

**Lorem Ipsum**

## Chapter 1 : A runaway reef

What is Lorem Ipsum?

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has survived not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged. It was popularised in the 1960s with the release of Letraset sheets containing Lorem Ipsum passages, and more recently with desktop publishing software like Aldus PageMaker including versions of Lorem Ipsum.

Why do we use it?

It is a long established fact that a reader will be distracted by the readable content of a page when looking at its layout. The point of using Lorem Ipsum is that it has a more-or-less normal distribution of letters, as opposed to using 'Content here, content here', making it look like readable English. Many desktop publishing packages and web page editors now use Lorem Ipsum as their default model text, and a search for 'lorem ipsum' will uncover many web sites still in their infancy. Various versions have evolved over the years, sometimes by accident, sometimes on purpose (injected humour and the like).

Where does it come from?

Contrary to popular belief, Lorem Ipsum is not simply random text. It has roots in a piece of classical Latin literature from 45 BC, making it over 2000 years old. Richard McClintock, a Latin professor at Hampden-Sydney College in Virginia, looked up one of the more obscure Latin words, consectetur, from a Lorem Ipsum passage, and going through the cites of the word in classical literature, discovered the undoubtable source. Lorem Ipsum comes from sections 1.10.32 and 1.10.33 of "de Finibus Bonorum et Malorum" (The Extremes of Good and Evil) by Cicero, written in 45 BC. This book is a treatise on the theory of ethics, very popular during the Renaissance. The first line of Lorem Ipsum, "Lorem ipsum dolor sit amet..", comes from a line in section 1.10.32.

The standard chunk of Lorem Ipsum used since the 1500s is reproduced below for those interested. Sections 1.10.32 and 1.10.33 from "de Finibus Bonorum et Malorum" by Cicero are also reproduced in their exact original form, accompanied by English versions from the 1914 translation by H. Rackham.

Where can I get some?

There are many variations of passages of Lorem Ipsum available, but the majority have suffered alteration in some form, by injected humour, or randomised words which don't look even slightly believable. If you are going to use a passage of Lorem Ipsum, you need to be sure there isn't anything embarrassing hidden in the middle of text. All the Lorem Ipsum generators on the Internet tend to repeat predefined chunks as necessary, making this the first true generator on the Internet. It uses a dictionary of over 200 Latin words, combined with a handful of model sentence structures, to generate Lorem Ipsum which looks reasonable. The generated Lorem Ipsum is therefore always free from repetition, injected humour, or non-characteristic words etc.

What is Lorem Ipsum?

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has survived not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged. It was popularised in the 1960s with the release of Letraset sheets containing Lorem Ipsum passages, and more recently with desktop publishing software like Aldus PageMaker including

versions of Lorem Ipsum.

Why do we use it?

It is a long established fact that a reader will be distracted by the readable content of a page when looking at its layout. The point of using Lorem Ipsum is that it has a more-or-less normal distribution of letters, as opposed to using 'Content here, content here', making it look like readable English. Many desktop publishing packages and web page editors now use Lorem Ipsum as their default model text, and a search for 'lorem ipsum' will uncover many web sites still in their infancy. Various versions have evolved over the years, sometimes by accident, sometimes on purpose (injected humour and the like).

Where does it come from?

Contrary to popular belief, Lorem Ipsum is not simply random text. It has roots in a piece of classical Latin literature from 45 BC, making it over 2000 years old. Richard McClintock, a Latin professor at Hampden-Sydney College in Virginia, looked up one of the more obscure Latin words, consectetur, from a Lorem Ipsum passage, and going through the cites of the word in classical literature, discovered the undoubtable source. Lorem Ipsum comes from sections 1.10.32 and 1.10.33 of "de Finibus Bonorum et Malorum" (The Extremes of Good and Evil) by Cicero, written in 45 BC. This book is a treatise on the theory of ethics, very popular during the Renaissance. The first line of Lorem Ipsum, "Lorem ipsum dolor sit amet..", comes from a line in section 1.10.32.

The standard chunk of Lorem Ipsum used since the 1500s is reproduced below for those interested. Sections 1.10.32 and 1.10.33 from "de Finibus Bonorum et Malorum" by Cicero are also reproduced in their exact original form, accompanied by English versions from the 1914 translation by H. Rackham.

Where can I get some?

There are many variations of passages of Lorem Ipsum available, but the majority have suffered alteration in some form, by injected humour, or randomised words which don't look even slightly believable. If you are going to use a passage of Lorem Ipsum, you need to be sure there isn't anything embarrassing hidden in the middle of text. All the Lorem Ipsum generators on the Internet tend to repeat predefined chunks as necessary, making this the first true generator on the Internet. It uses a dictionary of over 200 Latin words, combined with a handful of model sentence structures, to generate Lorem Ipsum which looks reasonable. The generated Lorem Ipsum is therefore always free from repetition, injected humour, or non-characteristic words etc.

*End of Chapter*

## Chapter 2 : the rpos and cons

### Unicode encoding model

Unicode and its parallel standard, the ISO/IEC 10646 Universal Character Set, together constitute a modern, unified character encoding. Rather than mapping characters directly to octets (bytes), they separately define what characters are available, corresponding natural numbers (code points), how those numbers are encoded as a series of fixed-size natural numbers (code units), and finally how those units are encoded as a stream of octets. The purpose of this decomposition is to establish a universal set of characters that can be encoded in a variety of ways.[8] To describe this model correctly requires more precise terms than "character set" and "character encoding." The terms used in the modern model follow:[8]

A character repertoire is the full set of abstract characters that a system supports. The repertoire may be closed, i.e. no additions are allowed without creating a new standard (as is the case with ASCII and most of the ISO-8859 series), or it may be open, allowing additions (as is the case with Unicode and to a limited extent the Windows code pages). The characters in a given repertoire reflect decisions that have been made about how to divide writing systems into basic information units. The basic variants of the Latin, Greek and Cyrillic alphabets can be broken down into letters, digits, punctuation, and a few special characters such as the space, which can all be arranged in simple linear sequences that are displayed in the same order they are read. But even with these alphabets, diacritics pose a complication: they can be regarded either as part of a single character containing a letter and diacritic (known as a precomposed character), or as separate characters. The former allows a far simpler text handling system but the latter allows any letter/diacritic combination to be used in text. Ligatures pose similar problems. Other writing systems, such as Arabic and Hebrew, are represented with more complex character repertoires due to the need to accommodate things like bidirectional text and glyphs that are joined together in different ways for different situations.

A coded character set (CCS) is a function that maps characters to code points (each code point represents one character). For example, in a given repertoire, the capital letter "A" in the Latin alphabet might be represented by the code point 65, the character "B" to 66, and so on. Multiple coded character sets may share the same repertoire; for example ISO/IEC 8859-1 and IBM code pages 037 and 500 all cover the same repertoire but map them to different code points.

A character encoding form (CEF) is the mapping of code points to code units to facilitate storage in a system that represents numbers as bit sequences of fixed length (i.e. practically any computer system). For example, a system that stores numeric information in 16-bit units can only directly represent code points 0 to 65,535 in each unit, but larger code points (say, 65,536 to 1.4 million) could be represented by using multiple 16-bit units. This correspondence is defined by a CEF.

Next, a character encoding scheme (CES) is the mapping of code units to a sequence of octets to facilitate storage on an octet-based file system or transmission over an octet-based network. Simple character encoding schemes include UTF-8, UTF-16BE, UTF-32BE, UTF-16LE or UTF-32LE; compound character encoding schemes, such as UTF-16, UTF-32 and ISO/IEC 2022, switch between several simple schemes by using byte order marks or escape sequences; compressing schemes try to minimise the number of bytes used per code unit (such as SCSU, BOCU, and Punycode).

Although UTF-32BE is a simpler CES, most systems working with Unicode use either UTF-8, which is backward compatible with fixed-width ASCII and maps Unicode code points to variable-width sequences of octets, or UTF-16BE, which is backward compatible with fixed-width UCS-2BE and maps Unicode code points to

variable-width sequences of 16-bit words. See comparison of Unicode encodings for a detailed discussion.

Finally, there may be a higher level protocol which supplies additional information to select the particular variant of a Unicode character, particularly where there are regional variants that have been 'unified' in Unicode as the same character. An example is the XML attribute `xml:lang`.

The Unicode model uses the term character map for historical systems which directly assign a sequence of characters to a sequence of bytes, covering all of CCS, CEF and CES layers.[8]

*End of Chapter*