لغات الشعوب الأصلية:

من الصفر إلى الرقمنة

الدليل لتضع لغتك على الإنترنت

TRANSLATION COMMONS

2022-2032 | INTERNATIONAL DECADE OF
Indigenous Languages

Translation Commons © 2023
هذا العمل مرخص بموجب الرخصة الدولية لكامنز الإبداعية إسناد 4.0

1. المقدمة 5
1.1 سلسلة من الصفر إلى الرقمنة 5
2. استعراض العملية 6
2.1. مسار العمل للوضع الحالي للغة. 6
الشكل 2: تحديد الوضع الحالي للغة 7
2.2. مسار عمل تطبيق التكنولوجيا 8
3. الوضع الحالي للغة 8
3.1. هل يستخدم المجتمع اللغة حاليًا؟ 8
3.2. هل اللغة مُعَدّة للاستخدام النشط من جانب المجتمع؟ 8
3.2.1. أعِد إحياء اللغة 8
3.3. هل اللغة في السجل العمومي؟ 10
3.4. هل اللغة مكتوبة؟ 10
3.4.1. طَوِّر الشكل الكتابي 10
3.4.2. وَثِّق اللغة 10
3.4.2.1. اللغة موثقة 10
3.4.2.2. اللغة غير موثقة 11
3.5. هل تستخدم اللغة نظام كتابة مُتسِق؟ 11
3.5.1. هل الحروف المستخدمة في كتابة اللغة مدعومة؟ 11
3.6. هل الكتابة مدعومة بمعيار قياسي ؟ 12
3.6.1. قدم اقتراحات حول الحروف 12
3.6.2. طَوّر المعيار القياسي 12
3.7. الشروع في التطبيق 12
4. سير عمل تطبيق تكنولوجيا اللغة 13
4.1 ملاحظة عن تكنولوجيا النص في النُظم الرقمية 13
4.2. تعريفات لتطبيق الدعم الرقمي 14
4.3. هل الرمز المعياري للغة متاح؟ 14
4.3.1. قدّم طلبًا للحصول على رمز اللغة 15
4.4. نعم: هل الترميز الموحد للخط متاحٌ؟ 15
4.4.1. إنشاء الخط 15
4.5. هل الخط متوفر على الأجهزة؟ 15
4.5.1. تثبيت يدوي أو اطلب من البائعين إضافة الدعم 16
4.6. هل يحتوي الجهاز على دعم الإدخال؟ 16

4.7. هل الإدخال مدعوم بتطبيقات تابعة لطرف ثالث؟ 16
4.7.1. طوّر طريقة للإدخال. 17
4.8. هل يحتوي الجهاز على دعم بيانات الترميز الموحد؟ 17
5. دعم إضافي للّغة 17
5.1. مصادر عامة للّغة 19
5.1.1. مصادر اللغة 19
5.1.2. مصادر الأداة 19
5.2. تكنولوجيا اللغة المتقدمة 19
6. مسرد 20
7. المراجع 21
7.1. إحياء اللغة 21
7.2. سجلات اللّغة 21
7.3. الترميز الموحد وترميز الخط 21
7.4. رموز اللغة 22
7.5. الخطوط 22
8. ملاحظات 22

لغات الشعوب الأصلية: من الصفر إلى الرقمنة.

المؤلفون: ديبورا دبليو أندرسون، لي كولنز، كريغ كورنيليوس وكريغ كامنجز
المراجعون والمساهمون: أندرو أوين، جوليا ني، لورنس وولف سونكين، آنا لويزا دايغنولت، جولي أندرسون، دانييل بورج أوديل.
التصميم والتسويق: ميتي عطار وجوانا بيم.
تنسيق المشروع: إيستر بيريز وجانيت ستيوارت

هذا العمل مرخص بموجب الرخصة الدولية لكامنز الإبداعية إسناد 4.0 Translation Commons © 2023

1. المقدمة

ترانسليشن كومنز (Translation Commons) هو مجتمع تطوعي غير ربحي يقدم الدعم لرقمنة اللغات ويوجه متخصصي اللغات ويوفر دورات وموارد للصناعات اللغوية.

أحد البرامج الأساسية ترانسليشن كومنز هو مبادرة رقمنة اللغة (LDI)، التي تسعى إلى توفير القدرات الرقمية للمجتمعات اللغوية التي ترغب بها. يوجد ما يقارب 6,000 لغة حول العالم بحضور رقمي قليل أو بدون حضور رقمي على الإطلاق. توفر مبادرة رقمنة اللغة خارطة طريق يمكن للمجتمع اتباعها لتحقيق رقمنة لغتهم.

عقدت ترانسليشن كومنز شراكة مع مبادرة السنة الدولية للغات الشعوب الأصلية عام 2019 وهي مبادرة قام بها اليونسكو لتركيز المزيد من الاهتمام على مجتمعات السكان الأصليين ورقمنة لغاتهم. يُعَدُّ دعم الوصول الرقمي إلى لغات الشعوب الأصلية ولغات الأقليات الأخرى جزءًا من مبادرة رقمنة اللغة لضمان مشاركة هذه المجتمعات في الأنشطة العالمية عبر الإنترنت والحصول على مزايا تطبيقات الكمبيوتر الحديثة بلغتهم الأم. إنَّ وضع التوجيهات وتزويد المجتمعات بالأدوات اللازمة لرقمنة مخطوطاتهم يُسَهِّلُ عملية تدوين لغاتهم في الإنترنت مع الحفاظ على استقلاليتهم. تقدم ترانسليشن كومنز دروسًا تعليمية وورش عمل بالإضافة إلى الإرشادات والتوجيهات كما تساعد المجتمعات في رقمنة اللغة من خلال تعريفهم بالخبراء المختصين الذين يوجهونهم خلال عملية توحيد المعايير.

1.1 سلسلة من الصفر إلى الرقمنة

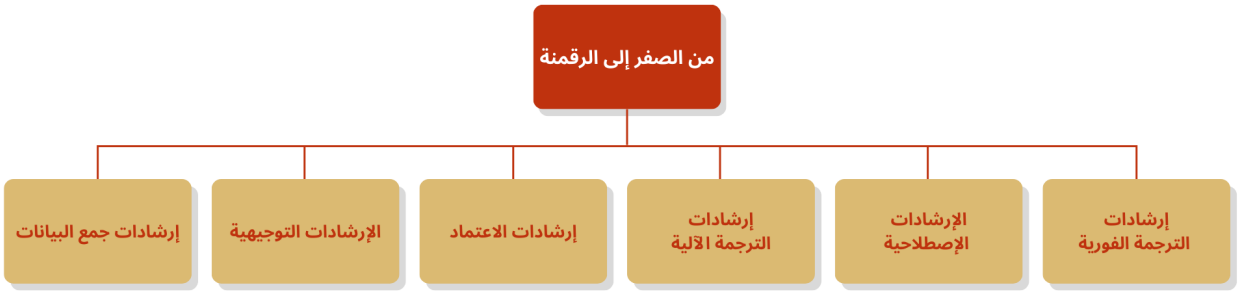
هذه الوثيقة هي واحدة من سلسلة من التوجيهات بعنوان من الصفر إلى الرقمنة، والتي تعالج كليًا تطبيقات رقمنة اللغة بصفة شاملة. إنّ مؤلفي هذه الإرشادات خبراء في تكنولوجيا اللغة واللغويات والجمهور المستهدف هو أي مجتمع لغوي يبحث عن كيفية استخدام لغته في الأنظمة الرقمية.

تعمل الرقمنة على نشر سبل التواصل في المجتمع اللغوي. راجع ملحق فوائد رقمنة اللغة للحصول على المزيد من التفاصيل حول فوائد رقمنة اللغة لمجتمعات السكان الأصليين وللعالم بأسره.

لمعرفة المزيد عن عملية رقمنة اللغة، انظر صفحة الويب الخاصة بمصادر ترانسليشن كومنز التي توفر معلومات إضافية عن مبادرة رقمنة اللغة والتي تتضمن الإرشادات والعروض التقديمية والفيديوهات وغيرها من الوثائق.

ستجد في الشكل 1 أدناه كل الإرشادات التي وُضعت لمساعدة مجتمعات السكان الأصليين.

الشكل 1: سلسلة من الصفر إلى الرقمنة



Translation Commons © 2023 هذا العمل مرخص بموجب الرخصة الدولية لكامنز الإبداعية إسناد 4.0

تشرح هذه الوثيقة بالتحديد كيفية تمكين برامج الهواتف المحمولة والكمبيوتر من دعم لغة مكتوبة. يسمح التطبيق الموصى به المتحدثين الأصليين للغة بالتواصل عبر الإنترنت ومشاركة المعارف والمستندات واستخدام البرامج والأجهزة التي قد يتعذر عليهم الوصول إليها.

الجماهير التي تستهدفها هذه الوثيقة هي:

- مجتمعات السكان الأصليين الذين يرغبون في جعل لغاتهم متاحة على الأجهزة المحمولة وأجهزة الكمبيوتر
- خبراء التكنولوجيا الذين يدعمون رقمنة لغة واحدة أو أكثر
- المنظمات التي تريد تمكين المجتمعات اللغوية

تهدف هذه الوثيقة إلى مساعدتك في تحديد الأدوات التي تحتاجها وكيف تستخدمها. كما يمكنها مساعدتك في استكشاف الأدوات المتاحة لاستخدام لغتك على الإنترنت.

عندما يسأل الأشخاص عن كيفية استخدام لغتهم على الإنترنت، هناك الكثير من الإجابات المحتملة. هناك عدة مستويات للتكنولوجيا التي يجب أن تؤخذ بعين الاعتبار تُسهل في استخدام اللغة عبر الإنترنت، وذلك في كل من التكنولوجيا المعتمدة على الإنترنت والتكنولوجيا المحمولة.

تتناول هذه الوثيقة استخدام المتحدثين والقراء والمبدعين للغات المكتوبة على الإنترنت. يتضمن ذلك المحادثات العادية والرسائل النصية والبريد الإلكتروني ووسائل التواصل الاجتماعي والمدونّات. والهدف هو مساعدة الأشخاص على تطوير المواقع الإلكترونية وتنويع المحتوى، مما يعزز التواصل مع المجتمعات المحلية ولغات الشتات في العالم. قد يوجد عدة أبجديات أو اصطلاحات للكتابة للغة مُعينة. تهدف هذه الوثيقة إلى تمكين الأشخاص من استخدام التقنيات المشروحة هنا بطرق تعزز من مكانة لغات الشعوب الأصلية والمفهوم العام لها وإمكانية استخدامها عالميًا، وليس الهدف منها شرح كيفية استخدام التقنيات الرقمية. في حين تُعَدّ طرق التوثيق الرسمية وقواعد النحو والقواميس مفيدة في الدراسات اللغوية والتوحيد القياسية إلا إنها ليست من متطلبات استخدام لغتك على الإنترنت.

تتألف الشراكة بين أصحاب المصلحة المتعددين من لجنة توجيهية للإشراف على التنفيذ، ومجموعات مختصة لتقديم المشورة الملائمة، بالإضافة إلى الشركاء المساهمين.

2. استعراض العملية

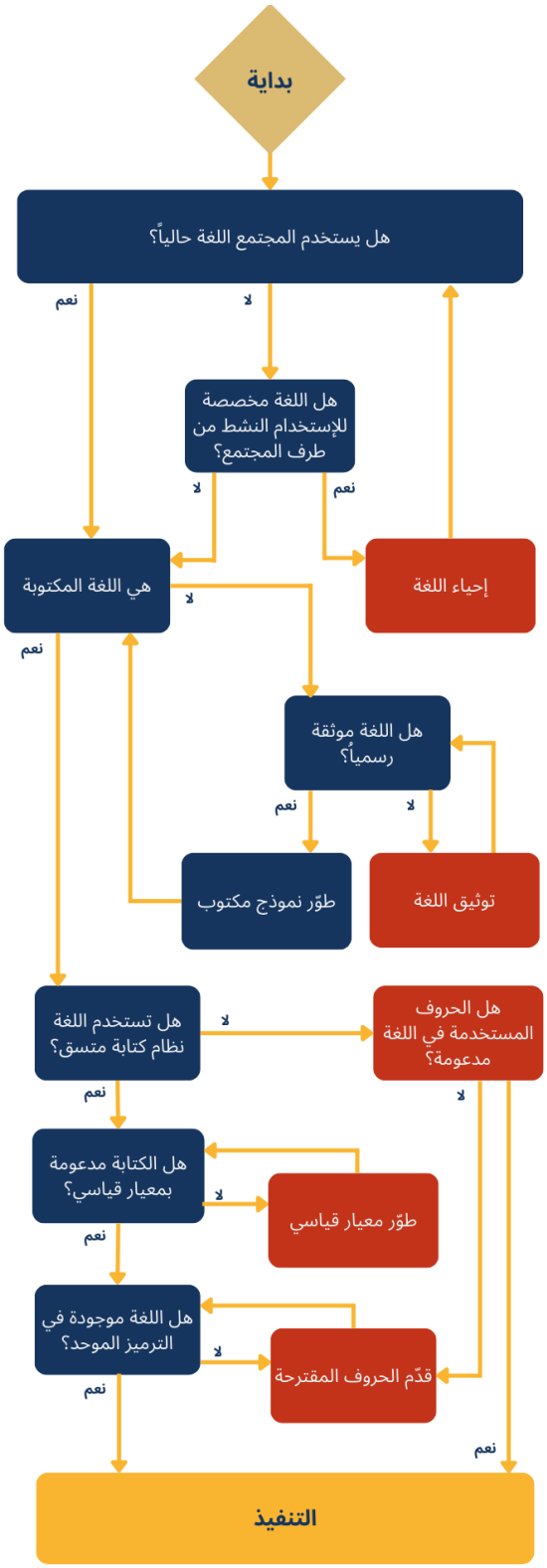
تقدم هذه الوثيقة طريقتين لسير العمل. تستخدم الطريقة الأولى لتحديد الوضع الحالي للّغة. بينما تستخدم الطريقة الثانية لتطوير حلول تكنولوجية لإتاحة الاستخدام الرقمي للّغة. تشير الخطوات في المخططات الانسيابية إلى فصول في الوثيقة حيث يوجد المزيد من المعلومات المفصلة. الخطوات استشارية فقط وبعضها يمكن تنفيذه في الوقت نفسه.

2.1. مسار العمل للوضع الحالي للغة.

يصف المسار (الشكل 2) الخطوات اللازمة لتحديد الوضع الحالي للّغة لإعدادها للاستخدام عبر الإنترنت.

Translation Commons © 2023 هذا العمل مرخص بموجب الرخصة الدولية لكامنز الإبداعية إسناد 4.0

الشكل 2: تحديد الوضع الحالي للغة

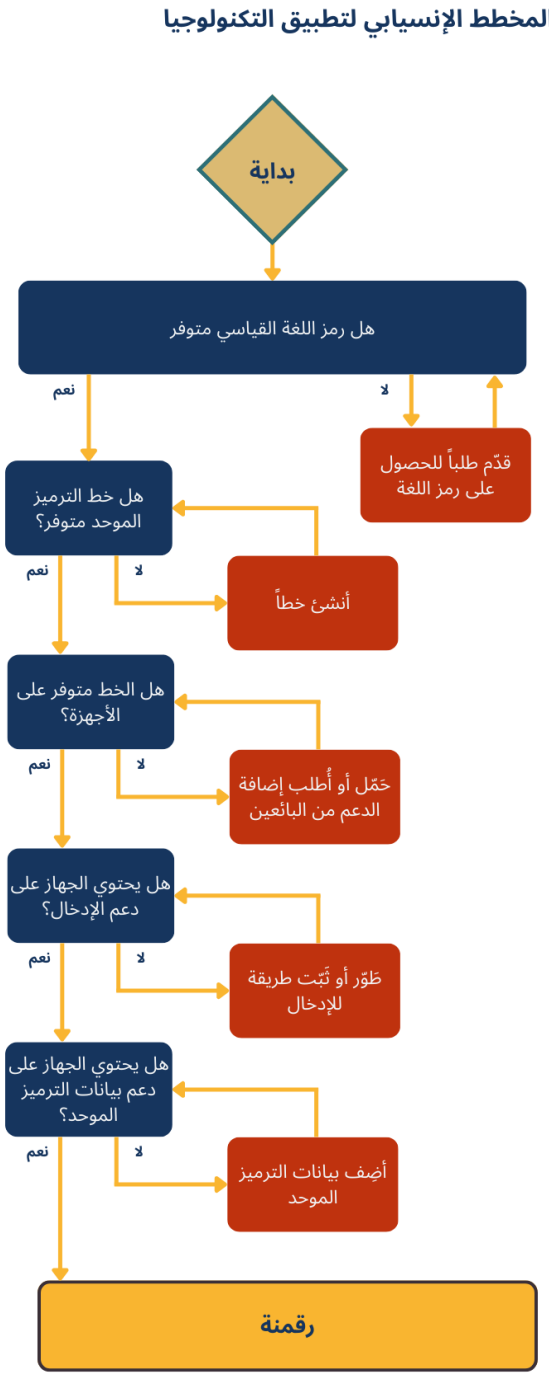


Translation Commons © 2023 هذا العمل مرخص بموجب الرخصة الدولية لكامنز الإبداعية إسناد 4.0

2.2. مسار عمل تطبيق التكنولوجيا

بعد تحديد حالة لغتك، استخدم مسار العمل هذا (الشكل 3) للبدء في تدوين لغتك على الإنترنت باستخدام التقنيات المتاحة.

الشكل 3: تطبيق التكنولوجيا



هذا العمل مرخص بموجب الرخصة الدولية

3. الوضع الحالي للغة

يصف هذا الجزء الخطوات في مسار عمل الوضع الحالي للغة. والهدف هو تحديد المستوى الحالي لدعم لغتك على الكمبيوتر والأجهزة المحمولة. كما يتضمن خطوات تالية مقترحة لمساعدتك على تدوين لغتك على الإنترنت. لا تحتاج لخلفية تقنية للإجابة على تلك الأسئلة.

3.1. هل يستخدم المجتمع اللغة حاليًا؟

نعم: هل اللغة مكتوبة؟
لا: هل اللغة مُعدّة للاستخدام النشط من جانب المجتمع؟

3.2. هل اللغة مُعَدّة للاستخدام النشط من جانب المجتمع؟

نعم: إحياء اللغة
لا: هل اللغة مكتوبة؟

3.2.1. أعِد إحياء اللغة

من الممكن إحياء اللغات، حتى التي ليس لها متحدثون حاليون من خلال التزام المجتمع وتنظيمه. يتضمن هذا الكثير من مصادر توثيق اللغة، كما يحتاج إلى مواد تعليمية ومعلمين لاستخدامها وجهد مكثف من المجتمع لمواصلة الجهود لعدة سنوات بل ربما عقود.

اتبعت نيوزيلندا منهجا لإحياء اللغة الماورية المعروفة أيضا باسم "تيريو" استند جزئيًا على منهج إحياء اللغة الويلزية في "ويلز" ووسائل الإعلام بلّغات السكان الأصليين والمدارس التأسيسية وما إلى ذلك. يُتّبع أيضا منهج مشابه مع اللغة الأيرلندية. تتضمن المشاريع الكثيرة حول العالم أمثلة أخرى منها:

- العبرية: مثال لإحياء لغة بعد عدة قرون من عدم التحدث بها.
- الشيروكي: تُكرس المجتمعات والحكومات جهودًا للمدارس الانغماسية أو مدارس التعليم المكثف، والتعليم المستمر وصولا إلى المدارس الابتدائية والثانوية وتدريب الكبار.

- موتسون: تطبق قبيلة "آماه موتسون" في "كاليفورنيا" برنامجًا لإعادة إحياء لغتهم.
- شاكما: لغة من لغات "بنغلاديش وشرق الهند". بدأ المجتمع في استخدام الأبجدية التاريخية في جهود التعليم ومحو الأمية.
- تونيكا: في عام 1948 أصبحت لغة الأجداد لقبيلة "تونيكا بيلوكسى" غير ناشطة ومنذ ذلك الحين أحيا المتحدثون عن التراث تلك اللغة عن طريق تدريب المتحدثين الجدد للتحدث بطلاقة وتشجيع آخرين للتسجيل في دورات انغماس اللغة.
- الكورنية: بدأ إحياء اللغة "الكورنية" أول مرة في أوائل القرن العشرين ولكن تسارعت وتيرة حركة تنمية اللغة في القرن الحادي والعشرين إذ استفاد متحدثوا الكورنية من المنتديات الموجودة على الإنترنت للعثور على بعضهم البعض واستخدام اللغة يوميا.

هذا العمل مرخص بموجب الرخصة الدولية لكامنز الإبداعية إسناد 4.0 Translation Commons © 2023

3.3. هل اللغة في السجل العمومي؟

إذا كانت اللغة في السجل العمومي أو قائمة اللغات.

نعم: هل اللغة مكتوبة؟
لا: هل الرمز اللغوي الموحد للّغة متوفر ؟

3.4. هل اللغة مكتوبة؟

هل لغتك لها شكل كتابي؟

نعم: وثّق اللغة
لا: طَوِّر الشكل الكتابي

3.4.1. طَوِّر الشكل الكتابي

اللغات غير المكتوبة خارج نطاق هذه الوثيقة. ومع ذلك مازال بإمكانك استخدام لغتك على الإنترنت عن طريق الموارد الصوتية وموارد الفيديو. توجد منظمات تقدم أساسًا الإرشاد والأدوات للّغات التي تعتمد على الشفوي. كما تُعدّ أقسام اللغويات الأكاديمية مصادر جيدة.

تكون المصادر الصوتية ومصادر الفيديو أكثر فائدة عندما يتمكن المستخدمون والباحثون من إيجادها. ينبغي أن تتضمن أي تسجيلات علامة أو رمزًا للّغة حتى تتمكن الفهرسة التلقائية من العثور على تلك البيانات. استخدم الرموز الموحدة للغة مثل BCP IETF-47.

3.4.2. وَثِّق اللغة

هل لغتك موثقة رسميًا، على سبيل المثال عن طريق القواعد النحوية أو القواميس أو الدراسات اللغوية؟

نعم: اللغة موثقة
لا: اللغة غير موثقة

3.4.2.1. اللغة موثقة

عندما توجد القواميس والقواعد النحوية ومعلومات لغوية أخرى، هل هي متاحة لمستخدمي اللغة؟ ابحث في الخيارات لتجعل هذه المعلومات متاحة ومفيدة للمجتمعات على نطاق واسع. قد يتضمن هذا إنشاء مصادر على الإنترنت ومواد تعليمية مناسبة للمدارس الانغماسية والتعليم الأساسي. والعمل على إسناد حقوق النشر لتلك المواد إلى المجتمعات اللغوية والحصول على أو إنشاء صيغة رقمية لتلك المعلومات بطريقة يمكن تحويلها إلى صيغة على الإنترنت حسب ما تراه المجتمعات مناسبًا.

عندما يتعذر الوصول لهذه المصادر، اتخذ خطوات لمشاركة الكتب والمستودعات الإلكترونية المتاحة عبر الإنترنت وغيرها من المعلومات عن اللّغة في المجتمعات التي تستخدمها.

Translation Commons © 2023 هذا العمل مرخص بموجب الرخصة الدولية لكامنز الإبداعية إسناد 4.0

3.4.2.2. اللغة غير موثقة

إذا لم توثق اللغة بعد، فقد يظل ممكنا استخدام اللغة شفويًا أو كتابة. إلا أن الدعم المتمثل في اقتراحات التدقيق الإملائي في والبحث والنص التنبؤي سيكون محدودًا.

3.5. هل تستخدم اللغة نظام كتابة مُتسِق؟

نظام الكتابة هو مجموعة من القواعد لاستخدام أبجدية أو أكثر لكتابة لغة معينة. تُكتب الكثير من اللغات بأكثر من أبجدية، على سبيل المثال، تكتب المجتمعات المختلفة اللغة "الصربية-الكرواتية" بالحروف السيريلية واللاتينية. وكذلك تُستخدم معظم نُظم الكتابة في أكثر من لغة واحدة. فمثلا، يمكن كتابة اللغة "البورمية والشان والمون" ولغات أخرى بالأبجدية الخاصة بـ"ميانمار".

هل تستخدم اللغة نظامًا واحدًا على الأقل للكتابة بأسلوب متسق متضمنا التهجئة؟

نعم: هل الكتابة مدعومة بمعيار؟
لا: هل الحروف المستخدمة في الكتابة مدعومة بالفعل؟

3.5.1. هل الحروف المستخدمة في كتابة اللغة مدعومة؟

إذا لم يكن هناك نظام متسق للكتابة، فقد تكون الكتابة غير رسمية ولا تستخدم تهجئة أو قواعد نحوية ثابتة (أو حتى تستخدم أكثر من أبجدية للكتابة). على الرغم من أنه يمكن إنشاء وعرض مثل هذا النص بالأدوات الموجودة، إلا أن أدوات التهجئة والنحو ستكون لها فائدة محدودة. في الحالات التي تتواجد فيها عدة نُظم من الكتابة المتضاربة، من المهم أن يدرك من هم خارج المجتمع اللغوي أن أساليب التهجئة تلك قد تمثل اهتمامات مختلفة في مجتمع اللغة.

قد تحتاج إلى المساعدة من اللغويين والأوساط التقنية:

- تستخدم مجتمعات كثيرة تهجئة مختلفة ومجموعة أحرف كثيرة لكتابة اللغات، بما في ذلك الاختلافات في اللهجات لذلك يجب أن يكون مطوروا أدوات اللغة على دراية بهذا الأمر، وأن يقدموا أكثر من مجموعة واحدة من الحروف وعلامات التشكيل واقتراحات التهجئة لمحاولة تلبية احتياجات جميع مجموعات المجتمع.
- يجب على المجتمع التقني البحث عن طرق لتحديد مثل تلك الاختلافات وجعلها قابلة للاكتشاف، وربما يطور طرقًا لتسهيل التحويل بين مثل هذه المتغيرات، حسب حاجة أعضاء المجتمع.
- إذا كانت الحروف المستخدمة في الكتابة مدعومة بالفعل على أجهزة الكمبيوتر بطريقة ما، فبإمكان المجتمع التقني العمل مع مجتمع اللغة لتحديد الخطوط ولوحات المفاتيح وتطويرهما.
- إذا كان عدم الاتساق يمثل عائقًا في استخدام اللغة، فبإمكان المعلمين المحليين وصانعي السياسات وقادة مجتمعات اللغة واللغويين وغيرهم المساعدة في تبني شكلًا أكثر اتساقًا للكتابة لاستخدامه على الإنترنت.

3.6. هل الكتابة مدعومة بمعيار قياسي؟

هل نظام الكتابة مدعوم بقواعد التهجئة ومجموعة القواعد النحوية سواء رسمية أو غير رسمية؟

إنَّ الكتابة غير الرسمية والاتصالات ممكنة حتى ولو لم يكن هناك معيارًا قياسيًا. إلا إن التهجئة غير القياسية والقواعد النحوية والمفردات أو الاختلافات الإقليمية الكبيرة قد تجعل الاستخدام الأكثر تطورًا صعبًا لوسائل التواصل الاجتماعي ومشاركة المستندات والعثور على مواقع إلكترونية والمعلومات باستخدام البحث عبر الإنترنت والأدوات الأخرى.

هذا العمل مرخص بموجب الرخصة الدولية لكامنز الإبداعية إسناد 4.0 Translation Commons © 2023

نعم: قَدّم اقتراحات حول الحروف
لا: طوّر المعيار القياسي

3.6.1. قدم اقتراحات حول الحروف

يمكن للتهجئة والنحو وعلامات الترقيم المتقاسمة وجوانب أخرى للغة المكتوبة أن تُحَسّن من قدرة أعضاء المجتمع على استخدام الخدمات والأدوات المتاحة عبر الإنترنت مثل اقتراحات الكتابة والنص التنبؤي والبحث على الإنترنت وأدوات الكتابة.

يُحدّد معيار الترميز الموحد الأبجدية وليس اللغات. يمكن استخدام أبجدية ما مثل اللاتينية لأكثر من لغة مثل اللغة الإنجليزية والسواحلية والإندونيسية وغيرها. إذا كانت حروف لغتك مدعومة بإمكانك المتابعة إلى سير عمل تكنولوجيا اللغة.

عندما يكون نظام الكتابة متاحًا وقيد الاستخدام ولكن حروفه ليست ضمن معيار الترميز الموحد، فإنه من المفيد توحيد الأحرف كي يمكن استخدام اللغة على الأجهزة المتصلة بالإنترنت بما فيها الهواتف المحمولة والحواسب المحمولة والمكتبية. بدون التوحيد، مازال بالإمكان مشاركة الملفات بين المستخدمين بنفس الخطوط وطرق الإدخال. إلا إنه رغم ذلك ستصبح الأدوات والخدمات على الإنترنت محدودة.

لتحقيق معيار الترميز الموحد:

1. تحقق ما إذا كانت الحروف مدرجة أو مدعومة في الترميز الموحد.
2. إذا كانت الحروف مفقودة، حضّر وقدم اقتراحًا.

3.6.2. طَوّر المعيار القياسي

قد يكون للغة أكثر من نظام للكتابة، ولكن لكي تُمَكّن الأشخاص من إيصال أفكارهم رقميا، ولكي يتماشى الطابع الرسمي مع الهدف الحالي للمجتمع، يجب تطوير إرشادات للتهجئة وعلامات الترقيم والنحو. يُحَسّن ذلك في لوحة المفاتيح لاقتراح خيارات كلمات وتصحيح التهجئة، كما يُمَكّن محركات البحث وغيرها من خدمات الإنترنت من إعطاء نتائج أكثر نفعًا.

3.7. الشروع في التطبيق

دعم الأحرف متاح في الترميز الموحد. توجد حروف نظام الكتابة في الأبجدية المدعومة. يمكنك الآن المتابعة إلى سير عمل تكنولوجيا اللغة.

Translation Commons © 2023 هذا العمل مرخص بموجب الرخصة الدولية لكامنز الإبداعية إسناد 4.0

4. سير عمل تطبيق تكنولوجيا اللغة

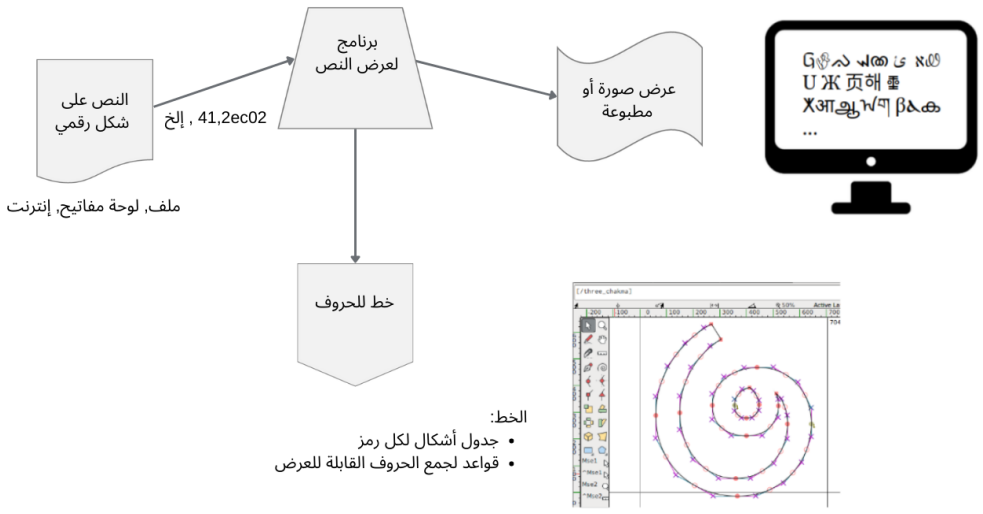
4.1 ملاحظة عن تكنولوجيا النص في النُظم الرقمية

يُخَزّن النص في الأجهزة الرقمية كأنماط من البتّات تُستخدم كنقطة ترميز. على سبيل المثال، نقطة الترميز U+0041 قد تمثل "A". يمكن إنشاء هذا الرمز باستخدام لوحة المفاتيح وتخزينه في ملف أو نقله إلى تطبيق آخر ويُعرض أو يُقدم عن طريق النُظم التي تفهم نظام الترميز المُستخدم للحروف.

يُعَيّن معيار الترميز معنى لكل من الرموز الممكنة. فمثلا: تُحدد ISO-8859-1 الحروف من 0 إلى 255، مما يعطي معنى لكل من نقاط الترميز تلك. الترميز الموحد هو معيار للحروف التي تتضمن عدة نُظم للكتابة، يُعَيّن قيمة فريدة لكل حرف لأكثر من 150 أبجدية مختلفة بما في ذلك اللاتينية والسيريلية والصينية والعربية والعبرية والديفاناغارية والتاميلية وميانمار وغيرها الكثير.

يمكن وضع النص المُكّون من ترميز محدد باستخدامٍ طريقة إدخال متوافقة ويُعرض باستخدام خط متوافق. تظهر هذه العملية في الرسم البياني أدناه (الشكل 4) مُوضحة أن مخططًا كالترميز الموحد قد يعرض حروف للغات كثيرة.

الشكل 4: يُظهر الترميز الموحد الحروف للكثير من اللغات



تستخدم عدة نظم من الكتابة الحروف الأساسية مع علامات التشكيل أو التعديل الاختيارية، مثل الحرف "e" مع تمييز حاد، مما يؤدي إلى الناتج "é". قد يحدد الترميز أنه يمكن اتّباع الحروف الأساسية من خلال دمج رموز التشكيل لإنتاج الحرف المدمج. في بعض الحالات، قد يشتمل الترميز على حروف مكونة مسبقًا باستخدام نقاط ترميز مفردة تتضمن كلًا من نقاط الترميز الأساسية والمُدمجة.

Translation Commons © 2023 هذا العمل مرخص بموجب الرخصة الدولية لكامنز الإبداعية إسناد 4.0

يجب أن تتضمن أي لغة مكتوبة مستخدمة في الأجهزة الرقمية ما يلي:

- نظام ترميز معياري للحروف، مثل الترميز الموحد.
- أنظمة التقديم بالإضافة إلى الخطوط للحروف المستخدمة في اللغة، و
- الأساليب والتطبيقات لتقديم نقاط الترميز للوسائط الإعلامية المطلوبة.

في هذه الوثيقة، نؤكد على معيار الترميز الموحد لأنه يُستخدم في جميع الأجهزة الرقمية الحديثة. وهو أيضًا الترميز الافتراضي للكثير من مصادر النص. يُبَسّط الترميز الموحد الكثير من الجهد المطلوب لوضع لغة على الإنترنت.

4.2. تعريفات لتطبيق الدعم الرقمي

يهدف هذا الجزء لمساعدتك على تطبيق الدعم الرقمي للغتكم المكتوبة. تستخدم الكثير من اللغات الخطوط والأبجدية الغير موحدة. وهذا لا يمنع استخدام اللغة على الإنترنت، ولكن يمكن أن يحد من مستوى الدعم المُتاح. تشرح الخطوات التالية جوانب الدعم الرقمي للّغات المكتوبة على الإنترنت.

تتضمن المتطلبات الأساسية لنظام الكتابة الرقمية:

- الحروف: حدّد مجموعة الحروف وعلامات التشكيل والروابط وعلامات الترقيم والأرقام والرموز التعبيرية والرموز الأخرى المستخدمة في تهجئة اللغة.
- محددّ هوية الأيزو (**ISO**): للغة والأبجدية. يمكن أن يتضمن هذا محددات إقليمية اختيارية مثلا *es-MX* للغة الإسبانية المستخدمة في "المكسيك".
- معيار الترميز: سواء معيار رسمي (الترميز الموحد مع التجميع والتصنيف ومجموعات الحروف) أو معيار غير رسمي (ترميز الخط).
- دعم الخط: المعلومات المطلوبة لإنشاء خط يتضمن كل أشكال الحروف المطلوبة:
  - تفاصيل عرض النص مثل مجموعات الارتباط والروابط والمجموعات المُدمجة (قد يكون هذا متوفرًا بالفعل في ملفات الترميز الموحد).
  - عينة النص لمصممي ومطوري الخط لاستخدامه في التجربة.

  قد يتطلب الأمر تحديثات لمحركات العرض أو لبرنامج آخر.
- طريقة الإدخال: طريقة لإدخال الحروف لكل الأجهزة. عادةً ما تكون لوحة مفاتيح فعلية أو على الشاشة.

تتوفر مصادر إضافية للّغة:

- يتضمن مستودع البيانات المحلية المشترك (CLDR) معلومات إضافية عن اللغة مثل التقويمات.
- رمز اللغة مثل BCP IETF -47.
- دعم تطبيق البرنامج لتجزئة الكلمات:
  - قواميس الكلمات المكتوبة بدون فواصل واضحة.
  - القواعد مثل المسافات وعلامات الترقيم.

Translation Commons © 2023 هذا العمل مرخص بموجب الرخصة الدولية لكامنز الإبداعية إسناد 4.0

4.3. هل الرمز المعياري للغة متاح؟

هل يوجد رمز مُوحد للّغة ومُتفق عليه؟

نعم: هل الترميز الموحد للخط متاحٌ؟
لا: قدّم طلبًا للحصول على رمز اللغة

4.3.1. قدّم طلبًا للحصول على رمز اللغة

إذا لم يكن رمز اللغة متاحًا، فستحتاج إلى إنشاء رمز مُوحد للّغة باستخدام متغير إقليمي اختياري أو الأبجدية أو كليهما. يمكن أن تؤدي علامات ورموز اللغة غير الموحدة إلى الالتباس وتحديد خاطئ لمصادر اللغة. اتبع توجيهات ISO-639 وIETF BCP-47.

4.4. نعم: هل الترميز الموحد للخط متاحٌ؟

هل الخط المتوافق مع الترميز المُوحد متوفر لأعضاء المجتمع لاستخدامه على الكمبيوتر المحمول والمكتبي؟

نعم: هل يتوفر الخط على الأجهزة؟
لا: أنشئ نوع الخط

4.4.1. إنشاء الخط

بعد إضافة الأبجدية إلى الترميز الموحد، قد لا يكون هناك أي خطوط بعد للترميز الموحد لدعم الحروف. في هذه الحالة، قد تحتاج إلى التعامل مع مصممي الخطوط والمجتمع التقني لإنشاء مثل هذه الخطوط لدعم النص في الأبجدية.

4.5. هل الخط متوفر على الأجهزة؟

هل الخط متوفر على أجهزة الكمبيوتر والأجهزة المحمولة الأخرى المتاحة للمجتمع؟ إذا كان الأمر كذلك، فسيصبح النص الموجود على المواقع الإلكترونية ومواقع التواصل الاجتماعي والتطبيقات الأخرى مقروءًا.

من الجدير بالذكر أن معظم أجهزة الكمبيوتر والمحمول ستعمل على نحو أفضل مع نص الترميز الموحد والخطوط المتوافقة وطرق الإدخال التي تُنتج نص الترميز الموحد.
تحقق من توفر الخط كجزء من خطوط نوتو (Noto) أو مواقع إلكترونية أخرى لتحميل واستخدام خطوط ترميز مُوحد.
لاحظ أن العديد من الأجهزة المحمولة قد لا تدعم مباشرة تحميل الخطوط أو تثبيتها. تسمح الحواسيب المكتبية والمحمولة بتثبيت الخطوط التي يمكن تحميلها.

قد تتطلب بعض التطبيقات مثل معالجات الكلمات والمتصفحات خطوة تكوين للتطبيق لعرض الخط في المستند أو صفحة الويب.

Translation Commons © 2023 هذا العمل مرخص بموجب الرخصة الدولية لكامنز الإبداعية إسناد 4.0

يمكن لمواقع الإنترنت استخدام خطوط الويب لعرض نص باستخدام خط يوفره موقع الويب. يتيح ذلك لصانعي المحتوى تحديد الخط المعين المستخدم، كما يجعل الموقع قابلًا للقراءة على الأجهزة، حتى إذا لم يكن الخط مثبتًا بالفعل على الجهاز. لاحظ أن خط الويب يعمل فقط مع الصفحات التي تجري تحديدًا دقيقًا له، ولا يؤدي هذا الاستخدام إلى تثبيت الخط بشكل دائم على الجهاز. تتوفر أحدث خطوط "نوتو" كخطوط ويب وهو أمر مفيد إذا لم يكن الجهاز يحتوي بالفعل على الإصدار المناسب من الخط أو لا يمكن تثبيته.

نعم: هل يحتوي الجهاز على دعم للإدخال؟
لا: تثبيت يدوي أو اطلب من البائعين إضافة الدعم

4.5.1. تثبيت يدوي أو اطلب من البائعين إضافة الدعم

قد لا تُدرَجُ نُظم الكتابة الحديثة الموحدة في الأجهزة المحمولة، حتى عندما يكون نوع الخط متاحًا بطريقة أخرى. أطلب من بائعي الأجهزة إدراج أنواع الخطوط المطلوبة للّغة.

البديل هو تثبيت خط ترميز مُوحّد يدويًا على الجهاز المحمول. يمكن أن يكون هذا فعالًا، ولكنه غالبًا يتطلب معرفة متخصصة.

تحذير: قد ينطوي تثبيت خط من الويب على جهاز محمول على مخاطر تتعلق بالأمان والخصوصية. كما قد يصبح ضمان الجهاز أو اتفاقيات الدعم غير صالحة في حال أجرِيَ هذا التثبيت.

4.6. هل يحتوي الجهاز على دعم الإدخال؟

سيحتاج مستخدمو اللغة إلى طريقة إدخال فعالة لإنشاء رسائل أو رسائل بريد إلكتروني أو تدوينات أو أي محتوى آخر بهذه اللغة. هل الجهاز (المحمول أو غير ذلك) يدعم بالفعل طريقة لإدخال اللغة؟

نعم: هل الإدخال مدعوم من أجهزة أو تطبيقات طرف ثالث؟
لا: طوّر طريقة الإدخال

4.7. هل الإدخال مدعوم بتطبيقات تابعة لطرف ثالث؟

هل دعم المدخلات مدمج في لوحات المفاتيح الاعتيادية للنظام؟ إذا كان كذلك، تعلم كيفية تمكين أو تثبيت دعم لوحة المفاتيح. هذا خاص بالجهاز، ولكن مصادر الإنترنت يمكن أن تساعد في هذا الإجراء.

يقدم بائعو الأجهزة المحمولة العديد من لوحات المفاتيح الموجودة على نظامي iOS وAndroid. على سبيل المثال:

- جي بورد (Gboard) لوحة مفاتيح جوجل للهاتف المحمول (https://support.google.com/gboard/answer/6380730?hl=en&co=GENIE.Platform=Android) تدعم الكثير من اللغات في IOS, Android.
- تحتوي أدوات إدخال غوغل (https://www.google.com/inputtools) على لوحات مفاتيح افتراضية للاستخدام على صفحات الويب في متصفح كروم على الحواسيب (وليس الهاتف المحمول).

Translation Commons © 2023 هذا العمل مرخص بموجب الرخصة الدولية لكامنز الإبداعية إسناد 4.0

- تتوفر العديد من تطبيقات الطرف الثالث لإدخال النص في متجر تطبيقات "أبل وجوجل".

لاحظ أن العديد من اللغات لديها بالفعل تخطيطات لوحة مفاتيح موصوفة على سبيل المثال في مستودع لوحات مفاتيح CLDR.

4.7.1. طوّر طريقة للإدخال

هناك العديد من الأدوات لتثبيت طرق الإدخال للكثير من اللغات وكذلك لتطوير لوحات مفاتيح جديدة. وهذا يتضمن:

- تتوفر تطبيقات لوحة المفاتيح من متجر "أبل أو جوجل" أو مصادر أخرى.
- كيمان (Keyman): أداة للوحات مفاتيح اللغة (https://keyman.com)
- لوحة مفاتيح مايكروسوفت بيلدر(Microsoft Keyboard Builder) (https://www.microsoft.com/en-us/download/details.aspx?id=102134).
- SIL Ukelele (https://scripts.sil.org/cms/scripts/page.php?site_id=nrsi&id=ukelele

تحذير: قد ينطوي تثبيت لوحة مفاتيح أو طريقة إدخال أخرى من الويب على مخاطر تتعلق بالأمان والخصوصية.

4.8. هل يحتوي الجهاز على دعم بيانات الترميز الموحد؟

هل بيانات الترميز الموحد متوفرة؟

نعم: تابع باستخدام لغتك.
لا: إذا لزم الأمر، انظر في إضافة بيانات اللغة الأساسية إلى CLDR. ويشمل هذا رمز واسم اللغة.
اطّلع على المزيد حول CLDR ضمن دعم اللغة الإضافي.

5. دعم إضافي للّغة

يغطي توحيد الترميز الموحد عدة من جوانب معالجة النص لنظام الكتابة واللغة. ومع دعم الخط والمدخلات ستنجح أشياء كثيرة. يتضمن ذلك معظم جوانب معالجة الكلمات الأساسية وجداول البيانات والبريد الإلكتروني.

الخطوات المذكورة هنا ما هي إلا البداية إذ يُوَفّر مستودع البيانات المحلية المشتركة (CLDR) "اللبنات الأساسية للبرمجيات لدعم لغات العالم" من خلال جمع معلومات مفيدة لمختلف المناطق (اللغة والبلد). يمكن أن توفر هذه البيانات أسماء اللغات والبلدان والشهور وأيام الأسبوع وغيرها من المعلومات. كما أنها تتيح التنسيق مع مراعاة الإعدادات المحلية للتاريخ والوقت والأرقام والمعلومات الأخرى ذات التنسيق الشائع.

على الرغم من أن بيانات CLDR ليست مطلوبة للتواصل النصي الأساسي في لغات السكان الأصليين، إلا أن هذه المعلومات تعزز وظائف اللغة. ستعمل جميع الأدوات تقريبًا مثل البريد الإلكتروني والرسائل النصية ووسائل التواصل الاجتماعي وما إلى ذلك بطريقة جيدة عند وجود الخطوط ولوحة المفاتيح.

يستخدم المبرمجون هذه المعلومات لإنشاء مخرجات للتطبيقات عبر الإنترنت في سياقات محددة مثل التقويمات المحلية وجداول البيانات والمخرجات الرقمية واختيارات القائمة وسياقات واجهة المستخدم الأخرى.

Translation Commons © 2023 هذا العمل مرخص بموجب الرخصة الدولية لكامنز الإبداعية إسناد 4.0

يخزن CLDR أيضًا معلومات لغة إضافية مثل الحروف المستخدمة في الكتابة وتخطيطات لوحة المفاتيح لإدخال النص. انظر لوحات مفاتيح CLDR لمزيد من المعلومات.

ومع ذلك، قد تكون هناك حاجة إلى بيانات إضافية وأدوات معالجة النصوص للحصول على دعم لغوي أكثر اكتمالًا في التطبيقات على الإنترنت والهاتف المحمول. فيما يلي بعض الأشياء اللازمة لتحقيق التكافؤ بين لغات السكان الأصليين واللغات المدعومة بالكامل:

تجزئة وتقسيم الكلمات: يلزم فصل النص من أجل التخطيط الصحيح للنص، وكذلك اختيار المستخدم لمجموعة مختارة من مجموعات غرافيم (وحدة خطية) والكلمات والجمل. بالنسبة للكثير من اللغات، لا تُوفّر إشارة واضحة لتوضيح حدود الكلمات، على سبيل المثال: المسافات أو علامات الترقيم. في هذه الحالات، هناك حاجة إلى بيانات القاموس أو الخوارزميات لتوفير مثل هذه المعلومات. للمزيد من المعلومات انظر: https://unicode-org.github.io/icu/userguide/boundaryanalysis.

- تقسيم السطر: تملك اللغات قواعد متنوعة تخص المواضع التي يمكن قطع النص بها وتحريكه إلى سطر جديد. على سبيل المثال يُحيط بالنص المقتبس حروف مختلفة خاصة باللغة التي تتبع قواعد مختلفة للفصل بين الكلمات والجمل. بالإضافة إلى ذلك، تعتمد المواضع التي يمكن فيها تقسيم سطر من النص على خصائص حروف الترميز الموحد في نظام الكتابة. كذلك الأرقام والعملات مثل 10$ سيلزم دمجها معا لتفادي الفهم الخاطئ للمعنى. تختلف قواعد علامات الترقيم باختلاف كل من اللغة والمنطقة. تقسيم السطر هو حالة خاصة من تحليل حدود الترميز الموحد (https://unicode-org.github.io/icu/userguide/boundaryanalysis/#line-break-boundary).

- تحديد لغة المحتوى النصي: يجب تسمية المستندات بوضوح برمز لغة أو معرّف آخر يصف اللغة البشرية للمستند. فعندما تتوفر هذه المعلومات في مستند، قد تستخدم الأدوات ذلك للعثور على معلومات مناسبة للمستخدمين على نحو أكثر فاعلية. في مستند متعدد اللغات، يمكن إرفاق الأقسام الفردية أو حتى الفقرات بلغة النص.
  إنّ استخدام الرموز الاعتيادية مثل BCP IETF-47 مهم جدًا (https://tools.ietf.org/html/bcp47)
  - تختلف آلية هذا التحديد بين التطبيقات، وقد يحتاج المستخدم إلى معرفة ما إذا تم وكيف يتم ذلك. لاحظ أن هذا التعريف قد يوفر معرفات من قائمة بدلًا من دعم أي رمز محتمل.
  - بالنسبة للمستندات والمواقع عبر الإنترنت، تُقدم لغة البرمجة HTML خاصية اللغة(https://www.w3schools.com/tags/att_global_lang.asp)
  لتسمية لغة أي مكون لـ HTML بوضوح. يجب أن تُؤخذ قيمة هذه الخاصية من مجموعات معيارية من مُعرّفات اللغة بدلاً من سلسلة عشوائية يحددها المستخدم. على سبيل المثال استخدم *rs* للّغة الصربية، *de* للّغة الألمانية، *zh-Hans* أو *zh-CN* (أو فقط *zh*) للّغة الصينية بطريقة الكتابة المبسطة.

- الكشف عن اللغة: غالبًا ما تقدم الخدمات والتطبيقات الأخرى عبر الإنترنت نتائج مفيدة وذات صلة عندما تكون لغة النص معروفة. بالنسبة للنص الذي لا يمكن تمييزه بوضوح فقد طُوّرَت كاشفات اللغة مثل cld2 (https://github.com/optimaize/language-detector) تُجري هذه الأدوات عادة تحليل إحصائي على حروف النص مما يتيح تحديدًا محتملاً للّغة البشرية. وذلك ضروري لأن معظم أنظمة الكتابة تستخدم لغات متعددة، على سبيل المثال الحروف اللاتينية للسواهيلية واللاكوتا والوارلبيري والفنلندية، والحروف السريالية للروسية والأوكرانية والكازاخية أو الكازاخستانية وميانمار للبورما والشان والمون وما إلى ذلك.

- القواميس لمعالجة الكلمات: تدعم معظم تطبيقات معالجة الكلمات صياغة النصوص الأساسية وتحريرها ومشاركتها وطباعتها. يستخدم النص التنبئي والتصحيح الإملائي والاقتراحات النحوية والأدوات الأخرى قوائم الكلمات مع تواتر تكرار البيانات والقواميس والبيانات اللغوية الأخرى. إنّ المرادفات والعبارات الاصطلاحية شائعة الاستخدام مفيدة أيضًا في أدوات مثل البحث عبر الإنترنت.

Translation Commons © 2023 هذا العمل مرخص بموجب الرخصة الدولية لكامنز الإبداعية إسناد 4.0

- أرقام غير آسكي **(Non-ASCII)**: العديد من النصوص لها أرقام مختلفة عن الأرقام الغربية. أمثلة تشمل: ميانمار وأدلام والعربية والفارسية. ومع ذلك، فالكثير من التطبيقات مثل جداول البيانات لا تفسر هذه الأرقام كقيم رقمية بل كقيم نصية. قد ينظر منفذو هذه التطبيقات في خصائص الترميز الموحد لهذه الحروف من أجل معالجتها كأرقام، ولكن هذا الدعم لا يُنفذ بانتظام (https://en.wikipedia.org/wiki/Numerals_in_Unicode).

- واجهات المستخدم المترجمة: في بعض التطبيقات خاصة للتعليم أو المعلومات بلغة المستخدم تحديدًا، قد يكون من المفيد ترجمة النص الذي يظهر في واجهة المستخدم (UI). مثلا، يمكن ترجمة عناصر القائمة لوظائف نظام التشغيل مثل "ابدأ" أو "فتح ملف". ومع ذلك، في كثير من الحالات يتعذر لمالك التطبيق توفير ترجمات للغات المجتمعات الصغيرة. عندما تكون واجهة المستخدم متاحة بلغة واحدة على الأقل من اللغات التي يمكن للمستخدم فهمها، تكون الواجهة المترجمة ذات قيمة مباشرة أقل.

- التعرف الضوئي على الحروف **(OCR)**: الكثير من اللغات لها قدر كبير من المؤلفات المكتوبة في الكتب وغيرها من الوثائق. يمكن استخدام التعرف الضوئي على الحروف في تحويل النص في تلك الوثائق إلى صيغة رقمية. يوجد مشاريع مفتوحة المصدر للتعرف الضوئي على الحروف (https://pdf.iskysoft.com/ocr-pdf/open-source-ocr.html)، كما يمكن توجيهها نحو نظم جديدة للكتابة. وجدير بالذكر أن نموذج اللغة الذي يحتوي على قوائم بالكلمات الشائعة يُحسّن بطريقة كبيرة من دقة أساليب التعرف الضوئي على الحروف.

5.1. مصادر عامة للّغة

هناك الكثير من المصادر المتاحة المجانية وبالاشتراك يمكن للمجتمعات استخدامها:

5.1.1. مصادر اللغة

- Panlex (https://panlex.org)
- Wikitongues (https://wikitongues.org)

5.1.2. مصادر الأداة

- SIL - Keyman
- أدوات الخط
- أدوات مفتوحة المصدر للقواميس

5.2. تكنولوجيا اللغة المتقدمة

تتطلب القدرات التالية كميات كبيرة جدًا من البيانات لتدريب أنظمة تعلم الآلة. لقد بدأت البرامج العامة المفتوحة المصدر تصبح متاحة، ولكن معظم العمل في هذا المجال يتم في البحث الأكاديمي أو منتجات الشركات. ومن غير المرجح أن تكون معظم هذه السمات متاحة لمعظم اللغات في المستقبل القريب.

- تحويل الكلام إلى نص: التعرف على الكلمات التي يتحدثها الشخص وتحويل الأصوات إلى نص الرسالة المنطوقة. هذا النص يمكن تحويلة إلى مستندات أو يمكن استخدامه للتحكم في التطبيقات والأجهزة.
- تحويل النص إلى كلام: إنتاج مخرج من الحديث الطبيعي من النص. يُعدّ هذا مفيدا لواجهات المستخدم بدون استخدام اليدين وللقراءة الآلية للبشر من المصادر النصية.

Translation Commons © 2023 هذا العمل مرخص بموجب الرخصة الدولية لكامنز الإبداعية إسناد 4.0

- كتابة الوسائط الصوتية الأكاديمية: مهم لتوثيق اللغة وخاصة في الدراسات اللغوية الأكاديمية. بدأت بعض المشاريع مفتوحة المصدر في تلبية هذه الحاجة بما في ذلك:
- النسخ المعجل للغويين: https://github.com/CoEDL/elpis
- الترجمة الآلية: يُعدّ تحويل النص من لغة بشرية إلى أخرى من أصعب المهام لأجهزة الكمبيوتر. يمكن للنُظم الحالية الترجمة بين مجموعة محدودة من اللغات المدعومة. ومع ذلك، فإن هذه النُظم عمومًا لا تفهم السياق وليست على مستوى الترجمة البشرية. إلا أنه مع التقنيات الجديدة لتعلم الآلة، تزداد جودة وموثوقية الترجمة الآلية بسرعة لأصحاب الشركات الكبيرة.

لا يتوفر دعم الترجمة الآلية للغات السكان الأصليين على نطاق واسع، ولكن بدأت تظهر جهود المصادر المفتوحة والجامعة. فعلى سبيل المثال www.apertium.org وهو أداة عبر الإنترنت تدعم جهود الترجمة الآلية للغات غير المهيمنة.

6. مسرد

أسكي (**ASCII**): المعيار القياسي الأمريكي لتبادل المعلومات: ترميز الحروف للاتصالات الإلكترونية.

BCP-47: وهو رمز *IETF* للتعرف على اللغات.

CLDR: مستودع البيانات المحلية المشتركة. يحتوي على معلومات إضافية للغة.

الحرف: أصغر مكون في اللغة المكتوبة وله قيمة دلالية فيشير إلى المعنى المجرد و/أو الشكل، وليس إلى شكل معين. المرجع: مسرد الترميز الموحد

نقطة الترميز: رقم يمثل حرفًا أو تنسيقًا معينًا.

علامة التشكيل: هي رمز يضاف إلى حرف أو رمز أساسي، عادة لتغيير صوته أو مدلوله اللفظي.

لغة خامدة: هي لغة ليس لها متحدثون حاليا.

الخط: هو من مجموعة من الصور الرمزية تُستخدم من أجل التمثيل المرئي لبيانات الحرف

الرمز: واحد من مجموعة من الرموز يمثل حرفًا مقروءًا بغرض الكتابة.

القواعد النحوية: هي القواعد التي تحكم تركيب اللغة الطبيعية.

IETF: فرقة عمل هندسة الإنترنت.

لغة الشعوب الأصلية: هي لغة أصلية لمنطقة معينة.

2019IYIL: 2019 السنة الدولية للغات الشعوب الأصلية.

إحياء اللغة: وقف انحدار أو إحياء لغة خامدة.

الوصلات الأبجدية: الدمج بين رمزين أو أكثر في رمز واحد.

لغة طبيعية: لغة تطورت طبيعيًا في البشر بشكل متفرد عن اللغات الرسمية مثل تلك المستخدمة في الحوسبة. تشمل اللغات الشفهية والمرئية والمرئية-اليدوية (الإشارة) واللغات المكتوبة.

نوتو: هي عائلة خطوط تضم أكثر من 100 خط فردي، تهدف معا لتغطية كل الأبجدية المشفرة في الترميز الموحد (يغطي في الوقت الحالي كل الأبجدية الموجودة في الترميز الموحد 6.0 وما قبله).

Translation Commons © 2023 هذا العمل مرخص بموجب الرخصة الدولية لكامنز الإبداعية إسناد 4.0

التهجئة: مجموعة من الطرق الثابتة لكتابة اللغة.

PUA: منطقة الاستخدام الخاص: مجموعة من نقاط الترميز في الترميز الموحد والتي لن يُعيّن حروف لها أبدا.

علامات الترقيم: المسافات والعلامات التي لا تنطق والتي تساعد على فهم النص.

الأبجدية: مجموعة من الحروف والعلامات الأخرى المكتوبة تستخدم لتمثيل المعلومات النصية في واحد أو أكثر من نُظم الكتابة.

ترانسليشن كومنز **(Translation Commons)**: هو مجتمع عبر الإنترنت منصة لمشاركة المعرفة اللغوية بحُرية.

يونسكو: منظمة الأمم المتحدة للتربية والعلم والثقافة.

الترميز الموحد: المعيار الأكثر استخدامًا الترميز الرقمي لحروف نُظم الكتابة في العالم.

نظام الكتابة**:** مجموعة من القواعد لاستخدام أبجدية أو أكثر لكتابة لغة معينة.

7. المراجع

7.1. إحياء اللغة

Routledge Handbook of Language Revitalization (Hinton, Huss, & Roche 2018)
The Green Book of Language Revitalization in Practice (Hinton & Hale 2001)
Language Documentation and Revitalization in Latin American Contexts (Perez-Baez, Rogers, & Roses Labrada 2016)
Developing Orthographies for Unwritten Languages (Cahill & Rice 2014)
http://cherokeepreservation.org/what-we-do/cultural-preservation/cherokee-language
https://language.cherokee.org
http://amahmutsun.org/language
https://rising.globalvoices.org/blog/2011/11/29/languages-online-activism-to-save-chakma-language

7.2. سجلات اللّغة

https://www.ethnologue.com
https://glottolog.org

7.3. الترميز الموحد وترميز الخط

https://unicode.org/main.html
https://unicode.org/standard/supported.html
https://unicode.org/standard/where
https://unicode.org/pending/proposals.html
https://unicode.org/glossary
https://linguistics.berkeley.edu/sei

Translation Commons © 2023 هذا العمل مرخص بموجب الرخصة الدولية لكامنز الإبداعية إسناد 4.0

يمكن أن تستخدم الخطوط المتخصصة مناهج غير معيارية للحروف، مثل مناطق الاستخدام الخاص (PUAs) أو نطاقات أخرى للحروف مثل "أسكي" أو اللغة العربية مع الرموز المخصصة لنقاط الترميز. وهذا يطلق عليه ترميز الخط. تُمكن هذه الأساليب غير المعيارية المستخدمين من رؤية الحروف التي يكتبونها، مع ذلك لن يتمكن الآخرين من رؤيتها ما لم يستخدموا نفس الخطوط. لن تتمكن أيضًا الخدمات والأدوات على الإنترنت من تفسير النص بنحو صحيح في ترميزات الخطوط هذه، لأن الترميز الأساسي لا يحتوي على معلومات عن المعنى الفعلي المقصود للحرف.

في حين أن النص الذي يعتمد على الترميز الموحد له مزايا مقارنة بالنص المشفر بالخط، قد يكون من الضروري استخدام خطوط متخصصة إلى أن يجري تضمين الحروف في الأبجدية في معيار الترميز الموحد. في هذه الحالة، يجب أن يستخدم الخط فقط مناطق الاستخدام الخاص من نطاق الترميز الموحد بدلاً من إعادة استخدام قيم الرمز المحجوزة بالفعل لأبجديات أخرى. هذا يمنع تداخل قيم الرمز ويسمح باستخدام أسهل للأبجديات الموجودة. يُعدّ تحويل هذا الخط إلى الترميز الموحد بعد توحيد الأبجدية أمرًا سهلاً نسبيًا بالنسبة لرموز منطقة الاستخدام الخاص، نظرًا لاستخدام رموز منطقة الاستخدام الخاص نفسها بانتظام.

7.4. رموز اللغة

https://en.wikipedia.org/wiki/List_of_ISO_639-1_codes

7.5. الخطوط

https://www.google.com/get/noto

تتضمن أدوات تطوير الخطوط:
- فونت فورج (FontForge)
- فونتلاب (FontLab)
- الرموز (Glyphs)
- بيردفونت (https://birdfont.org) (BirdFont)

جامعة ريدنغ لديها برنامج دراسات عليا في تصميم الخط (typefacedesign.net/) التي قد يكون طلابها قادرين على المساعدة في إنشاء خط ترميز موحد جديد.

يمكن لمطوري الخطوط التجارية إنشاء خطوط ترميز موحّدة جديدة.

8. ملاحظات

يقتصر نطاق هذا المستند على اللغات المكتوبة. وتشمل أشكال الاتصال الأخرى ما يلي:

- الأيقونات
- اللغات الشفهية
- اللغات المرئية اليدوية (الإشارة)
- اللغات المرئية

يمكن أيضًا كتابة اللغات التي لها أبجدية خاصة بها باستخدام أبجديات أخرى. على سبيل المثال:

Translation Commons © 2023 هذا العمل مرخص بموجب الرخصة الدولية لكامنز الإبداعية إسناد 4.0

- كانت اللغة التركية تُكتب في الأصل بالأبجدية العربية ولكنها الآن تُكتب بالأبجدية اللاتينية.
- يمكن كتابة اللغة الصينية بالأبجدية اللاتينية (بينيين).

لا تغطي هذه الوثيقة لهجات اللغة، ولكن هذه اللهجات تستحق أن تؤخذ بعين الاعتبار عند اقتراح معايير اللغة.

البحث اللغوي العالي الجودة له متطلبات خاصة:

- تحديد اللغة (من النص)
- التجزئة: تقسيم النص إلى كلمات
- تحديد أصول الكلمات كما في اللغة ذات عدة أشكال، على سبيل المثال: مسكن ومساكن وتسكين.

على الرغم من أن اللغات الشفهية خارج نطاق هذه الوثيقة، قد تكون المصادر التالية مفيدة:

- خرائط غوغل https://docs.google.com/forms/d/e/1FAIpQLSdphaDaz33syPoUDyTOTwwkaLWZx90zopUklha4uadfkUKG8A/viewform
- https://www.blog.google/products/earth/indigenous-speakers-share-their-languages-google-earth
- https://www.gerlingo.com
- Xtrans (https://www.ldc.upenn.edu/language-resources/tools/xtrans هي أداة نسخ من الجيل القادم متعددة الأنظمة واللغات ومتعددة القنوات تدعم النسخ اليدوي والتعليقات التوضيحية للتسجيلات الصوتية.

Translation Commons © 2023 هذا العمل مرخص بموجب الرخصة الدولية لكامنز الإبداعية إسناد 4.0