

# Rosette Core Library for Unicode

## Portable C++ Class Library for Multilingual Software Development

Rosette® Core Library for Unicode (RCLU) is a widely deployed solution for implementing Unicode compliance; character set transcoding; wide character handling; and basic linguistic transformations.

Before Unicode, supporting another language was like developing a new product. Each language or group of languages needed its own version. With Unicode, all modern written languages are supported under one code base, streamlining development and reducing the need to test multiple language versions.

With RCLU, programmers can quickly add Unicode support to their products. RCLU's extensive C++ classes and ISO C compliant, wide-character run-time library support flexible conversions between Unicode 5.0 and widely used legacy encodings.

### BENEFITS

- Despite the widespread adoption of the Unicode standard, vast amounts of data remain locked in "legacy" text encodings, and Unicode-compliant applications are expected to exchange data in legacy encodings. RCLU "transcodes" text between legacy encodings and Unicode reliably and efficiently.
- Available on a wide range of architectures and operating systems, including Microsoft Windows, Apple MacOS, Linux/Unix, and embedded environments.
- Provides a wide range of character transformations and character property classifiers that aid in functions such as data collation/sorting, normalization, searching, and formatting.

### FEATURES

- *Unicode Development Platform*  
Provides a powerful framework for implementing cross-platform Unicode applications.
- *Character Transcoding*  
Converts text between the Unicode 5.0 Standard and 168 legacy encodings to support the use of non-Unicode-enabled third-party applications or data feeds. Converts between the different Unicode "transformation formats" including UCS-2, UTF-8, and UTF-7. Converts between Unicode 5.0 and earlier versions of the Unicode Standard.
- *Character Classifications, Properties, and Transforms*  
Classifies Unicode characters according to 55 different properties (e.g. uppercase, lowercase, numeric, punctuation, whitespace, etc.) or 38 different types of script (Arabic, Cyrillic, Greek, Hebrew, Hiragana, Katakana, Latin, etc.) Provides a rich library of 37 Unicode string transforms, including cross-script transliteration functions.
- *Lightweight and High Performance*  
High execution speed and low memory requirements. Compact, efficient, and thread-safe. Dynamic loading of encoding tables helps minimize memory footprint.

### ABOUT UNICODE

The Unicode Standard is a major component in the globalization of e-business, as the marketplace demands technologies that enable seamless data interchange throughout international networks of suppliers, customers, and partners. Unicode is the default text representation in XML, an important open standard being rapidly adopted throughout e-business technology.

Unicode assigns every character a unique number, ensuring the same representation for text regardless of country, language, or operating system. As a result, computer programs written to its specifications can be used around the world without modification. Unicode-enabled programs can share textual data freely, without suffering the data loss that occurs with older systems.

Unicode is enabled in all modern Web browsers, operating systems, and Internet standards such as ECMAScript, HTML, Java, LDAP, and XML. For more information, visit [www.unicode.org](http://www.unicode.org).



Basis Technology is a member of the Unicode Consortium and has contributed to the development of the Unicode Standard for over ten years.

## CHARACTER ENCODINGS SUPPORTED

Arabic — CP10004	Hebrew — CP10005	Latin — ISO 8859-15
Arabic — CP1256	Hebrew — CP1255	Latin — ISO 8859-2
Arabic — CP20420	Hebrew — CP28598	Latin — MacRoman
Arabic — CP28596	Hebrew — CP38598	Latin — NextStep
Arabic — CP720	Hebrew — CP862	Latin — Adobe-Standard
Arabic — CP864	Hebrew — ISO 8859-8	Latin, Canadian French — CP863
Arabic — ISO 8859-6	Icelandic — CP10079	Latin, Central European — CP28592
Arabic — CP708	Icelandic — CP861	Latin, Central European — MacCentralEuropean
Baltic — CP1257	Icelandic — MacIcelandic	Latin, Eastern European — CP1250
Baltic — CP28594	Japanese — CCSID 1027	Latin, Esperanto — CP20905
Baltic — CP775	Japanese — CCSID 290	Latin, Portugese — CP860
Baltic — ISO 8859-4	Japanese — CCSID 930	Latin, Southeast European — ISO 8859-3
Baltic — ISO 8859-13	Japanese — CCSID 939	Latin, US English — ASCII
Celtic — ISO 8859-14	Japanese — CCSID 942	Latin, US English — CP037
Chinese, Simplified — CCSID 935	Japanese — CP10001	Latin, US English — CP1026
Chinese, Simplified — EUC-CN	Japanese — CP20290	Latin, US English — CP1252
Chinese, Simplified — GB2312	Japanese — CP21027	Latin, US English — CP20105
Chinese, Simplified — HZ-GB-2312	Japanese — EUC-JP	Latin, US English — CP437
Chinese, Simplified — CP936	Japanese — ISO-2022-JP	Latin, US English — CP500
Chinese, Simplified — MacChineseSimplified	Japanese — JIS_X_0201	Latin, US English — CP875
Chinese, Traditional — CCSID 937	Japanese — JIS_X_0208	Malayalam — CP10017
Chinese, Traditional — CNS-11643-1986	Japanese — MacJapanese	Nordic — CP865
Chinese, Traditional — CNS-11643-1992	Japanese — Shift-JISMS	Nordic — ISO 8859-10
Chinese, Traditional — EUC-TW	Japanese — Shift-JIS78	Romanian — MacRomanian
Chinese, Traditional — GB12345	Korean — CP10003	Slavic — CP852
Chinese, Traditional — Big5	Korean — CP1361	Symbol — Adobe-Symbol
Chinese, Traditional — CP10002	Korean — CP949	Symbol — Adobe-Zapf-Dingbats
Chinese, Traditional — CP950	Korean — EUC-KR	Symbol — CP10008
Chinese, Traditional — MacChineseTraditional	Korean — ISO-2022-KR	Symbol — MacDingbats
Croatian — MacCroatian	Korean — Johab	Symbol — MacSymbol
Cyrillic — CP10007	Korean — KS_C_5601-1987	Thai — CP20838
Cyrillic — CP1251	Korean — KS_C_5861-1992	Thai — CP874
Cyrillic — CP20866	Korean — MacKorean	Thai — ISO 8859-11
Cyrillic — CP20880	Latin — CP10000	Thai — MacThai
Cyrillic — CP21025	Latin — CP10029	Turkish — CP10081
Cyrillic — CP21866	Latin — CP10082	Turkish — CP1254
Cyrillic — CP28595	Latin — CCSID 1047	Turkish — CP28599
Cyrillic — CP855	Latin — CP20261	Turkish — CP857
Cyrillic — CP866	Latin — CP20269	Turkish — ISO 8859-9
Cyrillic — ISO 8859-5	Latin — CP20273	Turkish — MacTurkish
Cyrillic — MacCyrillic	Latin — CP20277	Ukrainian — MacUkrainian
Devanagari — MacDevanagari	Latin — CP20278	Unicode — BMP
Greek — CP10006	Latin — CP20280	Unicode — Java
Greek — CP1253	Latin — CP20284	Unicode — UCS2
Greek — CP20423	Latin — CP20285	Unicode — Unicode Big-endian
Greek — CP28597	Latin — CP20297	Unicode — Unicode Little-endian
Greek — CP737	Latin — CP20833	Unicode — Unicode11-UCS2
Greek — CP869	Latin — CP20871	Unicode — Unicode11-UTF7
Greek — ISO 8859-7	Latin — CP28591	Unicode — Unicode11-UTF8
Greek — MacGreek	Latin — CP28593	Unicode — UTF7
Gujarati — MacGujarati	Latin — CP850	Unicode — UTF8
Gurmukhi — CP10010	Latin — CP870	Unicode — UTF-EBCDIC
Gurmukhi — MacGurmukhi	Latin — ISO 8859-1	Vietnamese — CP1258

## SYSTEM PLATFORMS SUPPORTED

Software development kits (SDKs) and web services are available for the platforms listed below. Contact your sales representative for additional platform support.

AIX 6.1, PPC  
HP-UX 11i, IA64  
Linux CentOS 4.x/5.x, IA32/AMD64  
Linux Debian 5.x, IA32/AMD64  
Linux Red Hat 4.x/5.x, IA32/AMD64

Linux Ubuntu 10.x/11.x, IA32/AMD64  
MacOS  
Solaris 10, SPARC32/64, IA32/AMD64  
Windows XP/Vista/7, IA32/AMD64  
Windows Server 2003, 2008

VISIT [www.basistech.com](http://www.basistech.com) WRITE [info2011@basistech.com](mailto:info2011@basistech.com) CALL 617-386-2090

One Alewife Center  
Cambridge, MA 02140

171 Second Street  
San Francisco, CA 94105

2553 Dulles View Drive  
Herndon, VA 20171

9-6 Nibancho, Chiyoda-ku  
Tokyo 102-0084

