

# International Sample

## 1. Languages and Writing System

PDFreactor can layout texts in various languages and writing systems. It also contains default fonts for most of them, so providing fonts containing the required characters is not required. However, doing so is still recommended for more individualized documents.

Please note that some scripts require Java 17+ for best results.

The following samples show text from various languages with multiple kinds of writing systems using the default fonts:

### Arabic

نص حكيم له سر قاطع وذو شأن عظيم  
مكتوب على ثوب أخضر ومغلف بجلد  
أزرق.

*A wise text which has an absolute secret and great importance, written on a green cloth and covered with blue leather.*

### Chinese (simplified)

桃花潭水深千尺，不及汪伦送我情

*The waters of Peach Blossom Pool are a thousand feet deep, yet not as deep as Wang Lun's parting love for me.  
(Li Bai)*

### Danish

Quizdeltagerne spiste jordbær  
med fløde, mens cirkusklovnen  
Walther spillede på xylofon.

*The quiz contestants ate strawberry with cream while Walter the circus clown played the xylophone.*

### German

Victor jagt zwölf Boxkämpfer  
quer über den großen Sylter  
Deich.

*Victor chases twelve boxers across the great dam of Sylt.*

### Greek

Θέλει αρετή και τόλμη η  
ελευθερία.  
(Ανδρέας Κάλβος)

*Liberty requires virtue and mettle.  
(Andreas Kalvos)*

### Hebrew

דג סקרן שט לו בים זך אך לפתע  
פגש חבורה נחמדה שצצה כך.

*A curious fish sailed a clear sea, and suddenly found nice company that just popped up.*

### Japanese

いろはにほへとちりぬるをわか  
よたれそつねならむうゑのおく  
やまけふこえてあさきゆめみし  
ゑひもせず

*Even the blossoming flowers / Will eventually scatter / Who in this world / is unchanging? / The deep mountains of vanity-- / We cross them today / And we shall not see superficial dreams / Nor be deluded. (from Iroha-uta)*

### Khmer

ថ្ងៃច័ន្ទ ដល់ ថ្ងៃសុក្រ ម៉ោង ៨ ព្រឹក

*Monday to Friday, 8 AM*

### Russian

Съешь ещё этих мягких  
французских булок да выпей  
же чаю.

*Eat some more of these soft French buns and drink some tea.*

## 2. Bidi Text

Text can not only be strictly left-to-right or right-to-left, it can also contain phrases with an inherent direction opposite to the base direction. PDFreactor automatically handles such bi-direction (BiDi) content according to the Unicode specification, with no need for specific styles beyond the base direction (see below).

The following samples show an English text followed by translations that keep some English terms:

### Left-to-right sample - English text excerpt:

#### 5. XML is a family of technologies

XML 1.0 is the specification that defines what "tags" and "attributes" are. Beyond XML 1.0, "the XML family" is a growing set of modules that offer useful services to accomplish important and frequently demanded tasks. XLink describes a standard way to add hyperlinks to an XML file. XPointer is a syntax in development for pointing to parts of an XML document. An XPointer is a bit like a URL, but instead of pointing to documents on the Web, it points to pieces of data inside an XML file. CSS, the style sheet language, is applicable to XML as it is to HTML. XSL is the advanced language for expressing style sheets. It is based on XSLT, a transformation language used for rearranging, adding and deleting tags and attributes. The DOM is a standard set of function calls for manipulating XML (and HTML) files from a programming language. XML Schemas 1 and 2 help developers to precisely define the structures of their own XML-based formats. There are several more modules and tools available or under development. Keep an eye on W3C's technical reports page.

### Right-to-left sample - Arabic text excerpt:

#### 5. إكس إم إل مجموعة تكنولوجيات

XML 1.0 هي المواصفات التي تعرّف ماذا تكون "العلامات" و "الصفات". حول تلك المواصفات هناك مجموعة متنامية من الوحدات التي تعرّض خدمات مفيدة لإنجاز مهام كثيرة الطلب وهامة. Xlink يصف طريقة رسمية لإضافة وُصلة رابطة لملف إكس إم إل. XPointer و XFragments قواعد تركيب نحوية للإشارة إلى أجزاء من وثيقة إكس إم إل. إشارة XPointer تتشابه مع إشارة URL، لكنّ بدلاً من الإشارة إلى الوثائق على الويب، تشير إلى قطع صغيرة من البيانات داخل ملفّ إكس إم إل. CSS، لغة النمط والأسلوب، تنطبق على إكس إم إل كما تنطبق على إتش تي إم إل. XSL هي اللغة المتقدّمة للتعبير عن النمط والأسلوب. أساسها XSLT، لغة تحوّل استُخدمت لإعادة ترتيب أو لإضافة أو لحذف علامات أو صفات. DOM هي مجموعة رسمية لاستدعاءات الوظائف من أجل استغلال ملقات إكس إم إل (أو إتش تي إم إل) و ذلك من داخل لغة برمجة. مخططات إكس إم إل 1 و 2 (XML Schemas 1 and 2) تساعد المطورين بأن يُعرّفوا بدقّة هياكل أشكالهم الشّخصية المؤسّسة على لغة إكس إم إل. هناك عدة وحدات و أدوات متاحة أو تحت التطوير. أنظر إلى صفحة تقارير W3C التّقنية.

### Right-to-left sample - Hebrew text excerpt:

#### 5. XML הוא משפחה של טכנולוגיות

XML 1.0 הנו מפרט שמגדיר מהם "תגיות" ו"תכונות". מעבר לכך "משפחת XML" הינה מערך מודולים ההולך ומתרחב המציעים שירותים שמושיים כדי לבצע משימות חשובות ותכופות. Xlink מתאר כיצד להוסיף קישורים לקובץ XPointer. XML - XPointer הנו מעין URL, אך במקום להצביע על מסמכים ברשת, הוא מצביע על פיסת מידע בתוך הקובץ CSS. XML, שפה לעיצוב גליונות, זמינה ב- XML בדומה ל- HTML. XSL הינה שפה מתקדמת לעיצוב גליונות. השפה מבוססת על XSLT, שפה המשמשת לאירגון, הוספה ומחיקה של תגיות ותכונות. ה- DOM מערך פונקציות המשמשות למניפולציות בקבצי XML (ו- HTML) ע"י תכנות. פורמט הגדרות מבנה XML Schema 1 ו- 2. ישנם עוד מודולים וכלים מוכנים או בפיתוח. שים לב לדף הדיווחים הטכניים של W3C.

The text excerpts are taken from the original English and the translated Arabic and Hebrew versions of the W3C document "XML in 10 Points".

## 3. BiDi Layout

When the direction of the primary writing system of a document is right-to-left the base direction of the document should be set to RTL as well:

```
:root { direction: rtl; }
```

`direction` can also be set on any element and also affect its children.

In addition to text layout, `direction` also adjusts lists, tables and other structures, as demonstrated by the following juxtaposition:

### Paragraph:

Some text in a paragraph.

### :Paragraph

.Some text in a paragraph

### List:

- Item 1
- Item 2
- Item 3

### :List

- Item 1 •
- Item 2 •
- Item 3 •

### Table:

H1	H2	H3
A1	A2	A3
B1	B2	B3
C1	C2	C3

### :Table

H3	H2	H1
A3	A2	A1
B3	B2	B1
C3	C2	C1

### Multi-Column:

Line 1	Line 4	Line 6
Line 2	Line 5 has a break	Line 7
Line 3		Line 8

### :Multi-Column

Line 6	Line 4	Line 1
Line 7	Line 5 has a break	Line 2
Line 8		Line 3

### Flex:

Item 1	Item 2	Item 3	Item 4
--------	--------	--------	--------

### :Flex

Item 4	Item 3	Item 2	Item 1
--------	--------	--------	--------

### Grid:

Item 1	Item 2	Item 3
Item 4	Item 5	Item 6

### :Grid

Item 3	Item 2	Item 1
Item 6	Item 5	Item 4

## 4. Logical Properties and Values

Several horizontal positions and sizes can be specified depending on the text direction. The following sample uses the exact same HTML and CSS twice, except for different BiDi directions. The direction dependent styles and their effects are as follows:

The position of the title:

```
position: absolute;
top: -1em;
inset-inline-start: 2cm;          /* LTR: "left: 2em" / RTL: "right: 2em" */
```

The float value and margin of the image:

```
float: inline-start;           /* LTR: "float: left" / RTL: "float: right" */
margin-top: 0.2em;
margin-inline-end: 1em;       /* LTR: "margin-right: 1em" / RTL: "margin-left: 1em" */
```


The border and padding of the box:

```
border: 1pt none darkgrey;
border-top-style: solid;
padding-top: 0.5em;
border-inline-start-style: solid; /* LTR: "border-left-style..." / RTL: "border-right-style..." */
padding-inline-start: 1em;       /* LTR: "padding-left: 1em" / RTL: "padding-right: 1em" */
```

Resulting in:

### LTR

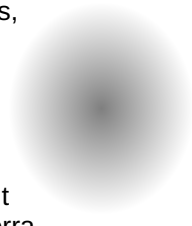
Title


 Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed lacinia ex eget nisl iaculis, vitae pellentesque sapien sollicitudin. Mauris iaculis, sem at laoreet tincidunt, ipsum justo placerat turpis, vitae finibus ipsum sapien ac sapien. Nullam dignissim finibus erat. Donec tincidunt scelerisque enim, et pharetra dolor blandit id. Fusce blandit sapien sollicitudin leo auctor, a mattis ex ullamcorper. Aenean hendrerit in sem et sodales. Nunc maximus sem non eros venenatis, cursus egestas mi imperdiet. Morbi in eros faucibus, ullamcorper sem eget, facilisis nunc. Vestibulum ut dignissim ligula, vel sagittis velit. Ut eget maximus leo. Aenean rhoncus euismod elementum. Nam velit orci, porttitor vel viverra scelerisque, accumsan sit amet urna. Quisque ut interdum dui. Suspendisse eu leo orci. Ut nec eros vel diam euismod consequat.

### RTL

Title

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed lacinia ex eget nisl iaculis, vitae pellentesque sapien sollicitudin. Mauris iaculis, sem at laoreet tincidunt, ipsum justo placerat turpis, vitae finibus ipsum sapien ac sapien. Nullam dignissim finibus erat. Donec tincidunt scelerisque enim, et pharetra dolor blandit id. Fusce blandit sapien sollicitudin leo auctor, a mattis ex ullamcorper. Aenean hendrerit in sem et sodales. Nunc maximus sem non eros venenatis, cursus egestas mi imperdiet. Morbi in eros faucibus, ullamcorper sem eget, facilisis nunc. Vestibulum ut dignissim ligula, vel sagittis velit. Ut eget maximus leo. Aenean rhoncus euismod elementum. Nam velit orci, porttitor vel viverra scelerisque, accumsan sit amet urna. Quisque ut interdum dui. Suspendisse eu leo orci. Ut nec eros vel .diam euismod consequat



## 5. International List Numbering

PDFReactor supports several international list style types, including the following:

1. decimal	二十一、 japanese-informal
02. padded decimal	೨೨. kannada
iii. lower-roman	又、 katakana
IV. upper-roman	ウ、 katakana-iroha
e. lower-alpha	ឃ៥. khmer
F. upper-alpha	໒໖. lao
V. arabic-indic	൨൭. malayalam
Ը. upper-armenian	୨୮. oriya
թ. lower-armenian	۲۹. persian
১০. bengali	叁拾、 simp-chinese-formal
১১. cambodian	三十一、 simp-chinese-informal
१२. devanagari	౩౨. telugu
Ⴍ. georgian	๓๓. thai
ξ. lower-greek	མ༤. tibetan
Ο. upper-greek	۳۵. urdu
૧૬. gujarati	##### footnote
ੴ. gurmukhi	thirty-seven spelled out English
そ、 hiragana-iroha	achtunddreißig spelled out German
て、 hiragana	trente-neuf spelled out French
貳拾、 japanese-formal	fortieth spelled out English ordinal

### Changing Counter-Style Based on "lang" Attribute

With the `:lang()` selector, different styles can be set based on the "lang" attribute that is set on an HTML element. In this example the differing "lang" attributes set, cause the spelled out counter style to change to the language set on the list element.

```
.langSelector:lang(de) {
  content: counter(itemCounter, spelledOutGerman);
}

.langSelector:lang(en) {
  content: counter(itemCounter, spelledOutEnglish);
}

.langSelector {
  content: counter(itemCounter);
}
```

#### English "lang" Attribute

one, two, three, four, five, six, seven, eight

#### German "lang" Attribute

eins, zwei, drei, vier, fünf, sechs, sieben, acht

#### Other "lang" Attribute

1, 2, 3, 4, 5, 6, 7, 8

## 6. Creating Custom List Style Types

PDFReactor supports creating custom list style types with the `@counter-style` rule. If the defined counter style is not valid, the style will fallback to the decimal counter-style. This counter style rule, for example, depicts dice:

```
@counter-style dice {
  system: additive;
  additive-symbols: 6 '\2685', 5 '\2684', 4 '\2683', 3 '\2682', 2 '\2681', 1 '\2680';
}
```

- ❶ dice 1
- ❷ dice 2
- ❸ dice 3
- ❹ dice 4
- ❺ dice 5
- ❻ dice 6
- ❼❶ dice 7
- ❼❷ dice 8
- ❼❸ dice 9

The `@counter-style` rule also allows you to extend already existing counter styles and change properties like suffix, prefix or the character used for negative values. For example the following counter style changes the suffix of the decimal counter style to ") ". The trailing space, to offset the list marker from the content, is necessary when the suffix is specified.

```
@counter-style decimal-suffix {
  system: extends decimal;
  suffix: ") ";
}
```

- 1) changed suffix
- 2) changed suffix
- 3) changed suffix
- 4) changed suffix

Another example for the extends system would be to spell out low values and fallback to decimal for higher ones. In English the fallback would be made after ten, while in German it would be made after 12. The following `@counter-style` rules show these two examples:

```
@counter-style spelledOutEnglish {
  system: extends -ro-spelled-out-en;
  range: 0 10;
  fallback: decimal;
}

@counter-style spelledOutGerman {
  system: extends -ro-spelled-out-de;
  range: 0 12;
  fallback: decimal;
}
```

### Spelled-Out English:

one, two, three, four, five, six, seven, eight, nine, ten, 11, 12, 13, 14, 15

### Spelled-Out German:

eins, zwei, drei, vier, fünf, sechs, sieben, acht, neun, zehn, elf, zwölf, 13, 14, 15

## 7. International Date Formats in JavaScript

PDFReactor supports date formatting via the `Intl` JavaScript object:

```
dtf = new Intl.DateTimeFormat("en-US", { month: "numeric", day: "numeric", weekday: "long" });
fd = dtf.format(new Date(2020, 0, 15)); // results in "Wednesday, 1/15"
```

The following table shows two sample date formats for various languages and variants:

Language	Country	Code	Short Date	Long Date
English	USA	en-US	1/15/2020, 6:23:45 PM	Wednesday, January 15, 2020 Anno Domini at 6:23:45 PM Greenwich Mean Time
English	Canada	en-CA	2020-01-15, 6:23:45 p.m.	Wednesday, January 15, 2020 Anno Domini at 6:23:45 p.m. Greenwich Mean Time
English	Australia	en-AU	15/01/2020, 6:23:45 pm	Wednesday 15 January 2020 Anno Domini at 6:23:45 pm Greenwich Mean Time
English	UK	en-GB	15/01/2020, 18:23:45	Wednesday, 15 January 2020 Anno Domini at 18:23:45 Greenwich Mean Time
Spanish	Spain	es-ES	15/1/2020, 18:23:45	miércoles, 15 de enero de 2020 después de Cristo, 18:23:45 (hora del meridiano de Greenwich)
Spanish	Mexico	es-MX	15/1/2020, 6:23:45 p.m.	miércoles, 15 de enero de 2020 después de Cristo, 6:23:45 p.m. hora del meridiano de Greenwich
Spanish	Argentina	es-AR	15/1/2020, 06:23:45	miércoles, 15 de enero de 2020 después de Cristo, 6:23:45 p. m. hora del meridiano de Greenwich
French	France	fr-FR	15/01/2020 18:23:45	mercredi 15 janvier 2020 après Jésus-Christ à 18:23:45 heure moyenne de Greenwich
Italian	Italy	it-IT	15/01/2020, 18:23:45	mercoledì 15 gennaio 2020 dopo Cristo alle ore 18:23:45 Ora del meridiano di Greenwich
German	Germany	de-DE	15.1.2020, 18:23:45	Mittwoch, 15. Januar 2020 n. Chr. um 18:23:45 Mittlere Greenwich-Zeit
German	Austria	de-AT	15.1.2020, 18:23:45	Mittwoch, 15. Jänner 2020 n. Chr. um 18:23:45 Mittlere Greenwich-Zeit
Dutch	Netherlands	nl-NL	15-1-2020, 18:23:45	woensdag 15 januari 2020 na Christus om 18:23:45 Greenwich Mean Time
Dutch	Belgium	nl-BE	15/1/2020, 18:23:45	woensdag 15 januari 2020 na Christus om 18:23:45 Greenwich Mean Time

Language	Country	Code	Short Date	Long Date
Danish	Denmark	da-DK	15.1.2020, 18.23.45	onsdag 15. januar 2020 efter Kristus kl. 18.23.45 GMT
Norwegian	Norway	nb-NO	15.1.2020, 18:23:45	onsdag 15. januar 2020 etter Kristus kl. 18:23:45 Greenwich middeltid
Portuguese	Brazil	pt-BR	15/01/2020, 18:23:45	quarta-feira, 15 de janeiro de 2020 depois de Cristo às 18:23:45 Horário do Meridiano de Greenwich
Portuguese	Portugal	pt-PT	15/01/2020, 18:23:45	quarta-feira, 15 de janeiro de 2020 depois de Cristo às 18:23:45 Hora de Greenwich
Russian	Russia	ru-RU	15.01.2020, 18:23:45	среда, 15 января 2020 г. от Рождества Христова в 18:23:45 Среднее время по Гринвичу
Chinese	China	zh-CN	2020/1/15 18:23:45	公元2020年1月15日星期三 格林尼治标准时间 18:23:45
Chinese	Hong Kong	zh-HK	15/1/2020 下午6:23:45	公元2020年1月15日星期三 下午6:23:45 [格林威治標準時間]
Chinese	Taiwan	zh-TW	2020/1/15 下午6:23:45	西元2020年1月15日 星期三 下午6:23:45 [格林威治標準時間]
Japanese	Japan	ja-JP	2020/1/15 18:23:45	西曆2020年1月15日水曜日 18時23分45秒 グリニッジ標準時
Hindi	India	hi-IN	15/1/2020, 6:23:45 pm	बुधवार, 15 जनवरी 2020 ईसवी सन को 6:23:45 pm ग्रीनविच मीन टाइम बजे

The Date used in the table is 2020-01-15T18:23:45.678Z.

The options are:

```

shortDate = {
    year: "numeric",
    month: "numeric",
    day: "numeric",

    hour: "numeric",
    minute: "numeric",
    second: "numeric"
};

longDate = {
    era: "long",
    year: "numeric",
    month: "long",
    day: "numeric",
    weekday: "long",
    hour: "numeric",
    minute: "numeric",
    second: "numeric",
    timeZoneName: "long"
};

```



## 8. Transliteration

Text can be transliterated, i.e. converted between writing systems to make it readable to the audience or to avoid characters not available in specific fonts. For the latter case further characters can also be simplified.

```
/* transliterate from other writing systems to Latin */
:root {
  -ro-text-replace: 'any' 'latn' transliterate;
}


/* avoid non-ASCII characters whenever possible */
:root {
  -ro-text-replace: 'any' 'latn' transliterate, 'any' 'ascii' transliterate;
}
```

Description	normal	'any' 'latn'	'any' 'ascii'	'de' 'ascii'	'any' 'latn' 'any' 'ascii'
German	Gänsefüßchen	Gänsefüßchen	Gansefusschen	Gaensefuesschen	Gansefusschen
Russian	русский язык	russkij âzyk	русский язык	русский язык	russkij azyk
Greek	Ελληνικά	Ellēniká	Ελληνικά	Ελληνικά	Ellenika
Japanese	日本語	rì bĕn yŭ	日本語	日本語	ri ben yu
Math	$6 \div 3 \times \frac{1}{4} = \frac{1}{2}$	$6 \div 3 \times \frac{1}{4} = \frac{1}{2}$	$6 / 3 * 1/4 = 1/2$	$6 / 3 * 1/4 = 1/2$	$6 / 3 * 1/4 = 1/2$
Symbols	© « ® »	© « ® »	(C) << (R) >>	(C) << (R) >>	(C) << (R) >>

## 9. Emojis

PDFreactor supports emojis, providing two default fonts, one for color emojis and one with monochrome versions. The default is color emojis, unless the character is covered by the text font. Either font can be specified, before or after the text font of your choice (here sans-serif and serif for simplicity):

	* 5 M © !! TM i	🕒 📺 😞 😊 ♀ 🍓 🐰 👍 👎 🛒
font-family: sans-serif, -ro-color-emoji	* 5 M © !! TM i	🕒 📺 😞 😊 ♀ 🍓 🐰 👍 👎 🛒
font-family: serif, -ro-color-emoji	* 5 M © !! TM i	🕒 📺 😞 😊 ♀ 🍓 🐰 👍 👎 🛒
font-family: -ro-color-emoji, sans-serif	* 5 M © !! TM i	🕒 📺 😞 😊 ♀ 🍓 🐰 👍 👎 🛒
font-family: -ro-color-emoji, serif	* 5 M © !! TM i	🕒 📺 😞 😊 ♀ 🍓 🐰 👍 👎 🛒
font-family: -ro-color-emoji	* 5 M © !! TM i	🕒 📺 😞 😊 ♀ 🍓 🐰 👍 👎 🛒
font-family: sans-serif, -ro-emoji	* 5 M © !! TM i	🕒 📺 😞 😊 ♀ 🍓 🐰 👍 👎 🛒
font-family: serif, -ro-emoji	* 5 M © !! TM i	🕒 📺 😞 😊 ♀ 🍓 🐰 👍 👎 🛒
font-family: -ro-emoji, sans-serif	* 5 M © !! TM i	🕒 📺 😞 😊 ♀ 🍓 🐰 👍 👎 🛒
font-family: -ro-emoji, serif	* 5 M © !! TM i	🕒 📺 😞 😊 ♀ 🍓 🐰 👍 👎 🛒
font-family: -ro-emoji	* 5 M © !! TM i	🕒 📺 😞 😊 ♀ 🍓 🐰 👍 👎 🛒

Complex emojis are also supported, like the following that consists of 10 HTML entities:  (&#x1f469;&#x1f3ff;&#x200d;&#x2764;&#xfe0f;&#x200d;&#x1f48b;&#x200d;&#x1f468;&#x1f3fb;)

## 10. Language-based Font Selection

The automatic font fallback of PDFreactor supports CJK characters for Japanese, Korean as well as Chinese in the variants Simplified, Traditional and Hongkong. When a language tag is specified for the element containing the text, the matching variant is automatically used. Please note that this functionality requires at least Java 9.

When the language tag contains a CJK script (Jpan, Kore, Hans, Hant), that is used. Otherwise the script is looked up for the language, potentially using the country as additional information. Finally the default language of the country, along with the country itself, is tried. The default is Japanese.

The only exception is Chinese (Hongkong), which is chosen when Chinese (Traditional) was detected and the country is Honkong.

The following list shows the 5 variations of a character, along with various example language tags that are mapped to the variant.

謁 謁

**Japanese:** lang="ja"

ja, ja-JP, ja-DE, und-JP, de-JP, und-Jpan, de-Jpan-DE

謁 謁

**Korean:** lang="ko"

ko, ko-KR, ko-DE, und-KR, de-KR, und-Kore, de-Kore-DE

謁 謁

**Chinese (Simplified):** lang="zh"

zh, zh-CN, zh-DE, und-CN, de-CN, und-Hans, de-Hans-DE

謁 謁

**Chinese (Traditional):** lang="zh-Hant"

zh-Hant, zh-TW, und-TW, de-TW, und-Hant, de-Hant-DE

謁 謁

**Chinese (Hongkong):** lang="zh-HK"

zh-HK, und-HK, de-HK, yue

The font-family `-ro-cjk-auto` is also mapped to the result of this detection.