The Unicode Guide to Software Internationalization

Background, Goals, Progress and Future Plans

Joel Sahleen, Spotify
Member, Unicode Community Engagement Team

Unicode Technology Workshop

November 8, 2023

UNi

UNICODE

Contents

1. Project Background
2. Purpose and Goals
3. Organization and Content
4. Website Demo
5. Publishing Process
6. Future Directions
7. Current Status

Project Background

What is the problem we are trying to solve?

Background

The global growth of the internet and the proliferation of internet-connected devices have increased the need for international software. International software is software that has been architected and written in a way that enables it to work *internationally* and support localization.

Context

The increased demand for international software has
created a corresponding demand for software developers
and localizers who understand internationalization; i.e., the
process of *making* software that works internationally and
supports localization.

Problem

Unfortunately, software internationalization is generally not taught in computer science departments and most localization programs focus on language and program management rather than development.

Consequences

As a result, most new developers have never learned how to internationalize an application, while most new localizers understand what internationalization is and why it is important, but not how it actually works.

UNICODE

Impact

Feature developers often end up having to reinvent the internationalization wheel and make things up as they go along. Meanwhile, localizers generally do not have the knowledge or experience to show developers how to do things in a more standardized, scalable way.

Significance

As software matures and the scope of localization expands, internationalization becomes more difficult and time consuming, leading to a deprioritization of internationalization work, an accumulation of internationalization technical debt, and ultimately, a degraded experience for international users.

Purpose and Goals

What are we planning to do? What do we want to achieve?

Hypothesis

The current deficit in internationalization knowledge and skills can be traced back to four main factors:

➔ A scarcity of qualified instructors
➔ An absence of well-defined curricula
➔ A lack of effective learning/teaching materials
➔ An increase in developers and localizers with non-traditional backgrounds

UNICODE

Opportunity

Most of these problems could be addressed by expert contributors working together to design, create, deploy and maintain an online guide to internationalization. The Unicode Consortium is an ideal space for this work to take place.

Objective 1: Authoritative

The guide should be a reliable, up-to-date source of information about internationalization that is based on industry standards and best practices.

UNICODE

Objective 2: Extensive

The guide should provide a comprehensive but high-level overview of internationalization that covers its most important aspects while supporting multiple self-directed learning paths and linking out to more detailed resources.

UNICODE

Objective 3: Accessible

The guide should be produced and written in a way that makes sense to individuals in a variety of software development and localization roles; not just hardcore engineers and linguists.

UNICODE

Objective 4: Engaging

The guide should provide an engaging learning experience that is well-structured and includes different types of content for different audiences and learning styles.

UNICODE

Objective 5: Discoverable

The guide should be cross-referenced, searchable and SEO-friendly, so that it is easy to navigate and you can find the information you need both inside the site and through external search engines.

Outcomes

1. The guide will be used in software development programs to help students learn how to internationalize their software.
2. The guide will be used in localization programs to help students learn how to work with software development teams.
3. The guide will serve as an online reference for working developers and localization professionals.
4. The guide will improve the practice of internationalization in the real world.
5. The guide will help make software more available for everyone.

Organization and Content

What will it actually look like? Why will it look that way?

Inspiration

The guide is intended to be a modern take on the traditional *edited volume* format. It employs the same publishing model with editors and contributors, while leveraging the interactive capabilities of the web.

Front Matter

The guide will have front matter that explains what the guide is about, why it was created and who should be contacted for further information.

Introduction

The guide itself should begin with a general introduction that summarizes the different parts of the guide and explains the purpose and function of each chapter.

Thematic Chapters

The bulk of the guide should consist of thematic chapters that focus on different aspects of internationalization. Each chapter should start with an overview and end with a review. The chapters should be independently produced but fit together to form a larger curriculum.

UNICODE

Interactive Pages

Each chapter should be made up of nested folders and pages. Each page should include a variety of information.

➔ A brief summary or overview of the content
➔ A clear statement of learning objectives for the page
➔ A short video lecture and/or screencast
➔ Supporting documentation and notes
➔ Working, copyable code samples, with highlighting
➔ Cross-references to other pages in the guide
➔ Links to outside resources
➔ Demo components and examples

UNICODE

Useful Appendices

The main content of the guide should be followed by a set appendices that focus on supporting topics.

➔ Managing an internationalization program
➔ Designing international software
➔ Building out a localization infrastructure

UNICODE

Demo

https://unicode-guide.vercel.app

Technology

The example guide is built using Next.js, the Nextra static site generation framework and a modified version of the Nextra documentation theme. It is currently hosted on Vercel, which offers free team accounts for not-for-profit organizations. (We can connect the guide to a unicode.org subdomain or move it to a new hosting provider.)

UNICODE

Features

➜ Static page generation
➜ Build time indexing and client-side search
➜ Accessibility and localization support
➜ Pages written in MDX with JSX components
➜ Automatic page and document TOCs

UNICODE

Publishing Process

How will we make this work? What are the next steps?

Education Working Group

The first step is to form a education working group made up of members of the Unicode Community Engagement Team. This working group will be responsible for defining the overall structure, style and format of the guide.

Invited Contributors

Guide contributors will be invited to produce each chapter in the guide based on expertise. Each contributor will receive a publication credit. Chapters can also be co-produced by multiple people, or in consultation with other experts.

External Reviewers

To ensure quality, each chapter will be reviewed by the education working group and external reviewers at multiple points in the production process. Contributors will also review each other's work for the sake of consistency across the guide.

Annual Editions

To keep the information in the guide up to date, new editions of the guide will be released each year. Most of the guide will consist of evergreen material that just needs to be updated. We can use quarterly updates to fix any errata.

Future Directions

What could be done in the future? What are the possibilities?

Localization

The Nextra framework supports localization of content. By combining this with translated subtitles from YouTube, the guide could support global reach and be made accessible to more people around the world. (Great opportunity for students.)

Monographs

If the guide proves successful, the Unicode Consortium could expand its online publishing to include more specialized website monographs that dive deeper into particular subjects. For example, there could be a published site devoted entirely to script encoding.

Journal

The same process used to produce the guide could also be used to publish a Unicode journal, at whatever frequency. The possibilities are endless!

Current Status

How far have we come? How is it going?

Proposal Submitted

A proposal for the guide was submitted to the Community Engagement Team earlier this year. It has been reviewed and feedback has been incorporated. All signs are good.

Website Proof of Concept

As you saw, an initial version of the actual site has been created in order to provide a clearer picture of what the guide will ultimately look like and determine if it makes sense to manage this on our own or use a third party education platform. Given our limited needs it seems we can manage it on our own.

Initial Organization

The demo site outlines a proposed structure for the guide that can inform further discussions. A lot of thought has been put into it already, but feedback is welcomed.

In Summary

➔ **Authoritative**. Produced by experts
➔ **Extensive**. Covers most aspects of i18n
➔ **Accessible**. Useful to a diverse range of people
➔ **Engaging**. Interactive and well-structured
➔ **Discoverable**. SEO and internal search.
➔ Coming Soon(ish) in 2024!

UNICODE

How you can help!

➔ Suggest topic ideas - i18nguide@unicode.org

➔ Join us for an unconference session to discuss the guide contents and publishing processes.

➔ Localization. (A great opportunity for students to get involved.)

➔ Reach out to education working group members with other questions or suggestions.

Thank you!

【 U+3010

シ U+30B7

メ U+30E1

ə U+10DB

☎ U+260E

ヂ U+3077

👽 U+1F47D

ق U+0642

☂ U+2602

个 U+4E2A

😌 U+1F60C

Ψ U+03A8

غ U+063A

ミ U+30DF

鏡 U+93E1

ε U+03B5

○ U+3007

♺ U+267A

ζ U+03B6

く U+304F

ß U+0D26

ৰ U+09F0