ"الحيادي". بعد ذلك يتم معالجة المحارف الواحد تلو الآخر، وذلك بتطبيق الخطوات (ص٢) إلى (ص٩). وإن مستويات التضمين المسموح بها هي من (٠١) ولغاية (٦١) فقط وذلك للالتزام بالمواصفات الدولية في هذا المجال.

(ص٢) كلما تواجه الخوارزمية حرف التضمين من اليمين إلى اليسار (RLE)، يتم تحديد أقل مستوى تضمين فردي بحيث يكون أكبر من المستوى الحالي. وهناك احتمالان:

(أ) إذا كان هذا المستوى الجديد مسموح به، تكون شفرة التضمين هذه مسموح بها. عندئذ يتم حفظ مستوى التضمين الحالي ووضع الإجبار (override status). ويتم جعل المستوى الحالي يساوي هذا المستوى الجديد، كما يتم جعل وضع الإجبار هو الحيادي (neutral).

(ب) إذا كان هذا المستوى الجديد غير مسموح به، فإن شفرة التضمين هذه يكون غير مسموح بها. وفي هذه الحالة لا يتم تغيير المستوى الحالي أو وضع الإجبار.

وعلى سبيل المثال، فإن المستوى صفر يصبح المستوى ١؛ والمستويات ١، ٢ تصبح المستوى ٣؛ والمستويات ٣، ٤ تصبح المستوى ٥؛؛ والمستويات ٥٩، ٦٠ تصبح المستوى ٦١؛ والمستوى أعلى من ٦٠، لا يتم إجراء أي تغيير.

(ص٣) كلما تواجه الخوارزمية حرف التضمين من اليسار إلى اليمين (LRE)، يتم تحديد أقل مستوى تضمين زوجي بحيث يكون أكبر من المستوى الحالي. وهناك احتمالان:

(ا) إذا كان هذا المستوى الجديد مسروح به، تكون شفرة التضمين هذه مسروح بها. عندئذ يتم حفظ مستوى التضمين الحالي ووضع الإجبار (override status). ويتم جعل المستوى الحالي يساوي هذا المستوى الجديد، كما يتم جعل وضع الإجبار هو الحيادي (neutral).

(ب) إذا كان هذا المستوى الجديد غير مسروح به، فإن شفرة التضمين هذه يكون غير مسروح بها. وفي هذه الحالة لا يتم تغيير المستوى الحالي أو وضع الإجبار.

وعلى سبيل المثال، فإن المستويات ،، ١ تصبح المستوى ٢؛ المستويات ٢، ٣ تصبح المستوى ٤؛ المستويات ٤، ٥ تصبح المستوى ٦؛……؛ المستويات ٥٨، ٥٩ تصبح المستوى ٦٠؛ والمستوى أعلى من ٥٩، لا يتم إجراء أي تغيير.

الإجبار الصريح للاتجاه

إن الإجبار الصريح للاتجاه يحدد مستوى التضمين بنفس الطريقة التي يتم بها تحديد المستوى بواسطة الشفرات الصريحة للتضمين. علاوة على ذلك، فإن الإجبار الصريح للاتجاه يغير أيضاً نوع اتجاه المحارف التي تتأثر لتصبح في الاتجاه المحدد صراحة. ويتم إجراء الخطوات الآتية في هذه الحالة:

(ص ٤) كلما تواجه الخوارزمية محرف جعل الاتجاه من اليمين إلى اليسار (Right - to- Left Override- RLO)، يتم تحديد أقل مستوى تضمين فردي بحيث يكون أكبر من المستوى الحالي.

(ا) إذا كان هذا المستوى الجديد مسروح به، تكون شفرة التضمين هذه مسروح بها. عندئذ يتم حفظ مستوى التضمين الحالي ووضع الإجبار. ويتم جعل المستوى الحالي يساوي هذا المستوى الجديد، كما يتم وضع الإجبار من اليمين إلى اليسار.

(ب) إذا كان هذا المستوى الجديد غير مسروح به، فإن شفرة التضمين هذه غير مسروح بها. وفي هذه الحالة، لا يتم تغيير المستوى الحالي أو وضع الإجبار.

(ص ٥) كلما تواجه الخوارزمية محرف جعل الاتجاه من اليسار إلى اليمين (Left-to-Right Override- LRO)، يتم تحديد أقل مستوى تضمين زوجي بحيث يكون أكبر من المستوى الحالي.

(ا) إذا كان هذا المستوى الجديد مسروح به، تكون شفرة التضمين هذه مسروح بها. عندئذ يتم حفظ مستوى التضمين الحالي ووضع الإجبار. ويتم جعل المستوى الحالي يساوي هذا المستوى الجديد، كما يتم وضع الإجبار من اليسار إلى اليمين.

(ب) إذا كان هذا المستوى الجديد غير مسروح به، فإن شفرة التضمين هذه غير مسروح بها. وفي هذه الحالة، لا يتم تغيير المستوى الحالي أو وضع الإجبار.

(ص ٦) باستثناء المحارف من نوع RLE، RLO، LRE، LRO، PDF، فإن جميع الأنواع الأخرى يكون تأثيرها كما يلي:

(ا) يتم تحديد مستوى المحرف الحالي مساوياً لمستوى التضمين الحالي

(ب) كلما كان وضع الاتجاه الإجباري غير "حيادي"، يتم جعل نوع اتجاه المحرف الحالي مساوياً لوضع الاتجاه الإجباري.

وإذا كان وضع الاتجاه الإجباري "حيادي"، فإن المغارف تبقى أنواعها العادلة، حيث يبقى نوع المغارف العربية (AL)، ويبقى نوع المغارف اللاتينية (L)، ويبقى نوع المغارف الحيادية (N)، وهكذا. وإذا كان وضع الاتجاه الإجباري هو (R)، فإن نوع اتجاه المغارف يصبح (R)، وإذا كان وضع الاتجاه الإجباري هو (L)، فإن نوع اتجاه المغارف يصبح (L).

نهاء التضمين والإجبار

إن شفرة واحدة (محرف تحكم واحد) يمكن أن تنهي مجال تأثير مغارف الاتجاه الصريحة، سواء كانت مغارف تضمين النص صراحة أو مغارف إجبار اتجاه النص. وإن جميع الشفرات والحالات التي تم حفظها يتم إلغاؤها عند انتهاء الفقرات.

(ص ٧) عند ورود محرف PDF (استعادة حالة الاتجاه لما كانت عليه)، يتم تحديد المحرف المناظر للتضمين أو إجبار الاتجاه. وإن وجد محرف مناظر، يتم استعادة آخر مستوى تضمين أو إجبار الاتجاه من الذاكرة التي حفظ بها. وعادة ما يتم حفظ هذه المعلومات في ذاكرة بحيث يتم استخراج العنصر الأخير أولاً (LAST-IN-FIRST-OUT).

(ص ٨) جميع المغارف الصريحة للتضمين أو إجبار الاتجاه ينتهي مفعولها في نهاية كل فقرة. وإن فواصل الفقرات غير مشمولة في مغارف التضمين.

(ص ٩) يتم إزالة جميع مغارف LRE، RLE، RLO، LRO، PDF، BN . وتتجدر الملاحظة أنه عند كتابة الخوارزمية، لا يشترط بالضرورة إزالة هذه المغارف فعلاً، بل يجب أن تعمل الخوارزمية وكان هذه المغارف غير موجودة لحين انتهاء تنفيذ الخوارزمية.

(ص ١٠) إن باقي خطوات الخوارزمية والقواعد المصاحبة لها، يتم إجراؤها على كل مجموعة مغارف لها نفس المستوى. ولكل مجموعة، يتم تحديد نوع بداية المجموعة (start-of-level-run sor) ونوع نهاية المجموعة (end-of-level-run eor)، والذي يكون إما L أو R. وإن هذا يعتمد على المستوى الأعلى من المستويين على جنبي المجموعة (جنب بداية ونهاية الفقرة، ويكون مستوى المجموعة "الأخرى" هو مستوى التضمين الأساسي). وإن كان المستوى الأعلى فردياً، يكون النوع R؛ وإلا فإن النوع يكون L.

وعلى سبيل المثال:

المجموعات:	المستويات:
<٣>	٢ → ٢ ← ١ → ٠٠٠ ١١١

- المجموعة ١ يكون لها مستوى ٠، ويكون sor هو L، ويكون eor هو R.
- المجموعة ٢ يكون لها مستوى ١، ويكون sor هو R، ويكون eor هو L.
- المجموعة ٣ تكون لها مستوى ٢، ويكون sor هو L، ويكون eor هو L.

ولكل مجموعتين متجاورتين، يكون eor للمجموعة الأولى هو نفسه sor للمجموعة الثانية.

تحديد الأنواع الضعيفة

يتم تحديد الأنواع الضعيفة لكل مجموعة في مستوى على حده. وعلى حواضن المجموعة الواحدة، عندما يلزم معرفة نوع المحرف على حافة مجموعة أخرى، فإن النوع المخصص لـ eor أو sor يتم استعماله. ويتم تحديد نوع علامات عدم الفراغ (nonspacing marks - NSM) بناء على المحارف السابقة لها.

(ص ١) يتم فحص كل عالمة عدم الفراغ (NSM) في المجموعة، ويتم تغيير نوع NSM ليصبح مثل نوع المحرف الذي يسبقها. وإن كانت NSM في بداية المجموعة، ستأخذ نوع مثل sor.

وبافتراض أن sor في المثال التالي لها النوع R، نجد أن

AL	NSM	NSM	\rightarrow	AL	AL	AL
sor	ASM			\rightarrow sor	R	

بعد ذلك يتم البحث في النص عن الأرقام. وفي هذه الخطوة يتم تغيير أنواع اتجاهات فاصل الأعداد الأوروبي، إنهاء الأعداد الأوروبي، وفاصل نص الأعداد العام ليصبح نص الأعداد الأوروبي؛ نص الأعداد العربي، أو نص محابد آخر. وقد يكون النص الذي نبحث فيه قد تم تغيير الأنواع بواسطة محارف الغاء الاتجاه، وفي هذه الحالة لا تظهر أنواع الأعداد.

(ص ٢) كلما تجد الخوارزمية عدد أوروبي، يتم البحث إلى الخلف حتى يتم إيجاد أول نوع قوي من الأنواع (R، L، AL، أو sor). وإن وجد نوع AL، يتم تغيير نوع العدد الأوروبي إلى العدد العربي. وعلى سبيل المثال:

AL	EN		\rightarrow	AL	AN	
AL	N	EN	\rightarrow	AL	N	AN
sor	N	EN	\rightarrow	sor	N	EN
L	N	EN	\rightarrow	L	N	EN
R	N	EN	\rightarrow	R	N	EN

(ص ٣) يتم تغيير جميع أنواع AL إلى نوع R

(ص ٤) إن وجد فاصل أوروبي وحيد بين عددين أوروبيين، يتم تغيير نوع الفاصل إلى عدد. وإن وجد فاصل عام وحيد بين عددين من نفس النوع، يتم تغيير نوع الفاصل ليصبح مثل نوع العددين. وعلى سبيل المثال:

EN	ES	EN	\rightarrow	EN	EN	EN
EN	CS	EN	\rightarrow	EN	EN	EN
AN	CS	AN	\rightarrow	AN	AN	AN

(ض٥) إن سلسلة الإنهاء الأوروبي التي تكون بجانب الأعداد الأوروبية يتم تغييرها إلى الأعداد الأوروبية. وعلى سبيل المثال:

ET	ET	EN	→ EN	EN	EN
EN	ET	ET	→ EN	EN	EN
AN	ET	EN	→ AN	EN	EN

(ض٦) فيما عدا ذلك، يتم تغيير محارف الفوائل والإنهاء لتصبح محارف حيادية أخرى (Other Neutral).

AN	ET		→ AN	ON	
L	ES	EN	→ L	ON	EN
EN	CS	AN	→ EN	ON	AN
ET	AN		→ ON	AN	

(ض٧) كلما تجد عدد أوروبي، يتم البحث إلى الخلف حتى يتم إيجاد أول نوع قوي من الأنواع (R، L، أو sor). وإن وجد نوع L، يتم تغيير نوع العدد الأوروبي ليصبح L.

L	N	EN	→ L	N	L
R	N	EN	→ R	N	EN

تحديد الأنواع الحيادية

يتم تحديد الأنواع الحيادية لكل مجموعة في مستوى على حدة. وعلى حواط المجموعة الواحدة، عندما يلزم معرفة نوع المحارف على حافة مجموعة أخرى، فإن النوع المخصص لـ eor أو sor يتم استعماله. وإن هذه المرحلة تحدد اتجاه المحارف الحيادية. وإن نتيجة هذه المرحلة أن جميع المحارف الحيادية تصبح إما R أو L. وبشكل عام، فإن المحارف الحيادية تأخذ اتجاه النص المحيط بها. وفي حال وجود تضارب، فإنها تأخذ اتجاه التضمين.

(ح١) إن سلسلة المحارف الحيادية تأخذ اتجاه النص القوي المحيط بها إذا كان هذا النص المحيط بالسلسلة من الجانبين له نفس الاتجاه. ويتم معاملة الأعداد الأوروبية والعربية كأنها R. ويتم استخدام sor و eor على جانبي المجموعة الواحدة. وعلى سبيل المثال:

R	N	R	→ R	R	R
L	N	L	→ L	L	L
R	N	AN	→ R	R	AN
AN	N	R	→ AN	R	R
R	N	EN	→ R	R	EN
EN	N	R	→ EN	R	R
EN	N	AN	→ EN	R	AN

وتجرد الملاحظة هنا أن الأعداد الأوروبيية التي تأتي بعد L يكون قد تم تغييرها إلى L قبل الوصول إلى هذه الخطوة ، وبالتالي فإن هذه الخطوة تعامل الأعداد الأوروبيية بعد R وكأنها R.

(ح) المحارف الحيادية الأخرى تأخذ نفس اتجاه التضمين (embedding direction). لفرض في المثال الآتي أن eor هي L، sor هي R: $e \rightarrow N$.

L	N	eor	\rightarrow L	L	eor
R	N	eor	\rightarrow R	e	eor
sor	N	L	\rightarrow sor	e	L
sor	N	R	\rightarrow sor	R	R

مثال: إن قائمة أعداد مفصولة بمحارف حيادية ومتضمنة في مجموعة ما ستظهر في ترتيب المجموعة.

العرض على الشاشة: إن القيم هي ١٢٣، ٤٥٦، ٧٨٩، واضح "he said"

التخزين في الذاكرة: " ح ض ١ او ٤٥٦، ٧٨٩ ي هـ م ي ق ل ا ن ! " he said

وفي هذه الحالة، فإن الفاصلة ومحرف الفراغ بين الأعداد تأخذ نفس اتجاه النص المحيط بها (من اليمين إلى اليسار)، متجاهلة الأعداد. وإن الفواصل لا تعتبر جزءا من العدد لأنها ليست محاطة بأرقام من كلا الجانبين.

وإن وجدت سلسلة محارف من اليسار إلى اليمين جنب العدد، فإن الأعداد الأوروبيية ستأخذ ذلك الاتجاه:

(أ) العرض على الشاشة: " إنها سيارة ٥٢٠ BMW، واضح." he said
 (ب) التخزين في الذاكرة: ". ح ض ١ او ٤٥٦، ٧٨٩ ي هـ ن ! " he said

تحديد المستويات الضمنية

في هذه المرحلة الأخيرة، يمكن زيادة مستوى تضمين النص، بناء على نوع المحرف. وإن النص ذو الاتجاه من اليمين إلى اليسار يجب أن يكون مستوى فرديا، بينما النص من اليسار إلى اليمين والنص العددي يجب أن يكون مستوى زوجيا. علاوة على ذلك، فإن النص العددي يجب أن يكون مستوى أكبر من مستوى الفقرة.

(ن ١) جميع المحارف التي لها مستوى تضمين زوجي (من اليسار إلى اليمين)، فإن المحارف التي نوعها R يزداد مستواها بمقدار ١، وإن المحارف التي نوعها AN أو EN يزداد مستواها بمقدار ٢.

(ن ٢) جميع المحارف التي لها مستوى تضمين فردي (من اليمين إلى اليسار)، فإن المحارف التي نوعها L، EN، أو AN يزداد مستواها بمقدار ١.

ويلخص الجدول رقم (٣) نتائج تحديد المستويات الضمنية.

الجدول رقم (٣): تحديد المستويات الضمنية

مستوى التضمين (Embedding Level)		النوع
فردي	زوجي	
EL+1	EL	L
EL	EL+1	R
EL+1	EL+2	AN
EL+1	EL+2	EN

ترتيب المستويات المحددة

تصف الخوارزمية التالية العملية المنطقية لتحديد الترتيب الصحيح لعرض النص، وتتضمن الخطوات

الآتية:

- (أ) يتم تحديد مستويات النص وفق خوارزمية ثنائية الاتجاه؛
- (ب) تم تحديد أشكال عرض المحارف حسب السياق (مع الأخذ بعين الاعتبار مستويات التضمين لعكس اتجاه المحارف)؛
- (ج) يتم احتساب عرض المحارف (بالترتيب المنطقي لها) لتحديد نهايات الأسطر؛
- (د) لكل سطر، يتم تطبيق القواعد (١م) إلى (٤م) لترتيب المحارف في ذلك السطر؛
- (ه) يتم عرض الأشكال المناظرة للمحارف في السطر وفق ترتيبها.

(١م) في كل سطر، يتم جعل مستوى التضمين للمحارف التالية مثل مستوى التضمين للفقرة:

- (أ) فواصل الأجزاء،
- (ب) فواصل الفقرات
- (ج) أي سلسلة محارف من نوع الفراغ الأبيض (whitespace) التي تسبق فاصل الجزء أو فاصل الفقرة،
- (د) أي سلسلة محارف من نوع الفراغ الأبيض في نهاية السطر.

وتجرد ملاحظة أن أنواع المحارف المستخدمة هنا هي الأنواع الأصلية، وليس الأنواع التي تم تعديلها في المرحلة السابقة. وحيث أن فاصل الفقرة ينهي السطر، فإنه يمكن وجود فاصل واحد فقط في السطر، ويكون هذا في نهاية السطر.

(٢م) ابتداء من أعلى مستوى في النص إلى أقل مستوى فردي في كل سطر، يتم عكس اتجاه أي سلسلة متصلة من المحارف يساوي مستواها أو يزيد على أقل مستوى فردي في السطر.

ويمكن لهذه العملية أن تعكس اتجاه عدد من السلال الفرعية. وتوضح الأمثلة الأربع الآتية هذه القاعدة.

car	means	. ة را .	التخزين في الذاكرة:
000 0	0 0 0 0	0 1 1 1 1 1	المستويات :
	0	1	
car	means	. سيا رة .	عكس اتجاه المستوى ١

car	ي ن ع ت	. ة ر ا ي س	التخزين في الذاكرة:
٢٢٢	١ ١١١	١ ١١١١	المستويات :
rac	ي ن ع ت	. ة ر ا ي س	عكس اتجاه المستوى ١
سيارة.	تعني	car	عكس اتجاه المستويات ٢،١

(م٣) عندما يتم وضع علامات (مثل علامات التشكيل) على محارف اللغة العربية (أو أي محارف أخرى يكُون اتجاهها الأساسي من اليمين إلى اليسار)، فإن هذه العلامات في هذه الخطوة تسبق المحارف المصاحبة لها. وإن كانت ببرامج العرض تتوقع أن تظهر هذه العلامات بعد المحارف المصاحبة لها عند العرض بالشكل النهائي، فإن ترتيب هذه العلامات والمحارف المصاحبة لها يجب أن يعكس.

(م٤) المحرف الذي له خاصية المرأة (mirrored property)، يجب عرض شكل المرأة الخاص به إذا كان اتجاه المحرف هو R. وعلى سبيل المثال فإن محرف قوس الفتح باللغة اللاتينية (" ")، عندما يكون المستوى الخاص به زوجي، يجب عرضه بهذا الشكل ("")، ويتم عرض صورة المرأة له (" ") عندما يكون المستوى الخاص به فردي.

استخدام خوارزمية الترئية الأساسية

بناء على وصف الخوارزمية كما ورد أعلاه وتحديد النوع الضمني للمحارف وطرق تحديد أنواع واتجاهات المحارف الحيادية والعديدية، فإن ترتيب المحارف الضمني يعطي عادة الترتيب الصحيح للنصوص دون أي جهد إضافي. إلا أنه يوجد حالات خاصة، مثلاً عندما تبدأ الفقرة ذات الاتجاه من - اليمين - إلى - اليسار بمحارف ذات الخاصية من - اليمين - إلى - اليمين، أو عندما يتم تداخل النصوص باتجاهين، أو عند ورود محرف من النوع الضعيف في حافة سلسلة لها اتجاه من - اليمين - إلى - اليسار. وفي هذه الحالات، يلزم وضع محارف إضافية لتحديد الاتجاه للحصول على الترتيب الصحيح. وإن إدخال أرقام الأجزاء في المستودعات قد يلزمهها محارف إجبار الاتجاه. وإن أكثر حالة قد تسبب مشاكل تنشأ عند وجود محرف محيد على حافة نص متضمن في نص آخر. ويمكن معالجة هذا الوضع بتحديد مستوى النص المتضمن. وعلى سبيل المثال، إذا كان النص الآتي له مستوى صفر، تكون النتيجة كما يلي:

- التخزين في الذكرة: . " ! ام ين طع أ " he said and expired.
- العرض على الشاشة: . " ! أعطني ماء " he said and expired.

وإذا أردنا جعل علامة التعجب جزءاً من النص باللغة العربية، فإن المستخدم يمكنه وضع علامة لجعل النص "أعطني ماء !" متضمناً باللغة العربية داخل النص باللغة الإنجليزية، ويتم تخزين النص كالتالي:

he said "<RLE >! ام ين طع أ >" , and expired.

كما توجد طريقة أسهل ألا وهي وضع محرف تحديد الاتجاه من - اليمين - إلى - اليسار بعد علامة التعجب، ويتم تخزين النص كالتالي:

he said "<RLM >! ام ين طع أ " , and expired.

وتكون نتائج العرض على الشاشة في الحالتين الأخيرتين كما يلي:

he said " ! أعطني ماء " , and expired.

المواصفات ذات العلاقة

إن المواصفات التي تتعلق بتشغير المحارف العربية هي:

- (أ) المواصفة العربية 449 ، وهي الأساس لجمع الموصفات اللاحقة، إذ يتم تمثيل كل حرف بسبعة عناصر ثنائية؛
- (ب) مواصفة إكما رقم 114 ، وهي مبنية على المعاصفة العربية 449؛
- (ج) المعاصفة العربية 708 ، والتي تحتوي على المحارف العربية واللاتينية، حيث يتم تمثيل كل حرف بثمانية عناصر ثنائية؛
- (د) المعاصفة الدولية إيزو 8859/6 ، وهي مبنية على المعاصفة العربية رقم 708؛
- (ه) المعاصفة الدولية إيزو 10646 التي تحتوي على جميع المحارف العالمية، ويتم تمثيل كل حرف بعد (32) عنصر ثالثي؛
- (و) المعاصفة الدولية يوني코드، والتي تشبه مواصفة إيزو 10646.

وإن المحارف العربية الموجودة في هذه الورقة هي نفسها تلك الموجودة في الموصفات الدوليتين إيزو 10646 ويوني코드.

مبادئ التشفير

إن مجموعة المحارف العربية المشفرة محددة تماماً، وهي التي وردت في المعاصفة العربية 449. وتخصص شفرة واحدة لكل حرف عربي مهما تعددت أشكاله المرئية والمطبوعة. وت تكون الكلمة من مجموعة من هذه الحروف. وإن الشكل الذي يكتب في المعاصفة هو شكل الحرف العربي عندما يظهر منفرداً (أي غير متصل) في الكلمة. وتم اختيار هذا الشكل لسهولة إظهار الحرف، ولا يؤثر ذلك على كيفية إظهاره في النصوص المرئية والمطبوعة.

علامات الترقيم

إن علامات الترقيم المستخدمة باللغة العربية واللغة اللاتينية لا تعطى شفرات مختلفة. أما علامات الترقيم التي لها شكل مختلف كلها عندما تستخدم في النصوص العربية، فلا بد من إعطائها شفرات مختلفة، وهذه العلامات هي:

- (أ) الفاصلة العربية (،)
- (ب) الفاصلة المنقوطة (؛)
- (ج) علامة الاستفهام (؟)
- (د) علامة النسبة المئوية (%)
- (ه) علامة الضرب (×)
- (و) علامة القسمة (÷)
- (ز) وغيرها.

أما بالنسبة لعلامات الترقيم المزدوجة (مثل الأقواس)، فتعطى نفس الشفرة لقوس الفتح بغض النظر عن اللغة، وكذلك تعطى نفس الشفرة لقوس الإغلاق، ويتم اختيار الشكل المناسب حسب اتجاه الكتابة (من اليمين إلى اليسار أو العكس).

شفرة الفاصل والواصل

يتم تخصيص شفتين ليتمكن المستخدم من فصل الحروف عن بعضها أو وصلها:

- (١) محرف الفاصل بعرض صفر (Zero Width Non-Joiner – ZWNJ)
- (٢) محرف الواصل بعرض صفر (Zero Width Joiner – ZWJ)

وإن استخدام محرف الفاصل بين حرفين يمنع توصيلهما مع بعض عند ترئية النص. وللوضيح استخدام هذين المحرفين، نعطي الأمثلة الآتية المبينة في الأشكال ٤-٢.

الشكل رقم (٢) – استخدام محرف الواصل

التخزين في الذاكرة (الاتجاه من اليسار إلى اليمين) ـ ZWJ ـ ـ ـ	ـ ـ ـ ـ ZWNJ ـ ـ ـ
ـ ـ ـ ـ ـ ZWNJ ـ ـ ـ	ـ ـ ـ ـ ـ ـ تغيير الاتجاه (يصبح من اليمين إلى اليسار)
ـ ـ ـ ـ ـ ـ ـ ـ وصل الحروف مع بعضها	ـ ـ ـ ـ ـ ـ ـ ـ

الشكل رقم (٣) – استخدام محرف الفاصل

التخزين في الذاكرة (الاتجاه من اليسار إلى اليمين) ـ ZWNJ ـ ـ ـ	ـ ـ ـ ـ ZWNJ ـ ـ ـ
ـ ـ ـ ـ ـ ZWNJ ـ ـ ـ	ـ ـ ـ ـ ـ ـ تغيير الاتجاه (يصبح من اليمين إلى اليسار)
ـ ـ ـ ـ ـ ـ ـ ـ وصل الحروف مع بعضها	ـ ـ ـ ـ ـ ـ ـ ـ

الشكل رقم (٤) – استخدام محرفي الواصل والفاصل

التخزين في الذاكرة (الاتجاه من اليسار إلى اليمين) ـ ZWNJ ـ ـ ـ	ـ ـ ـ ـ ZWNJ ـ ـ ـ
ـ ـ ـ ـ ـ ZWNJ ـ ـ ـ	ـ ـ ـ ـ ـ ـ تغيير الاتجاه (يصبح من اليمين إلى اليسار)
ـ ـ ـ ـ ـ ـ ـ ـ وصل الحروف مع بعضها	ـ ـ ـ ـ ـ ـ ـ ـ

حركات التشكيل

تعتبر الحركات مهمة في اللغة العربية وتوضع فوق الحروف الساكنة لتحريكها، وكثيراً ما يكون نفس الكلمة أكثر من معنى بتغيير حركات التشكيل فوق واحد أو أكثر من حروفها. وإن ورود محرف من حركات التشكيل وإظهاره جنب دائرة منقطة في هذه الورقة يعني بأن هذه الحركة سيتم وضعها (وفق عملية معينة) مصاحبة لحرف الذي يسبقها مباشرة في سلسلة المحارف في النص المعنى.

الأرقام العربية والأوروبية

إن أسماء الأشكال المختلفة للأرقام العشرية يختلف كثيراً بين لغات العالم. وإن نظام العد العشري قد تم تبنيه في الدول العربية ويستخدم الأرقام (٠، ١، ٢، ٣، ... الخ) والتي أصبحت فيما بعد تسمى بالأرقام العربية. ثم تبنت أوروبا الأرقام العشرية من العالم العربي، وتم تعديل الأرقام بشكل جذري حيث أصبحت الأرقام هي ٠، ١، ٢، ٣، ... الخ، والتي سميت بالأرقام الأوروبية.

وستستخدم الأرقام العربية في الوقت الحاضر في معظم الدول العربية في الشرق الأوسط، باستثناء دول شمال إفريقيا التي تستخدم الأرقام الأوروبية. وعلى كل الأحوال فإن معنى ومدلول الأرقام العشرية هو نفسه في جميع الدول العربية، وأن الشكل فقط هو الذي يختلف. وتحتوي هذه الورقة على الأرقام بشكليها العربي والأوروبي.

الحروف الإضافية للغات الأخرى

تستخدم الحروف العربية لكتابية عدد من اللغات الأخرى، ومنها اللغة الفارسية ولغة الأوردو. وحتى تكون هذه الورقة متناسبة مع الموصفات العالمية في هذا المجال، يحتوي جدول الشفرة في هذه الورقة على الحروف العربية وبعض الحروف الأخرى التي تستخدم في عدد من اللغات العالمية التي تستخدم أحرف اللغة العربية بشكل أساسي.

علامات ترقيم القرآن الكريم

تستخدم بعض المحارف في القرآن الكريم لعدد من الأغراض منها تسهيل لفظ الكلمات والدلالة على أماكن الوقف أثناء التلاوة وغير ذلك. وتحتوي جداول الشفرة في هذه الورقة على علامات الترقيم المستخدمة في القرآن الكريم.

الحد الأدنى لمتطلبات عرض النصوص

إن طبيعة اللغة العربية واتصال الحروف مع بعضها البعض في الكلمة الواحدة يضع بعض المتطلبات لعرض (ترئية أو طباعة) النصوص. إن هذا يتطلب إجراء عمليات معينة للتحويل من الترتيب المنطقي للمحارف في الذاكرة (أو وسائل التخزين الأخرى) إلى الترتيب الطبيعي الملائم للعرض على شاشة الحاسوب أو للطباعة. ومن جهة أخرى، فإنه من الضروري أن تخذل عملية التحويل هذه شكل الحرف الملائم للعرض وذلك حسب موقع الحرف في الكلمة والحرف الذي يسبقه والحرف الذي يليه. ونبين فيما يلي الحد الأدنى من القواعد التي تضمن أن يكون للنص نفس الشكل تماماً عند مرسل المعلومات والمستقبل لهذه المعلومات عبر شبكة الإنترنت.

اتصال الحروف في الكلمات

يتم عرض (أو ترئية) كل حرف في اللغة العربية بوحد من عدة أشكال محتملة لهذا الحرف. ويتم تحديد الشكل المناسب حسب مجموعة الاتصال لهذا الحرف وكذلك مجموعات الاتصال للحروف المجاورة له. وينتمي كل حرف في اللغة العربية إلى واحدة فقط من المجموعات المبينة في الجدول رقم (٤).

ويمكن إعادة تقسيم المجموعات الواردة في الجدول رقم (٤) إلى مجموعات شاملة كما هو مبين في الجدول رقم (٥)، بحيث يتم استخدام هذه المجموعات في قواعد الاتصال للحروف التي سيتم ذكرها لاحقاً.

قواعد الاتصال

هناك عدة قواعد لوصف إمكانية اتصال الحروف العربية مع بعضها وأشكال الحروف عند ترئيتها (عرضها) في النصوص المطبوعة والمرئية. إن موقع الحروف في الأمثلة المبينة أدناه هي الواقع التي سيظهر بها الحرف بعد معالجة خوارزمية الاتجاه للنص المعنى، وظهوره هذا النص على الشاشة. وعند تنفيذ هذه القواعد في أنظمة معالجة البيانات وتراسل المعطيات، يستحسن القيد بهذه القواعد وتنفيذها في الترتيب

المنطقى المبين أدناه وذلك قبل تتنفيذ خوارزمية الاتجاه الثنائى الازمة لمعالجة اللغة العربية (أنظر المواصفة الدولية (UNICODE).

ويبين الجدول رقم (٦) الاحتمالات الأربع لشكل حرف ما حسب اتصاله مع الحروف الأخرى.
وعندما يرمز X إلى الحرف، فإن الأشكال المختلفة لهذا الحرف تكون كما هو موضح في الجدول رقم (٦).

الجدول رقم (٤) : المجموعات الأساسية للحروف حسب اتصالها في الكلمة

مجموعة الاتصال	رمز المجموعة	أعضاء المجموعة
الاتصال من اليمين Right-Joining	R	الألف ، والدال ، والذال ، والراء ، والزين
الاتصال من اليسار Left-Joining	L	لا يوجد
الاتصال من الجهتين Dual-Joining	D	باء ، والتاء ، والثاء ، والجيم ، ...
مبسب الاتصال Join-Causing	C	حرف الاتصال ذو عرض صفر (Zero Width Joiner) ، وحرف التطويل. وتميز هذه المحارف من محارف الاتصال من الجهتين بأنها لا يتغير شكلها.
غير متصلة Non-Joining	U	حرف عدم الاتصال ذو عرض صفر (Zero Width Non-Joiner) وجميع محارف الفسحة (الفراغ) ، والهمزة ، والأرقام ، وعلامات التقىط ، والحروف غير العربية ، وغير ذلك.
الشفافة Transparent	T	جميع علامات التشكيل ، وعلامات التوضيب ، بما في ذلك الفتحة ، والضمة ، والكسرة ، والسكون ، والضمة ، والكسرتان ، والفتحتان ، والشدة ، والألف فوق الكلمة ، وعلامة من اليمين إلى اليسار ، وغير ذلك.

الجدول رقم (٥) : المجموعات الشاملة للحروف حسب اتصالها في الكلمة

مجموعة الاتصال	رمز المجموعة	المجموعات الشاملة لمجموعات الاتصال من الجهتين
مبسب الاتصال من اليمين Right Join-Causing		المجموعة الشاملة لمجموعات الاتصال من اليمين والاتصال من اليسار ، ومبسب الاتصال.
مبسب الاتصال من اليسار Left Join-Causing		المجموعة الشاملة لمجموعات الاتصال من الجهتين والاتصال من اليمين ، ومبسب الاتصال.

الجدول رقم (٦) : أشكال الحروف العربية عند الترئية

الشكل	الرمز	الوصف
حرف منفصل (Nominal)	Xn	يظهر الحرف بشكله المنفصل كما هو مبين في جدول الشفرة
حرف متصل من اليمين (Right)	Xr	يظهر الحرف متصلة من اليمين ، ويمكن استخدام هذا الشكل لمجموعات الحروف التي تتصل من اليمين أو من الجهتين
حرف متصل من اليسار (Left)	Xl	يظهر الحرف متصلة من اليسار ، ويمكن استخدام هذا الشكل لمجموعات الحروف التي تتصل من اليسار أو من الجهتين
حرف متصل من الجهتين (Medial)	Xm	يظهر الحرف متصلة من الجهة اليمنى والجهة اليسرى ، ويمكن استخدام هذا الشكل لمجموعة الحروف التي تتصل من الجهتين فقط لا غير .

أما قواعد اتصال الحروف فهي مبنية أدناه، مع ملاحظة أن كلمة يمين ويسار تعني أيضا قبل وبعد على التوالي، حيث أن اتجاه اللغة العربية من اليمين إلى اليسار.

القاعدة رقم (١):

الحروف الشفافة لا تؤثر على خصائص اتصال الحروف الرئيسية. وعلى سبيل المثال:

$$\begin{array}{c} - \text{ميم (N)} + \text{شدة (N)} + \text{لام (N)} \leftarrow \text{ميم (R)} + \text{شدة (N)} + \text{لام (L)} \\ - \text{م} + \text{ل} \leftarrow \text{م} + \text{ل} \leftarrow \text{لم} \end{array}$$

وتجر الملاحظة هنا أن الحروف الإنجليزية بين قوسين تحدد شكل الحرف في جدول الشفرة أولاً، ومن ثم شكل الحرف بعد تطبيق القاعدة وخوارزمية إظهار الحروف. علاوة على ذلك فإن ترتيب الحروف أعلاه هو حسب ترتيبها في الذاكرة عندما تكون الحروف منفصلة (أي أن ترتيبها من اليسار إلى اليمين)، وعند إظهارها في كلمة واحدة يعكس اتجاه الحروف لتصبح من اليمين إلى اليسار.

القاعدة رقم (٢):

الحرف الذي له قابلية الاتصال من اليمين عندما يكون على يمينه محرف آخر من مجموعة مسبب الاتصال من اليمين، سيأخذ شكل الاتصال من اليمين (Xr) .

وعلى سبيل المثال:

$$\begin{array}{c} - \text{ألف (N)} + \text{محرف التطويل (N)} \leftarrow \text{ألف (R)} + \text{محرف التطويل (N)} \\ - \text{ا} + \text{ـ} \leftarrow \text{ـ} \leftarrow \text{ـ} \end{array}$$

القاعدة رقم (٣):

الحرف الذي قابلية الاتصال من اليسار عندما يكون على يساره محرف آخر من مجموعة مسبب الاتصال من اليسار ، سيأخذ شكل الاتصال من اليسار (Xl). .

القاعدة رقم (٤):

الحرف الذي له قابلية الاتصال من الجهتين عندما يكون على يمينه محرف آخر من مجموعة مسبب الاتصال من اليمين ويكون على يساره محرف آخر من مجموعة مسبب الاتصال من اليسار ، سيأخذ شكل الاتصال من الجهتين (Xm) . وعلى سبيل المثال:

$$\begin{array}{c} - \text{محرف التطويل (N)} + \text{ميم (N)} + \text{محرف التطويل (N)} \leftarrow \text{محرف التطويل (N)} + \text{ميم (M)} \\ - \text{ـ} + \text{ـ} \leftarrow \text{ـ} + \text{ـ} \leftarrow \text{ـ} \end{array}$$

القاعدة رقم (٥) :

الحرف الذي له قابلية الاتصال من الجهتين عندما يكون على يمينه حرف آخر من مجموعة مسبب الاتصال من اليمين، ولا يكون على يساره حرف آخر من مجموعة مسبب الاتصال من اليسار، سيأخذ شكل الاتصال من اليمين (Xr). وعلى سبيل المثال:

$$- \text{ ميم (N)} + \text{ حرف التطويل (N)} \leftarrow \text{ ميم (R)} + \text{ حرف التطويل (N)} \\ - \text{ م} + - \leftarrow \text{ م} + - = \text{ م}$$

القاعدة رقم (٦) :

الحرف الذي له قابلية الاتصال من الجهتين عندما يكون على يساره حرف من مجموعة سبب الاتصال من اليسار ولا يكون على يمينه حرف آخر من مجموعة مسبب الاتصال من اليمين، سيأخذ شكل الاتصال من اليسار (I). وعلى سبيل المثال:

$$- \text{ حرف التطويل (N)} + \text{ ميم (N)} \leftarrow \text{ حرف التطويل (N)} + \text{ ميم (L)} \\ - \text{ م} + - \leftarrow - \text{ م} + - = - \text{ م}$$

القاعدة رقم (٧) :

إذا لم تتطبق أي من القواعد الواردة أعلاه على حرف ما ، فسيأخذ الشكل المنفصل (Xn)

الحروف المجتمعة (أو المرتبطة)

هناك نوعان من الحروف المجتمعة (Ligatures) : الأول إجباري بغض النظر عن نوع الكتابة، والثاني اختياري حسب أنواع الكتابة المستخدمة. وتحدد هذه المواصفة خصائص النوع الأول فقط و القواعد الخاصة بكتابته، ولا تتعرض لنوع الثاني الذي هو اختياري .

وعند وصف هذا النوع من الحروف المجتمعة، فإن ما ينطبق على حرف الألف ينطبق أيضا على الحروف الآتية المدة فوق الألف، والهمزة فوق الألف، والهمزة تحت الألف، الخ...

في القواعد الآتية، عندما تشير X، Y إلى محارف ، فإن أشكال الحروف المجتمعة التي تضم هذين المحرفين تكون كما هو مبين في الجدول رقم (٧) أدناه.

الجدول رقم (٧) : مصطلحات كتابة الحروف المجتمعة

الرمز	الوصف
(X.Y)n	الحروف المجتمعة بشكلها الاعتيادي المنفرد nominal التي تمثل مجموعة Xr مع Yl .
(X.Y)r	الحروف المجتمعة المتصلة من اليمين (Right-Joining) التي تمثل مجموعة Xr مع Ym .
I	الحروف المجتمعة المتصلة من اليسار (Left-Joining) التي تمثل مجموعة Xm مع Xl
(X.Y)m	الحروف المجتمعة المتصلة من الجهتين (Dual- or Medial- Joining) التي تمثل مجموعة Xm مع Ym

القاعدة رقم (١): المحارف الشفافة لا تؤثر على طريقة اتصال المحارف الرئيسية. وعلى سبيل المثال: **ألف (R) + فتحة (N) + لام (L) ← لام-ألف (N) + فتحة (N)**

القاعدة رقم (٢): أي سلسلة بها ألف (R) على اليسار ولام (M) على اليمين ستتشكل الحروف المجتمعية لام-ألف (L). وعلى سبيل المثال: $\underline{ا} + \underline{م} \rightarrow \underline{ل}$ (وليس $\underline{ا}$)

القاعدة رقم (٣): أي سلسلة بها ألف (R) على اليسار ولام (L) على اليمين ستتشكل الحروف المجتمعية لام-ألف (N). وعلى سبيل المثال : لـ + لـ ← لا (وليس لـ)

أما فيما يتعلق بالحروف المجتمعية الاختيارية، فهناك عدد كبير منها و يمكن استخدامه في التطبيقات المختلفة. وتخصص هذه الموصفة شفرات لبعض هذه المجموعات. وعلى جميع الأحوال، لا يجوز استخدام شفرات هذه المجموعات في تبادل المعلومات بشكل عام .ومن جهة أخرى ، ليس من الضروري أن تتيح طرق الكتابة المختلفة طباعة أو ترئية جميع هذه المجموعات من الحروف المجتمعية، كما أن هذه المجموعات لا تمثل جميع ما تتيحه طرق الكتابة المختلفة.

إن أنظمة العرض والترئية تستخدم محارف إضافية لتحديد الشكل ووضع الحرف في مكانه المناسب. وعلى سبيل المثال، فإن وضع حركات التشكيل في مكانها المناسب بواسطة الخوارزمية المختصة بذلك سيعطي مظهراً أفضل للكلمة. كما أن تتميق اللغة العربية يتم إجراؤه عادةً بزيادة طول الكلمة (بوضع محارف التطويل) وليس بزيادة المسافات بين الكلمات. وإن المكان المناسب والمحارف التطويل داخل الكلمة يعتمد على نوع الكتابة وبرامج الترئية المستخدمة. أخيراً، فإن بعض الأنظمة المتقدمة للترئية تخذل أشكالاً مختلفة للحروف (مثل أشكال حرف الكاف، لزيادة طول الكلمة تمهدًا لتمييق النص وجعل جميع الأسطر لها نفس الطول).

أنواع اتصال الحروف العربية

تبين الجداول أرقام ٨، ٩، ١٠ المحارف العربية والقابلة للاتصال إما من اليمين أو من الجهتين (اليمين واليسار). وجميع المحارف العربية الأخرى (باستثناء محرف النطويل) غير قابلة للاتصال إطلاقاً. وإن معظم المحارف العربية الإضافية ليست إلا أشكالاً مختلفة للمحارف العربية الأساسية.

الجدول رقم (٨): الحروف العربية القابلة للاتصال من الجهتين

الجدول رقم (٨) - (تابع)

المجموعة	Xn	Xr	Xm	Xl	الصاد	ص ص ص	المحارف الأخرى التي لها أشكال شبيهة
الطاء	ط	ط	ط	ط	ـ	ـ	جميع الحروف التي يشبه شكلها الطاء مثل الظاء
العين	ع	ع	ـ	ـ	ـ	ـ	جميع الحروف التي يشبه شكلها العين مثل الغين
الفاء	ف	ف	ـ	ـ	ـ	ـ	جميع الحروف التي تعتمد على شكل الفاء
القاف	ق	ق	ـ	ـ	ـ	ـ	جميع الحروف التي تعتمد على شكل القاف
الميم	م	ـ	ـ	ـ	ـ	ـ	جميع الحروف التي تعتمد على شكل الميم
الهاء	ه	ـ	ـ	ـ	ـ	ـ	جميع الحروف التي تعتمد على شكل الهاء
الكاف	ك	ـ	ـ	ـ	ـ	ـ	جميع الحروف التي تعتمد على شكل الكاف
اللام	ل	ـ	ـ	ـ	ـ	ـ	جميع الحروف التي تعتمد على شكل اللام

تظهر بعض المحارف في نهاية الكلمات فقط ولا تظهر في منتصف الكلمات. ومن هذه المحارف التاء المربوطة، والألف المقصورة، والضمنان. وعندما تكون هذه المحارف متصلة (مثل التاء المربوطة)، فإنها تصنف بأنها قابلة للاتصال من اليمين. وإن المحارف غير القابلة للاتصال، أو لا تسبب الإتصال (مثل الضمنان)، فإنها تصنف محارف شفافة. إن المحارف المبينة في الجداول (٨، ٩، ١٠) قد قسمت إلى مجموعات حسب شكل المحارف وليس حسب الترتيب الأبجائي.

الجدول رقم (٩): الحروف العربية القابلة للاتصال من اليمين

المجموعة	Xn	Xr	المحارف الأخرى التي لها أشكال شبيهة
الألف	ـ	ـ	جميع الحروف التي تعتمد على شكل الألف
الواو	ـ	ـ	جميع الحروف التي تعتمد على شكل الواو
ال DAL	ـ	ـ	جميع الحروف التي يشبه شكلها الدال، مثل الذال
الراء	ـ	ـ	جميع الحروف التي يشبه شكلها الراء، مثل الزين
التاء المربوطة	ـ	ـ	جميع الحروف التي تعتمد على شكل التاء المربوطة

الجدول رقم (١٠): أنواع الاتصال الأخرى للحروف العربية

الصنف	المحارف في هذا الصنف
مسبب الاتصال	حرف الاتصال ذو طول صفر (Zero-Width Joiner-200D) ومحرف التطويل (0640)
مسبب الانفصال	حرف الانفصال ذو طول صفر (Zero-Width Non Joiner-200C) ، وجميع محارف الفسحة (الفراغ) غير التي ذكرت في الجدولين رقم ٥، ٦، والهمزة (0621) ، والأرقام، وعلامات التقىط، والحرروف غير العربية، الخ
الشفافة	جميع علامات التشكيل، بما في ذلك الفتحتان (064B) وغيرها، والمدة من فوق (065) والهمزة من فوق (0654)، والهمزة من أسفل (0655)، والألف من فوق (0670)، والعلامات القرائية، وعلامة من اليمين إلى اليسار (200F)، وغير ذلك.

جدول الشفرة للمحارف الأساسية

تبين الموصفة الدولية إيزو 10646 ومواصفة يونيكود المحارف العربية الأساسية والشفرات المخصصة لها، كما تبين المحارف الإضافية للمحارف العربية والتي تستخدم في عدد من اللغات العالمية مثل اللغة الفارسية ولغة الأوردو وغيرها. إن المحارف العربية الأساسية مبنية في الموضع 060C ولغاية 0671، وتتضمن هذه الموضع جميع الحروف العربية الأساسية، وأشكال الهمزة المختلفة، وعلامات التقطيع، وعلامات الشكل، والأرقام الهندية (٠، ١، ٢، ٣، ٤، ٥، ٦، ٧، ٨، ٩). كما تتضمن الموصفات محارف أخرى تستخدم في العديد من اللغات المختلفة التي تستخدم الحروف العربية بصورة رئيسية، ومن هذه اللغات المختلفة: البلوشي، والكمبوري، والказاخ، والسندي، والباشتوي، والفارسية، والأوردو، والبربرية، والكردية، وغير ذلك.

التعامل مع اللغات المختلفة

هناك عدة موصفات دولية لتحديد كيفية ترتيب المعلومات المنقولة عبر شبكة الإنترنت ليتم عرضها على الشاشة بطريقة موحدة. وهناك عدد من الموصفات الدولية في هذا المجال مثل موصفة (HTML) Hyper Text Markup Language التي تم تطويرها واعتمادها من قبل الائتلاف العالمي للشبكة العنكبوتية (World Wide Web Consortium). وتتضمن الموصفات الدولية عدة خصائص للتعامل مع لغات العالم المختلفة وتمكنها من العرض بطريقة صحيحة. وإنه من الضروري تصميم الموقع الإلكتروني على شبكة الإنترنت وتوسيع محتواها باللغات المختلفة بما في ذلك اللغة العربية بشكل يتوافق مع الموصفات الدولية من جهة، كما يمكن من نقل التطبيقات واستخدامها على الأنواع المختلفة من الأجهزة والبرامج من جهة أخرى.

استخدام مجموعات المحارف

إن مجموعات المحارف المسموح باستخدامها في الوثائق المختلفة التي يتم تبادلها عبر شبكة الإنترنت تقوم بتحديدها هيئة تخصيص أرقام الإنترنت.

ووظائف شبكة الإنترنت العالمية للمصلحة العامة في كافة أنحاء العالم.

يتم تحديد مجموعة المحارف المستخدمة في وثيقة ما ضمن الحقل المخصص في الموصفة المستخدمة (مثل حقل HEAD في HTML)، وهذا يحدد للبرنامج الذي يعرض الوثيقة كيفية تشفير النص الذي تحتويه هذه الوثيقة. على سبيل المثال، يمكن تحديد مجموعة المحارف المستخدمة، ووضعها مع معلومات أخرى في عنوان وثيقة HTML كما يبين المثال الآتي:

```
>META HTTP-EQUIV="Content-Type" Content="text/html; charset=iso-8859-6"<
```

ويحدد هذا المثال مجموعة المحارف المستخدمة بأنها مجموعة المحارف العربية/اللاتينية ذات الثمانية عناصر المعرفة في الموصفة الدولية إيزو ٦-٨٨٥٩. وإن مجموعات المحارف الآتية عادة ما تستخدم في موقع الإنترنت التي يمكن أن تصممها الدول العربية (أو التطبيقات المستخدمة في هذه الدول):

- إيزو ١-٨٨٥٩ (iso-8859-1): مجموعة المحارف اللاتينية المشفرة ذات الثمانية عناصر
- إيزو ٦-٨٨٥٩ (iso-8859-6): مجموعة المحارف العربية/اللاتينية المشفرة ذات الثمانية عناصر

- المواصفة العربية رقم ٧٠٨ (asmo-٧٠٨) : مجموعة المحارف العربية / اللاتينية المشفرة ذات الثمانية عناصر

وتتجدر الاشارة هنا أن أي جدول شفرة يمكن أن يحتوي على عدد محدود فقط من المحارف. وعندما يحتوي موقع الانترنت على محارف موجودة في جدول الشفرة، فيجب إعلام برنامج الترئية (browser) عن هذه المحارف وذلك بتعريفها في وثيقة HTML مثلاً كما هو مبين في تلك المواصفة، وذلك حتى يتمكن برنامج الترئية من إظهار هذه المحارف بطريقة صحيحة.

وإذا تم استخدام مواصفة XML على سبيل المثال لتبادل نصوص تحتوي على اللغة العربية، يكون الاستخدام كما يبين المثال الآتي:

```
>/p> Following is an example in Arabic language.>p xml:lang="en"< -  
- ة س و ك ع م ة ي ب ر ع ل ا ن ي ز خ ت م ت ي >p xml:lang="ar"<  
>/p<
```

وإن الرمز "en" يعني اللغة الإنجليزية والرمز "ar" يعني اللغة العربية وذلك وفقاً للمواصفة الدولية إيزو ٦٣٩ . وتتجدر الملاحظة هنا أن اللغة العربية تظهر في المثال أعلاه وفق تخزينها في الذاكرة، حيث يتم تخزين الأحرف من اليسار إلى اليمين، وإن خوارزمية الترئية تعكس ذلك عند إظهار النص على الشاشة، حيث يظهر النص في المثال أعلاه كالتالي "يتم تخزين اللغة العربية معكوسه".

ونؤكد في هذا المجال أن المواصفات العالمية (مثـل HTML و XML) تتيح إمكانية تعريف اللغة وذلك المساعدة فقط في تحليل السياق والنص، مثل استخدام المعاجم لتصحيح الأخطاء. ولا تستخدم خاصية تعريف اللغة مطلقاً في تحديد اتجاه عرض النص المحتوى في وثيقة HTML أو XML أو غيرها، إذ تتولى هذه المهمة خوارزمية الترئية الأساسية مستخدمة محارف التحكم كما ورد سابقاً في هذه الورقة.

REFERENCES

1. Unicode 3.1: The Unicode Standard, Ver. 3.1.0, The Unicode Consortium. The Unicode Standard, Ver. 3.0, The Unicode Consortium, Reading, Mass: Addison-Wesley Developers Press, 2000, as amended by Unicode Standard Annex #27: Unicode 3.1. <http://www.unicode.org/unicode/>
2. ISO/IEC 10646: Information technology – Universal Multiple-Octet Coded Character Set (UCS) – Part 1: Architecture and Basic Multilingual Plane. International Organization for Standardization (ISO), Geneva, Switzerland, 2000.
3. ISO 8859/6: Information technology – 8-bit Arabic/Latin coded character set for information interchange. International Organization for Standardization (ISO), Geneva, Switzerland.
4. ISO 639: Codes for the representation of names of languages. International Organization for Standardization (ISO), Geneva, Switzerland, 1988.
5. ISO 3166: Codes for the representation of names of countries and their subdivisions – Part 1: Country Codes. International Organization for Standardization (ISO), Geneva, Switzerland, 1997.
6. IANA, Official Names for Character Sets, ed. Keld Simonsen et al. Internet Assigned Number Authority. See <ftp://ftp.isi.edu/in-notes/iana/assignments/character-sets>.
7. IETF RFC 1766: Tags for the Identification of Languages, ed. H. Alvestrand. Internet Engineering Task Force, 1995.
8. W3C XML: Extensible Markup Language (XML) 1.0, ed. Tim Bray et al., World Wide Web Consortium, 1998.
9. W3C XHTML: The Extensible Hyper Text Markup Language (XHTML) 1.0, World Wide Web Consortium, 2000.