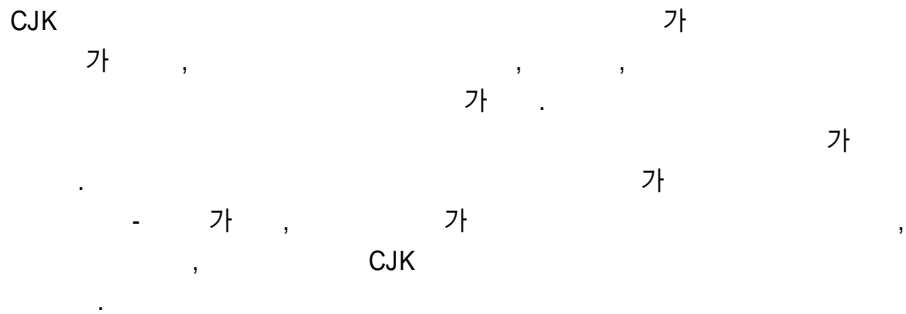


Support on Ideograph Characters Search of Unicode Based Information System

(SoYoung Yoon)*



ABSTRACT

Unicode Han ideograph character set differed from the our principle of the phonetic value ordering in that it followed the principle of KangXi radical-stroke ordering of the characters. Therefore, information system should support ideograph search on precise analysis of materials which consist of korean character (hangul) and ideograph character (hanja). History Information system has been maintaining Hanja(Chinese Character) to Hangul Dictionary, Terminology Dictionary for composition, borrowing, non-ideographic principles, Variant Forms Dictionary, and Recently discovered Chinese Characters List.

: , , , 가,
 unicode, ideograph character search, character presentation, hangul phonetic value, variant forms of chinese character

* (syoon@moe.go.kr)

1. 5.1.0 가

, , 21 , 가

가 IT 가 CJK

, 가 가 가 가

가 가 가 가 가

가 가 가 가 가

가 가 가 가 가

가 가 가 가 가

1990 가 가 가

(Unicode) 가 가 가

1996 v.2.0 가 가 가

, 2006 v.5.0 2007 가 가 가

2.

2.1

가

1980

가

. 1991

ISO/IEC 10646 (UCS; Universal Multiple-Octet Coded Character Set)

가

1985

가

1987

ISO -2022

‘ KS C 5601(1998

KS X 1001)’

2,350 ,

4,888 , 1,128 ,

470

가

. KS C 5601

가

1991

1,930 1,677

가 ‘ KS C 5657(2001 KS X 1002)’

가

가 가

가 가

1991 KS C 5601 가

1992 ISO/IEC JTC1/SC2
BMP(Basi Multilingual Plane;) 11,172

240 (90 , 66 , 82 , 2) 가 240

1995 가

2.0 11,172 BMP 가 , 가 (, (東國正韻 (蒙古語, . 2007). 式) (漢語),)

- 240 : U+1100 U+11F9(Hangul Jamo) 90 , 66 , 82 , 2
- 94 : U+3131 U+318E(Hangul Compatibility Jamo) 가
- 11,172 : U+AC00 U+D7A3(Hangul Syllables) 19 , 21 , 27 11,172 가 가 가 가 가1100 11FF

D7A3 , AC00 < : 34 >
 가 Hangun Jamo: U+115A U+115E
 Hangun Jamo Extended - A: U+A960
 가 U+A97C
 가 < : 28 >
 Hangun Jamo: U+11A3 U+11A7
 Hangun Jamo Extended - B: U+D7B0
 U+D7C6
 , 가 < : 55 >
 가 Hangun Jamo: U+11FA U+11FF
 Hangun Jamo Extended - B: U+D7CB
 U+D7FB

2.2

(CJK Unified Ideographs)

(, , , ,)
 . 2007). <
 > 117 가 BMP
 가 , 'CJK'
 가 가
 'CJKV'
 65,536
 4
 ISO/IEC JTC1/SC2/WG2
 50 <
 > 117 가 BMP 가 가 .
 가 가
 <
 > < > 117
 BMP

	V1.0.0	V1.0.1	V1.1	V2.0	V2.1	V3.0	V3.1	V3.2	V4.0	V4.1	V5.0
Alphabets, Symbols	4,734	4,728	6,290	6,491	6,493	10,210	11,798	12,753	13,973	15,117	16,486
Han (URO)		20,902	20,902	20,902	20,902	20,902	20,902	20,902	20,902	20,902	20,902
Han (URO Extension)										22	22
Han Extension A						6,582	6,582	6,582	6,582	6,582	6,582
Han Extension B							42,711	42,711	42,711	42,711	42,711
Han Compatibility		302	302	302	302	302	844	903	903	1,009	1,009
Subtotal Han		21,204	21,204	21,204	21,204	27,786	71,039	71,098	71,098	71,226	71,226
Hangul Syllables	2,350	2,350	6,656	11,172	11,172	11,172	11,172	11,172	11,172	11,172	11,172
Graphic Characters	7,084	28,282	34,150	38,867	38,869	49,168	94,009	95,023	96,243	97,515	98,884
Format Characters	12	12	18	18	18	26	131	133	139	140	140
Graphic + Format	7,096	28,294	34,168	38,885	38,887	49,194	94,140	95,156	96,382	97,655	99,024
Controls	65	65	65	65	65	65	65	65	65	65	65
Private Use	5,632	6,144	6,400	137,468	137,468	137,468	137,468	137,468	137,468	137,468	137,468
Total Assigned	12,793	34,503	40,633	176,418	176,420	186,727	231,673	232,689	233,915	235,188	236,557
Surrogate Code Points				2,048	2,048	2,048	2,048	2,048	2,048	2,048	2,048
Noncharacters	2	2	2	34	34	34	66	66	66	66	66
Total Designated	12,795	34,505	40,635	178,500	178,502	188,809	233,787	234,803	236,029	237,302	238,671
Reserved Code Points	52,741	31,031	24,901	935,612	935,610	925,303	880,325	879,309	878,083	876,810	875,441

< 1>

(The Unicode Consortium. 2007. p.1101)

가 가

(BMP) 16 가

20,902 A(Ext.A) 6,582 , (. 2007). 가

B(Ext.B) 42,711 , C(Ext.C) 4,219 CJK

가 . 7 18,065

가 가

가

20,902 가

B

가

가

가 가

가 가

- BMP
- CJK : 15,390
- CJK Ext.A : 1,836
- CJK : 268
- CJK Ext.B : 166
- CJK Ext.C : 405

KSC5601-1987

가 가

가

가 . 가
刻() 街(가) 2 街()가
刻() 6 行() 街(가)

가 가 가2 가
가 가 가
' (車) ' ,
KSC5601 가

(車)
KSC5601

가 ,
,

가
가 가

3.1 - 가

, , , 가

3.

가가

가가2
가가

가

2

가 가

' 가'

- 가

< 1 >

				가
BMP(Basic Multi-lingual Plane)	U+4E00(一) ~ U+9FA5(韻)	20,902	20,902	· 가 8,760
A	U+3400(北) ~ U+4DB5(龔)	6,582	27,484	· 가
(KSX)	U+F900(峴) ~ U+FA0B(鄺)	268	27,752	· KSC 5601 · 288 가
(Big 5)	U+FA0C(紅) ~ U+FA0D(馥)	2	27,754	· 가
(IBM 32)	U+FA0E(婁) ~ U+FA2D(瞿)	32	27,786	· 가 · U+FA2E, U+FA2F
(JIS X)	U+FA30(侮) ~ U+FA6A(賴)	59	27,845	· 가

	A	B	C	D	E	F	G	H
1	자형	코드	대표 음1	대표 음2	대표 음3	대표 음4	대표 음5	대표 음6
737	嬭	U+36E4	의	이				
738	媧	U+36E5	담					
739	媧	U+36E6	담	남				
740	媧	U+36E7	담					
741	媧	U+36E8	강					
742	媧	U+36E9	축					
743	媧	U+36EA	영					
744	媧	U+36EB	석					
745	媧	U+36EC	목	육				
746	媧	U+36ED	석					
747	媧	U+36EE	수					
748	媧	U+36EF	면	반	찬			
749	媧	U+36F0	혼					
750	媧	U+36F1	외					
751	媧	U+36F2	발					
752	媧	U+36F3	예					
753	媧	U+36F4	뇌					

< 2 > - 가

가 가 3.2

가 가 가 가 가
 (. 2005).
 < 1> 가 가
 가 , 3.1 (三一運動), 1 (正一品)

가 가
 < 2> 가 가
 가

	A	B	C	D	G	H	I	J	L
1	한자점 연계	연계 종기	대표어	대표어 빈도수	검색어	역동점 색결과	Navor점 색결과	국문점 색결과	
2192	학	학	學	2	유학				學
2193	갑	조	平三갑二	2	平三갑二	0	-	0	平三
2194	갑	조	干三갑二	2	간삼조이	0	-	1	干三
2195	갑	조	大갑	2	대조	0	-	0	大
2196	갑	초	召兒	3	초이	971	-	4057	召兒
2197	갑	초	召兒	3	召兒	1	-	3	召兒
2198	갑	초	召兒	3	召兒	1	-	2	召兒
2199	학	조	玉練	2	유조				玉練
2200	학	조	玉練	2	유조				玉練
2201	갑	조이	召兒	2	초이	0	-	0	召兒
2202	갑	조이	召兒	2	초이	0	-	0	召兒
2203	甲	수	副南率	2	부남수				副南
2204	甲	수	副南率	2	부남수				副南
2205	甲	수	南率	2	남수				南
2206	甲	수	南率	2	남수				南
2207	甲	수	南率	2	남수				南
2208	甲	수	南率	2	남수				南
2209	학	희	麗賓	4	유희				麗賓



< 4>

	A	B	C	D	E	F	G
I	한자음언어	대표어		대표어 빈도수	검출	확정 어휘 수	
170	이두	白乎所	3	7	TRUE	7	사오 비
171	이두	白乎所	3	7	TRUE	7	여포오 비
172	이두	白乎所	3	7	TRUE	7	음신 비
173	이두	白乎所	3	7	TRUE	7	수호비
174	이두	白乎所	3	7	TRUE	7	수호비
175	이두	白乎所	3	4	FALSE	7	白乎所
176	이두	白乎所	3	4	FALSE	7	사오 비
177	이두	白乎所	3	4	FALSE	7	수호비
178	이두	白乎所	3	4	FALSE	7	수호비
179	이두	不檢在	3	5	FALSE	7	不檢在
180	이두	不檢在	3	5	FALSE	7	아닌
181	이두	不檢在	3	5	FALSE	7	아닌 것
182	이두	不檢在	3	5	FALSE	7	아닌디건
183	이두	不檢在	3	5	FALSE	7	아닌지건
184	이두	顯可只	3	4	FALSE	7	로부디
185	이두	顯可只	3	4	FALSE	7	顯可只
186	이두	顯可只	3	4	FALSE	7	순합적
187	이두	顯可只	3	4	FALSE	7	조차도
188	이두	顯音可	3	3	FALSE	7	顯音可
189	이두	顯音可	3	3	FALSE	7	수호비

< 5>

< 4> . (召
 史)
 가
 (. <
 2005). 5>
 2005 5,680 , 2,631
 2006 가 3.3
 4,380
 9,309 ,
 가 (,)
 < 3> ,
 () , 가
 3.0(Extension A)
 27,484
 () ,

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	연번	대표	부수	총획	이체1	구분	이체2	구분	이체3	구분	이체4	구분	이체5
1037	1036	從	彳	11	從	속	從*	속	從	동	從	본	
1038	1037	復	彳	12									
1039	1038	微	彳	13	微	동							
1040	1039	德	彳	15	德	속	惠	표	惠	속	惠	동	惠
1041	1040	微	彳	15	微	동							
1042	1041	微	彳	15	微	간							
1043	1042	微	彳	16									
1044	1043	忉	心	5									
1045	1044	忉	心	6									
1046	1045	忉	心	6	忉	동							
1047	1046	忌	心	7									
1048	1047	忘	心	7	忘								
1049	1048	恕	心	7	恕	고							
1050	1049	忍	心	7	忍	동							
1051	1050	忱	心	7	忱	속							
1052	1051	忉	心	8	忉	보							



< 7>

가 가
가 가
가

< 6>

가 가
가 가

(2005).

< 7>

3.4

가 가

가

가 , 讀) ,

： 1

가 (固有漢字) , 1

가 𠵼(𠵼)·𠵼(𠵼)·𠵼(𠵼)·𠵼(𠵼)·𠵼(𠵼)

造字 ：

(造字原理) (國音字)

(形聲) (造字) ‘ 釗()·串()·卜()·歛()’

‘ 가 (假借) ‘ 召(棗)·干(薑)’ 가

： 가

‘ 澹()·羣()·番()·憚()’ (國義字) , ‘ 遷[]·評[]’

(形部) (聲部)

가 ‘ 가 DB

(, 2004).

(終聲)

가 : (吏 가 ,

終聲	漢字	造字例
ㄴ	隱(卬)/分/叱	𠵼, 𠵼, 𠵼, 𠵼
ㄹ	乙	𠵼, 𠵼, 𠵼, 𠵼, 𠵼, 𠵼
ㅁ	音/味	𠵼, 𠵼
ㅂ	邑(巴)/業/立	𠵼, 𠵼, 𠵼, 𠵼
ㅅ	叱	𠵼, 𠵼, 𠵼, 𠵼, 𠵼, 𠵼
ㅇ	同	𠵼, 𠵼

‘ ㄹ ’ ‘ 加應·末應·無應 ’ ‘ 應 ’

‘ ㅁ ’ ‘ 古勿·今勿 ’ ‘ 勿 ’

가 .
 가
 1) ‘隱(丁)/分/叱’ ‘ ㄴ ’
 , 2) ‘乙’ ‘ ㄹ ’
 , 3) ‘音/味’ ‘ ㅁ ’ , 4)
 ‘邑(巴)/業/立’ ‘ ㅂ ’ , 5)
 ‘叱’ ‘ ㅅ ’ , 6) ‘同/應’
 ‘ ㅍ ’ 가 .
 가 ,

가 .
 가
 가
 ‘叱’ ‘ ㅅ ’
 造字 ‘ ㄱ ’
 , ‘ 龔(,) ’ ; ‘ 龔(,) ’ ;
 ‘ 龔(,) ’
 . ‘ 宋加龔(가
) ’ ‘ 龔() ’
 .

가 가 ,
 (2005).
 1)
 . 誤讀
 .
 .
 .

2) SuperCJK Ver.14.0
 .
 .
 ±1-2
 3) IRG ‘Annex S’
 . Annex S(IRGN951)
 .
 .
 .
 4) Ext.C1
 Ext.C2 가
 . Ext.C1 2002 4
 IRG
 . Ext.C1
 C2
 가
 5)
 . , ,
 . ‘ (IRG N881) ’
 6)
 2006 5,980
 , 가

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
	발행번호	발서코드	기관명	제정년도	형태	출판 서명	출판 물공명	부서 정보	파일명	중문단어	발행 위치	차별	비고	부수번호
1	85.국공간 .8341	유정리.3047	국공간학회 학술구회	2000	역사문화재 지표	9명	국공간교섭 사건(유정리) 제	204368	발	대륙에는 陸 1中, 半島鮮 1東洋 ★(東/北) 北, 物子 1中기 魯賢 오기	국공간 8341	漕	漢/北	005
7	88.경북대 .8033	전북대.3035	전북대학교 학술논문	2000	고문서		1850년에 慶 禮館 刊印된 에 1卷은 古 新가 舊新刊 서기후주	8800297.A 81	8800297.A and	口也. 伊豆國本居十 五年前. 一列五枚書/乙文年二十 四	8800297.A 81	毛	漢/乙	005
8	13.한글본 .8421	한글본_기타 분류.8307	한글학술집 연구회	2000	조선조역사 발행지표		제.A.24045 .001	제.A.24045 .001.83000 1	a.kma.2404 5.001.and	上士集後漢書子集★(東/ 北)Onke type="reader">제 名量/	제.A.24045 .001.83000 1	鹿	漢/乙	005
9	13.한글본 .8154	한글본_공공 문헌.8314	한글학술집 연구회	2000	한국공공문 헌지표		제.A.20118 .001	제.A.20118 .001.83051 1	a.kma.2011 8.000.and	第7卷(漢)子集下及一先史書★(中 +東)<Onke type="reader">제 其量漢書中集	제.A.20118 .001.83051 1	矜	中+東	000
10	88.경북대 .8034	전북대.3024	전북대학교 학술논문	2000	고문서			2	18592_and1	1750년에 高 麗館 刊印된 에 1卷은 韓 의 古新刊 書 기후주	2	古	中/乙	008
11	90.민중 .8318	민중.8028	한국문학수 련회	2000	한국문학수 련지표	이재우 2008 韓國文學 集(漢)	제.A.20040 .001	제.A.20040 .001.83001 2	a.kma.2004 0.001.and	次Onke>稿子上/等主 第7卷(漢)子集上Onke type="reader">次/其量 韓文集(漢)第7卷上下★(中/東/ 北)Onke type="reader">集	제.A.20040 .001.83001 2	藝	中/東/北	008
13	80.국경 .8033	1307.303	국사편찬위 원회	2000	한국군현대 인물지표	대한민국인 물지표 1년		11000_01	88.8012.80 0.000	本集(中+北)中+北	88.8012.80 0.000	他	中+北	009
14	90.민중 .8026	민중.8026	한국문학수 련회	2000	한국문학수 련지표	이재우 2008 韓國文學 集(漢)	제.A.20040 .001	제.A.20040 .001.83001 2	a.kma.2004 0.001.and	次Onke>稿子上/等主 第7卷(漢)子集上Onke type="reader">次/其量 韓文集(漢)第7卷上下★(中/東/ 北)Onke type="reader">集	제.A.20040 .001.83001 2	人	中+北	009
15	90.민중 .8026	민중.8026	한국문학수 련회	2000	한국문학수 련지표	이재우 2008 韓國文學 集(漢)	제.A.20040 .001	제.A.20040 .001.83001 2	a.kma.2004 0.001.and	次Onke>稿子上/等主 第7卷(漢)子集上Onke type="reader">次/其量 韓文集(漢)第7卷上下★(中/東/ 北)Onke type="reader">集	제.A.20040 .001.83001 2	此	中+北	009

< 8>

가

< 8>

4.

가

가

가

가

가

가

가

가

가

가

가

가 F9

가

가

가

, F9

가

F9

가

. 2005. . 2005. 21

. 2003.

“ KS

”

2000. . 2001. “

” 8

. 2004. 23 - 28.

KRISTAL - 2002. “

15 : 36 - 40. , 1995.

. 2004. “ Allen J.D., Joe Baker, Richard Cook,

” 2004 Mark Davis, Michael Everson,

. 2007. “ Asmus Freytag, Jonh H. Jenkins,

” 8 Mike Ksar, Rick McGowan, Lisa

19 : 1 - 3. Moore, Eric Muller, Markus

2004 “ Scherer, Michel Suignard, and

() KS X 1001 ; Ken Whistler. 2007The Unicode

. Standard 5.0, Unicode

. 2004. “ 固有漢字 Consortium, Addison - Wesley.

(1) ” 8 Korpela, Jukka K. 2006. Unicode

14 : 7 - 8. Explained O'Reilly.

. 2007. “

” 8

34 2 : 169 -

177.

. 2005. “ DB

” 8

15 : 2 - 4.

