

نظام حاسوبي لرومنة الأسماء العربية

د. عبد الملك السلطان

جامعة الملك سعود

د. منصور الغامدي وخالد الحقييل

مدينة الملك عبد العزيز للعلوم والتقنية

صالح الصبي

شركة موباييلي

٢١٤

الكلمات المفتاحية: الأسماء العربية، النقل الكتابي، الرومنة، الحروف، نظام الكتابة.

ملخص: تشكل كتابة الأسماء العربية بالحرف الروماني معضلة ظلت لسنوات عديدة بدون حل. ويرجع أحد أسباب وجود المشكلة إلى عدم وجود معيار موحد للرومنة يمكن الاتفاق عليه. تقدم هذه الورقة توصيفا لنظام حاسوبي تم بناؤه لرومنة الأسماء العربية، وهو قابل للتنفيذ على المستوى العربي على جميع الأنظمة مستخدما الحروف الرومانية المتوفرة على الأجهزة الحاسوبية والطابعات ويمكن قراءته بسهولة من قبل القارئ العادي. يتناول هذا البحث خلفيات مشكلة رومنة الأسماء العربية وطرائق الحل الممكنة وأنموذج الرومنة الذي تم اعتماده ووصفا تقنيا لتصميم وتنفيذ نظام الرومنة.

مقدمة

اللغات الأجنبية، وذلك لاستخدام هذا النوع من كتابة الأسماء في المستندات الرسمية وبطاقات التعامل كجوازات السفر والبطاقات المصرفية وتذاكر السفر وغيرها، وفي حالة اختلاف كتابة الاسم الواحد للشخص فإنه يترتب على ذلك ما قد يظهر للمتعامل مع الاسم على أنه ازدواجية لهوية الشخص. يهتم هذا البحث العديد من القطاعات، والتي منها وزارة الداخلية ووزارة الخارجية والجهات التي تتعامل بالحرف الروماني عند كتابة الأسماء العربية كشرركات الطيران ووكالات السفر والسياحة والمستشفيات والبنوك. ولن تقتصر الاستفادة من هذا البحث على الجهات السابق ذكرها بل وسيستفيد منه المواطن العادي الذي سيجد أن هناك نظاما موحدا لكتابة الاسم العربي بالحرف الروماني، وكذلك وجود نظام حاسوبي يتولى عملية النقل الكتابي أليا بمجرد كتابة الاسم بالحرف العربي. لا يجد الكاتب العربي صعوبة تذكر في كتابة أي اسم عربي بالحرف العربي حتى لو كان يسمعه للمرة الأولى. ويعود ذلك لعاملين: الأول، أن الكتابة العربية تقوم على النطق غالبا؛ بمعنى أن الاسم يكتب كما ينطق وأن العلاقة بين التمثيل الصوتي (الفونيم phoneme) التمثيل الكتابي (الغرافيم grapheme) هي علاقة متطابقة إلى حد بعيد - وهذا عكس لغات أخرى كالإنجليزية، على سبيل المثال. العامل الثاني، أن الكتابة العربية المعاصرة تقتصر على الحروف التي غالبا ما تمثل الصوامت consonants دون علامات التشكيل التي غالبا ما تمثل الصوائت vowels

كتابة الاسم العربي بالحرف الروماني وهو ما يعرف بالنقل الكتابي transliteration؛ إذ نجد تبايناً واضحاً عند كتابة الأسماء العربية بحروف اللغات التي تستخدم الحرف الروماني في الكتابة. فعلى سبيل المثال، أظهرت نتيجة البحث باستخدام محرك البحث Google أن الاسم العربي المشهور "محمد" يظهر اختلافا جليا في الحروف المستخدمة لرومنته (الجدول ١). فرغم قدم استعمال هذا الاسم وكثرة استخدامه إلا أنه يكتب بطرق متعددة ولكل منها شيوع في الاستخدام مما يبرز التباين في رومنة الأسماء العربية.

جدول ١ :
تكرار النقل
الكتابي للاسم
"محمد"
بالحرف

التكرار	النقل الكتابي	التسلسل
2.280.000	Muhammad	[1]
2.000.000	Mohammed	[2]
1.600.000	Mohamed	[3]
1.150.000	Mohammad	[4]
388.000	Muhammed	[5]
264.000	Mohamad	[6]
69.100	Muhamed	[7]
44.600	Muhamad	[8]

Google



ويعود السبب لهذا التباين في النقل الكتابي للأسماء العربية بالحرف الروماني لعدة عوامل، منها:

نظام حاسوبي لرومنة الأسماء العربية

٢١٥

(١) اختلاف النظام الصوتي sound system العربي عن نظيره في اللغات التي تستخدم الحرف الروماني؛ فعلى سبيل المثال، هناك ٢١ فونيمًا في العربية ليس لها نظير في الإنجليزية .

(٢) لم تكن للمستشرقين والرحالة الغربيين الأوائل الذين كتبوا عن العالم العربي دراية دقيقة باللغة العربية وأصواتها. مما جعلهم يعتمدون في النقل الكتابي للأسماء العربية على نطق المتحدثين باللهجات العامية في تلك العصور ثم اجتهاداتهم الشخصية. لذا نجد كاتب كـ "لورنس" يكتب مدينة "جدة" ٢٥ مرة: Jeddah، و٦ مرات أخرى: Jidda، ومرة واحدة يكتبها: Jedda في الكتاب نفسه الذي نشره للمرة الأولى عام ١٩٢٦م، وجميعها تختلف عن الاسم المعتمد في المملكة العربية السعودية وهو Jaddah. هذا الاختلاف في كتابة الأسماء العربية استمر بعد ذلك بسبب عدم وجود نظام للنقل الكتابي بين الحرف العربي والحرف الروماني؛ فيذكر "لورنس" في الصفحة ٩ من المرجع السابق ما نصه: "لا يمكن نقل الأسماء العربية إلى الإنجليزية نقلًا دقيقًا، بسبب صوامتها التي تختلف عن صوامتها (يقصد صوامت اللغة الإنجليزية) وصوائتها التي كصوائتها تختلف من مكان إلى آخر".

(٣) أن اسم العلم العربي قد يأتي مفردًا كما في "سليمان" وقد يأتي مركبًا كما في "أبو داود" و "عبد الله" مع وجود مسافة بين الكلمتين المكونة للاسم المركب، وهذا يزيد من صعوبة الرومنة؛ إذ نجد الاسم العربي المركب يرومن مرة بكلمة واحدة وأحيانًا بكلمتين منفصلتين وبأشكال مختلفة:
Abdu Allah, Abdu-allah, Abdullah.

(٤) عند كتابة الأسماء العربية بالحرف العربي فإنه غالبًا ما تكتب بالحروف دون علامات التشكيل مما يجعل بعض الأسماء موقع لبس في نطقها كما في "حميد" الذي يمكن أن ينطق "حميد" و "حميد" و "سعد" و "سعد". هذا الاختلاف ينتقل إلى الحرف الروماني فيجعل الاسم الواحد يكتب بأكثر من طريقة.

(٥) كثرة معايير النقل الكتابي وتباينها؛ فهناك نظام مكتبة الكونغرس، ونظام الأمم المتحدة، ونظام دائرة المعارف الإسلامية، ونظام قاموس فهر، ونظام المنظمة العالمية للمقاييس ISO. من أجل ذلك عُقدت الندوة العلمية الأولى حول توحيد معايير النقل الكتابي: الأبعاد الأمنية، بجامعة نايف العربية للعلوم الأمنية عام ١٤٢٤هـ. وأوصت باستخدام نموذج مقترح لكتابة الأسماء العربية بالحرف الروماني.

والنقل الكتابي لأسماء الأعلام لا يقتصر على الأسماء العربية، فهو قائم بين اللغات باختلاف أصولها ونظمها الصوتية وأساليبها في الكتابة. إلا أن النقل الكتابي بين اللغات التي تستخدم الحرف نفسه في الكتابة عادة ما يكون أسير منه عندما تختلف حروف كتابتها. فالنقل الكتابي أقل تعقيدًا بين الفرنسية والألمانية والإنجليزية والفرنلندية، على سبيل المثال، لأن جميعها تستخدم الحرف الروماني. وغالبًا ما تكون كتابة اسم العلم هي نفسها في أي من هذه اللغات، فنجد اسمًا مثل: Antonio يكتب كما هو في أي من هذه اللغات سألفة الذكر. إلا أن الأمر يصبح أكثر تعقيدًا عندما يختلف الحرف والنظام الصوتي كما هو قائم بين الإنجليزية والصينية، وكما هو كذلك بين العربية والإنجليزية. أما في اليابانية فرغم اختلاف نظمها الكتابي عن النظام الروماني إلا أن النقل الكتابي للأسماء التي تكتب بالحرف الروماني إلى اليابانية يكون باستخدام ألفبائية خاصة بالنقل الكتابي غير تلك التي تستخدم في كتابة اللغة اليابانية، وهذا يسهل إلى حد ما عملية النقل الكتابي. وتختلف نظم كتابة اللغات، فنجد أن الكتابة العربية فونيمية، حيث لكل صوت (فونيم) رمز كتابي (قرايفم). إلا أننا نجد لغات أخرى كالصينية لها نظام مقطعي syllabic أي أن لكل مقطع رمز خاص به. فعلى سبيل المثال، لكل مقطع من المقاطع التالية رمز مستقل: /كـ/، /كـ/، /كـ/، ويقاس على ذلك الاحتمالات الأخرى المكونة للمقاطع من صوامت وصوائت مما يجعل عدد القرايفمات يصل إلى ما يقرب من ستة آلاف رمز وهذا يزيد من تعقيد النقل الكتابي بين هذين النوعين من اللغات إلا إذا استحدثت ألفبائية لغرض النقل الكتابي كما هو بالنسبة لليابانية.

وقد نهج الباحثون طرائق شتى في عملية النقل الكتابي تختلف بناء على اللغات التي يتعاملون معها والأسلوب الذي يروونه أنسب للخروج بما يروونه حلولًا ناجعة لما بين أيديهم من بيانات لأسماء أعلام أو مصطلحات. فقام "جارموتوي فونين" وآخرون باستخدام القوانين التحويلية transformation rules في عملية النقل الكتابي لأسماء الأعلام والمصطلحات بين لغات جميعها يستخدم الحرف الروماني كالألمانية والأسبانية والسويدية والفرنلندية والفرنسية إلى الإنجليزية. حيث تستخلص القوانين التحويلية أليًا من القواميس. فعلى سبيل المثال، ia في الأسبانية يقابلها y في الإنجليزية كما في "علم الجسد" somatologia و somatology. على التوالي. فيكون القانون التحويلي هنا هو: ia ← y ولم تكن نسبة الدقة باستخدام هذه الطريقة في النقل الكتابي عالية. فعلى سبيل المثال، نسبة النقل الكتابي الصحيح من الفنلندية إلى الإنجليزية ٧٢٪. ويعود السبب في ذلك إلى كثرة الحالات التي ليس لها قانون ثابت.

واستخدم باحثون آخرون القوانين التي تتولى تحويل القرايفمات في لغة كالصينية إلى رموز صوتية ثم تحويلها بعد ذلك إلى حروف رومانية. فالقوانين هنا ضرورية حتى يكون النقل الكتابي أقرب ما يكون لنطق الصوت بلغته الأصلية، وحتى تفك رموز الكتابة

الصينية التي تقوم على الكتابة المقطعية ولا تعتمد على الألفبائية كما هي الحال بالنسبة للحرف الروماني.

ومن طرق النقل الكتابي الإحصاء الذي يقوم على عامل الاحتمالية probability. وهذه الطريقة كسابقتها تعتمد على قواعد البيانات، إلا أنها في هذه الحالة، يكون هناك أكثر من شكل للنقل الكتابي للكلمة الواحدة. وتعتمد هذه الطريقة على الشكل الأكثر شيوعاً. وتكون قاعدة البيانات إما قواميس إلكترونية أو استخدام محركات البحث على شبكة الإنترنت. فعلى سبيل المثال، الاحتمالية المرشحة للنقل الكتابي للاسم "محمد"، كما في الجدول ١، هي Muhammad، فاحتماليتها ٢٩٪، بينما احتمالية الكلمة التي تليها Mohammed هي ٢٥٪. ويعاب على هذه الطريقة أن الكلمة الأكثر شيوعاً قد لا تكون هي الصحيحة. فعلى سبيل المثال كلمة "مكة" وردت ١٦٠,٠٠٠ مرة بالشكل Makkah عند البحث عنها في Google بينما وردت ١,١٥٠,٠٠٠ مرة بالشكل Mecca، رغم أن الرومنة الأولى هي الصحيحة والمعتمدة رسمياً.

ومن طرق رومنة الأسماء العربية القائمة الآن ما تعتمد على قواعد البيانات أي أنه يتم إدخال الاسم يدوياً بالحرفين العربي والروماني للمرة الأولى وعند وروده فيما بعد فإن النظام يستدعي الاسم الذي سبق تخزينه في قاعدة البيانات. وفي هذه الحالة فإنه لا بد أن يكون مدخل الاسم على دراية بقواعد رومنة الأسماء العربية وعلى دراية كذلك بنطق الاسم. إلا أن هذا النوع من النظم يعتمد على اجتهاد مدخل البيانات مما يجعل الرومنة عرضة للخطأ إضافة إلى أنها لا تفرق بين الأسماء عند تطابقها في الحروف واختلافها في التشكيل وفي كل مرة يرد اسم جديد فلا بد من إضافته إلى قاعدة البيانات.

عالج (Sinha) أسماء الأعلام على أنها كلمات مجهولة unknowns وعرفها بأنها أحد أنواع الكلمات التي لا يحويها المعجم؛ بمعنى أن الأسماء من الكلمات التي ليس لها معنى ومن ثم لا بد من معالجتها عند الترجمة بأن تنقل كتابياً ولا تترجم. ففي مرحلة التحليل الصريح للكلمات أثناء عملية الترجمة الآلية يستبعد المحلل الصريح الكلمات التي ليس لها جذور في قاعدة بيانات اللغة وتعامل هذه الكلمات على أنها أسماء تنقل كتابياً ولا تترجم.

وقد استخدمت برمجيات مختلفة لتنفيذ عملية النقل الكتابي؛ ومنها Hidden Markov Model التي استخدمت في تحويل حروف أسماء الأعلام من الإنجليزية إلى الكورية التي تتطابق فيها الأصوات مع الحروف. كما استخدمت الشبكة العصبية الصناعية artificial neural networks في تشكيل الأسماء العربية تمهيداً لرومنتها.

تصميم وبرمجة نظام الرومنة

هناك طريقتان على الأقل لرومنة الأسماء، هما:

الطريقة الأولى: النقل الكتابي الدقيق للحروف العربية، وذلك بإيجاد رمز بالحرف الروماني لكل رمز يستخدم في الكتابة العربية بما في ذلك علامات التشكيل إضافة إلى الحروف، فيمكن بذلك الرجوع لأي اسم عربي بمجرد معرفة كتابته بالحرف الروماني. فعلى سبيل المثال، يكتب الاسم "سَمَر" هكذا: Samar. هنا لكل رمز عربي ما يقابله من الروماني: س = a، ' = s، م = m، ر = r.

إلا أن المعضلة تكمن في وجود ٢١ رمزا كتابياً في العربية ليس لها نظير في الحرف الروماني. فالاسم "عَصْمَاء" لا يمكن نقله كتابياً إلى الحرف الروماني لأن هناك أربعة رموز في هذا الاسم ليس لها مقابل وهي: ع، ص، ألف المد، ء. مما يحتم في هذه الحالة استخدام رموز جديدة تضاف إلى رموز الحرف الروماني. وقد اجتهد الباحثون في إيجاد رموز إضافية لتغطية النقص في الحرف الروماني فعمدوا إلى استخدام عدة وسائل منها:

- استخدام الحروف الصغيرة والكبيرة، فيرمز لـ "س" بـ S ويرمز لـ "ص" بـ S وهكذا بالنسبة لبقية الحروف التي ليس لها نظير في الحرف الروماني.

- استخدام رموز جديدة، مثل ' لتدل على "ع".

- وضع خط تحت الحروف، كما في t للدلالة على "ط" وللتفريق بينه وبين t الذي يرمز للحرف "ت".

- استخدام الحروف التي ليس لها مقابل في العربية مثل X لتقابل "ح" و v لتقابل "ذ".

ورغم شيوع هذه الطريقة عند المتخصصين والدارسين إلا أنه يؤخذ عليها صعوبة التذكر والاستخدام عند الناس العاديين وذلك عند تعبئة بياناتهم الشخصية في الأوراق الرسمية الأجنبية. إضافة إلى ذلك، وهذا الأهم، أن بعض النظم الحاسوبية والطابعات لا تتعامل معها. فعلى سبيل المثال، نظم الطيران العالمية لا تستخدم الحروف الرومانية الصغيرة ولا تستخدم الخط تحت الحرف أو الشحطة " " في الاسم، كما هي عند كتابة الاسم "العباد" Al-Abbad. وتكمن الأهمية هنا في أن الأسماء الآن أصبحت مدخلة في الحاسب الآلي وتعالج حاسوبياً وإذا لم تكن متوافقة معه فإنها لن تستخدم، إذ لا تكمن المشكلة في الحاسب الشخصي والطابعات المكتبية ولكن في النظم الحاسوبية العالمية التي تتعلق بقواعد البيانات الكبيرة كما هو في شركات الخطوط الجوية ومراكز المعلومات.

وسبق وأن عمل نظام حاسوبي لرومنة الأسماء العربية مبني على هذا النوع من الربط بين الحرف العربي والروماني إلا أن هذا النظام

يقوم بتمثيل كافة الحروف العربية مما يجعله يضيف رموزاً أخرى غير تلك التي في الحرف الروماني، ومن ذلك "h" ، "t" ، " " ، التي تقابل الهمزة والحاء والطاء، على التوالي. ورغم أن هذا النظام يفيد في حالة النقل الكتابي الدقيق للأسماء العربية إلا أنه لا يمكن تطبيقه في السجلات والوثائق الرسمية للأسباب المذكورة أعلاه.

الطريقة الثانية: دمج الرموز العربية لتقابل الرموز الرومانية؛ وذلك باستخدام رمز واحد بالحرف الروماني ليقابل أكثر من رمز في الحرف العربي. فدمج "ص" و "س" في الحرف s، ودمج "ت" و "ط" في t، ودمج ألف المد والفتحة في a. وتحذف بعض الرموز العربية مثل "ع" في "علي" Ali. وهذه الطريقة هي الأكثر شيوعاً في الاستخدام بين الناس، إضافة إلى إمكانية تطبيقها واستخدامها في النظم الحاسوبية باختلاف برمجياتها وقواعد بياناتها، إلا أنه يعيبها أنها غير عكسية (لا يمكن عمل التحويل العكسي: من روماني إلى عربي).

وقد تم اعتماد الطريقة الثانية (أنظر الملحق أ)، وهو ما يميز هذا البحث عن غيره من الأبحاث والبرامج السابقة الذكر. وقد تم تصميم نظام حاسوبي مبني على هذا مكون من أربعة أجزاء (برامج)، هي:
برنامج مصحح قائمة الأسماء.
برنامج تهيئة قاعدة البيانات.
برنامج الرومنة (محرك الرومنة).
برنامج الواجهة لنظام ويندوز (كمثال لأحد الأنظمة المشهورة).
ولكل من هذه البرامج وظيفة أساسية سيتم التحدث عنها بالتفصيل فيما يلي:

(أ) مصحح قائمة الأسماء

ينبغي نظام الرومنة على فرضية أن هناك مجموعة من الأسماء العربية التي تم إدخالها يدوياً بطريقة محددة، وقد تم بالفعل إدخال ما يربو على ٧٠,٠٠٠ اسم عربي مضبوط بالشكل. وظيفه هذا البرنامج الأساسية هي فحص الأسماء المدخلة والتأكد من خلوها من الأخطاء الكتابية. في حالة وجود خطأ، فإن البرنامج يمكن المستخدم من تصحيح الخطأ وإعادة إدخاله من جديد. لذا، فإن فائدة هذا البرنامج هو التخفيف من عملية الاتصال بين مدخل البيانات وبين منشئ قاعدة البيانات حيث أنه عند تهيئة قاعدة البيانات، كما سيتم شرحه لاحقاً، فإن البرنامج لا يقوم بإدخال الأسماء التي بها أخطاء وعندئذ، يجب الاتصال بمدخل البيانات حتى يقوم بالتصحيح.

يستقبل البرنامج ملفاً نصياً مكوناً من أسطر، كل سطر يحوي على أصل الاسم في البداية وبعده التشاكيل الخاصة به وكل اسم يفصل عن الاسم الذي بجانبه بـ TAB، حيث يقوم البرنامج بفحص الملف بحثاً عن أخطاء. هذه الأخطاء هي في الغالب بسبب الكتابة أو بسبب مخالفة قواعد التشكيل في اللغة العربية. ملحق ب يبين قائمة بالأخطاء التي يقوم البرنامج بالتدقيق عليها.

بعد الانتهاء من عملية التصحيح، فإن نفس الملف الذي تم اختياره للتصحيح هو الملف المصحح الذي يتم إدخاله على البرنامج الثاني (برنامج تهيئة قاعدة البيانات).

(ب) برنامج تهيئة قاعدة البيانات

يقوم هذا البرنامج بإنشاء قاعدة البيانات للأسماء بصيغة XML، وتتم هذه العملية باستخدام برنامج مخصص لذلك يقوم باستقبال ملف نصي (منتهي باللاحقة "txt") ومن ثم يقوم البرنامج بإنشاء قاعدة البيانات وتنسيقها بشكل مناسب، كما يقوم بإنشاء ملف يحتوي على قائمة بالأسماء التي لم يتمكن من إدخالها في قاعدة البيانات مع ذكر سبب عدم الإدخال.

مدخلات ومخرجات البرنامج

مدخلات البرنامج عبارة عن أسماء أعلام عربية مذكورة ومؤنثة تكون مرتبة في الملف النصي الذي يستقبله البرنامج ومنسقة بتنسيق خاص؛ حيث يجب أن تكون الأسماء مرتبة في الملف نصي بناءً على الترتيب الأبجدي للحروف. كما يجب أن يكون كل سطر في الملف مختصاً باسم واحد فقط مع التشاكيل المتاحة له؛ حيث يكون أول اسم في السطر هو الاسم الأساس بدون تشكيل وباقي الأسماء في نفس السطر هي التشاكيل الممكنة للاسم الأساسي. كما يجب أن يفصل بين الأسماء في نفس السطر باستخدام (TAB)، كما في المثال التالي:

العِيدَان العِيدَان العِيدَان

ويمكن الحصول على هذا الملف باستخدام أغلب أنظمة قواعد البيانات حيث أنها تدعم تحويل البيانات في قاعدة البيانات إلى صيغة ملف نصي.

و نظراً لأن بعض قواعد البيانات لا تدعم التشكيل العربي، فقد تم الاستعاضة عن علامات الضبط بالشكل برموز مقابلة لها كما هو موضح في الجدول ٢:

الرمز المكافئ	حرف التشكيل
0	"_"
1	"َ"
2	"ُ"
3	"ِ"
4	"ْ"
5	"ّ"
6	"ّّ"
7	"ّّّ"
8	"ّّّّ"

جدول ٢ :
الرموز المكافئة
لحروف التشكيل

٢١٨

و كما هو ملاحظ في جدول ٢، فإن علامات التنوين قد تم استبعادها نظراً لأن الأسماء لا يمكن أن تشكل بالتنوين.

طريقة عمل البرنامج

يحتاج المستخدم إلى تشغيل هذا البرنامج مرة واحدة فقط لبناء قاعدة البيانات الأساسية التي تستخدم بواسطة البرنامج الرئيسي ”محرك الرومنة“.

يقوم البرنامج بقراءة الملف النصي وتحليل محتوياته بحيث يكون كل سطر على حدة، فيقوم بإزالة الفراغات في أول وآخر السطر حتى يكون السطر مبتدئاً بحرف ومنتهاً بحرف أو شدة.

يقوم البرنامج في البداية بالتأكد من الكلمة الأساس في السطر (الكلمة الأولى) حيث يجب أن تكون مكونة من حروف فقط أو حروف وفراغات في حالة الأسماء المركبة، إذا كان الاسم متبوعاً بالتشكيل الممكنة له، فإن البرنامج يقوم بإنشاء صف جديد في قاعدة البيانات يحتوي في الخلية الأولى على الاسم الأساس. ثم يبدأ البرنامج بتحليل باقي الأسماء على نفس السطر، حيث يقوم بمقارنة الاسم بعد حذف التشكيل مؤقتاً بالاسم الأساس بدون تشكيل بالكلمة الأساس للتأكد من عدم وجود أخطاء في الإدخال ومن ثم يقوم البرنامج بتحويل حروف التشكيل إلى الرموز المكافئة لها؛ ويقوم بكتابة تسلسل الرموز في الخلية التالية على الصف نفسه، وبهذا فإنه يتم الاحتفاظ بالاسم الأساس وعدد التشكيل الممكنة للاسم وسرد لتلك التشكيل كما هو موضح في الشكل ١، وهو ناتج عن حفظ البيانات الموجودة في الشكل 4 في قاعدة البيانات التي ينتجها البرنامج. أما إذا كان الاسم غير متبوع بالتشكيل الممكنة له فإن البرنامج يقوم بتجاهل هذا الاسم وكتابة هذا الخطأ في ملف الأخطاء.

العنيدان العنيدان العنيدان

```
<rawname name="العنيدان" count="2">
<variations>00214100</variations>
<variations>00130000</variations>
</rawname>
```

شكل ١ :
ترتيب الاسم
مع التشكيل
الخاصة به في
قاعدة البيانات

ويلاحظ من الشكل أن الرموز ”00214100“ تكافئ التشكيل الخاص بالاسم ”العنيدان“ حسب الرموز المكافئة في الجدول ٢. بعد أن ينهي البرنامج تحليل جميع السطور في الملف النصي يتم حفظها في ملف بامتداد ”XML“؛ ويقوم بإنشاء ملف نصي يحوي الأسماء التي تحتوي على أخطاء ولم يتمكن البرنامج من التعامل معها.

بالنسبة للأخطاء التي يتم الكشف عنها في هذا البرنامج هي نفس الأخطاء في برنامج مصحح قائمة الأسماء (البرنامج الأول) الموجودة في الملحق ب، بالإضافة إلى الكشف عن الأسماء المكررة في قاعدة البيانات وتبنيه المستخدم على ذلك حفاظاً على صحة قاعدة البيانات.

(ج) محرك الرومنة

يقوم البرنامج باستقبال اسم مشكل ومن ثم يعمل على تحليله وإخراج الرومنة الخاصة بهذا الاسم، إذا كان الاسم خالياً من

الأخطاء، كما يقوم البرنامج بإخراج رسالة الخطأ المناسبة في حالة حدوثه. كما يقوم البرنامج برومنة جميع الأسماء الموجودة على قاعدة البيانات وإخراجها ويقوم أيضا بإضافة اسم جديد أو تشكيل جديد على قاعدة البيانات.

مدخلات ومخرجات البرنامج

يقوم البرنامج باستقبال الأمر ومن ثم تحليله وإخراج الناتج. فمن الأوامر التي يستقبلها البرنامج رومنة الاسم العربي حيث يقوم البرنامج باستقبال الاسم مشكلاً وإخراج الرومنة الخاصة بذلك الاسم وفقاً للقواعد المعتبرة للرومنة، إذا كان الاسم صحيحاً. وحتى يعمل البرنامج بشكل سليم يجب أن تكون قيم الترميز (Encoding Values) الخاصة بالحروف العربية والحركات العربية مجهزة مسبقاً (سيتم بيانه لاحقاً).

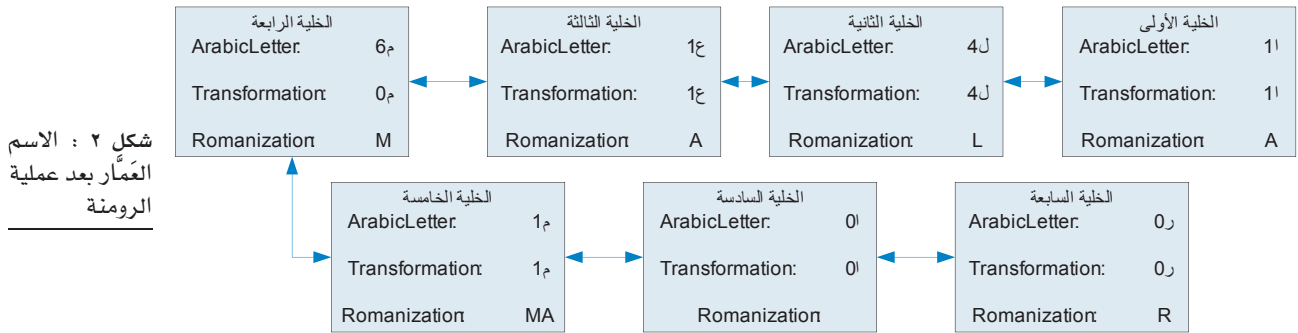
البرنامج يعمل على موجه الأوامر (Command Prompt) حيث يتم تشغيل البرنامج وتوفير المعطيات المناسبة (Command Arguments) لعمله، مثل الاسم، ملف قيم الترميز ونوع العملية المطلوبة. وقد تم تصميم البرنامج للعمل على موجه الأوامر حتى نضمن للبرنامج بأن يعمل على معظم الأنظمة المستخدمة حالياً مثل أنظمة ويندوز، لينكس، ماكنتوش وغيرها من الأنظمة. ولتسهيل التعامل مع البرنامج فقد تم تصميم واجهة لكل نظام للتعامل مع محرك الرومنة.

ومن الأمثلة على المعطيات التي يمكن بواسطتها تشغيل برنامج الرومنة:

- لرومنة الاسم "صالح"
- C:\>Romanization.exe -encoding Windows.code -صالح
- لرومنة جميع الأسماء في قاعدة البيانات
- C:\>Romanization.exe -encoding Linux.code -all
- لإضافة الاسم "مسيلمة" إلى قاعدة البيانات
- C:\>Romanization.exe -add مَسَيْلَمَة

تراكيب البيانات المستخدمة في محرك الرومنة

حتى يتم فهم طريقة عمل البرنامج، يجب أولاً شرح تراكيب البيانات (Data Structure) المستخدمة أثناء عمل البرنامج. في الشكل ٢ نجد شكل تراكيب البيانات بعد رومنة الاسم "العمار".



كما هو ملاحظ في الشكل ٢، فإن تراكيب البيانات على شكل سلسلة من الخلايا، وكل خلية تحتوي معلومة متعلقة بحرف من الحروف في الاسم. كل خلية تحتوي على ثلاث معلومات. المعلومة الأولى هي الحرف العربي (Arabic Letter) وتحتوي على الحرف العربي من الاسم دون أي تغيير فيه مع التشكيل المرافق للحرف. وكما هو ملاحظ فإن التشكيل عبارة عن نفس الأرقام الموجود في جدول ٢. المعلومة الثانية الموجودة في كل خلية هي التحويل الذي يطرأ على الحرف أثناء عملية الرومنة (Transformation) مثل تغيير التشكيل كما في الخلية الرابعة في الشكل. والسبب في عدم التغيير على المعلومة الأولى هو أننا نريد الاحتفاظ بالحرف الأصلي لحاجتنا له أثناء الرومنة. المعلومة الثالثة والأخيرة هي الحرف بعد رومنته (Romanization) ويحتوي على الحرف الروماني المقابل للحرف العربي والتشكيل التابع له. من الممكن أن تكون هذه الخلية خالية كما في الخلية السادسة أو على أكثر من حرف كما في الخلية الخامسة، أو على حرف واحد كما في بقية الخلايا.

طريقة عمل البرنامج

يقوم البرنامج بوظيفتين هما الرومنة (-encoding) وإضافة اسم جديد لقاعدة البيانات (-add).
أ. الرومنة "encoding":

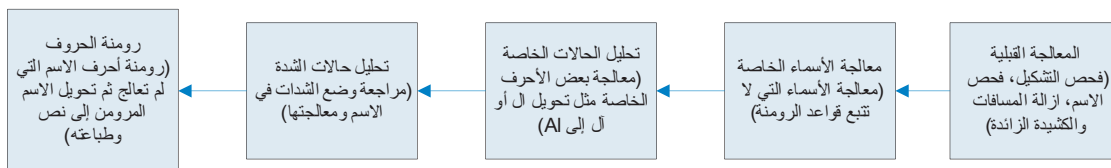
ويكون متبوعاً بالملف الذي يحتوي على قيم الترميز (في حالة عدم وجوده يتم استخدام الترميز الافتراضي)، ومن ثم الاسم

المطلوب رومنته أو متبوعاً بكلمة all- لرومنة جميع الأسماء المخزنة في قاعدة البيانات.

ب. إضافة اسم جديد لقاعدة البيانات "add-":

ويكون متبوعاً بالاسم المشكل المراد إضافته إلى قاعدة البيانات. ولإضافة الاسم على القاعدة فإن البرنامج يقوم أولاً بالبحث عن الاسم، فإذا وجدته ولم يكن هناك التشكيل المراد إضافته، فإن البرنامج يقوم بإضافة التشكيل. أما إذا لم يجد الاسم فإنه يتم إضافته كاسم جديد.

عندما يستقبل البرنامج الاسم يقوم أولاً بتحويل الحروف العربية والحركات إلى رموزها المكافئة بناء على القواعد المتوفرة في ملفات الترميز. ثم يقوم بعد ذلك بفصل التشكيل عن الاسم. الخطوات التالية تتم فيها عملية رومنة الاسم واستبدال الحروف والحركات العربية بمقابلاتها الرومانية. يتم ذلك على خمس مراحل متتالية (كما في الشكل ٣):



٢٢٠

شكل ٣ :

ترتيب الاسم
مع التشكيل
الخاصة به في
قاعدة البيانات

١. المعالجة القبلية Preprocessing

يتم في هذه المرحلة التأكد من أن الاسم سليم من الأخطاء في التشكيل وفي تركيب الاسم، كما يتم إزالة المسافات الزائدة والكثيثة " (تطويل الحرف). في حالة وجود أخطاء فإنه يتم إيقاف عمل البرنامج بعد هذه المرحلة مباشرة. أما في حالة عدم وجود أخطاء فإنه يتم إدخال الاسم مع التشكيل في تراكيب البيانات.

٢. معالجة الأسماء الخاصة

في هذه المرحلة يتم معالجة بعض الأسماء الخاصة التي لا تتبع قواعد الرومنة. حيث أن البرنامج يقوم بقراءة الملف Exceptions.txt فإذا كان الاسم موجوداً في هذا الملف فإنه يقوم بتحويله حتى تتم الرومنة بشكل صحيح. ومن الأمثلة لهذه الأسماء اسم "طه" حيث يقوم البرنامج بتحويله إلى "طَاهَا".

٣. تحليل الحالات الخاصة

يتم في هذه المرحلة تحويل الحالات الخاصة مثل "آل" و "آل" المستخدمة في تعريف الأسماء إلى مقابلاتها وفقاً لقواعد الرومنة.

٤. تحليل حالات الشدة

يتم في هذه المرحلة استبدال الشدة الموجودة في الاسم بحالاتها المختلفة إلى مكافئاتها في الاسم العربي. التغيير في هذه المرحلة يؤثر على الاسم العربي فقط ويظهر أثر هذا التعديل في عملية الرومنة. يتم التعامل مع الشدة بناء على الحرف الذي تقع عليه، ويمكن تقسيم حالات الحروف مع الشدة إلى قسمين رئيسيين:

القسم الأول: حروف لا يمكن أن تشكل بالشدة، وهي: 'أ'، 'إ'، 'ي'
القسم الثاني: حروف يمكن أن تشكل بالشدة وتتنقسم إلى مجموعتين:

أ. الحروف: 'ب'، 'ت'، 'ج'، 'ح'، 'د'، 'ر'، 'ز'، 'س'، 'ص'، 'ط'، 'ف'، 'ق'، 'ك'، 'ل'، 'م'، 'ن'، 'ه'، 'و'، 'ي'
وفيها يتم تكرار الحرف مع إزالة الشدة، وتكون الحركة المرافقة للشدة على الحرف الثاني.

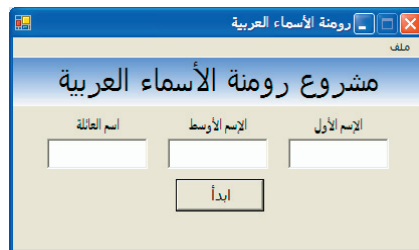
ب. الحروف: 'ث'، 'خ'، 'ذ'، 'ش'، 'ظ'، 'ض'، 'ع'، 'أ'، 'ؤ'، 'ئ'، 'ء'، 'إ'، 'ع'
وفيها يتم إزالة الشدة وتكون الحركة المرافقة للشدة على الحرف نفسه.

٥. رومنة الحروف

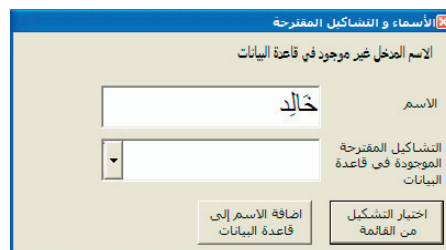
يتم في هذه المرحلة رومنة الحروف. حيث يرومن كل حرف بناء على الحركة المرافقة له، ويتم التعامل مع الاسم في هذه المرحلة على أنه مجموعة من الحروف المتتابعة نتجت عن عمليات المعالجة السابقة. بعد ذلك يتم استرجاع الاسم من تراكيب البيانات وتحويله إلى نص من جديد، ويتم طباعة الاسم بعد تجهيزه لعملية الطباعة بالتنسيق المناسب. الشكل ٤ يوضح محتويات نظام الرومنة وكيفية ترابط أجزائه مع بعضها البعض وكيفية عملها.

البرنامج يحتوي على الواجهة الأساسية في الشكل ٥ ويحتوي على ثلاث مواضع لكتابة الاسم. موضع للاسم الأول وموضع للاسم الأوسط وموضع للاسم العائلة. والسبب في فصل الأسماء عن بعضها حتى يتم التفرقة بين الأسماء المنفصلة والأسماء المركبة. بعد تعبئة هذه الأماكن، يستطيع المستخدم الضغط على زر "ابدأ" حتى يقوم البرنامج برومنة الأسماء. إذا كان الاسم المراد رومنته غير موجود في قاعدة البيانات، فإن البرنامج يقوم بعرض النافذة الموجودة في الشكل ٦. حيث يقوم البرنامج بعرض مجموعة من الاقتراحات لتصحيح الاسم المدخل. إذا كانت الخيارات غير مناسبة، فإن للمستخدم الخيار في إضافة الاسم إلى قاعدة البيانات.

شكل ٥ :
الواجهة
الأساسية
لبرنامج ويندوز



٢٢٢

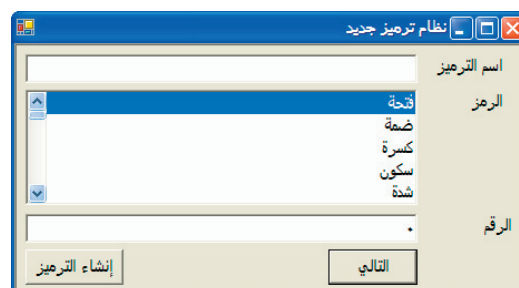


شكل ٦ : إضافة
الاسم على
قاعدة البيانات



الاسم	الرومنة
إبّداء	Ibtidaa
إبّسام	Ibtisam
إبّكار	Ibtikar
إبّناغو	Ibtanaghu
إبّهاج	Ibthaj
إبّثال	Ibthah
إبّتوري	Obturiy
إبّتيل	Ibtiyl
إبّراhib	Ibrahib

شكل ٧ : رومنة
جميع الأسماء

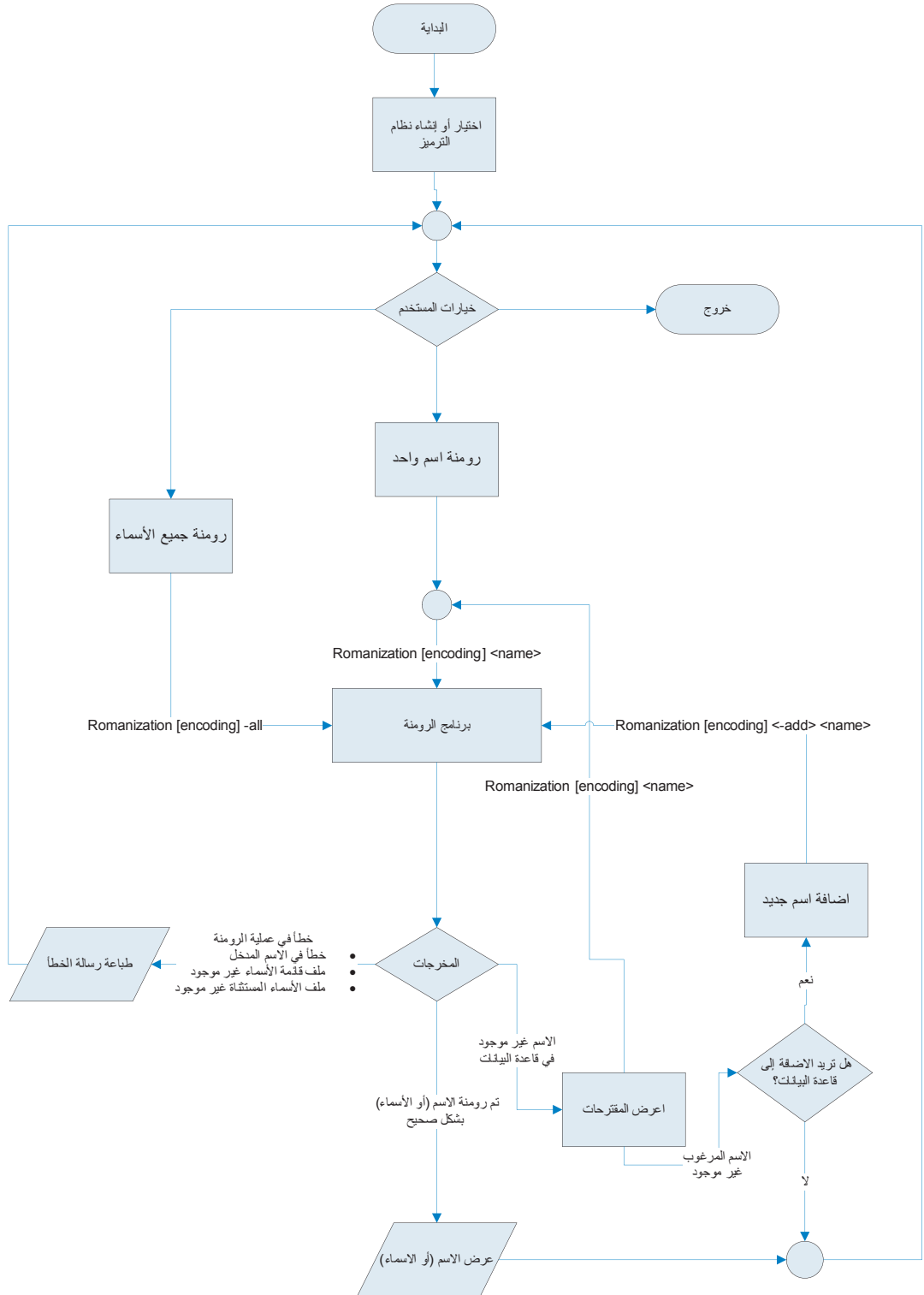


شكل ٨ : إنشاء
نظام ترميز
جديد

بالنسبة لنافذة رومنة جميع الأسماء فإنه يتم عرض جميع الأسماء في قائمة، وبجانب كل اسم، توجد الرومنة لذلك الاسم. وفوق هذه القائمة يمكن أن تجد عدد الأسماء في قاعدة البيانات التي تم رومنتها. كما يمكن حفظ هذه الأسماء على ملف.

أما نافذة إنشاء نظام ترميز جديد، فإنها تتيح للمستخدم إنشاء ملف ترميز جديد حتى يتم قراءة الحروف العربية بشكل صحيح. وبعد إنشاء النظام سيتم حفظه في مجلد ملفات الترميز الأخرى.

الشكل ٩ يبين محتويات واجهة نظام ويندوز وكيفية ترابطها وعملها.



شكل ٩ :
تركيب برنامج
واجهة ويندوز

الخاتمة

في هذا البحث تم اقتراح أنموذج لرومنة الأسماء العربية يتوافق مع متطلبات معظم الأنظمة والأجهزة الحاسوبية وملحقاتها، كما يمكن التعامل معه بسهولة من قبل الأفراد العاديين. بناءً على هذا الأنموذج تم تصميم وتنفيذ برنامج حاسوبي يقوم بعملية الرومنة بشكل آلي. ولاختبار هذا النظام فقد تم إدخال ما يربو على ٧٠,٠٠٠ اسم عربي مضبوط بالشكل. من الأعمال المستقبلية لهذا المشروع هو عمل واجهات مناسبة لتنظيم تشغيل أخرى مثل يونكس ولينكس وماكنتوش و آي بي إم وغيرها. نأمل أن نكون قد وفقنا لإصدار نظام فعّال تستفيد منه كافة القطاعات ذات العلاقة سواء داخل المملكة العربية السعودية أو في الدول العربية الأخرى.

المراجع

- الشمسان، أبو أوس إبراهيم. أسماء الناس في المملكة العربية السعودية. مجلة جامعة الملك سعود: الآداب. ٩: ٢٣١-٢٤٥. ١٤١٧هـ.
- الشمسان، أبو أوس إبراهيم. تباين كتابة الأسماء العربية في الحروف والتشكيل: صورته وأسبابه. الندوة العلمية الأولى حول توحيد معايير النقل الكتابي لأسماء الأعلام العربية: الأبعاد الأمنية. ٩-٤٥. ١٤٢٤هـ.
- معجم أسماء العرب، موسوعة السلطان قابوس لأسماء العرب، ط١ (مسقط: جامعة السلطان قابوس، ١٩٩١م).
- صالح، محمود بن إسماعيل. رومنة الأسماء العربية في المراجع الأكاديمية. الندوة العلمية الأولى حول توحيد معايير النقل الكتابي لأسماء الأعلام العربية: الأبعاد الأمنية. ١٧٥-١٩٥. ١٤٢٤هـ.
- المهنا، فيصل بن محمد. من الأبجدية العربية إلى الأبجدية الرومانية: تناظر صوتي. الندوة العلمية الأولى حول توحيد معايير النقل الكتابي لأسماء الأعلام العربية: الأبعاد الأمنية. ٢٦٥-٢٩٩. ١٤٢٤هـ.
- العاني، دحام إسماعيل وفايز بن عبد الله الحرقان ومنصور بن محمد الغامدي ومحمد بن إبراهيم الكنهل. آلية لتوظيف الشبكة العالمية (الإنترنت) في رصد المصطلح العلمي وتعريبه وضبطه ونشره. المؤتمر الثالث لمجمع اللغة العربية. دمشق. ٢٥-٢٨ شعبان ١٤٢٥هـ.
- الغامدي، منصور بن محمد. خوارزميات لرومنة الأسماء العربية. ١٤٢٥. مجلة جامعة الملك سعود: علوم الحاسب والمعلومات. المجلد السابع عشر. ١-٢٧.
- الغامدي، منصور بن محمد. ألفبائية صوتية دولية تقوم على الحرف العربي، الندوة الثالثة لأفاق البحث العلمي والتطوير التكنولوجي في العالم العربي. الرياض. ١٤٢٥.
- الغامدي، منصور بن محمد. الألفبائية الصوتية الدولية والحرف الروماني، الندوة العالمية الأولى حول: توحيد معايير النقل الكتابي لأسماء الأعلام العربية: الأبعاد الأمنية. الرياض. ١٤٢٤. ٢٤٣ - ٢٦٤.
- توصيات الندوة العلمية الأولى حول توحيد معايير النقل الكتابي لأسماء الأعلام العربية: الأبعاد الأمنية. أكاديمية نايف العربية للعلوم الأمنية. الرياض. ١٤٢٤هـ. ٣٨٣.
- Lawrence, T. E. Seven pillars of wisdom. Doubleday. Reprinted in July ١٩٩١.
- القتم، عبد الله. الأنماط المتبعة في كتابة الأسماء الشخصية. توحيد معايير النقل الكتابي لأسماء الأعلام العربية: الأبعاد الأمنية. أكاديمية نايف العربية للعلوم الأمنية. الرياض. ٩٧-١١٦. ١٤٢٤هـ.
- القتم، عبد الله. كتابة الأسماء العربية بالحروف اللاتينية. مطبوعات جامعة الكويت. الكويت. ١٩٩٥م.
- صالح، محمود بن إسماعيل. رومنة الأسماء العربية في المراجع الأكاديمية. توحيد معايير النقل الكتابي بأسماء الأعلام العربية: الأبعاد الأمنية. ١٧٥-١٩٥. ١٤٢٤هـ.
- Toivonen, Jarmo, Ari Pirkola, Heikki Keskustalo, Kari Visala, and Kalervo Jarvelin. Translating cross-lingual spelling variants using transformation rules. Information Processing and Management. in press. 2004.
- Virga, Paola and Sanjeev Khudanpur. Transliteration of Proper Names in Cross-Language Applications. In Proceedings of SIGIR 2003. Toronto, Canada.
- Meng, H.M.; Wai-Kit Lo; Berlin Chen; Tang, K. Automatic Speech Recognition and Understanding. ASRU '01. IEEE Workshop on, 9-13 Dec. 2001, 311 - 314, 2001.
- Knight, k. and J. Graehl. Machine transliteration. In Proceedings of ACL, 1997.
- Stalls, B. and K. Knight. Translating names and technical terms in Arabic text. In Proceedings of the COLING/ACL Workshop on Computational Approaches to Semitic Languages. 1998.
- Abduljaleel, Nasreen. Statistical transliteration for english-arabic cross language information retrieval. In the proceedings of TICIKM. New Orleans, USA. 2003.
- المحتسب، حسني عبد الغني. نظام آلي لكتابة الأسماء العربية بحروف إنجليزية: نحو معايير موحدة. توحيد معايير النقل الكتابي بأسماء الأعلام العربية: 175. 195. 1424هـ- الأبعاد الأمنية.
- Sinha, R.M.K. Dealing with unknown in machine translation. IEEE International Conference on Systems, Man and Cybernetics. Volume: 2. 940 - 944 vol.2. 2001.
- Jung, S. Y., S. Hung and E. Paek. An English to Korean transliteration model of extended Markov window. In proceedings of COLING. Saarbrucken, Germany. 2000.
- Arbabi, M., S. M. Fischthal, V. C. Cheng and E. Bart. Algorithms for Arabic names transliteration. IBM Journal of Research and Development. 38: 183-193. 1994.
- UNEGN Working Group on Romanization Systems. REPORT ON THE CURRENT STATUS OF UNITED NATIONS ROMANIZATION SYSTEMS FOR GEOGRAPHICAL NAMES. Arabic Version 2.2. January 2003

ملحق أ أنموذج رومنة أسماء الأعلام العربية

يخصص في الأوراق الرسمية مساحة خاصة لكل من: الاسم الأول، واسم الأب، واسم الجد، واللقب. ولا تستخدم كلمات الفصل بين الأسماء: "ابن، بنت، ولد". مثال:

الاسم	الأب	الجد	اللقب
سالم	علي	شمس الدين	بلال

First Name	Father	G. Father	Last Name
Salim	Ali	Shamsuldin	Bilal

يعامل الاسم المركب والكنية معاملة الاسم المفرد، وذلك بدمج الكلمات المكونة له في كلمة واحدة مرومنة. ويراعى عند إدخال الاسم المركب بالحرف العربي أن تكون هناك مسافة بين الكلمات المكونة له لتوظيفها في برمجات الرومنة الآلية. ويشكل الحرف الأخير في كلمات الاسم المركب عدا الأخيرة.

٢٢٥

مثال: أبو قيس ← Abuqays
شرف الدين ← Sharafuldin

ملحوظة: تحذف الألف من ال التعريف في وسط الأسماء المركبة (أنظر الجدول: ج).
تعامل ال التعريف حسب كتابتها وليس حسب نطقها، ففي حالة اللام الشمسية تهمل الشدة الواقعة على الحرف الذي يلي ال التعريف والنتيجة عن إدغام اللام الشمسية في الحرف الذي يليها، حيث أن الإدغام - في جميع حالاته - لم يؤخذ بعين الاعتبار في النقل الكتابي وقد أبقى على لام ال التعريف كما هي دون إدغام:

مثال: السَّالِم ← Alsalim
والمُسْلِم ← Almuslim

تعامل آل بنفس طريقة ال التعريف:

مثال: آل سالم و السَّالِم ← Alsalim

الحروف المشددة إذا كان:

تناظرها: 1 مقابل 1 (انظر الجدول: أ)، فتكرر:

مثال: بَسَّام ← Bassam

تناظرها: 1 مقابل 2 (أنظر الجدول: ب)، فتبقى دون تكرار:

مثال: بَشَّار ← Bashar

تناظرها: 1 مقابل 0 (أنظر الجدول: ج)، فلا تكرر حيث أنها لا تكتب.

مثال: دَعَّاس ← Daas

ليس هناك فرق بين الحروف الرومانية الصغيرة والكبيرة في تمثيل الأسماء العربية، إلا أنه إتباعا لما هو متعارف عليه في كتابة أسماء الأعلام بالحرف الروماني فإن جميع حروف الاسم تكون صغيرة ما عدا الحرف الأول فيكون كبيرا.

التسلسل	الحرف		أمثلة بالحرف	
	العربي	الروماني	العربي	الروماني
1	ء ¹	a	رَدَّء	Rida
2	إ ²	e	إِبْدَاء	Eida
3	آ	a	آدَم	Adam
4	ب	b	بَدْر	Badr
5	ت	t	تَمَّام	Tammam
6	ج	J	جَابِر	Jabir
7	ح	h	حَامِد	Hamid
8	د	d	دَحَّام	Dahham
9	ر	R	رِيَّاض	Riyadh
10	ز	z	زَيْد	Zayd
11	س	S	سَعِيد	Said

التسلسل	الحرف		أمثلة بالحرف	
	العربي	الروماني	العربي	الروماني
13	ط	T	طَارِق	Tariq
14	ع ^١	A	دَرَع	Dira
15	ع ^٢	E	عِيد	Eid
16	ع ^٣	O	عُود	Oud
17	ف	F	فَالِح	Falih
18	ق	Q	قُنْبُر	Qunbur
19	ك	k	كَلَيْب	Kulayb
20	ل	L	لُجَيْن	Lujayn
21	م	M	مُنَى	Muna
22	ن	N	نَعْمَة	Nimah
23	هـ	H	هَاشِم	Hashim
24	هـ ^٤	H	هَيْة	Hibah
25	هـ ^٥	T	هَيْة الله	Hibatullah
27	و ^٦	W	وَلِيد	Walid
28	ي ^٥	Y	يَحْيَى	Yahya
29	الفتحة	a	سَعْد	Sad
30	الضمة	u	زُهَيْر	Zuhayr
31	الضمة ^٧	o	عُمَر	Omar
32	الكسرة	i	زَيَاد	Ziyad

٢٢٦

جدول أ :
تناظرا مقابل ١

١. في نهاية الكلمة ومسبوق بحرف ساكن.
٢. في بداية الكلمة ومتبوع بياء المد.
٣. في بداية الكلمة ومتبوع بواو المد.
٤. عندما تكون في نهاية الاسم.
٥. عندما تكون في نهاية كلمة متبوعة بكلمة أخرى في اسم مركب.
٦. حرف صامت (ليس من حروف المد، أنظر الجدول: ج).
٧. عندما تأتي الضمة في بداية الاسم، إذ أنه عند حذف الهمزة أو العين المضمومة من بداية الاسم تصبح الضمة هي الحرف الروماني الأول وعليه تحول من u إلى o. وكذلك عند تنالي ضمتين نتجتا عن حذف همزة أو عين مضمومة ومسبوقه بضمة فإن الضمة الأولى تتحول من u إلى o.

	الحرف		أمثلة بالحرف	
	العربي	الروماني	العربي	الروماني
1	ث	th	ثَابِت	Thabit
2	خ	kh	خَالِد	Khalid
3	ذ	th	ذَعَار	Thiar
4	ش	sh	شَمْس	Shams
5	ض	dh	ضَيْف الله	Dhayfullah
6	ظ	th	ظَلِيَّان	Thabyan
7	غ	gh	غَالِب	Ghalib

جدول ب :
تناظرا مقابل ٢

التسلسل	الحرف		أمثلة بالحرف	
	العربي	الروماني	العربي	الروماني
1	ا	∅	سَاهِر	Sahir
2	أ	∅	أَحْمَد	Ahmad
3	إ	∅	إِبْرَاهِيم	Ibrahim
4	ؤ	∅	لَوْلُؤَة	Luluah
5	ئ	∅	رَائِد	Raid
6	وا	∅	لُهَوف	Lahuf
7	يا	∅	أَنْبِيس	Anis
8	ء ²	∅	أَسْمَاء	Asma
9	ع ²	∅	مَعْن	Man
10	ى (الألف المقصورة)	∅	لَيْلَى	Layla

١. حرف مد ، أو ألف ال التعريف في وسط الأسماء المركبة.
٢. متحرك ليس في بداية الكلمة أو ساكن مسبق بحرف متحرك أو بحرف مد.

ملحق ب الأخطاء في الأسماء والتي يقوم النظام بالتدقيق فيها

المسببات	رسالة الخطأ
إذا كان لا يوجد على السطر سوى الاسم الأساس فقط دون أية تشاكيل أخرى للاسم.	لا يوجد تشاكيل للاسم
إذا كان عدد الحروف في الاسم المشكل بعد إزالة حروف التشكيل لا يساوي عدد الحروف في الاسم الأساس.	عدد الأحرف في الاسم المشكل غير متساوي مع أصل الاسم
إذا كان حرف الألف من "ال" التعريف ... التشكيل.	حرف الألف جزء من "ال" التعريف ... التشكيل على الحرف غير صحيح
إذا كان حرف الألف مسبقاً بفتحة فإنه لا يمكن تشكيله.	التشكيل على الحرف غير صحيح لأن حرف الألف يجب أن يسبق بفتحة
لا يمكن تشكيل حرف الألف إذا كان في وسط أو في آخر الكلمة.	حرف الألف لا يمكن تشكيله
حرف المد "ا" لا يقبل التشكيل.	حرف المد "ا" لا يمكن أن يشكل
يجب أن يكون حرف التشكيل مسبقاً بحرف من الحروف العربية، وإلا فإنه يكون غير صحيح وفقاً لقواعد الكتابة باللغة العربية.	يجب أن تكون حركات التشكيل مسبوقة بحروف
لا يقبل الحرف العربي أن يكون مشكلاً بأكثر من حركة من حركات التشكيل إلا في حالة الشدة مع حركة أخرى.	لا يمكن أن تكون هناك حركتان متتاليتان إلا في حالة الشدة
آخر الاسم لا يقبل التشكيل إلا بالشدة فقط أي حركة أخرى تعد غير متوافقة مع قواعد التشكيل العربي.	من غير الممكن أن تكون هناك حركة غير الشدة فقط على آخر الكلمة
الشدة في منتصف الاسم لا يمكن أن تأتي منفردة دون أن تكون متبوعة بحركة أخرى.	يجب أن تكون الشدة متبوعة بفتحة أو ضمة أو كسرة فقط، إذا كانت في منتصف الكلمة
الاسم العربي يتكون من حروف متتابعة يفصل بينها فراغ واحد فقط في حالة الاسم المركب.	لا يمكن استخدام غير الحروف والحركات في كتابة الاسم العربي
جميع الحروف العربية يجب تشكيلها ماعدا: الألف غير المهموزة، الألف الممدودة، الواو في وسط أو آخر الاسم، الألف المقصورة، الياء في وسط أو آخر الكلمة، اللام في ال التعريف.	يجب تشكيل الحرف ... في الاسم

هناك خطأ واحد يتم تصحيحه بشكل آلي عند وجوده، ألا وهو لفظ الجلالة "الله". فعند ورود اللام الثانية بدون حركة، فإن البرنامج يقوم بوضع الشدة والفتحة عليها بحيث تصبح "الله" ومن ثم وضعها في قاعدة البيانات.