
Enhancing Chinese character normalization in Primo with the HKIUG TSVCC mapping table

Alan Ng

Systems Librarian

The Chinese University of Hong Kong Library

Agenda

- ❖ Primo out-of-box character normalization
- ❖ Background on CJK normalization
- ❖ HKIUG TSVCC mapping table
- ❖ Implementing TSVCC on Primo

About CUHK Library

- ❖ established in 1963
- ❖ 7 branches
- ❖ 200 staff
- ❖ 260K current patrons
- ❖ 130K journal subscriptions, 4.5M ebooks, 2.5M printed volumes
- ❖ special collections includes from oracles bones, Chinese rare books, modern Chinese literary archives ...

character normalization

- ❖ different people type differently
- ❖ normal to expect “Apple” will have the exact results from “aPPLE”, “ApPle”, “appLE” ...
- ❖ before indexing, Primo will first “clean up” (normalize) the data to its lower case (e.g. A → a, B → b ...)
- ❖ Primo FE will do the same for the search term typed by users, to get a match with the index

Primo out-of-box normalizations

- ❖ Primo provides OTB normalizations for different languages at:
- ❖ `/exlibris/primop4_1/ng/jaguar/home/profile/analysis/specialCharacters/CharConversion/OTB/OTB`
- ❖ e.g.
- ❖ latin languages (`non_cjk_unicode_normalization.txt`)
- ❖ CJK (`cjk_unicode_trad_to_simp_normalization.txt`)

OTB CJK normalization table

- ❖ 2700+ entries
- ❖ mainly for mapping Traditional Chinese into its Simplified form
- ❖ assume it is a 1:1 mapping, Simplified Chinese being the “lowercase” like the English language
- ❖ But in fact, Simplified Chinese is only one kind of variant form for Chinese character
- ❖ other variant forms (ideograph) of the same character need to be cover as well

extract of the OTB table

```
otb_cjk_unicode.txt
1  !!! This is a system file - PLEASE DO NOT CHANGE!
2  !!! For user-defined conversions, please use user_defined_unicode_normalization.txt
3  !!! Format of this file: <FROM> <tabulation> <TO>
4  !!!TRADITIONAL TO SIMPLIFIED!!!
5  !!!
6  !!!
7  4E1F    4E22
8  4E26    5E76
9  4E7E    5E72
10 4E82    4E71
11 4E99    4E98
12 4E9E    4E9A
13 4F15    592B
14 4F47    4F2B
15 4F48    5E03
16 4F54    5360
17 4F75    5E76
18 4F86    6765
19 4F96    4ED1
20 4F9A    5F87
21 4FB6    4FA3
22 4FB7    5C40
23 4FC1    4FE3
24 4FC2    7CFB
25 4FD4    4F23
```

background on CJK

- ❖ Traditional Chinese characters have been used since as early as 2nd centuryBC (Han Dynasty, 漢朝)
 - ❖ used by people in Taiwan, Hong Kong and Macau
- ❖ Simplified Chinese characters were introduced by PRC government during 1950's
 - ❖ used by people in PRC, SE Asia countries e.g. Singapore and Malaysia
- ❖ “Traditional Chinese” is a retronym

background on CJK

- ❖ in the Chinese language:
 - ❖ “Traditional Chinese” is 正體字 OR 繁體字, meaning formal form, complicated format
 - ❖ “Simplified Chinese” is 簡體字, meaning simplified form

background on CJK

- ❖ Japanese adopted Chinese characters in ~5th century AD, called it Kanji (かんじ, 漢字)
- ❖ Korean adopted Chinese characters in ~4th century AD, called it Hanja (한자, 漢字)
- ❖ both Japanese and Korean have created variant forms for the Chinese characters (glyph), making new ideograms

CJK in Unicode character table

- ❖ Unicode table has different segments for CJK
- ❖ 2E80 - 2EFF (CJK radicals supplement)
- ❖ 3000 - 303F (CJK symbols and punctuation)
- ❖ 31C0 - 31EF (CJK strokes)
- ❖ 3200 - 32FF (enclosed CJK letters and months)
- ❖ 3400 - 4DBF (CJK unified ideographs extension A)
- ❖ 4E00 - 9FFF (CJK unified ideographs)
- ❖ F900 - FAFF (CJK compatibility ideographs)

multiple-mapped nature of CJK

- ❖ for the same character, it can be the simplified form of two Traditional characters, having 2 meanings
 - ❖ e.g. 干 (U+5E72) is simplified form of both 乾 (U+4E7E) and 幹 (U+5E79), meaning “dry” and “take an action”
- ❖ for the same character, it can be the simplified form of another Traditional character, but also being the traditional form for another meaning
 - ❖ e.g. 松 (U+677E) as Traditional form means “pine tree”, is also the simplified form for 鬆 (U+9B06), meaning “soft”

preference by different countries

- ❖ the Chinese people recognize 裴 (U+88F4) as the standard form, while Korean people use 裴 (U+88F5)
- ❖ Chinese people use 澀 (U+6F80) as Traditional form and 涩 (U+6DA9) as simplified form, while Japanese people refer 澁 (U+6E0B) for the exact same character

Chinese inputting methods

- ❖ Chinese characters can be inputted by various means
 - ❖ Pinyin, Cangjie, Zhuyin, Romanji (Japanese), HNC Romaja (Korean)
 - ❖ handwriting on smartphones and tablets
- ❖ no way to control the resulting Chinese character (ideograph) retrieved by these inputting methods
- ❖ no way to control how users key in search terms, as well as the metadata keyed in by catalogers

spot the difference

- ❖ 内 (U+5167) | 内 (U+5185), meaning “inside”
- ❖ 别 (U+5225) | 别 (U+522B), meaning “another”
- ❖ 夠 (U+5920) | 够 (U+591F), meaning “enough”
- ❖ 步 (U+6B65) | 步 (U+6B69), meaning “pace”
- ❖ 瘦 (U+75E9) | 瘦 (U+7626), meaning “thin”
- ❖ 黑 (U+9ED1) | 黑 (U+9ED2), meaning “black”

Ideograph and Glyph

- ❖ these sample characters are NOT the same glyph presented in different styles (think of fonts)
- ❖ they are different ideograms, having individual code-points in the Unicode table
- ❖ these ideographs are variant forms of the “same” Chinese character
- ❖ the OTB Primo CJK normalization table is not sufficient
- ❖ need to make logical groups of ideograph based on these complicated relationship

HKIUG TSVCC mapping table

- ❖ HKIUG is a local user group for Millennium ILS, including all government funded university libraries, and other institutions
- ❖ in 2004, they attempted to compile a list which could indicate all the possible relationships among these groups Chinese ideograph
- ❖ TSCVV stands for “Traditional, Simplified, Variant Chinese Characters”

HKIUG TSVCC mapping table

- ❖ the list was used as a workaround in the Millennium OPAC and its search index.
- ❖ it has been a hard-coded customization for Millennium users in Hong Kong, needs to be done by III
- ❖ Until now, Millennium and Sierra still do not have an official way to handle customer-defined normalization mapping, like Primo does

HKIUG TSVCC mapping table

- ❖ the list was completed in 2006
- ❖ the list has consolidated relationships found in:
 - ❖ EACC (LC's East Asian Character Code Tables)
 - ❖ EACC-to-Unicode mapping
 - ❖ kSimplifiedVariant and kZVariant (Unicode consortium's UniHan database)
- ❖ entries defined by HKIUG (via original cataloging of Chinese material)

implement TSVCC on Primo

- ❖ analyze and compare the OTB table
(`cjk_unicode_trad_to_simp_normalization.txt`) with the TSVCC table
- ❖ consolidate all mapping into:
 - ❖ `/exlibris/primo/p4_1/ng/jaguar/home/profile/analysis/specialCharacters/CharConversion/CUSTOMER/CUSTOMER/user_defined_unicode_normalization.txt`
- ❖ 3490+ ideograph groups,
- ❖ 4600+ entries, up from original 2700+ entries
- ❖ delete the index (`clean_indexes.sh`) and run a full re-index
(`Indexing_and_Didumean_and_Hotswapping`)

extract of the resulting table

| | | | |
|----|------|------|---------------------------|
| 78 | 4E79 | 5E72 | # 軋 --> 干 (1, tsvcc) |
| 79 | 4E7E | 5E72 | # 乾 --> 干 (1, otb, tsvcc) |
| 80 | 4E81 | 5E72 | # 乾 --> 干 (1, tsvcc) |
| 81 | 5E79 | 5E72 | # 幹 --> 干 (1, otb, tsvcc) |
| 82 | 69A6 | 5E72 | # 榦 --> 干 (1, otb, tsvcc) |
| 83 | 6E0B | 6DA9 | # 洩 --> 澀 (1, tsvcc) |
| 84 | 6F80 | 6DA9 | # 澀 --> 澀 (1, otb, tsvcc) |
| 85 | 6F81 | 6DA9 | # 澁 --> 澀 (1, tsvcc) |
| 86 | 6FC7 | 6DA9 | # 瀦 --> 澀 (1, otb, tsvcc) |
| 87 | 88F4 | 88F5 | # 裴 --> 裴 (pure tsvcc) |

three sample ideograph groups

- ❖ **RED** is the group representative in OTB
- ❖ **BLACK** is from OTB “cjk_unicode_trad_to_simp_normalization.txt”
- ❖ **BLUE** is from TSVCC table (new entries)
 - ❖ 涩 (U+6DA9) | 洩 (U+6E0B) | 澀 (U+6F80) | 澁 (U+6F81)
| 濇 (U+6FC7) | 漚 (U+7012)
 - ❖ 裴 (U+88F4) | 裴 (U+88F5)
 - ❖ 干 (U+5E72) | 軋 (U+4E79) | 乾 (U+4E7E) | 乾 (U+4E81)
| 幹 (U+5E79) | 榦 (U+69A6)

case 1: BEFORE adding TSVCC

❖ search term: 涩谷 (12 results, OTB normalized)

The screenshot shows a web browser window titled "Easy Search - 涩谷". The address bar contains the URL: easysearch.lib.cuhk.edu.hk/primo_library/libweb/action/search.do?fn=search&ct=search&initialSearch=true&mode. The page features a navigation menu with "Books", "Articles", "Books & Articles", and "Extended Search". The search term "涩谷" is entered in the search box, and the "Go" button is highlighted. The search results are displayed in a list format, sorted by "Relevance".

Results 1 - 10 of 12 for Entire Collection Sorted by: Relevance 1 2 +

- 涩谷克彦 = Shibuya Kazihiko. [electronic resource]**
涩谷克彦. 朱钧,
第1版.. 南宁 : 广西美术出版社 2003
● Online access
E-Book
View Online Details Reviews & Tags Findit@CUHK
- 回归涩谷**
Hui gui Segu / Shanben Xingjiu ; Xiang Xuan yi
山本幸久, 1966- ; Yukihiya Yamamoto 1966- 向轩; Xuan Xiang
第1版.; Di 1 ban.. 南京 : 南京大学出版社; Nan jing : Nan jing da xue chu ban she 2010
● Holding: CC PL877.5.A63 S4512 2010
Book
Locations Details Reviews & Tags Findit@CUHK Virtual Browse
- 解开心里秘密 [electronic resource] / 涩谷昌三著**
涩谷昌三.
第1版.. China : 正向出版社 2000
● Online access
E-Book
View Online Details Reviews & Tags Findit@CUHK

The left sidebar contains filters for "Show only" (Full Text Online (3)), "Refine My Results" (Resource Type: Book (9), E-Book (3)), "Subject" (United States (1), Psychology, Industrial (1), Death (1), Probabilities (1), Psychology (1)), and "Author" (Shibuya, S (3), 涩谷昌三 (2), Nakahama, T (1), Suzuki, K (1)).

case 1: BEFORE adding TSVCC

❖ search term: 濳谷 (12 results, OTB normalized)

The screenshot shows a web browser window titled "Easy Search - 濳谷". The address bar contains the URL: easysearch.lib.cuhk.edu.hk/primo_library/libweb/action/search.do?fn=search&ct=search&initialSearch=true&mode. The browser's toolbar includes icons for Apple, Yahoo!, Google Maps, YouTube, Wikipedia, News, and Popular. The page header features the "EASYSEARCH" logo with the tagline "e-resources discovery at your fingertip", the CUHK crest, and a search icon. Navigation links include "Guest", "e-Shelf", "My Account", and "Sign in". Below the header, there are tabs for "Books", "Articles", "Books & Articles", and "Extended Search". The search input field contains "濳谷" and the dropdown menu is set to "Entire Collection". A "Go" button and "Advanced Search" / "Browse Search" links are visible. The search results are displayed in a list format, sorted by "Relevance". The first three results are:

- 濳谷克彦 = Shibuya Kazihiko. [electronic resource]**
濳谷克彦. 朱锔,
第1版.. 南宁 : 广西美术出版社 2003
● Online access
E-Book
View Online Details Reviews & Tags Findit@CUHK
- 回归濳谷**
Hui gui Segu / Shanben Xingjiu ; Xiang Xuan yi
山本幸久, 1966- ; Yukihiisa Yamamoto 1966- 向轩; Xuan Xiang
第1版.; Di 1 ban.. 南京 : 南京大学出版社; Nan jing : Nan jing da xue chu ban she 2010
● Holding: CC PL877.5.A63 S4512 2010
Book
Locations Details Reviews & Tags Findit@CUHK Virtual Browse
- 解开心里秘密 [electronic resource] / 濳谷昌三著**
濳谷昌三.
第1版.. China : 正向出版社 2000
● Online access
E-Book
View Online Details Reviews & Tags Findit@CUHK

The left sidebar contains a "Refine My Results" section with filters for "Resource Type" (Book (9), E-Book (3)), "Subject" (United States (1), Psychology, Industrial (1), Death (1), Probabilities (1), Psychology (1)), and "Author" (Shibuya, S (3), 濳谷昌三 (2), Nakahama, T (1), Suzuki, K (1)).

case 1: BEFORE adding TSVCC

❖ search term: 渋谷 (33 results, NOT normalized)

The screenshot shows a web browser window with the URL `easysearch.lib.cuhk.edu.hk/primo_library/libweb/action/search.do?sessionid=E0089F5EE7436C169DC6115684A2E`. The page title is "Easy Search - 渋谷". The search bar contains the term "渋谷" and the dropdown menu is set to "Entire Collection". The search results are displayed in a list format, sorted by Relevance. The first three results are:

- 伊藤若冲 : 渋谷区立松濤美術館特別陳列.**
Itô Jakuchû : Shibuya Kuritsu Shôtô Bijutsukan tokubetsu chinretsu
伊藤若冲, 1716-1800. ; Jakuchû Itô 1716-1800. 渋谷区立松濤美術館; Shibuya Kuritsu Shôtô Bijutsukan.
東京 : 渋谷区立松濤美術館; Tôkyô : Shibuya Kuritsu Shôtô Bijutsukan 1981
● Holding: NA ND1059.I777 A4 1981
- 「特別展」三雲祥之助 : 1989年8月8日-9月24日渋谷区立松濤美術館**
"Tokubetsuten" Mikumo Shônosuke : 1989-nen 8-gatsu 8-nichi-9-gatsu 24-nichi Shibuya Kuritsu Shôtô Bijutsukan / [henshû Shibuya Kuritsu Shôtô Bijutsukan]
三雲祥之助, 1902-1982. ; Shônosuke Mikumo 1902-1982. 渋谷区立松濤美術館; Shibuya Kuritsu Shôtô Bijutsukan.
東京 : 渋谷区立松濤美術館; Tôkyô : Shibuya Kuritsu Shôtô Bijutsukan 1989
● Holding: NA ND1059.M4764 A4 1989
- タイ・ベトナムの古陶磁 : 特別展 : 昭和63年8月9日-9月25日・渋谷区立松濤美術館.**
Tai Betonamu no kotôji : tokubetsuten : Shôwa 63-nen 8-gatsu 9-nichi-9-gatsu 25-nichi, Shibuya Kuritsu Shôtô Bijutsukan
渋谷区立松濤美術館; Shibuya Kuritsu Shôtô Bijutsukan.
東京 : 渋谷区立松濤美術館; Tôkyô : Shibuya Kuritsu Shôtô Bijutsukan Shôwa 63 [1988
● Holding: NA NK4156.6.A1 T35 1988

The left sidebar contains navigation and filtering options:

- rss
- Save page to e-Shelf
- Refine My Results**
- Resource Type
 - Book (30)
 - DVD (2)
 - Journal (1)
 - More options
- Subject
 - Painting, Chinese (6)
 - Hashimoto, Sueyoshi, 1902- (5)
 - Painting (4)
 - Motion pictures, Japanese (2)
 - Maeda, Kanji, 1896-1930 (1)
 - More options
- Author
 - Shibuya Kuritsu Shôtô Bijutsukan (20)
 - 渋谷勝己 (2)
 - Itô (2)
 - Shibuya, S (2)
 - Shibuya, M (2)
 - More options
- Collection
 - Audio-visual Materials (2)
 - Periodicals (1)

case 1: BEFORE adding TSVCC

- ❖ search term: 澁谷 (8 results, NOT normalized)

The screenshot shows a web browser window with the URL `easysearch.lib.cuhk.edu.hk/primo_library/libweb/action/search.do?fn=search&ct=search&initialSearch=true&mode`. The page title is "Easy Search - 澁谷". The search bar contains the term "澁谷" and the search button is labeled "Go". The results are sorted by "Relevance" and show 8 results for the "Entire Collection".

Refine My Results

- Resource Type
 - Book (7)
 - Videocassette (1)
 - More options
- Subject
 - Psychology (2)
 - Skyscrapers (1)
 - Office buildings (1)
 - Nakano, Keishō, 1899-1974 (1)
 - Seikei Gakuen (1)
 - More options
- Author
 - Shibuya, S (2)
 - Shibuya, M (2)
 - Lin, Z (1)
 - Taniguchi, H (1)
 - Nakano, K (1)
 - More options
- Library
 - University Library (3)
 - Architecture Library (2)
 - New Asia College Library (1)

8 Results for Entire Collection Sorted by: Relevance

- Book**
 - ☆ 特別展中野恵祥：板金の造形 = Keishō Nakano
Tokubetsuten Nakano Keishō : itagane no zōkei = Keishō Nakano / [Shibuya Kuritsu Shōtō Bijutsukan]
中野恵祥, 1899-1974. ; Keishō Nakano 1899-1974. 澁谷区立松濤美術館; Shibuya Kuritsu Shōtō Bijutsukan.
東京 : 澁谷区立松濤美術館; Tōkyō : Shibuya Kuritsu Shōtō Bijutsukan 1992?
● Holding: NA NK6498.N35 A4 1992
- Videocassette**
 - ☆ さるとかに ; はくちょうの王子
Saru to kani ; Hakucho no oji [videorecording] / kikaku seisaku Shin'ei Kabushiki Gaisha
新映株式会社; 田口眞; 池多孝春; 納谷六朗; Makoto Taguchi; Takaharu Iketa; Rokuro Naya;
Shin'ei Kabushiki Kaisha.
東京 : 新映株式会社; Tokyo : Shin'ei Kabushiki Gaisha 1986?
● Holding: UC Video-VHS PN1979.K3 B523 1986 v.3
- Book**
 - ☆ 恩師の面影.
Onshi no omokage
澁谷光長. ; Mitsunaga Shibuya
東京 : 成蹊学園; Tōkyō : Seikei Gakuen Shōwa 11 [1936
● Holding: CC LG275.T648 S45

case 1 : AFTER adding TSVCC

❖ search term: 澁 (71 results, normalized)

The screenshot shows a web browser window titled "Easy Search - 澁". The address bar contains the URL: pmstg.lib.cuhk.edu.hk/primo_library/libweb/action/search.do?fn=search&ct=search&initialSearch=true&mode=Bas. The page features a purple header with the "EASY SEARCH" logo and navigation links like "Find Databases", "Library Location Codes", "Help", and "Language: English". Below the header, there are tabs for "Books", "Articles", "Books & Articles", "Search", "Extended Search", "E-Theses", and "test-tab". The search term "澁" is entered in the search box, and the "Go" button is highlighted. The results section shows "Results 1 - 10 of 71 for Entire Collection" sorted by "Relevance". Three results are visible, each with an "E-Book" icon, a star icon, and a "Like" button with a count of 0. The first result is "涩泽荣一传 [electronic resource] / (日) 幸田露伴著; 余炳跃译." with details: "幸田露伴, 1867-1947. 余炳跃, 第1版.. 上海: 学林出版社 1992". The second result is "近代中日两国企业家比较研究: 张謇与涩泽荣 [electronic resource] / 周见著." with details: "周见, 第1版.. 北京: 中国社会科学出版社 2004". The third result is "涩涩伊哥儿之食人蝙蝠 [electronic resource] / 涩涩伊哥儿著." with details: "涩涩伊哥儿, 第1版.. 沈阳: 辽宁教育出版社 2005". On the left side, there are sections for "Show only" (Full Text Online (30)), "Refine My Results" (Subject: Chinese literature (4), Tokugawa, Yoshinobu, 1837-1913 (3), Confucian ethics (2), Shibusawa, Eiichi, 1840-1931 (2), 方歌 (2)), and "Author" (Shibusawa, E (5), 高桥由美 (3), 涩川育由 (3), 澁谷勝己 (2), 涩泽荣一 (2)).

case 1 : AFTER adding TSVCC

❖ search term: 涩 (71 results, normalized)

The screenshot shows a web browser window titled "Easy Search - 涩". The address bar contains the URL: pmstg.lib.cuhk.edu.hk/primo_library/libweb/action/search.do?fn=search&ct=search&initialSearch=true&mode=Bas. The page features a purple header with the "EASY SEARCH" logo and navigation links like "Find Databases", "Library Location Codes", "Help", and "Language: English". Below the header is a search bar with the term "涩" entered, and a "Go" button. The search results are displayed in a list format, showing the first 10 of 71 results. The results are sorted by Relevance. The first three results are:

- 1. 涩泽荣一传 [electronic resource] / (日) 幸田露伴著; 余炳跃译. 幸田露伴, 1867-1947. 余炳跃, 第1版.. 上海: 学林出版社 1992. E-Book. Online access.
- 2. 近代中日两国企业家比较研究: 张謇与涩泽荣 [electronic resource] / 周见著. 周见. 第1版.. 北京: 中国社会科学出版社 2004. E-Book. Online access.
- 3. 涩涩伊哥儿之食人蝙蝠 [electronic resource] / 涩涩伊哥儿著. 涩涩伊哥儿. 第1版.. 沈阳: 辽宁教育出版社 2005. E-Book. Online access.

Each result includes a "View Online" link, "Details", "Reviews & Tags", and "Findit@CUHK" options. There are also "Like" buttons for each result.

On the left side of the page, there is a sidebar with the following sections:

- rss**: Add page to e-Shelf
- Show only**: Full Text Online (30)
- Refine My Results**:
 - Subject**: Chinese literature (4), Tokugawa, Yoshinobu, 1837-1913 (3), Confucian ethics (2), Shibusawa, Eiichi, 1840-1931 (2), 方歌 (2). More options v
 - Author**: Shibusawa, E (5), 高桥由美 (3), 涩川育由 (3), 波谷勝己 (2), 涩泽荣一 (2).

case 1 : AFTER adding TSVCC

❖ search term: 濳 (71 results, normalized)

The screenshot shows a web browser window titled "Easy Search - 濳". The address bar contains the URL: pmstg.lib.cuhk.edu.hk/primo_library/libweb/action/search.do?fn=search&ct=search&initialSearch=true&mode=Bas. The browser's toolbar includes navigation buttons and a search bar with the term "濳".

The website header features the "EASY SEARCH" logo with the tagline "e-resources discovery at your fingertip". Navigation links include "Find Databases", "Library Location Codes", "Help", and "Language: English".

The search interface shows the search term "濳" in the search box, with a "Go" button and a dropdown menu set to "Entire Collection". There are also links for "Advanced Search" and "Browse Search".

The search results are displayed as follows:

- Results 1 - 10 of 71 for Entire Collection** (Sorted by: Relevance)
- Result 1:** 濳泽荣一传 [electronic resource] / (日) 幸田露伴著; 余炳跃译. 幸田露伴, 1867-1947. 余炳跃, 第1版.. 上海: 学林出版社 1992. **Online access**. E-Book. Includes "View Online", "Details", "Reviews & Tags", and "Findit@CUHK" links.
- Result 2:** 近代中日两国企业家比较研究: 张謇与濳泽荣 [electronic resource] / 周见著. 周见, 第1版.. 北京: 中国社会科学出版社 2004. **Online access**. E-Book. Includes "View Online", "Details", "Reviews & Tags", and "Findit@CUHK" links.
- Result 3:** 濳濳伊哥儿之食人蝙蝠 [electronic resource] / 濳濳伊哥儿著. 濳濳伊哥儿, 第1版.. 沈阳: 辽宁教育出版社 2005. **Online access**. E-Book. Includes "View Online", "Details", "Reviews & Tags", and "Findit@CUHK" links.

On the left side, there are navigation options: "rss", "Add page to e-Shelf", "Show only Full Text Online (30)", and "Refine My Results" with filters for Subject (Chinese literature, Tokugawa, Yoshinobu, Confucian ethics, Shibusawa, Eilichi, 方歌) and Author (Shibusawa, E, 高桥由美, 濳川育由, 波谷勝己, 濳泽荣一).

case 1 : AFTER adding TSVCC

❖ search term: 澁 (71 results, normalized)

The screenshot shows a web browser window titled "Easy Search - 澁". The address bar contains the URL: pmstg.lib.cuhk.edu.hk/primo_library/libweb/action/search.do?fn=search&ct=search&initialSearch=true&mode=Bas. The page features a purple banner with the "EASY SEARCH" logo and the tagline "e-resources discovery at your fingertip". Below the banner, there are navigation tabs for "Books", "Articles", "Books & Articles", "Search", "Extended Search", "E-Theses", and "test-tab". The search term "澁" is entered in the search box, and the "Go" button is highlighted. The search results are displayed in a list format, sorted by "Relevance". The first three results are:

- 澁泽荣一传 [electronic resource] / (日) 幸田露伴著 ; 余炳跃译.**
幸田露伴, 1867-1947. 余炳跃,
第1版.. 上海 : 学林出版社 1992
● Online access
- 近代中日两国企业家比较研究 : 张謇与涩泽荣 [electronic resource] / 周见著.**
周见.
第1版.. 北京 : 中国社会科学出版社 2004
● Online access
- 涩涩伊哥儿之食人蝙蝠 [electronic resource] / 涩涩伊哥儿著.**
涩涩伊哥儿.
第1版.. 沈阳 : 辽宁教育出版社 2005
● Online access

Each result includes a book icon, a "Like" button with a count of 0, and links for "View Online", "Details", "Reviews & Tags", and "Findit@CUHK". On the left side of the page, there is a sidebar with options for "rss", "Add page to e-Shelf", "Show only" (Full Text Online (30)), and "Refine My Results" (Subject: Chinese literature (4), Tokugawa, Yoshinobu, 1837-1913 (3), Confucian ethics (2), Shibusawa, Eiichi, 1840-1931 (2), 方歌 (2); Author: Shibusawa, E (5), 高桥由美 (3), 涩川育由 (3), 浪谷勝己 (2), 涩泽荣一 (2)).

case 2 : BEFORE adding TSVCC

❖ search term: 裴 (1757 results, OTB normalized)

The screenshot shows a web browser window titled "Easy Search - 裴". The address bar displays the URL: easysearch.lib.cuhk.edu.hk/primo_library/libweb/action/search.do?fn=search&ct=search&initialSearch=true&mode. The page features a purple banner with the "EASY SEARCH" logo and the tagline "e-resources discovery at your fingertip". Navigation links include "Books", "Articles", "Books & Articles", and "Extended Search". The search term "裴" is entered in the search box, and the results are sorted by "Relevance".

Results 1 - 10 of 1,757 for Entire Collection. Sorted by: Relevance. 1 2 3 4 5 →

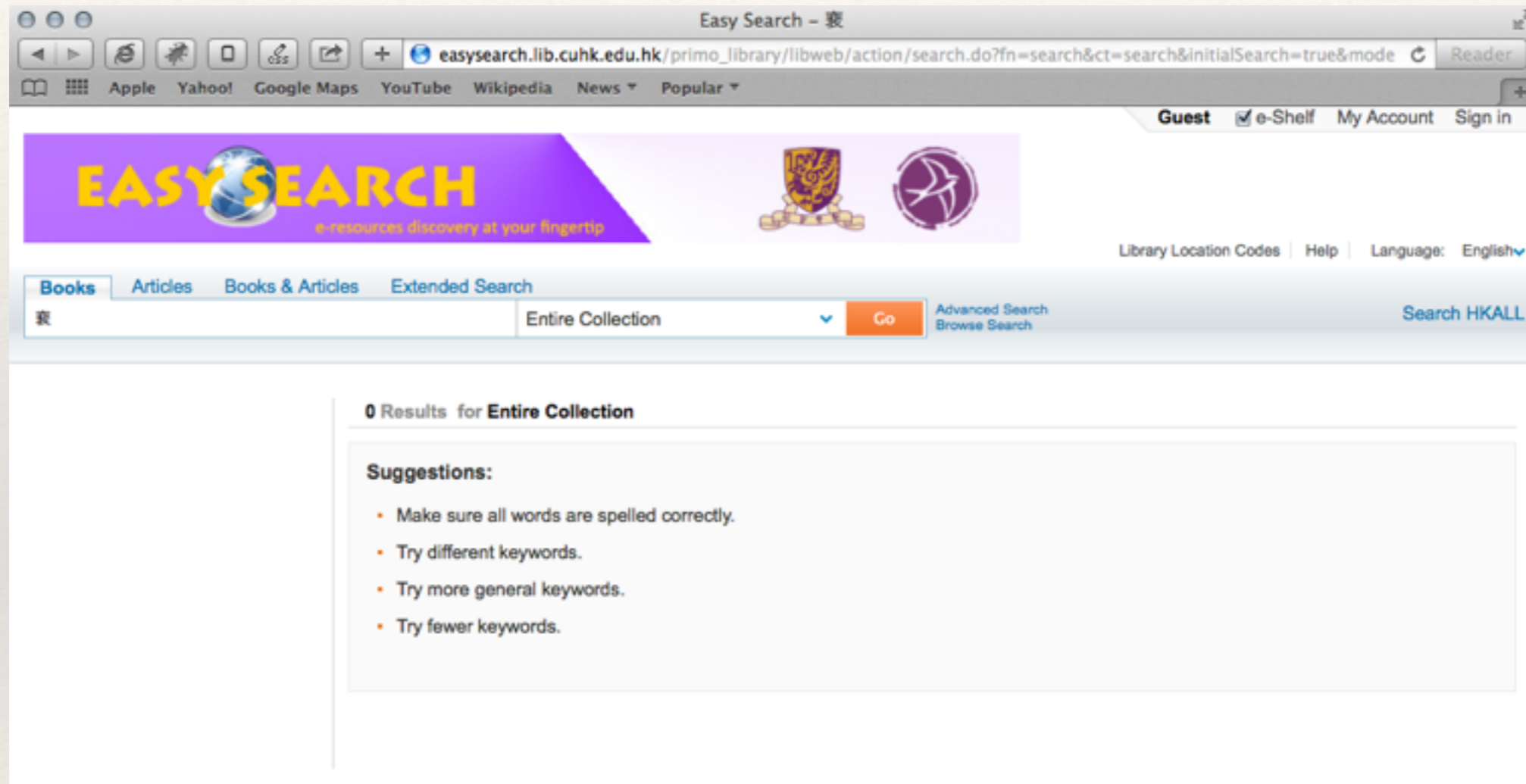
- 裴文中科学论文集 [electronic resource] / 裴文中著**
裴文中, 1904-1982.
第1版.. 北京 : 科学出版社 1990
● Online access
E-Book
View Online Details Reviews & Tags Findit@CUHK
- 将相文武 风流千古 : 裴氏人物志传 [electronic resource] / 西江编撰校注**
西江.
第1版.. 北京 : 中国社会科学出版社 2001
● Online access
E-Book
View Online Details Reviews & Tags Findit@CUHK
- 裴光祿年譜**
Pei guang lu nian pu / Pei Shiqi deng ji ; Xu Jia bian
裴士麒.; Shiqi Pei 徐嘉; Jia Xu
第1版.; Di 1 ban.. 北京 : 北京圖書館出版社; Beijing : Beijing tu shu guan chu ban she 1999
● Holding: UL DS734 .B385 v.165
Book
Locations Details Reviews & Tags Findit@CUHK Virtual Browse

Left sidebar options:

- rss
- Save page to e-Shelf
- Show only: Full Text Online (979)
- Refine My Results:
 - Resource Type: E-Book (937), Book (740), DVD (19), Videocassette (16), microform (9)
 - Subject: China (100), Chinese literature (32), Chinese language (23), Chinese poetry (22), Chinese fiction (22)
 - Author: Apabi (Corporation) (63)

case 2 : BEFORE adding TSVCC

- ❖ search term: 褻 (0 result, NOT normalized)



The screenshot shows a web browser window titled "Easy Search - 褻". The address bar contains the URL: easysearch.lib.cuhk.edu.hk/primo_library/libweb/action/search.do?fn=search&ct=search&initialSearch=true&mode. The browser's toolbar includes navigation buttons and a search bar with the text "褻".

The website header features the "EASY SEARCH" logo with the tagline "e-resources discovery at your fingertip". To the right of the logo are the logos of the Chinese University of Hong Kong and the library. The header also includes user options: "Guest", "e-Shelf", "My Account", and "Sign in".

The search interface includes a navigation menu with "Books", "Articles", "Books & Articles", and "Extended Search". The search bar contains the text "褻" and a dropdown menu set to "Entire Collection". A "Go" button is present next to the search bar. To the right of the search bar are links for "Advanced Search" and "Browse Search", and a "Search HKALL" button.

The search results section displays "0 Results for Entire Collection". Below this, a "Suggestions:" box contains the following advice:

- Make sure all words are spelled correctly.
- Try different keywords.
- Try more general keywords.
- Try fewer keywords.

case 2 : AFTER adding TSVCC

❖ search term: 裴 (9 results, normalized)

The screenshot shows a web browser window titled "Easy Search - 裴". The address bar contains the URL: pmstg.lib.cuhk.edu.hk/primo_library/libweb/action/search.do?fn=search&ct=search&initialSearch=true&mode=Bas. The page features a navigation menu with options like "Books", "Articles", "Books & Articles", "Search", "Extended Search", "E-Theses", and "test-tab". The search term "裴" is entered in the search box, and the "Go" button is highlighted. The search results are displayed under the heading "9 Results for Entire Collection" and are sorted by "Relevance".

The results list includes:

- 宣室志; 裴传奇**
张读, 852年进士. 萧逸. 田松青. 裴, 9th cent.
第1版. 上海: 上海古籍出版社, 2012.
● Online access
- 王安忆小说主题研究**
裴艳艳
中国戏剧出版社, 2012
● Online access
- 椭圆与超椭圆曲线公钥密码的理论实现**
王学理 裴定一
第1版. 科学出版社, 2006
● Online access

Each result entry includes a book icon, a star icon, and a "Like" button with a count of 0. Below each entry are links for "Locations", "Details", "Reviews & Tags", and "Findit@CUHK".

On the left side of the page, there is a "Refine My Results" section with filters for "Subject" (Philosophy, Confucian (1), Corporation law (1), Neo-Confucianism (1), China. Gong si fa (2005) (1), China (1)) and "Author" (罗冬娥 (1), 裴 (1), Pae, C (1), Sima, Z (1), 法律出版社. 专业出版编委会 (1)).

case 2 : AFTER adding TSVCC

❖ search term: 裴 (9 results, normalized)

The screenshot shows a web browser window titled "Easy Search - 裴". The address bar contains the URL: pmstg.lib.cuhk.edu.hk/primo_library/libweb/action/search.do?fn=search&ct=search&initialSearch=true&mode=Bas. The page features a navigation menu with options like "Books", "Articles", "Books & Articles", "Search", "Extended Search", "E-Theses", and "test-tab". The search bar contains the term "裴" and the "Entire Collection" dropdown is selected. The search results are displayed in a list format, sorted by "Relevance".

9 Results for Entire Collection Sorted by: Relevance

- 宣室志; 裴传奇**
张读, 852年进士. 萧逸. 田松青. 裴, 9th cent.
第1版. 上海: 上海古籍出版社, 2012.
● Online access
- 王安忆小说主题研究**
裴艳艳
中国戏剧出版社, 2012
● Online access
- 椭圆与超椭圆曲线公钥密码的理论与实现**
王学理 裴定一
第1版. 科学出版社, 2006
● Online access

On the left side, there is a "Refine My Results" section with filters for Subject, Author, and Library. The "Subject" filter includes "Philosophy, Confucian (1)", "Corporation law (1)", "Neo-Confucianism (1)", and "China. Gong si fa (2005) (1)". The "Author" filter includes "罗冬娥 (1)", "裴 (1)", "Pae, C (1)", and "Sima, Z (1)". The "Library" filter includes "New Asia College Library (1)", "University Library (1)", and "Lee Onn Wei Law Library (1)".

points of interest

- ❖ make sure not group the characters in loop
 - ❖ CORRECT:
 - ❖ $B \rightarrow A$
 - ❖ $C \rightarrow A$
 - ❖ $D \rightarrow A$
 - ❖ $E \rightarrow A$
 - ❖ NOT SO SURE:
 - ❖ $B \rightarrow A$
 - ❖ $C \rightarrow A$
 - ❖ $A \rightarrow D$
 - ❖ $D \rightarrow E$
- ❖ make sure all OTB entries appear in the user-defined table, not sure how Primo take precedence among the files

Reference (1)

- ❖ The Unicode consortium

- ❖ <http://www.unicode.org/reports/tr38/index.html>

- ❖ <http://www.unicode.org/charts/unihangridindex.html>

- ❖ <http://www.unicode.org/charts/unihan.htm>

- ❖ Unicode character table

- ❖ <http://unicode-table.com/en>

- ❖ Wikipedia

- ❖ http://en.wikipedia.org/wiki/Traditional_Chinese_characters

- ❖ <http://en.wikipedia.org/wiki/Ideogram>

- ❖ <http://en.wikipedia.org/wiki/Glyph>

Reference (2)

- ❖ Hong Kong Innovative User Group TSVCC table
 - ❖ http://hkiug.ln.edu.hk/unicode/hkiug_tsvcc_table-UnicodeVersion-1.0.html
- ❖ Library of Congress EACC
 - ❖ <http://www.loc.gov/marc/specifications/specchareacc.html>

Q & A

Thank you!

Alan Ng (alanng@cuhk.edu.hk)