

Internal Use 内部使用

CJK 汉字拆分

CJK Hanzi Decomposed

项目数据文件

V3.12

张轴材 主编

北京书同文数字化技术有限公司
二〇〇八年四月二十五日

CJK 汉字拆分文件

CJK Decomposed

说明文件

DATE: 2007-09-06~2008-04-25

Version: v 3.12

Source: Zhang Zhoucai 张轴材 joe.zhang@unihan.com.cn

Status: Internal Use 内部使用

编者注: 这是 2007 年夏天 NHN 项目文件。应当前 CJK Cool Hanzi 项目需要, 重新编排打印成册存档。与 V3.1 相比, 数据库内容没有变化。但是相关的汉字构件集有新的文档, Refer To : **CJKcomponentSet080425**。2008-04-25

DatabaseVersion : 200709061800 (FileName: CJKDecomposed200709061800.xls)

EUDC Version : 20070827

DATE: 2007-09-06 18:30

By : Tricy Zhang, UniHan

CJKDecomposed 使用说明

一、提交内容

1. CJK 汉字拆分数据库 CJKDecomposed200709061800.xls
2. CJK 汉字单字拆分 xml 表达: 共 20902 个 XML 文件
(Out of The Document)
3. EUDC 区域拆分组件图像表达:
共 222 个 GIF 文件 (存放在子目录 Images 下)
EUDCcomponent.xls (EUDC 区域构件对应的汉字图像)
4. CJKDecomposed 使用说明: CJKDecomposed 使用说明 200709061830.doc

二、分项介绍

1. 以 CJKcomponent 为汉字构件, 见 CJKDecomposed200709061800.xls
2. EUDCcomponent 在数据库中同时用代码和图像表示, 图像的规格是:
 - font style : Simsun
 - font color : #333333
 - font size : 30px
 - GIF 格式
3. CJK 汉字拆分数据库说明:
 - (1) 记录数: 20902 条

- Basic 该字的基本结构。字符串类型

Basic 内容	结构表示	Basic 内容	结构表示
	左右结构		右包围
	上下结构		全包围
	左上包围		左中右
	左下包围		左包围
	上中下		下包围
	上包围		嵌套
X	独体字或不宜再分解的构件		

- IDS：汉字结构串 Ideographic Description Sequence

- IDS=空，亦表示为独体字
- 结构符总是置前，对应构件在后，
 - ◆ 从上到下
 - ◆ 从外到里
 - ◆ 从左到右
 - ◆ 从主到次（镶嵌时）排列。
- 对某些汉字肯定会有多种 IDS 表示，目前只选一种

举例：

- (1) 蓀：  艹 繇 代表上下结构由 艹 和 繇 组成
- (2) 亂：   𠂔 マ  内 又 L 总的为左右结构。左边由上中下结构的 𠂔 マ 和 内 上包围 又 组成，右边由 L 组成

- WholeIDS：将 IDS 中的非 CJKcomponent 构件字的 IDS 代入。
 - ImgInWholeIDS：该 WholeIDS 记录中 EUDC（U+E000~U+E0DE）构件字的图像文件名称序列（可能不只一个）
- 举例：

丝 WholeIDS 字段内容为 𠂇𠂇𠂇𠂇一,

ImgInIDS 字段内容为(3)E055.gif(4)E055.gif

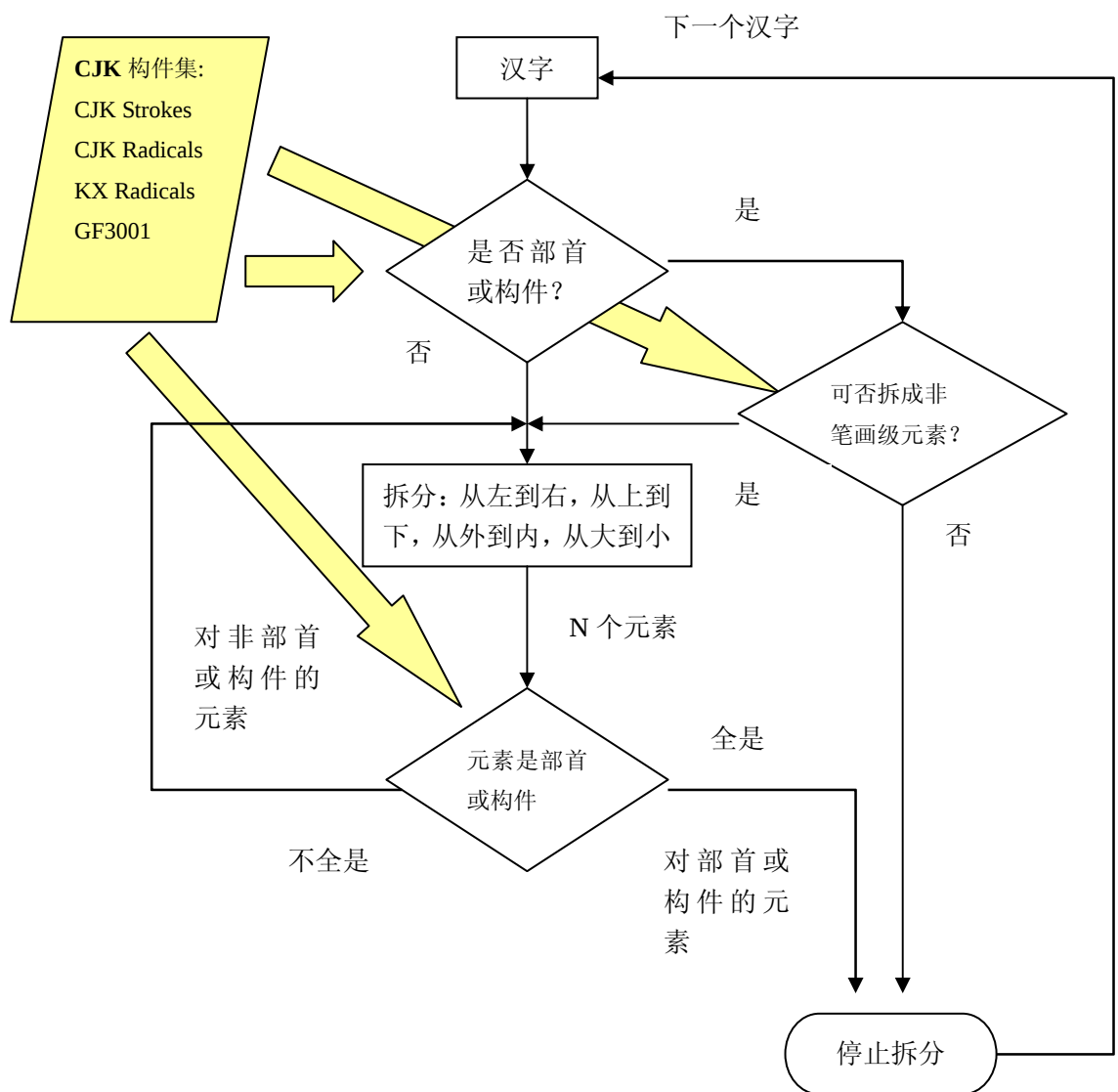
其中:

𠂇 𠂇 为 EUDC 构件字 , 并且在 WholeIDS 内容字串的第三位第四位, E055.gif 表示在 Images 子目录下对应的该代码对应汉字的字迹图像, E055 表示该字 Unicode 代码

2. CJKdecomposed (1st Depth Decomposed) , 遵循下列拆分规则:

<Decomposing Rule>

- A. If an element /component has no meaning, and ALL basic components composing it have no meaning, either, then stop decomposing the element.
- B. If an element /component has no meaning, but SOME basic components composing it have meaning, then decompose the element..
- C. If an element/ component has meaning, then stop decomposing it.
- D. If an element/ component is a CJK Radical or Kangxi Radical., then stop decomposing.
- E. If a Hanzi itself is a CJK Radical or Kangxi Radical, and more than one basic components defined by GF3001 compose the radical, then decompose the Hanzi.



一 4E00	✓ x		丘 4E18	✓ x		丰 4E30	✓ 𠂔	𠂔三 𠂔三
丁 4E01	✓ 𠂔	日一 日一	丙 4E19		日一内 日一内	𠂔 4E31	x	
𠂔 4E02	✓ 𠂔	日一 日一	业 4E1A	✓ x		串 4E32	✓ 𠂔	𠂔月 𠂔日口口
七 4E03	✓ 𠂔	𠂔一 𠂔一	从 4E1B		日从一 日𠂔人人一	弗 4E33	✓ 𠂔	𠂔月 𠂔日口口
上 4E04	✓ 𠂔	日 一 日 一	东 4E1C	✓ x		临 4E34		𠂔 日 𠂔 𠂔 日 𠂔
𠂔 4E05	✓ 𠂔	日一 日一	丝 4E1D		日𠂔 𠂔 𠂔 日𠂔 𠂔 𠂔	𠂔 4E35		日业日 𠂔 日业日 𠂔
𠂔 4E06	✓ x		丞 4E1E		日丞一 日𠂔了 𠂔一	、 4E36	✓ x	
万 4E07	✓ 𠂔	日一 𠂔 日一 𠂔	丢 4E1F		日一去 日一 日土 𠂔	、 4E37	✓ x	
丈 4E08	✓ 𠂔	𠂔一 𠂔 𠂔一 𠂔	北 4E20		日北一 日𠂔 𠂔 𠂔一	丸 4E38	✓ 𠂔	𠂔九、 𠂔九、
三 4E09	✓ 𠂔	日一 二 日一 二	𠂔 4E21		𠂔而 𠂔而	丹 4E39	✓ 𠂔	𠂔月 𠂔 𠂔月 𠂔
上 4E0A	✓ 𠂔	日 𠂔 一 日 𠂔 一	丢 4E22		日 𠂔 去 日 𠂔 日土 𠂔	为 4E3A	✓ 𠂔	𠂔 𠂔 𠂔 𠂔、 𠂔 𠂔 𠂔 𠂔、
下 4E0B	✓ 𠂔	日一 𠂔 日一 𠂔	𠂔 4E23	✓ x		主 4E3B		日、王 日、日一 土
𠂔 4E0C	✓ 𠂔	日一 𠂔 日一 𠂔	两 4E24	✓ 𠂔	日一 𠂔 𠂔 从 日一 𠂔 𠂔 𠂔 𠂔 𠂔	井 4E3C	𠂔 𠂔	𠂔井、 𠂔井、
不 4E0D	✓ x		𠂔 4E25		日亚 日亚	𠂔 4E3D		日一 𠂔 𠂔 𠂔、 日一 𠂔 𠂔 𠂔、
与 4E0E		𠂔与一 𠂔与一	並 4E26		日 𠂔 亚 日 𠂔 亚	举 4E3E		日兴 𠂔 日 𠂔 𠂔 一 八 𠂔
𠂔 4E0F	✓ x		𠂔 4E27		日𠂔 𠂔 𠂔 日𠂔 𠂔 𠂔	𠂔 4E3F	✓ x	
𠂔 4E10	✓ 𠂔	日下 𠂔 日下 𠂔	𠂔 4E28	✓ x		𠂔 4E40		
丑 4E11	✓ 𠂔	日𠂔 𠂔 𠂔 日𠂔 𠂔 𠂔	𠂔 4E29	✓ x		𠂔 4E41	✓ x	
𠂔 4E12		日 𠂔 一 日 𠂔 𠂔、一	个 4E2A	✓ 𠂔	日人 日人	𠂔 4E42	✓ 𠂔	𠂔 𠂔 𠂔 𠂔 𠂔 𠂔
专 4E13	✓ x		𠂔 4E2B	✓ 𠂔	日 𠂔 𠂔 日 𠂔 𠂔	乃 4E43	✓ x	
且 4E14	✓ x		𠂔 4E2C	✓ x		𠂔 4E44	✓ x	
𠂔 4E15		日 𠂔 一 日 𠂔 𠂔 一	中 4E2D	✓ 𠂔	𠂔 𠂔 𠂔 𠂔	久 4E45	✓ x	
世 4E16	✓ 𠂔	𠂔 𠂔 𠂔 𠂔 𠂔 𠂔	𠂔 4E2E		𠂔 𠂔 𠂔 𠂔 𠂔 𠂔	久 4E46		𠂔 𠂔 𠂔 𠂔 𠂔 𠂔
𠂔 4E17	✓ x		𠂔 4E2F	✓ 𠂔	𠂔 𠂔 𠂔 𠂔 𠂔 𠂔	𠂔 4E47		日 𠂔 七 日 𠂔 七

么	日ノム	习	√ 日ノ?	𪛗	日也母
4E48	日 日ノム	4E60	日 日ノ?	4E78	日 日也母
义	日、ノ	乡	√ 日ノ	𪛗	日卓シ
4E49	日 日、ノ	4E61	日 日ノ	4E79	日 日卓シ
𪛗	日ノ	𪛗	日山シ	𪛗	日所乙
4E4A	日 日ノ	4E62	日 日山シ	4E7A	日 日戸斤乙
之	√	𪛗	日玄シ	𪛗	日於乙
4E4B	x	4E63	日 日玄シ	4E7B	日 日方日人、乙
𪛗	日ノ一	𪛗	日下乙	𪛗	日注乙
4E4C	日 日ノ一	4E64	日 日下乙	4E7C	日 日ノ 日、日一土乙
𪛗	√	𪛗	日ノ乙	𪛗	日者乙
4E4D	x	4E65	日 日ノノ乙	4E7D	日 日ノ日乙
乎	√	𪛗	√	𪛗	日卓 乞
4E4E	x	4E66	x	4E7E	日 日卓 日ノ乙
𪛗	日ノ之	𪛗	日斗乙	𪛗	日 日ノ系シ
4E4F	日 日ノ之	4E67	日 日斗乙	4E7F	日 日 日ノ系シ
𪛗	√	𪛗	日合シ	𪛗	√ 日ノ
4E50	x	4E68	日 日合シ	4E80	日 日ノ
𪛗	√	𪛗	日占シ	𪛗	日 日十干乞
4E51	x	4E69	日 日占シ	4E81	日 日十干日ノ乙
𪛗	日ノ	𪛗	日乙	𪛗	日 日ノ内又シ
4E52	日 日ノ	4E6A	日 日乙	4E82	日 日 日ノ内又シ
𪛗	日ノ、	𪛗	日加乙	𪛗	日 日ノ
4E53	日 日ノ、	4E6B	日 日加力口乙	4E83	日 日 日ノ
𪛗	日ノ	𪛗	日巨乙	𪛗	日 日ノ
4E54	日 日ノ	4E6C	日 日巨乙	4E84	日 日 日ノ
𪛗	日ノ	𪛗	日石乙	𪛗	√
4E55	日 日ノ	4E6D	日 日石乙	4E85	x
𪛗	日ノ	𪛗	日卯乙	𪛗	√ 日ノ
4E56	日 日ノ	4E6E	日 日卯ノ乙	4E86	日 日ノ
𪛗	√ 日ノ	𪛗	日乎乙	𪛗	日ノ
4E57	日 日ノ	4E6F	日 日乎乙	4E87	日 日ノ
𪛗	日ノ	𪛗	日ノ	𪛗	√ 日ノ
4E58	日 日ノ	4E70	日 日ノ	4E88	日 日ノ
乙	√	𪛗	日舌シ	𪛗	日ノ
4E59	x	4E71	日 日舌シ	4E89	日 日ノ
𪛗	√	𪛗	日次乙	𪛗	√ 日ノ
4E5A	x	4E72	日 日次乙	4E8A	日 日ノ
𪛗	√	𪛗	日孚シ	𪛗	√ 日ノ
4E5B	x	4E73	日 日孚シ	4E8B	日 日ノ
𪛗	√	𪛗	日折乙	𪛗	√
4E5C	x	4E74	日 日折斤乙	4E8C	x
九	√ 日ノ	𪛗	日辛シ	𪛗	日二ノ
4E5D	日 日ノ	4E75	日 日辛シ	4E8D	日 日二ノ
𪛗	日ノ乙	𪛗	日甫乙	𪛗	√ 日二ノ
4E5E	日 日ノ乙	4E76	日 日甫乙	4E8E	日 日二ノ
也	√ 日ノ	𪛗	日沙乙	𪛗	日一ノ
4E5F	日 日ノ	4E77	日 日沙ノ乙	4E8F	日 日一ノ