



## Rosette® 言語処理プラットフォーム v7.2仕様

2010年6月 リリース

Rosette® 形態素解析システム (RBL)

	Rosette 形態素解析システム (RBL)							REX	RNI	RNT
	分かち書き	基本化	複合語分解	品詞タグづけ	文の区切り検出	名詞句抽出	読み	固有表現抽出	人名・地名照合	人名・地名翻訳
日本語	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
中国語(簡体字/繁体字)	Y	n/a	Y	Y	Y	Y	Y	Y	Y	Y
韓国語	Y	Y	Y	Y	Y	N	N	Y	Y	Y
アラビア語	Y	Y	n/a	Y	Y	Y	N	Y	Y	Y
パシュトゥ語	Y	N	N	N	Y	N	N	Y	Y	Y
近代ベルシア語	Y	Y	n/a	N	Y	N	N	Y	Y	Y
ウルドゥ語	Y	Y	n/a	N	Y	N	N	Y	Y	Y
英語	Y	Y	n/a	Y	Y	Y	n/a	Y	Y	n/a
フランス語	Y	Y	n/a	Y	Y	Y	N	Y	N	N
イタリア語	Y	Y	n/a	Y	Y	Y	N	Y	N	N
ドイツ語	Y	Y	Y	Y	Y	Y	N	Y	N	N
スペイン語	Y	Y	n/a	Y	Y	Y	N	Y	N	N
オランダ語	Y	Y	Y	Y	Y	Y	N	Y	N	N
ポルトガル語	Y	Y	n/a	Y	Y	Y	N	Y	N	N
ロシア語	Y	Y	n/a	Y	Y	N	N	Y	N	Y
チェコ語	Y	Y	n/a	Y	Y	N	N	Y	N	N
ギリシャ語	Y	Y	n/a	Y	Y	N	N	Y	N	N
ポーランド語	Y	Y	n/a	Y	Y	N	N	Y	N	N
ハンガリー語	Y	Y	Y	Y	Y	N	N	Y	N	N

n/a — この言語には該当せず

# Rosette® 固有抽出表現システム (REX)

	統計的 / ガゼティア								正規表現												
	人名	組織名	場所	都市名	施設	宗教	国籍	役職	クレジットカード番号	距離	Eメール	緯度 / 経度	通貨	数値	住民登録番号	電話番号	URL	UTM図法	日付	時間	
日本語	S	S/G	S	--	--	G	G	S	R	R	R	R	R	R	R	R	R	R	R	R	R
中国語(簡/繁体字)	S	S/G	S	--	--	G	G	S	R	R	R	R	R	R	R	R	R	R	R	R	R
韓国語	S	S/G	S	S	S	G	G	S	R	R	R	R	R	--	R	R	R	R	R	R	R
アラビア語	S	S/G	S/G	--	--	G	G	S	R	R	R	R	R	R	R	R	R	R	R	R	R
近代ペルシア語	S	S/G	S	--	--	G	G	G	R	R	R	R	R	R	R	R	R	R	R	R	R
ウルドゥ語	S	S/G	S	S	S	S/G	S/G	G	R	--	R	--	R	--	R	R	R	R	--	--	--
パシュトゥ語	S	S/G	S	--	--	--	--	S	R	R	R	R	R	R	R	R	R	R	R	R	R
英語	S	S/G	S/G	--	--	G	G	S	R	R	R	R	R	R	R	R	R	R	R	R	R
英語(大文字)	S	S/G	S/G	--	--	G	G	S	R	R	R	R	R	R	R	R	R	R	R	R	R
フランス語	S	S/G	S	--	--	--	--	S	R	R	R	R	R	R	R	R	R	R	R	R	R
イタリア語	S	S/G	S	--	--	--	--	S	R	R	R	R	R	R	R	R	R	R	R	R	R
ドイツ語	S	S/G	S	--	--	--	--	S	R	R	R	R	R	R	R	R	R	R	R	R	R
スペイン語	S	S/G	S	--	--	--	--	S	R	R	R	R	R	R	R	R	R	R	R	R	R
オランダ語	S	S/G	S	--	--	--	--	G	R	R	R	R	R	R	R	R	R	R	R	R	R
ポルトガル語	--	G	--	--	--	--	--	G	R	R	R	R	R	R	R	R	R	R	R	R	R
ハンガリー語	--	G	--	--	--	--	--	--	R	--	R	R	R	--	R	R	R	R	R	R	--
チェコ語	--	G	--	--	--	--	--	--	R	--	R	R	R	--	R	R	R	R	R	R	--
ギリシャ語	--	G	--	--	--	--	--	--	R	--	R	R	R	--	R	R	R	R	R	R	--
ロシア語	S	S/G	S	--	--	G	G	S/G	R	R	R	R	R	R	R	R	R	R	R	R	R
ポーランド語	--	G	--	--	--	--	--	--	R	--	R	R	R	--	R	R	R	R	R	R	--

S 統計的モデル  
G ガゼティア  
R 正規表現ルール

# Rosette® 言語判別システム (RLI)

言語 56, plus 8 Latin-script variants  
 言語と文字コードのペア 194  
 文字コード 44

対応言語	文字コード (UTF-8は全部の言語をサポート)	文字コード数	文字コード
アルバニア語	Windows-1252, ISO-8859-1	3	Big5
アラビア語(アラビア文字)	Windows-1256, Windows-720, ISO-8859-6	4	EUC-JP
アラビア語(ラテン文字)	Windows-1252, ISO-8859-1, Windows-1256	4	EUC-KR
ベンガル語	ISCII-Bengali	2	GB18030
ブルガリア語	Windows-1251, KOI8-R, ISO-8859-5	4	GB2312
カタロニア語	Windows-1252, ISO-8859-1	3	HZ-GB-2312
中国語(簡体字)	HZ-GB-2312, GB18030, GB2312, ISO-2022-CN	5	IBM866
中国語(繁体字)	Big5	2	ISCII-Bengali
クロアチア語	Windows-1250, ISO-8859-2	3	ISCII-Devanagari
チェコ語	Windows-1250, ISO-8859-2	3	ISCII-Gujarati
デンマーク語	Windows-1252, ISO-8859-1	3	ISCII-Kannada
オランダ語	Windows-1252, ISO-8859-1	3	ISCII-Malayalam
英語	Windows-1252, ISO-8859-1	3	ISCII-Tamil
英語 (大文字)	Windows-1252, ISO-8859-1	3	ISCII-Telugu
エストニア語	Windows-1257, ISO-8859-13	3	ISO-2022-CN
フィンランド語	Windows-1252, ISO-8859-1	3	ISO-2022-JP
フランス語	Windows-1252, ISO-8859-1	3	ISO-2022-KR
ドイツ語	Windows-1252, ISO-8859-1	3	ISO-8859-1
ギリシャ語	Windows-1253, ISO-8859-7	3	ISO-8859-2
グジャラート語	ISCII-Gujarati	2	ISO-8859-5
ヘブライ語	Windows-1255, ISO-8859-8	3	ISO-8859-6
ヒンディー語	ISCII-Devanagari	2	ISO-8859-7
ハンガリー語	Windows-1250, ISO-8859-2	3	ISO-8859-8
アイスランド語	Windows-1252, ISO-8859-1	3	ISO-8859-9
インドネシア語	Windows-1252, ISO-8859-1	3	ISO-8859-13
イタリア語	Windows-1252, ISO-8859-1	3	KOI8-R
日本語	EUC-JP, Shift_JIS-2004, Shift_JIS, ISO-2022-JP	5	Shift_JIS
カナダ語	ISCII-Kannada	2	Shift_JIS-2004
韓国語	EUC-KR, ISO-2022-KR	3	TCVN
クルド語(アラビア文字)	Windows-1256	2	Utf-8
クルド語(ラテン文字)	Windows-1252, ISO-8859-1, Windows-1256	4	VIQR
ラトビア語	Windows-1257, ISO-8859-13	3	VISCII
リトアニア語	Windows-1257, ISO-8859-13	3	VNI
マケドニア語	Windows-1251, ISO8859-5	3	VPS
マレー語	Windows-1252, ISO-8859-1	3	Windows-1250
マラーヤラム語	ISCII-Malayalam	2	Windows-1251
ノルウエー語	Windows-1252, ISO-8859-1	3	Windows-1252
バシトゥ語(アラビア文字)	Windows-1256	2	Windows-1253
バシトゥ語(ラテン文字)	Windows-1252, ISO-8859-1, Windows-1256	4	Windows-1254
近代ベルシア語(アラビア文字)	Windows-1256	2	Windows-1255
近代ベルシア語(ラテン文字)	Windows-1252, ISO-8859-1, Windows-1256	4	Windows-1256
ポーランド語	Windows-1250, ISO-8859-2	3	Windows-1257
ポルトガル語	Windows-1252, ISO-8859-1	3	Windows-720
ルーマニア語	Windows-1250, ISO-8859-2	3	Windows-874
ロシア語	Windows-1251, KOI8-R, ISO-8859-5, IBM866, x-mac-cyrillic	6	x-mac-cyrillic
セルビア語(キリル文字)	Windows-1251, ISO-8859-5	3	
セルビア語(ラテン文字)	Windows-1250, ISO-8859-2	3	
スロバキア語	Windows-1250, ISO-8859-2	3	
スロベニア語	Windows-1250, ISO-8859-2	3	
ソマリ語	Windows-1252, ISO-8859-1	3	
スペイン語	Windows-1252, ISO-8859-1	3	
スウェーデン語	Windows-1252, ISO-8859-1	3	
タガログ語	Windows-1252, ISO-8859-1	3	
タミル語	ISCII-Tamil	2	
テルグ語	ISCII-Telugu	2	
タイ語	Windows-874	2	
トルコ語	Windows-1254, ISO-8859-9	3	
ウクライナ語	Windows-1251, KOI8-R, ISO-8859-5	4	
ウルドゥ語(アラビア文字)	Windows-1256	2	
ウルドゥ語(ラテン文字)	Windows-1252, ISO-8859-1, Windows-1256	4	
ウズベキスタン語(キリル文字)	Windows-1251, KOI8-R, ISO-8859-5	4	
ウズベキスタン語(ラテン文字)	Windows-1251	2	
ベトナム語	TCVN, VIQR, VISCII, VPS, VNI	6	

言語と文字コードのペア 194

## 対応ハードウェアおよびOS

OS	バージョン	アーキテクチャ	コンパイラ	バージョン
AIX	6.1	PowerPC 32-bit	xlc	
AIX	6.1	PowerPC 64-bit	xlc	
HP-UX	11i (V3)	IPF 32-bit	aCC	
HP-UX	11i (V3)	IPF 64-bit	aCC	
Mac OS	10.5 (Darwin 9)	gcc	aCC	4.0
Linux Red Hat	ES 3.0	IA32	gcc	3.2
Linux Red Hat	ES 4.0, 5.0	IA32	gcc	3.4.4
Linux Red Hat	ES 4.0, 5.0	AMD64	gcc	3.4
Solaris	10	AMD64	CC	5.8
Solaris	10	IA32	CC	5.8
Solaris	10	SPARC32	CC	5.8
Solaris	10	SPARC64	CC	5.8
Solaris	9	SPARC32	CC	5.8
Solaris	9	SPARC64	CC	5.8
Windows 32-bit	XP, 2003, Vista, Server2008, 7	IA32	Visual C++	8.0
Windows 64-bit	XP, 2003, Vista, Server2008, 7	AMD64	Visual C++	8.0

### SDK Distribution

C++ platform  
 DLL implementation  
 Java JNI and native Java APIs  
 .NET (Windows)