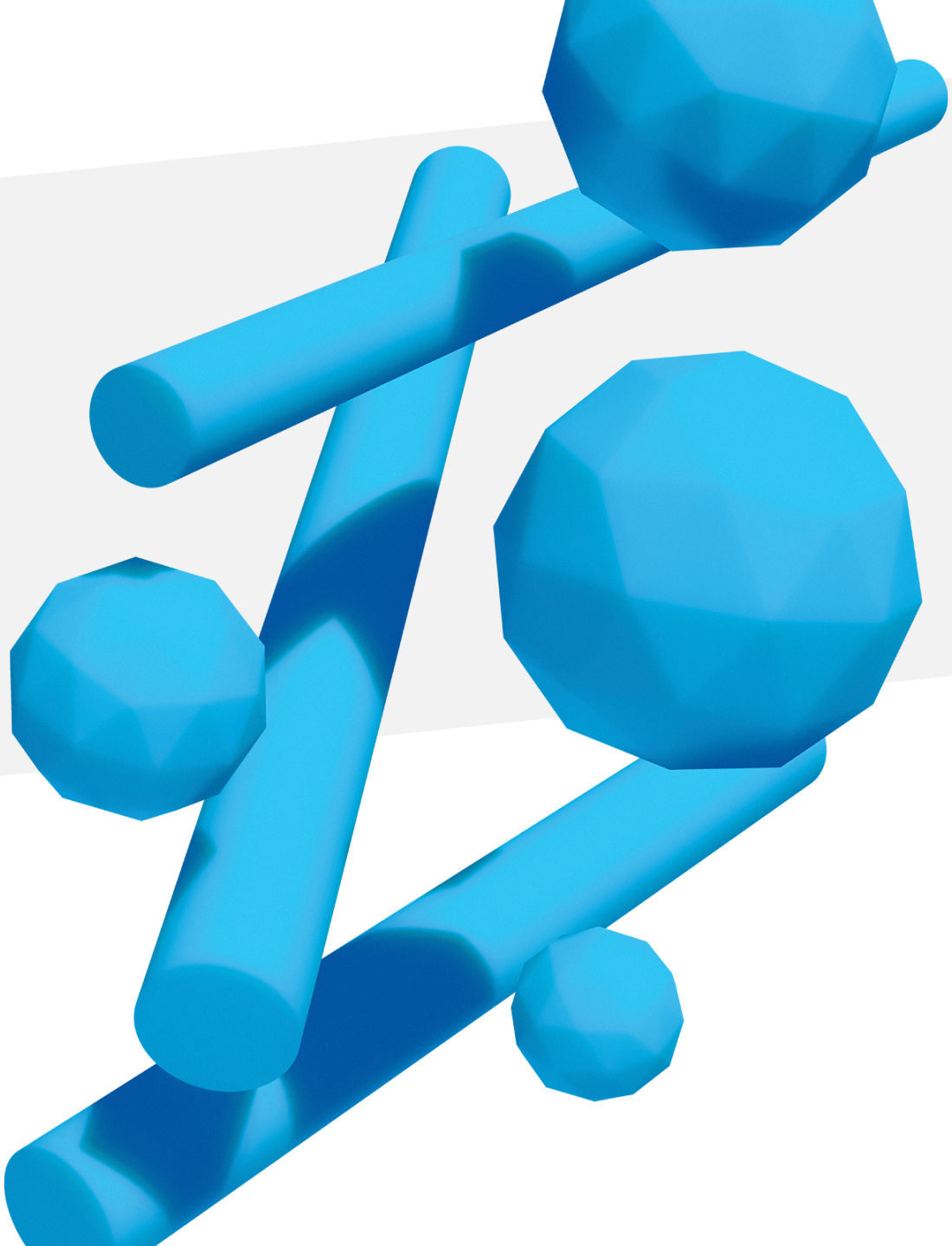
instantiations

instantiations

UnicodeString Implementations



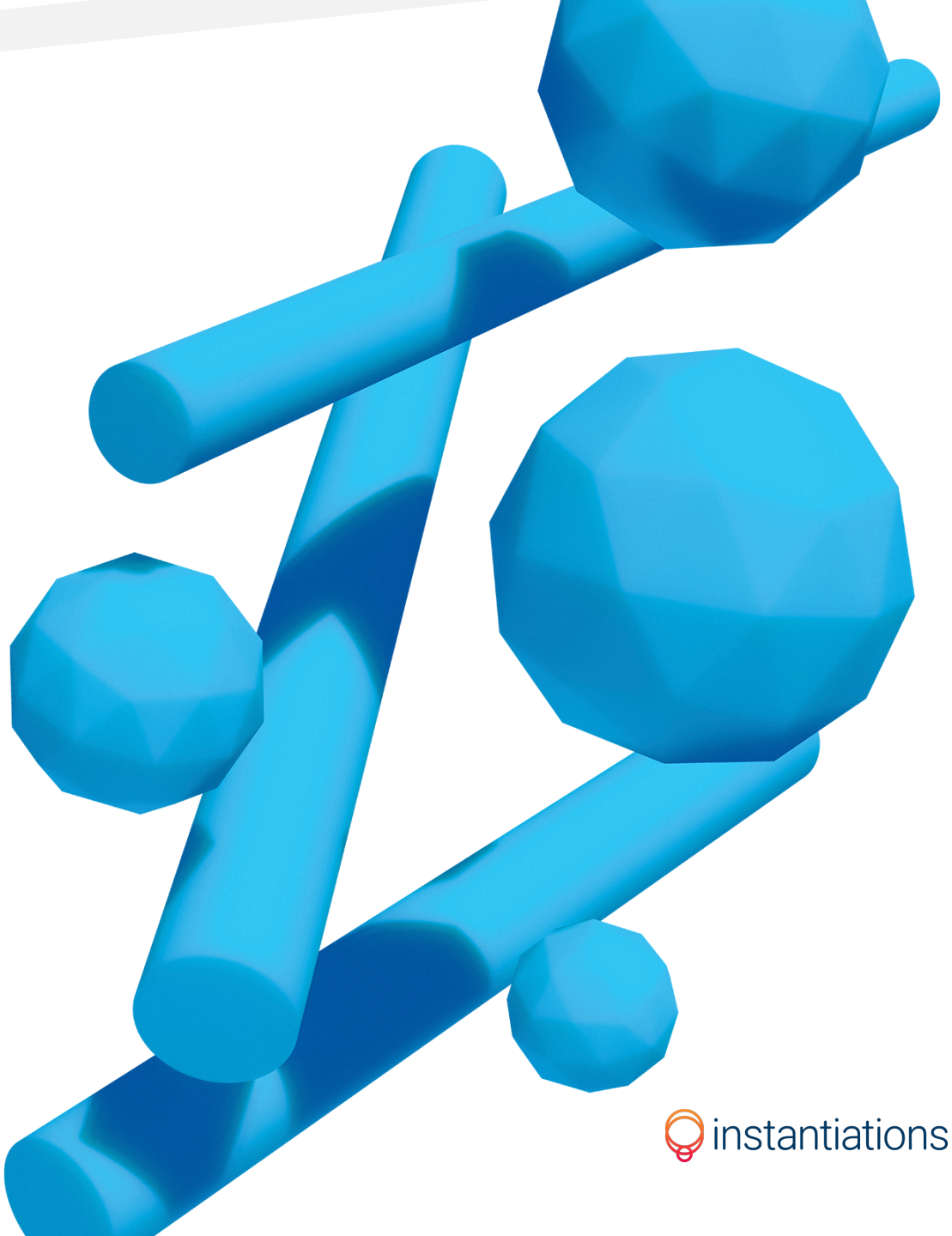
Henrik Sperre Johansen

Developer

hjohansen@instantiations.com

Topics

- Unicode
- Implementations Overview
- Implementation Details



instantiations

Unicode

- Character set
- Encoding
- Normalization
- Equality
- And more!

UnicodeString Overview

Dialect	Encoding	Base unit	Built on	Bytestring
Pharo	UTF-32	Codepoint	Squeak v3	Latin 1
Cuis	UTF-8	Codepoint	svenvc/ UTF8String	Latin 9
VAST	UTF-8	Grapheme	Rust / Swift	System locale

Equality

- Precomposed vs combining marks
- Multiple combining marks

- Normalized forms
- Case-insensitivity

Hashing

- Consistent with equality
- Consistent with single-byte strings

UnicodeString hash	Ascii	Byte	Unicode
Pharo	stringHash:	stringHash:	stringHash:
Cuis	primitiveStringHash + skip optimisation	Convert to byteString Then same as ascii	primitiveStringHash + skip optimisation
VAST	StringHash primitive	Convert to EsString Then same as ascii	seahash

Rendering

- Pharo - FreeType2 library + embedded fonts
- Cuis – VectorCanvas + freetype parser + bundled fonts
- VAST - System rendering

instantiations

Questions?

Henrik Sperre Johansen

Developer

hjohansen@instantiations.com

Contact

General Inquiry
info@instantiations.com

Sales
sales@instantiations.com

VAST Support Portal
vast-support.instantiations.com

North America, Toll Free
855 476 2558

International
+1 503 263 0058

© Instantiations, Inc. All rights reserved. 'Instantiations' and the 'intersecting circle design' are registered trademarks of Instantiations, Inc. in the United States. All product names, trademarks, and registered trademarks are property of their respective owners. Company, product, and service names not owned by Instantiations are used for identification purposes only. Use of these names, trademarks, and brands does not imply endorsement.