

# KANJIDIC Project

---

## Contents

---

### The KANJIDIC Project

- Introduction
- Content & Format
- Radical and Stroke Counting Rules
  - Radicals
  - Other Stroke Patterns
- Kanji Dictionary Search Codes
  - SKIP Codes
  - De Roo Codes
  - Four Corner Codes
- Proposing Changes
- Kanji Information Sites
- Legacy Documentation
- Copyright and Permissions
- History

# The KANJIDIC Project

*(Note that this page is in the process of being rewritten, so be patient with any aspects that seem incomplete.)*

## Introduction

---

The KANJIDIC project, which began in 1991, has the goal of compiling and distributing comprehensive information on the kanji used in Japanese text processing. It covers the 13,108 kanji in three main Japanese standards:

- JIS X 0208-1998 ([https://en.wikipedia.org/wiki/JIS\\_X\\_0208](https://en.wikipedia.org/wiki/JIS_X_0208)), which includes 6,355 kanji.
- JIS X 0212-1990 ([https://en.wikipedia.org/wiki/JIS\\_X\\_0212](https://en.wikipedia.org/wiki/JIS_X_0212)), which includes extra 5,801 kanji
- JIS X 0213-2012 ([https://en.wikipedia.org/wiki/JIS\\_X\\_0213](https://en.wikipedia.org/wiki/JIS_X_0213)), which extends JIS X 0208, overlaps with some of JIS X 0212, and adds 952 additional kanji.

Three data files are distributed by this project:

- the KANJIDIC2 file, which is in XML format and Unicode/UTF-8 (<https://en.wikipedia.org/wiki/UTF-8>) coding, and contains information about all 13,108 kanji. ([download \(http://www.edrdg.org/kanjidic/kanjidic2.xml.gz\)](http://www.edrdg.org/kanjidic/kanjidic2.xml.gz))
- the KANJIDIC file, which is in EUC-JP ([https://en.wikipedia.org/wiki/Extended\\_UNIX\\_Code#EUC-JP](https://en.wikipedia.org/wiki/Extended_UNIX_Code#EUC-JP)) coding and covers the 6,355 kanji in JIS X 0208. ([download \(http://www.edrdg.org/kanjidic/kanjidic.gz\)](http://www.edrdg.org/kanjidic/kanjidic.gz))

- the KANJD212 file, which also is in EUC-JP coding and covers the 5,801 kanji in JIS X 0212. ([download](http://www.edrdg.org/kanjidic/kanjd212.gz) (<http://www.edrdg.org/kanjidic/kanjd212.gz>))

## Content & Format

---

The database and distributed data files contain an entry for each of the kanji, with each entry containing a number of fields of data about the kanji. The data is described in the following table. The format of the distributed files as follows:

- the KANJI and KANJD212 files are text files with one line per kanji and the information fields separated by spaces. The format of each line is:
  - the kanji itself followed by the hexadecimal form of the JIS *ku-ten* coding, e.g. "亜 3021" (the decimal *ku-ten* code is 16-01);
  - information fields beginning with one or two-letter codes as per the table below. For example "S10" indicates a stroke count of 10;
  - the Japanese readings of the kanji. ON readings (音読み) are generally in *katakana* and KUN readings (訓読み) in *hiragana*. An exception is the set of *kokujī* for measurements such as centimetres, where the reading is in *katakana*. Hyphens are used to indicate prefixes/suffixes, and '.' indicates the portion of the reading that is *okurigana*. There may be several classes of reading fields, with ordinary readings first, followed by members of the other classes, if any. The current other classes, and their tagging, are:
    - where the kanji has special *nanori* (i.e. name) readings, these are preceded the marker "T1";
    - where the kanji is a radical, and the radical name is not already a reading, the radical name is preceded the marker "T2".
  - the meanings (usually in English). Each field begins with an open brace '{' and ends at the next close brace '}'.
- the KANJI file is in XML and is structured according to its DTD ([http://www.edrdg.org/kanjidic/kanjidic2\\_dtdh.html](http://www.edrdg.org/kanjidic/kanjidic2_dtdh.html)) (Document Type Definition). The DTD contains extensive annotations and is intended to be the primary documentation for the file. This sample (<http://www.edrdg.org/kanjidic/kd2examph.html>) illustrates the structure of a typical entry. Information fields are grouped by type within entities such as <dic\_number> and <query\_code>, with specific values indicated by an attribute code. For example the kanji 亜 has the number 43 in the original Nelson kanji dictionary and 81 in the New Nelson. This is recorded in the XML file as:

```

<dic_number>

  <dic_ref dr_type="nelson_c">43</dic_ref>
  <dic_ref dr_type="nelson_n">81</dic_ref>
  ...
</dic_number>

```

## Kanjidic Information Fields

Field	Kanjidic Code (if any)	Group Entity	Entity plus Attribute(s) (if any)	Comment
Kanji	none	literal		
JIS code-point	none	codepoint	cp_value cp_type="jis208" (or "jis212" or "jis213")	e.g. 亜 is "3021" in KANJIDIC and "1-16-01" in KANJIDIC2
Unicode code-point	U	codepoint	cp_value cp_type="ucs"	
Radical (Classical) (See Note 1 below)	B/C	radical	rad_value rad_type="classical"	Where Nelson uses the classical radical this has a "B" code, otherwise it has a "C" code
Radical (Nelson)	B	radical	rad_value rad_type="nelson_c"	
Grade	G	misc	grade	<p>The "grade" of the kanji.</p> <ul style="list-style-type: none"> <li>- G1 to G6 indicates the grade level as specified by the Japanese Ministry of Education for kanji that are to be taught in elementary school (1026 Kanji). These are sometimes called the <i>kyōiku</i> (education) kanji and are part of the set of <i>jōyō</i> (daily use) kanji;</li> <li>- G8 indicates the remaining <i>jōyō</i> kanji that are to be taught in secondary school (additional 1130 Kanji);</li> <li>- G9 and G10 indicate <i>jinmeiyō</i> ("for use in names") kanji which in addition to the <i>jōyō</i> kanji are approved for use in family name registers and other official documents. G9 (649 kanji, of which</li> </ul>

				640 are in KANJI indicates the kanji is a "regular" name kanji, and G10 (212 kanji of which 130 are in KANJI) indicates the kanji is a variant of a <i>jōyō</i> kanji.
Stroke count	S	misc	stroke_count	The stroke count of the kanji. If more than one, the first is considered the accepted count, while subsequent ones are common miscounts. (See the section later in this document on counting strokes for some of the rules applied especially to radicals.)
Frequency-of-use ranking	F	misc	freq	The 2,501 most-used characters have a ranking which expresses the relative frequency of occurrence of a character in modern Japanese. The data is based on an analysis of word frequencies in the Mainichi Shimbun over 4 years by Alexandre Girardi. Note: (a) these frequencies are biased towards words and kanji used in newspaper articles, and (b) the relative frequencies for the last few hundred kanji so graded is quite imprecise.
Variant JIS 0208 kanji	XJ0	misc	variant var_type="jis208"	Code-point of a similar or related kanji. (In the kanjidic file the JIS hex code is used and in the XML file the

				equivalent "1-nn-nn" kuten code is used.)
Variant JIS 0212 kanji	XJ1	misc	variant var_type="jis212"	Code-point of a similar or related kanji. (In the kanjidic file the JIS hex code is used and in the XML file the equivalent "1-nn-nn" kuten code is used.)
Variant JIS 0213 kanji	XJ2	misc	variant var_type="jis213"	Code-point of a similar or related kanji. (In the kanjidic file the plane number (P: 1 or 2) plus the JIS hex code is used, and in the XML file the equivalent "P-nn-nn" kuten code is used.)
Variant kanji (De Roo index)	XJD	misc	variant var_type="deroo"	Code-point of a similar or related kanji.
Variant kanji (NJECD index)	XH	misc	variant var_type="halpern_njecd"	Code-point of a similar or related kanji.
Variant kanji (S&H index)	XI	misc	variant var_type="s_h"	Code-point of a similar or related kanji.
Variant kanji (Nelson index)	XN	misc	variant var_type="nelson_c"	Code-point of a similar or related kanji.
Variant kanji (O'Neill index)	XO	misc	variant var_type="oneill"	Code-point of a similar or related kanji.
Radical name(s)	none	misc	rad_name	The name of the radical in <i>hiragana</i> . In the KANJI DIC edition these are placed after the readings and preceded by the "T2" tag.
JLPT Level	J	misc	jlpt	The pre-2010 level of the Japanese Language Proficiency Test (JLPT) in which the kanji occurs (1-4). Note that the JLPT test levels changed in 2010, with a new 5-level system (N1 to N5) being introduced. No official kanji lists are available for the

				new levels. The new levels are regarded as being similar to the old levels except that the old level 2 is now divided between N2 and N3, and the old levels 3 and 4 are now N4 and N5.
Nelson (Classic) number	N	dic_number	dic_ref dr_type="nelson_c"	The index number in "The Modern Reader's Japanese-English Character Dictionary", edited by Andrew Nelson. If not present, the character is not in Nelson, or is considered to be a non-standard version, in which case it may have a variant. Note that many kanji glyphs currently used are what Nelson described as "non-standard".
Nelson (New) number	V	dic_number	dic_ref dr_type="nelson_n"	The index number in "The New Nelson Japanese-English Character Dictionary", edited by John Haig.
NJECD number	H	dic_number	dic_ref dr_type="halpern_njecd"	The index number in the "New Japanese-English Character Dictionary" (1990), edited by Jack Halpern.
Kodansha Kanji Dictionary number	DP	dic_number	dic_ref dr_type="halpern_kkd"	The index numbers used by Jack Halpern in the "Kodansha Kanji Dictionary" (2013), which is the revised version of the "New Japanese-English Kanji Dictionary" of 1990.
Kanji Learners Dictionary number	DK	dic_number	dic_ref dr_type="halpern_kkld"	The index numbers used by Jack Halpern in the "Kanji Learners Dictionary", published by Kodansha in 1999.

Kanji Learners Dictionary number (2nd ed)	DL	dic_number	dic_ref dr_type="halpern_kkld_2ed"	The index numbers used by Jack Halpern in the 2nd edition of the "Kanji Learners Dictionary", published by Kodansha in 2013.
Remembering The Kanji number	L	dic_number	dic_ref dr_type="heisig"	The index number used in "Remembering The Kanji" by James Heisig.
Remembering The Kanji number (6th ed)	DN	dic_number	dic_ref dr_type="heisig6"	The index number used in "Remembering The Kanji, 6th Edition" by James Heisig.
Gakken number	K	dic_number	dic_ref dr_type="gakken"	The index number in the Gakken Kanji Dictionary ("A New Dictionary of Kanji Usage"). Some of the numbers relate to the list at the back of the book, jouyou kanji not contained in the dictionary, and various historical tables at the end.
O'Neill's Japanese Names number	O	dic_number	dic_ref dr_type="oneill_names"	The index number in "Japanese Names", by P.G. O'Neill. (Weatherhill, 1972) (Note: some of the numbers end with 'A'.)
O'Neill's Essential Kanji number	DO	dic_number	dic_ref dr_type="oneill_kk"	The index numbers used in P.G. O'Neill's "Essential Kanji".
Morohashi number	MN/MP	dic_number	dic_ref dr_type="moro" m_vol m_page	The index number and volume.page respectively of the kanji in the 13-volume Morohashi Daikanwajiten. A terminal 'P' in the number, e.g. 4879P, indicates that it is 4879' in the original. In some 500 cases, the number is terminated with an 'X', to indicate that

				the kanji in Morohashi has a close, but not identical, glyph to the form in the JIS X 0208 standard. In the XML the volume and page are attribute values.
Henshall number	E	dic_number	dic_ref dr_type="henshall"	The index number used in "A Guide To Remembering Japanese Characters" by Kenneth G. Henshall.
Kanji & Kana number	IN	dic_number	dic_ref dr_type="sh_kk"	The index number used in Spahn & Hadamitzky's "Kanji & Kana", 2nd edition (Tuttle).
Kanji & Kana number (2011 ed)	DA	dic_number	dic_ref dr_type="sh_kk2"	The index number used in 2011 edition of Spahn & Hadamitzky's "Kanji & Kana".
Sakade number	DS	dic_number	dic_ref dr_type="sakade"	The index numbers used in the early editions of "A Guide To Reading and Writing Japanese", edited by Florence Sakade.
Japanese Kanji Flashcards number	DF	dic_number	dic_ref dr_type="jf_cards"	The index numbers used in the "Japanese Kanji Flashcards", by Max Hodges and Tomoko Okazaki (White Rabbit Press).
Henshall Guide number	DH	dic_number	dic_ref dr_type="henshall3"	The index numbers used in the 3rd edition of "A Guide To Reading and Writing Japanese" edited by Ken Henshall et al.
Tuttle Kanji Cards number	DT	dic_number	dic_ref dr_type="tutt_cards"	The index numbers used in the Tuttle Kanji Cards, compiled by Alexander Kask.
Crowley number	DC	dic_number	dic_ref dr_type="crowley"	The index numbers used in "The Kanji Way"

				to Japanese Language Power" by Dale Crowley.
Kanji in Context number	DJ	dic_number	dic_ref dr_type="kanji_in_context"	The index numbers used in the "Kanji in Context" by Nishiguchi and Kono.
Kodansha Compact Kanji Guide number	DG	dic_number	dic_ref dr_type="kodansha_compact"	The index numbers used in the "Kodansha Compact Kanji Guide".
Japanese For Busy People number	DB	dic_number	dic_ref dr_type="busy_people"	The index numbers used in "Japanese For Busy People" vols I-III, published by the AJLT. The codes are the volume.chapter.
Maniette number	DM	dic_number	dic_ref dr_type="maniette"	The numbers in Yves Maniette's "Les Kanjis dans la tête", the French adaptation of Heisig's "Remembering The Kanji".
SKIP code	P	query_code	q_code qc_type="skip"	The SKIP (System of Kanji Indexing by Patterns) developed by Jack Halpern. The code is of the form "I-m-n". See <a href="#">SKIP Codes</a> section for more information.
S&H descriptor	I	query_code	q_code qc_type="sh_desc"	The index code in "The Kanji Dictionary" (Tuttle 1996), by Spahn & Hadamitzky. It is the form nxnn.n, e.g. 3k11.2, where the kanji has 3 strokes in the identifying radical, it is radical "k" in the S&H classification system, there are 11 other strokes, and it is the 2nd kanji in the 3k11 sequence.
Four Corner code	Q	query_code	q_code qc_type="four_corner"	The Four Corner code for the kanji. See the <a href="#">Four Corner codes</a>

				section for more information.
De Roo code	DR	query_code	q_code qc_type="deroo"	The codes developed by Father Joseph De Roo, and published in his book "2001 Kanji" (Bonjinsha). See the <a href="#">De Roo Codes</a> section for more information.
Misclassification code	ZPP	query_code	q_code qc_type="skip" skip_misclass="posn"	SKIP misclassification by position.
Misclassification code	ZSP	query_code	q_code qc_type="skip" skip_misclass="stroke_count"	SKIP misclassification by stroke count.
Misclassification code	ZBP	query_code	q_code qc_type="skip" skip_misclass="stroke_and_posn"	SKIP misclassification by both position and stroke count.
Misclassification code	ZRP	query_code	q_code qc_type="skip" skip_misclass="stroke_diff"	SKIP misclassification by differing opinions on stroke counts.
Chinese reading	Y	rmgroup	reading r_type="pinyin"	The PinYin (Chinese) reading of the kanji.
Korean reading (romanized)	W	rmgroup	reading r_type="korean_r"	The Korean reading of the kanji in the (Republic of Korea) Ministry of Education style.
Korean reading (hangul)	not included	rmgroup	reading r_type="korean_h"	The Korean reading of the kanji in the hangul script.
Vietnamese reading (chữ quốc ngữ)	not included	rmgroup	reading r_type="vietnam"	The Vietnamese reading of the kanji in chữ quốc ngữ.
Japanese on reading (katakana)	none	rmgroup	reading r_type="ja_on"	In the KANJI edition the readings are placed between the information fields and the meanings.
Japanese kun reading (usu. hiragana)	none	rmgroup	reading r_type="ja_kun"	
Meanings	none	rmgroup	meaning m_lang="xx"	The kanji meaning(s). For languages other than English the m_lang attribute is used with two-letter

				ISO 639-1 language codes. In the KANJI edition the meanings are placed at the end of the line.
Name reading(s) ( <i>hiragana</i> )	T1		nanori	The readings only associated with named-entities. In the KANJI edition the first of these is preceded by the "T1" tag.

Note 1: For the sake of consistency the classical radical is the one indicated in the JIS漢字字典 (日本規格協会).

## Radical and Stroke Counting Rules

---

These rules apply to:

1. the stroke-counts themselves;
2. the stroke counts in the SKIP codes. Where this results in a SKIP which differs from that in the NJECD, or in the non-NJECD SKIPS provided by Jack Halpern, the Jack Halpern version is included prefixed with "ZR".

## Radicals

The radicals listed below are ones where there are differing approaches to the counting of radicals in the various references. The stroke counting in this file does not strictly follow any reference, but tends to more aligned to Halpern.

1. B54 ENYOU - 𠂔. Traditionally counted as 3 strokes, but more recently often counted as 2. S&H count this as 2; Nelson, Halpern, Koujien, etc, count it as 3. I treat it as 3.
2. B97 URI - 瓜. Traditionally counted as 5 strokes, as the middle portion looks like a katakana ホ. Modern glyphs invariably make it look like 6 strokes. Nelson says it is 5 strokes. Halpern does too, but then counts the shape as 6 in other kanji. Koujien says 6, as do S&H. I treat it as 6.
3. B113 SHIMESU e.g. 礼, is counted as 4 strokes in that form, and 5 strokes in its older form, 祀. 18 kanji are in the 4-stroke form and 20 are in the 5-stroke form. (Nelson and S&H count it as 4; Halpern counts it as 4 or 5. [See Note 1.] )
4. B131 SHIN/KERAJ 臣. Counted as 7 (Nelson counts it as 6, Halpern as 7 (in the book), and S&H as both for different kanji.)
5. B136 MAI ASHI 夂. Counted as 7 (traditionally counted as 6, in accordance with the older writing of 'ヰ'. Nelson counts as 6, S&H as 7, and Halpern as 7 for 常用 and 人名用漢字 and 6 for the rest.) Note this is also applied to counting 終 and for kanji with the 韋 pattern.
6. B140 KUSA-KANMURI e.g. 苛 always counted as 3 strokes (Halpern counts this 4 strokes for the (mostly level 2) kanji where the older form is often printed.) Note that this has been carried through to kanji where this element is not the indexing radical, such as 膩.
7. B162 SHIN-NYUU e.g. 遙 or 逢 counted as 3 or 4 strokes. (Nelson and S&H count it as 2 strokes, and Halpern as either 3 or 4.) [See Note 1 below.]
8. B163 OOZATOZUKIRI & B170 KOZATO-HEN 邦 and 阮 always counted as 3 strokes (Nelson and S&H count it as 2, Halpern as 3.) This also applies where it appears mid-kanji, such as in 檻.

9. B184 SHOKU HEN 食, 飢, etc. is counted as 8 strokes in the 飢 form, and as 9 strokes in the 飴 and 餐 forms. (Nelson and S&H count it as 8 strokes, and Halpern as 8 or 9.) [See Note 1. below.]
10. B199 MUGI 麦 always counted as 7 strokes, except for 麥 & 麥 where it is counted as 11. (Nelson and Halpern do the same, and S&H avoid treating it as a radical, but count it as 12 in the remainder.)
11. The ROO or OI radical (老) has a variant consisting of the top 4 strokes. For example, it is in 者. Traditionally, this variant had an extra dot, and was counted as 5 strokes. I'm counting it as 4 throughout.

## Other Stroke Patterns

1. While the pattern 𠂔 is a 6-stroke radical, the top half of 婁 is made up of three distinct parts totalling 8 strokes. Note that this also is the case with 嫂, 漢, 艸 and 瘦 despite the simplification in the JIS glyphs.
2. 牙 (KIBA HEN) is a problem. It is classically counted as 4 strokes, but these days has a flick that makes it effectively 5. Halpern, Nelson and S&H usually have it as 5 strokes, so I'm standardizing on that.
3. Another little horror is 死 (MU or NASHI), which is classically counted as 4 strokes. The most common variant has 5 strokes, but looks like 6. Halpern, S&H and the Classical Nelson count this as 4 strokes, and the New Nelson as 5. I'm making it 5 too.
4. The JUU or ASHIATO radical is at the bottom of 禽 and 禹. It is traditionally counted as 5 strokes, although sometimes it looks like 4. I'm using 5 throughout.
5. A related shape is 𠂔, as in 瓜, 孤, 弧, etc. This is sometimes counted as two strokes (both Nelsons) and sometimes as three strokes (Halpern, S&H). Classically it is regarded as two strokes. I am using 6 strokes for 瓜.
6. The pattern to the left of 敝, which appears in several kanji, e.g. 幣 and 譬, has 8 strokes. (There are 3 strokes at the top as in 尚.)
7. The "east" pattern (東) has 8 strokes. There is an older form in which there are two strokes in the box (凵). It is counted as 8 strokes here in the 東 form (e.g. 講) and 9 in the 凵 form, as in 講.
8. The pattern at the bottom of 雋 is counted as 4 strokes in modern dictionaries, although traditionally it was 5.
9. The pattern 卷, which appears in several kanji, is counted as 9 strokes. Several dictionaries count it as either 8 or 9.
10. The pattern on the left of 収 is variously handled as 2 strokes or 3 strokes. As more recent dictionaries make it 4, I will do so too.
11. The 夂 pattern has 3 and 4-stroke versions, and sometimes the glyphs can be confusing as to which is used. In the 繖 kanji, for example, it is traditionally counted as 3, but Spahn & Hadamitzky count it as 4 and the Nelsons include both.

Note: The JIS X 0208-1990 standard does not formally specify the precise glyphs used for kanji, however the glyphs it uses in the published version have become de facto standards for many font compilations. In the published standard, for several kanji, e.g. 辿/辻, 礼/祀, 飢/飴, the JIS level one kanji use the simpler form, and the Level 2 kanji use the older more complex form. Just to make matters worse, many fonts for JIS X 0208 kanji are based on the bit-maps specified in JIS X 9051-1984 standard, which defines the 16x16 patterns for JIS X 0208-1983 characters. According to Ken Lunde: "This standard was not very good, and JSA is no longer supporting it." Anyway, JIS X 9051-1984 had the simpler form for all these bushu in both Levels 1 and 2, as well as having simplifications of kanji like 濁. Thus, as the font foundries have freedom to choose whichever glyphs they like, what you see on your screen may well not agree with these rules. All the rules in this appendix relate to the glyphs as published in the JIS X 0208-1990 standard, and as appearing in font compilations based on them.

# Kanji Dictionary Search Codes

---

## SKIP Codes

The System of Kanji Indexing by Patterns (SKIP) is a scheme for the classification and rapid retrieval of Chinese characters on the basis of geometrical patterns. Developed by Jack Halpern, it first appeared in the New Japanese-English Character Dictionary (Kenkyusha, Tokyo 1990; NTC, Chicago 1993), and in successor publications such as the "Kanji Learners Dictionary" (Kodansha 1999,2011) and the "Kodansha Kanji Dictionary" (2013). A description of the coding system is available (<http://www.edrdg.org/wwwjdic/SKIP.html>).

As examples, 割 has a SKIP code of 1-10-2, indicating it is divided into left-right portions with 10 strokes at the left and 2 at the right. 度 has a SKIP code of 度 indicating it has a 3-stroke enclosure with 6 strokes inside it.

## De Roo Codes

The De Roo codes were developed by Father Joseph De Roo, and published in his book "2001 Kanji" (Bonjinsha). They are based on the shapes observed at the top and bottom of the character. A detailed description (<http://www.edrdg.org/wwwjdic/deroo.html>) is available.

As an example, 亜 has a code of 3273 indicating that the top of the kanji is pattern number 32 (兀) and the bottom pattern number 73 (horizontal line with two vertical strokes above it).

## Four Corner Codes

The Four Corner coding system was invented by Wang Chen in 1928, it has since then been widely used in dictionaries in China and Japan for classifying kanji and hanzi. In China it is losing popularity in favour of Pinyin ordering. Some Japanese dictionaries, such as the Morohashi Daikanwajiten have a Four Corner Index. An overview (<http://www.edrdg.org/wwwjdic/FOURCORNER.html>) of the coding system is available. In some cases a character may have two of these codes, as it is can be little ambiguous, and Morohashi has some kanji coded differently from their traditional Chinese codes. The coding system indexes characters according to the shapes at the corners.

## Proposing Changes

---

There is currently no online access to the database the holds the KANJIDIC contents (the information is mostly quite static.) Anyone wishing to propose a change to the data for a kanji, e.g. add or change a reading will need to email Jim Breen at [jimbreen@gmail.com](mailto:jimbreen@gmail.com).

## Kanji Information Sites

---

*(Being expanded)*

- Jim's Kanji Information Page (<http://nihongo.monash.edu/kanjiinfo.html>).
- The Kanji alive (<https://kanjialive.com/>) site at the University of Chicago.
- The Kanjidipedia (<https://www.kanjidipedia.jp/>) sit (mostly in Japanese).

## Legacy Documentation

---

The current Wiki page was compiled from several older documents, which are no longer being maintained. They are still available for historical purposes. They are:

- a basic home page about KANJIDIC2 ([http://www.edrdg.org/kanjidic/kanjd2index\\_legacy.html](http://www.edrdg.org/kanjidic/kanjd2index_legacy.html));
- an overview page about the KANJIDIC2 structure ([http://www.edrdg.org/kanjidic/kanjidic2\\_ov\\_legacy.html](http://www.edrdg.org/kanjidic/kanjidic2_ov_legacy.html));
- an overview page about KANJIDIC and KANJD212 ([http://www.edrdg.org/kanjidic/kanjidic\\_legacy.html](http://www.edrdg.org/kanjidic/kanjidic_legacy.html));
- the original KANJIDIC ([http://www.edrdg.org/kanjidic/kanjidic\\_doc\\_legacy.html](http://www.edrdg.org/kanjidic/kanjidic_doc_legacy.html)) documentation;
- the original KANJD212 ([http://www.edrdg.org/kanjidic/kanjd212\\_doc\\_legacy.html](http://www.edrdg.org/kanjidic/kanjd212_doc_legacy.html)) documentation.

## Copyright and Permissions

---

The KANJIDIC project files are released under a Creative Commons Attribution-ShareAlike Licence (V3.0). See the [EDRDG General Doctionary Licence Statement](http://www.edrdg.org/edrdg/licence.html) (<http://www.edrdg.org/edrdg/licence.html>) for details.

For the most part the information provided in the project's files is in the public domain. Information relating to the sequence numbers of kanji in published dictionaries is not considered to be subject to copyright. Descriptor and other search codes are considered to be the intellectusl policy of the developers. With regard to the codes included in the KANJIDIC files:

- in 2014 the SKIP codes were placed by Jack Halpern under a under a CC-SA licence. See [this page](http://www.kanji.org/kanji/dictionaries/skip_permission.htm) ([http://www.kanji.org/kanji/dictionaries/skip\\_permission.htm](http://www.kanji.org/kanji/dictionaries/skip_permission.htm)) for his announcement. It is now under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 Unported Licence (<https://creativecommons.org/licenses/by-nc-sa/3.0/>).
- Fr De Roo provided written permission for the De Roo codes to be included in KANJIDIC.
- the Spahn and Hadamitzky descriptor codes were kindly supplied by Mark Spahn for inclusion in KANJIDIC.

## History

---

*(some comments by Jim Breen)*

KANJIDIC began around 1991 as two files: jis1detl.lst and jis2detl.lst, which were later merged into a single file.

The first file was compiled initially from the file "kinfo.dat" supplied by Stephen Chung, who in turn compiled his file from a file prepared by Mike Erickson. I originally added about 1900 "meanings" by James Heisig keyed in by Kevin Moore from the book "Remembering The Kanji". I later added the meanings from Rik Smoody's files, compiled when he was working for Sony in Japan. These appear to have been based on Nelson.

The second file was compiled from a complete JIS2 list with Bushu and stroke counts kindly supplied to me by Jon Crossley, to which I added Nelson numbers, yomikata and meanings extracted from Rik Smoody's file.

Theresa Martin was an early assister with this file, particularly with tracking down and correcting many mistranscribed yomikata (the old zu/dzu, oo/ou, ji/dji, etc. problems).

Jeffrey Friedl did a major overhaul in September-October 1992, in which he added the original frequency rankings, Halpern codes, SKIP patterns, updated the grading ("G" fields) to reflect the modern Jouyou lists, corrected radical numbers, corrected stroke counts and readings to fall in line with modern usage.

Magnus Halldorsson corrected some erroneous Halpern numbers, and provided them for a lot of the radicals. He provided the list of Heisig indices, which he originally compiled himself, then verified and expanded using lists from Richard Walters and Antti Karttunen. He also passed on to me the list of Gakken indices compiled by Antti Karttunen.

Lee Collins provided the Unicode mappings.

Iain Sinclair has provided the yomikata, meanings and S&H indices of many of the obscure JIS2 kanji.

Christian Wittern, a Sinologist working at Kyoto University, sent me a monster file prepared by Dr Urs App from Hanazono College. From this I have extracted the Four Corner and Morohashi information. Christian also provided the original Pinyin details, which were later replaced. I am very grateful for these significant contributions.

In March 1994 the Morohashi indices were proof-read and corrected by Christian.

Alfredo Pinochet supplied all the Henshall numbers.

Ingar Holst has provided considerable assistance in regularizing the Bnnn and Cnnn radical classifications to remove some errors that were in the original JIS2 file, and to make it all conform to Nelson's classification.

In mid-1993 I withdrew the SKIP codes from the distributed file as it appeared that their presence violated Jack Halpern's copyright on these codes. Jeffrey Friedl contacted Jack about this, and Jack obtained permission from his publisher for the codes to be included subject (initially) to copyright and usage restrictions. In March 1994 the Halpern indices and SKIP codes were checked against an extract from Jack's files, and the "Z" mis-classification codes added, again from his files. Jack has also made a lot of useful comments and suggestions about the content and format of the file. I am most grateful to Jack for his permission and assistance, and also to Jeffrey for making the contact.

In May 1995, a number of updates took place. Jeffrey Friedl established contact with James Heisig, and obtained a further set of his indices. I contacted Mark Spahn (via the "honyaku" mailing list) and he kindly provided most of the missing S&H descriptors, and Jack Halpern released to me the SKIP codes of the kanji not in the New Japanese-English Character Dictionary. For all this material I am most grateful.

In August 1995, I added the O'Neill index numbers. These were compiled by Jenny Nazak, David Rosenfeld and myself. Thanks to Jenny & David for their assistance.

In January and February 1996 the Morohashi numbers were checked thoroughly against two important sources: a file of Unicode-Morohashi data (Uni2Dict) which was prepared by Koichi Yasuoka from the allocation in the JIS X 0221 standard, and the review draft of the proposed revision of the JIS X 0208 standard, which was prepared by the INSTAC Committee, and made available in a text file, thus enabling comparisons. All the mismatches between the three files were examined against the Morohashi text, and extensive corrections made to all three files. I am grateful to Koichi Yasuoka and Masayuki Toyoshima for their considerable assistance in this task.

In March 1996 the Korean readings were added. They were provided by Dr Charles Muller, then of Toyo Gakuen University, to whom I am most grateful. Chuck's compilation of Korean readings is extremely thorough and scholarly, and I am pleased to be able to incorporate them.

In April 1996 the readings of all the kanji were compared with those in the JIS X 0208 draft, and a number of corrections and additions made.

In May 1996 I carried out a "unification" of the readings of the KANJIDIC and KANJD212 files, wherein all the readings of the "itaiji" were brought into line. The identification of these itaiji was drawn from a file posted to the fj.kanji group by Taichi Kawabata (kawabata@is.s.u-tokyo.ac.jp), which was compiled at the ETL from the itaiji identification in the JIS X 0208 and JIS X 0212 standards. I corrected a few errors, and added some extra sets which were indicated in the JIS X 0208-1996 draft.

In July 1996 the Pinyin details were completely replaced by a new set. The original Pinyin were from an earlier compilation by Christian Wittern, and contained many errors. Two more reliable sources had become available: the Uni2Pinyin file compiled by Koichi Yasuoka, which is based in part on the TONEPY.tit by Yongguang Zhang; and the PYCHAR set of readings of Big5 hanzi compiled by Christian Wittern. The Pinyin currently in the KANJIDIC file is a combination of the two, following the order in the Uni2Pinyin file.

In August 1996 I corrected a few more missing and erroneous Nelson numbers, using a massive Nelson list prepared by Wolfgang Cronrath. He also flagged the kokuji, so I added these to the readings fields as "{{(kokuji)}".

Also in August 1996 I deleted the handful of former "XJxxxx" cross-references, and replaced them with a much more comprehensive set, so that they now represent all the recognized "itaiji". The file I used for this was the corrected itaiji file mentioned above.

In April 1997 I corrected a large number of bushu codes. Many of these had been identified as errors by Jean-Luc Leger who analyzed and examined all the Nelson bushu. I also identified and added a large number of missing Cnnn codes.

Also in April 1997 I added the S&H "Kanji & Kana" indices. These had been keyed by Olivier Galibert (Olivier.Galibert@mines.u-nancy.fr). (There must be an outbreak of kanji interest on Nancy.)

In February 1998, the long-awaited inclusion of the "New Nelson" numbers took place. I had been waiting for the editor of the New Nelson, John Haig, to supply a list (as he had agreed some years before), but in the meantime, Jean-Luc Leger keyed a list, so they are now available.

Also between December 1997 and February 1998 a large number of Level 2 kanji had their stroke counts corrected to bring them into line with the counting principles used in the Level 1 kanji. This usually aligned the counts with those used in the New Nelson and in S&H. Appendix E of this document was amended to reflect this. The leg-work in tracking this material down was done by Wolfgang Cronrath.

During December 1998 & Jan 1999 I updated the stroke counts of many of the Level 2 kanji, using an analysis of them carried out by Wolfgang Cronrath. I also added the De Roo codes, which had been keyed by Jasmin Blanchette, who also typed the explanatory material. I contacted Fr De Roo in Tokyo who readily agreed to the inclusion of the codes.

The extension of the S&H Kana & Kanji numbers to the 2nd edition was done by Enrique Sanchez Rosa.

The Hangul versions of the Korean readings (which only appear in the XML version) were provided by Francis Bond and Kyonghee Paik.

I did the Tuttle card numbers myself.

James Rose provided the numbers from Crowley's "The Kanji Way to Japanese Language Power", Sakade's "A Guide To Reading and Writing Japanese", and also for that book's 3rd Edition edited by Henshall, Seeley & De Groot.

The "Kodansha's compact Kanji guide" codes were provided by Richard Fremmerlid.

The "Kanji in Context" codes were provided by Randy Foreman.

The Spanish kanji meanings (which appear in the XML format, and may also appear in special versions of KANJIDIC) were compiled by Francisco Gutierrez and provided by Gabriel Sanroman.

Alain Thierion translated the meanings of the kanji into French, and also provided the Maniette numbers.

Andrew Slater provided updates to the JLPT numbers, and additional numbers for the Japanese Flashcards series.

Retrieved from "[https://www.edrdg.org/wiki/index.php?title=KANJIDIC\\_Project&oldid=894](https://www.edrdg.org/wiki/index.php?title=KANJIDIC_Project&oldid=894)"

**This page was last edited on 22 October 2020, at 02:17.**