한자漢字검색 텔레그램봇

PyCon APAC 2016 Day1(2016.08.13) - Lightning Talk
성대현(DaeHyun Sung/成大鉉/成大铉/ソン·デヒョン)

Sponsor of CJK

Unicode Consortium ✔ @unicode · 4월 20일   번역 보기 ⊕

성대현(成大鉉 DaeHyun Sung)(@studioego)is now a Bronze sponsor of CJK CJK-5FB7!#UnicodeSponsor unicode.org/consortium/ado…

OFFICIAL BRONZE SPONSOR · UNICODE CONSORTIUM · ADOPT A CHARACTER

德

↩ ↻ 4 ♥ 2 •••

https://twitter.com/unicode/status/722829746233409536

德 성대현(成大鉉, DaeHyun Sung) (@studioego)

http://www.unicode.org/consortium/adopted-characters.html#b5FB7

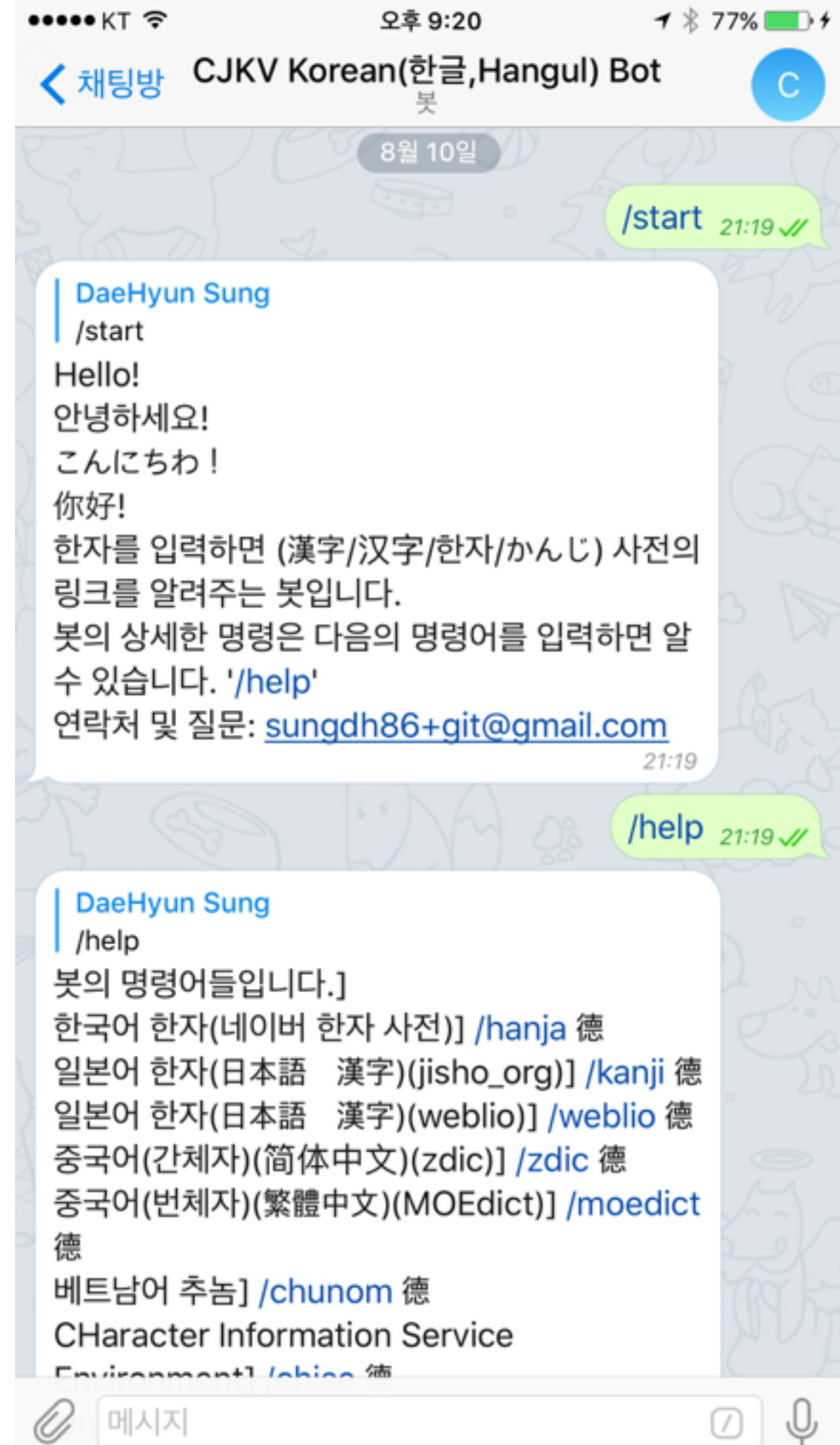# 만들어보게 된 계기

- Python이 뭘까?
- Telegram bot?
- 파알못(파이썬을 알지 못함)이지만 한번 만들어볼까?

Telegram Bot

- Telegram Bot (English)
  - @CJKVBot
    https://telegram.me/cjkvBot
- 텔레그램 봇 (한국어)
  - @CJKV_korean_bot
    https://telegram.me/CJKV_korean_bot

이 봇은 무엇을 할 수 있나요?

Unihan(CJKV, Unified Chinese-Japanese-Korean-Vietnamese Chinese characters) data info bot

8월 10일

/start   21:19 ✓✓

DaeHyun Sung
/start
Hello!
안녕하세요!
こんにちわ！
你好!
This is Chinese Character(漢字/汉字/한자/かんじ) dictinary link bot.
Detail options usage enter '/help'
Contract & Question:
sungdh86+git@gmail.com
     21:19

/help   21:19 ✓✓

DaeHyun Sung
/help
This bot's Command Option]
Korean(한국어) Hanja(Naver Hanja Dictionary)] /hanja 德
Japanese kanji(日本語　漢字)(jisho_org)]

8월 10일

/start   21:19 ✓✓

DaeHyun Sung
/start
Hello!
안녕하세요!
こんにちわ！
你好!
한자를 입력하면 (漢字/汉字/한자/かんじ) 사전의 링크를 알려주는 봇입니다.
봇의 상세한 명령은 다음의 명령어를 입력하면 알 수 있습니다. '/help'
연락처 및 질문: sungdh86+git@gmail.com
     21:19

/help   21:19 ✓✓

DaeHyun Sung
/help
봇의 명령어들입니다.]
한국어 한자(네이버 한자 사전)] /hanja 德
일본어 한자(日本語　漢字)(jisho_org)] /kanji 德
일본어 한자(日本語　漢字)(weblio)] /weblio 德
중국어(간체자)(简体中文)(zdic)] /zdic 德
중국어(번체자)(繁體中文)(MOEdict)] /moedict 德
베트남어 추놈] /chunom 德
CHaracter Information Service
[Environment] /chica 德

한자(漢字)

- Chinese Character
- 漢字/汉字
- ㄏㄢˋ ㄗˋ/hànzì
- 한자
- かんじ
- hán tự

ㄏㄢˋ ㄗˋ ㄨㄣˊ ㄏㄨㄚˋ ㄑㄩㄢ
🇹🇼 漢字文化圈

hàn zì wén huà quān
🇨🇳 汉字文化圈

hon³ zi⁶ man⁴ faa³ hyun¹
🇭🇰🇲🇴 漢字文化圈

한 자 문 화 권
🇰🇷🇰🇵 漢字文化圈
han ja mun hwa gwon

vòng văn hóa chữ hán
🇻🇳 綾文化𡨸漢

かん じ ぶん か けん
🇯🇵 漢字文化圈
kan ji bun ka ken

Reference:
https://commons.wikimedia.org/wiki/File:漢字文化圈％EF％BC％8F汉字文化圈_·_한자_문화권_·_Vòng_văn_hóa_chữ_Hán_·_漢字文化圈.jpg

漢字사전 링크 조사

- **한국어 한자사전(漢字辭典, 玉篇, Korean-Chinese Character Dictionary)**
  - 네이버 한자사전 http://hanja.naver.com/
- **일본어 사전(日本語, Japanese Dictionary)**
  - Weblio http://ejje.weblio.jp
  - Jisho http://jisho.org
- **베트남어 사전(Vietnamese Chữ Nôm Dictionary)**
  - Chu nom http://www.chunom.org

漢字사전 링크 조사

- **중국어 사전(漢語/汉语, Chinese Dictionary)**
  - ZDict (简体中文) http://www.zdic.net
  - MOE Dict (**M**inistry **O**f **E**ducation, Taiwan's Chinese Dictionary) (繁體中文) https://www.moedict.tw

漢字사전 링크 조사

- **Glyph Wiki**
  - English http://en.glyphwiki.org
  - 한국어(Korean) http://ko.glyphwiki.org
  - 日本語(Japanese): http://glyphwiki.org
  - 简体中文(Simplified Chinese) http://zhs.glyphwiki.org
  - 繁體中文(Traditional Chinese) http://zht.glyphwiki.org

漢字사전 링크 조사

- **Unicode Consortium**
  - Unihan database lookup
  - http://www.unicode.org/charts/unihan.html
- **Chise (CHaracter Information Service Environment)**
  - http://www.chise.org/ids-find

Unicode

- Unihan Database
  - http://www.unicode.org/charts/unihan.html
- Code Point
  - 德 - U+5FB7
    - dak1, 덕, dé, OSHIE, TOKU, đức
  - 成 - U+6210
    - cing4 seng4 sing4, 성, chéng, NARU TAIRAGERU NASU, SEI JOU, thành

U+5FB7
Chinese Character:德|CodePoint: U+5FB7|field: kCantonese|content: dak1
kCantonese = CodePoint: U+5FB7|field: kCantonese|content: dak1
U+5FB7
Chinese Character:德|CodePoint: U+5FB7|field: kDefinition|content: ethics, morality, virtue
kDefinition = CodePoint: U+5FB7|field: kDefinition|content: ethics, morality, virtue
U+5FB7
Chinese Character:德|CodePoint: U+5FB7|field: kHangul|content: 덕
kHangul = CodePoint: U+5FB7|field: kHangul|content: 덕
U+5FB7
Chinese Character:德|CodePoint: U+5FB7|field: kHanyuPinlu|content: dé(97)
kHanyuPinlu = CodePoint: U+5FB7|field: kHanyuPinlu|content: dé(97)
U+5FB7
Chinese Character:德|CodePoint: U+5FB7|field: kHanyuPinyin|content: 20841.170:dé
kHanyuPinyin = CodePoint: U+5FB7|field: kHanyuPinyin|content: 20841.170:dé
U+5FB7
Chinese Character:德|CodePoint: U+5FB7|field: kJapaneseKun|content: OSHIE
kJapaneseKun = CodePoint: U+5FB7|field: kJapaneseKun|content: OSHIE
U+5FB7
Chinese Character:德|CodePoint: U+5FB7|field: kJapaneseOn|content: TOKU
kJapaneseOn = CodePoint: U+5FB7|field: kJapaneseOn|content: TOKU
U+5FB7
Chinese Character:德|CodePoint: U+5FB7|field: kKorean|content: TEK
kKorean = CodePoint: U+5FB7|field: kKorean|content: TEK
U+5FB7
Chinese Character:德|CodePoint: U+5FB7|field: kMandarin|content: dé
kMandarin = CodePoint: U+5FB7|field: kMandarin|content: dé
U+5FB7
Chinese Character:德|CodePoint: U+5FB7|field: kTang|content: dək
kTang = CodePoint: U+5FB7|field: kTang|content: dək
U+5FB7
Chinese Character:德|CodePoint: U+5FB7|field: kVietnamese|content: đức
kVietnamese = CodePoint: U+5FB7|field: kVietnamese|content: đức
U+5FB7
Chinese Character:德|CodePoint: U+5FB7|field: kXHC1983|content: 0224.010:dé
kXHC1983 = CodePoint: U+5FB7|field: kXHC1983|content: 0224.010:dé

U+6210

Chinese Character:成|CodePoint: U+6210|field: kCantonese|content: cing4 seng4 sing4

kCantonese = CodePoint: U+6210|field: kCantonese|content: cing4 seng4 sing4

U+6210

Chinese Character:成|CodePoint: U+6210|field: kDefinition|content: completed, finished, fixed

kDefinition = CodePoint: U+6210|field: kDefinition|content: completed, finished, fixed

U+6210

Chinese Character:成|CodePoint: U+6210|field: kHangul|content: 성

kHangul = CodePoint: U+6210|field: kHangul|content: 성

U+6210

Chinese Character:成|CodePoint: U+6210|field: kHanyuPinlu|content: chéng(4462) cheng(17)

kHanyuPinlu = CodePoint: U+6210|field: kHanyuPinlu|content: chéng(4462) cheng(17)

U+6210

Chinese Character:成|CodePoint: U+6210|field: kHanyuPinyin|content: 21399.010:chéng

kHanyuPinyin = CodePoint: U+6210|field: kHanyuPinyin|content: 21399.010:chéng

U+6210

Chinese Character:成|CodePoint: U+6210|field: kJapaneseKun|content: NARU TAIRAGERU NASU

kJapaneseKun = CodePoint: U+6210|field: kJapaneseKun|content: NARU TAIRAGERU NASU

U+6210

Chinese Character:成|CodePoint: U+6210|field: kJapaneseOn|content: SEI JOU

kJapaneseOn = CodePoint: U+6210|field: kJapaneseOn|content: SEI JOU

U+6210

Chinese Character:成|CodePoint: U+6210|field: kKorean|content: SENG

kKorean = CodePoint: U+6210|field: kKorean|content: SENG

U+6210

Chinese Character:成|CodePoint: U+6210|field: kMandarin|content: chéng

kMandarin = CodePoint: U+6210|field: kMandarin|content: chéng

U+6210

Chinese Character:成|CodePoint: U+6210|field: kTang|content: *zhiɛng

kTang = CodePoint: U+6210|field: kTang|content: *zhiɛng

U+6210

Chinese Character:成|CodePoint: U+6210|field: kVietnamese|content: thành

kVietnamese = CodePoint: U+6210|field: kVietnamese|content: thành

U+6210

Chinese Character:成|CodePoint: U+6210|field: kXHC1983|content: 0135.080:chéng

kXHC1983 = CodePoint: U+6210|field: kXHC1983|content: 0135.080:chéng

Dictionary URL Patterns

- Patterns
  - Example) 德
    - Chinese Character
      - https://www.moedict.tw/德
      - http://hanja.naver.com/hanja?q=德
    - Code Point
      - http://www.zdic.net/z/jd/?u=5fb7
      - http://www.unicode.org/cgi-bin/GetUnihanData.pl?codepoint=5FB7

Development

- Python3
- Heroku (Free Dyno)
- Telegram python API
  - https://github.com/eternnoir/pyTelegramBotAPI

Create Telegram Bot

BotFather ✔
bot

DaeHyun   ✓✓ 25.06.16
/newbot

BotFather   25.06.16
Alright, a new bot. How are we going to call it? Please choose a name for your bot.

DaeHyun   ✓✓ 25.06.16
CJKV Korean(한글,Hangul) Bot

BotFather   25.06.16
Good. Now let's choose a username for your bot. It must end in `bot`. Like this, for example: TetrisBot or tetris_bot.

DaeHyun   ✓✓ 25.06.16
CJKV_korean_bot

BotFather   25.06.16
Done! Congratulations on your new bot. You will find it at telegram.me/CJKV_korean_bot. You can now add a description, about section and profile picture for your bot, see /help for a list of commands. By the way, when you've finished creating your cool bot, ping our Bot Support if you want a better username for it. Just make sure the bot is fully operational before you do this.

Get a Bot Token: xxxx2134:1234:xxx (example)

Result

DaeHyun
/moedict 德

CJKVBot

DaeHyun Sung
/moedict 德

德
CJK UNIFIED IDEOGRAPH-5FB7
https://www.moedict.tw/德

www.moedict.tw
德 - 萌典
人類共同遵循的規範。品行、作風。恩澤、恩惠。心意、信念。德意志聯邦共和國的簡稱。姓。如漢代西域有德若國，後人以國為姓，簡姓德。感激、感恩。好的、善的。

DaeHyun
/hanja 韓

CJKVBot

DaeHyun Sung
/hanja 韓

韓
CJK UNIFIED IDEOGRAPH-97D3
http://hanja.naver.com/hanja?q=韓

Q&A, Reference

- Twitter @studioego

- sungdh86+git at gmail.com

- Github
  https://github.com/studioego/CJKV/tree/master/
  unihan-dict-link

- CJKV Information Processing, 2nd Edition
  Written by Ken Lunde