

Transcript: Technical aspects and usage in Korean language

Park Shinjo
from KAIST, Korea
2008. 8. 10



I am

- "Park" is family name; "Shinjo" is given name
 - It is in Korean order
- Freshmen in KAIST
- Korean KDE translator
- One of admins of KAIST FTP
 - Mirroring several F/OSS projects



Contents

- Technical aspect of Transcript
- Current usage of Transcript
- Application of Transcript in Korean language
- Examples
- Summary

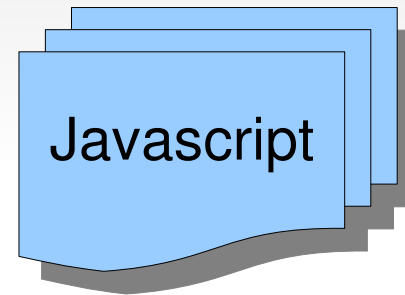
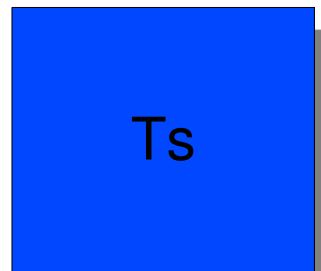
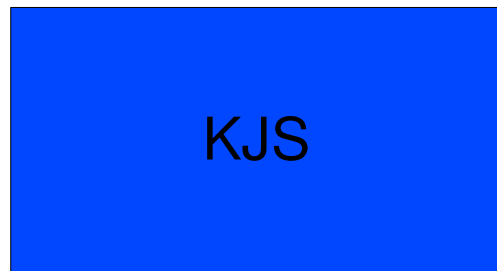
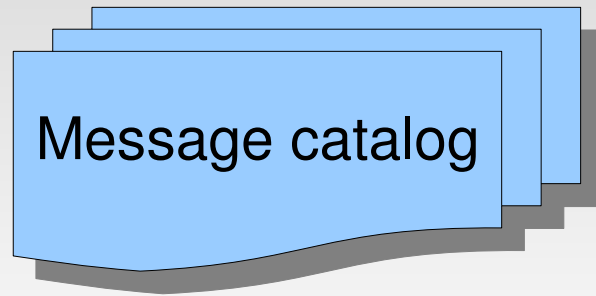


Transcript?

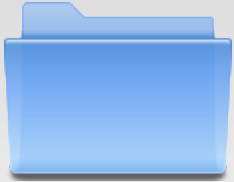
- New feature in KDE 4
- Solve language-specific situation
- Javascript-aided programmable translation
- Separated role; translator write script



Base structure



Repository locations



`l10n-kde4/xx`



`scripts/ (module) / (pofile) /script.js`



`messages/ (module) /message.po`



`/usr/share/locale/xx`



`LC_SCRIPTS/ (pofile) /script.js`



`LC_MESSAGES/message.mo`



Functions

- Function binding
 - `setcall, hascall, load`
- Message
 - `msgid, msgctxt, msgstrf, fallback`
- Property map
 - `getProp, setProp`
- Configuration
 - `getConfString, getConfBool, getConfNumber`



Property map

- Used in some languages
- Key-value pair of available properties
- Stored in `.pmap` file

Property map file

```
# example.pmap
=:
change:перемена
:nom=перемена
:gen=перемены
:dat=перемене
:acc=перемену
::
```

Javascript file

```
Ts.loadProps("example");

Ts.getProp("change", "gen");
// Returns "перемены"

Ts.getProp("перемена",
"gen");
// Also returns "перемены"
```



Configuration

- User configuration by `~/ .transcriptrc`

Configuration file

```
# Settings for Korean.  
[ko]  
serve-kimchi = True  
shout-count = 3
```

Javascript file

```
if (Ts.getConfBool("serve-kimchi")) {  
    // Returns 'true'  
    serveKimchi();  
}  
  
var j = Ts.getConfInt("shout-count");  
for (i=0;i<j;i++) {  
    // Returns '3'  
    shout();  
}
```



A simple example

- Neverland people shout one's name for greeting.

Without transcript

```
# neverness_login.cpp:10
msgid "Hello, %1!"
msgstr "Heelyy, %1!"
```

With transcript

```
# neverness_login.cpp:10
msgid "Hello, %1!"
msgstr "Heelyy, %1!"
"|||"
"Heelyy, $[shout %1]!"
```

Javascript file

```
function capitalize (str) {
    return str.toUpperCase();
}

Ts.setcall("shout", capitalize);
```

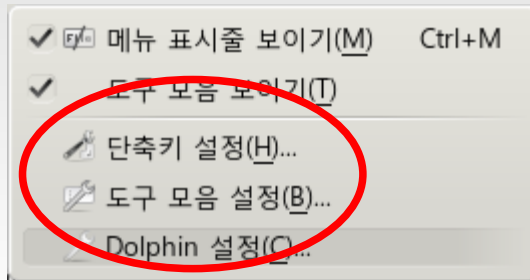
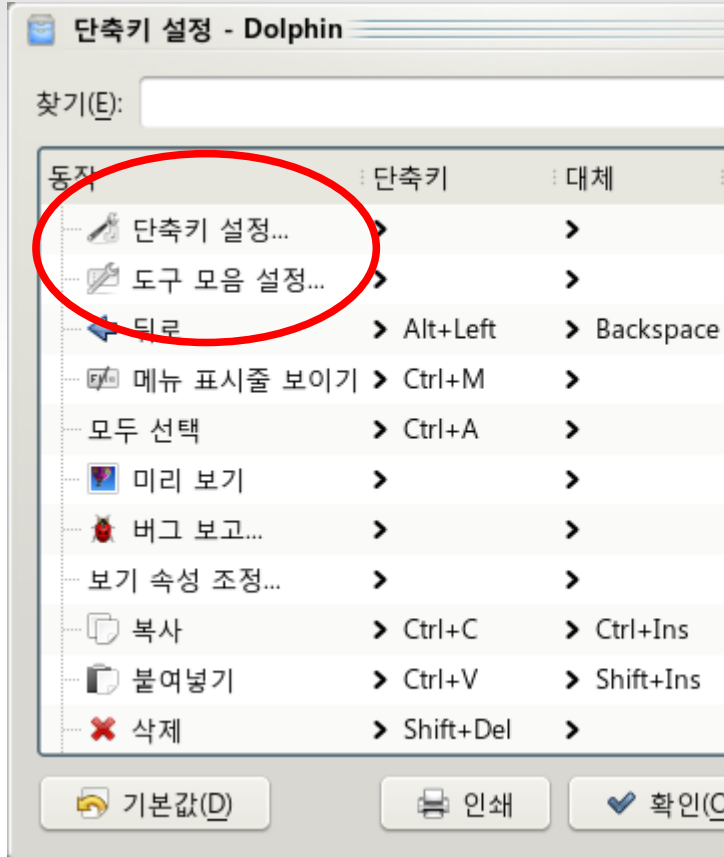


Currently used in

- CJK accelerator keys
- Polish, Serbian cases
- Korean postpositions



CJK accelerator keys



```
stripAccRx = /\((&|&amp;);)?[a-zA-Z0-9]\)/;  
function stripAccel(phrase)  
{  
    return phrase.replace(stripAccRx, "");  
}  
Ts.setcall("~stripAccel", stripAccel);
```



What to do

Without script | / | With script



Write translation script



Test it



조사 (Postposition)

- Dependant to other words
- Roles
 - Parts of speech
 - Special meaning of word
- Examples
 - 나는 어제 시계를 샀다 . (I bought a watch yesterday.)

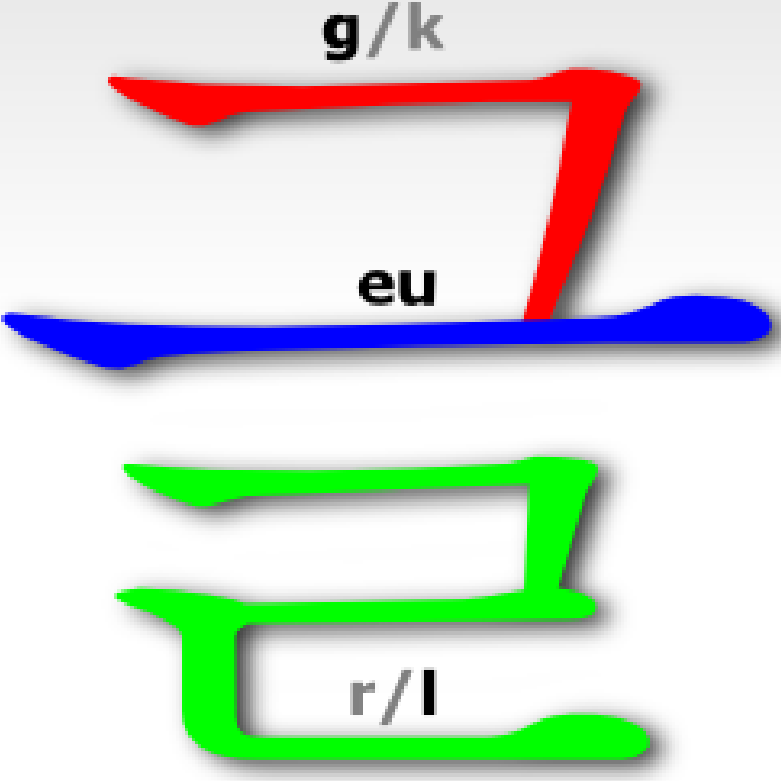


What's changing

- 은 (는) : subject
- 을 (를) : object
- 이 (가) : complement
- 와 (과) : conjunctive
- (으) 로 , (이) 에 게 , etc.



Structure of Hangeul



How postposition changes?

- Depends on previous character
- Existence and kind of final consonant
 - Most cases: existing, non-existing
 - (으), (이): existing, non-existing and final is ㄹ
- Example
 - 철수 and 을 (를): 수 hasn't final consonant, 를
 - 미영 and 을 (를): 영 has final consonant, 을



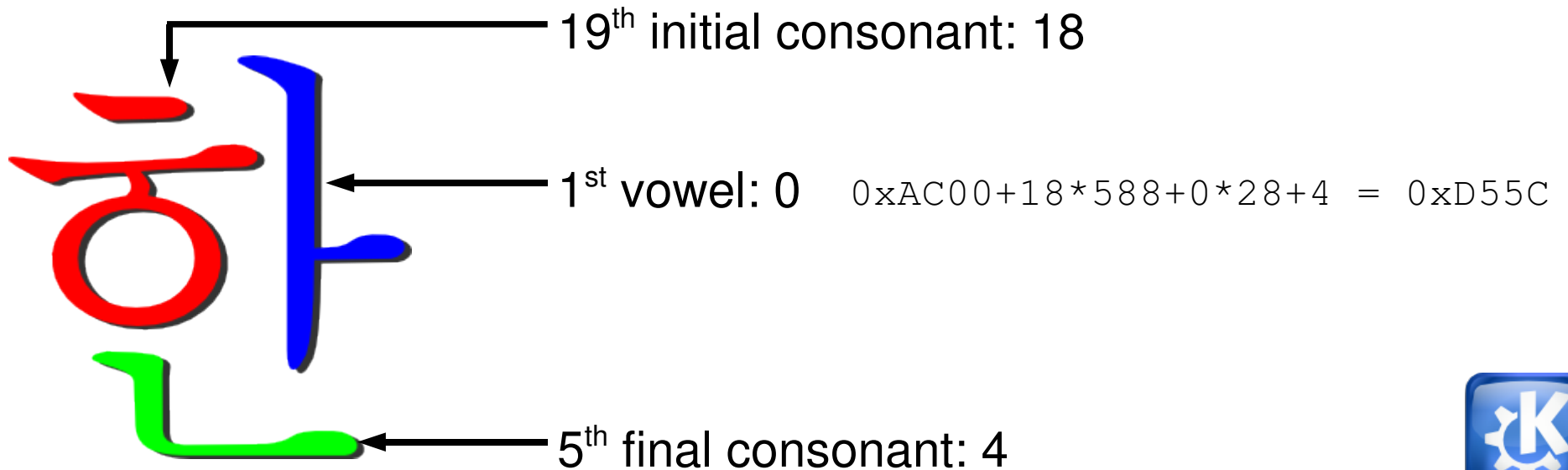
Hangeul in Unicode

- U+AC00~U+D7FF, U+1100~U+11FF
- AC00: Hangeul in composed form
- 1100: Each jamo is decomposed; making a whole character by combining
- Prior is most frequently used



U+AC00 block

- 19 initial consonant
- 21 vowel
- 28 final consonant
- $(\text{initial}) \times 588 + (\text{medial}) \times 28 + (\text{final}) + \text{AC00h}$



Getting finals

```
// *** Detect the provided word is Hangeul.
function isHangeul(code)
{
    if (code >= 0xac00 && code <= 0xd7af) return true;
    else return false;
}
// *** Detect the phrase has 종성 (Final jamo).
function getFinalJamoType(phrase, rieul)
{
    var code = phrase.charCodeAt(phrase.length - 1);
    if (!isHangeul(code)) return 2;
    code -= 0xac00;
    code %= 28;
    if (rieul && code == 8) return 0;
    switch (code) {
    case 0:
        return 0;
    default:
        return 1;
    }
}
```



Displaying postpositions

```
lstJosa = [ ["가", "이", "이 (가)", false],  
            ["를", "을", "을 (를)", false],  
            ["는", "은", "은 (는)", false],  
            ["와", "과", "과 (와)", false],  
            ["", "이", "(이)", false],  
            ["", "으", "(으)", true] ];
```

```
function attachJosa(type, phrase, marker)  
{  
    var i = getFinalJamoType(phrase, lstJosa[type][3]);  
    if (marker) phrase = marker + phrase + marker;  
    return phrase + lstJosa[type][i];  
}
```

```
function attachIGa(phrase, marker)  
{  
    return attachJosa(0, phrase, marker);  
}
```

```
Ts.setCall("이가", attachIGa);
```



Application to po files

- Aimed to drop-in replacement of placeholder

Without transcript:

```
#: src/toplevel.cpp:259
#, kde-format
msgid "%1 is now ready!"
msgstr "%1 이 ( 가 ) 준비되었습니다 !"
```

With transcript:

```
#: src/toplevel.cpp:259
#, kde-format
msgid "%1 is now ready!"
msgstr "%1 이 ( 가 ) 준비되었습니다 !"
" | / | "
"$[ 이가 %1 ] 준비되었습니다 !"
```

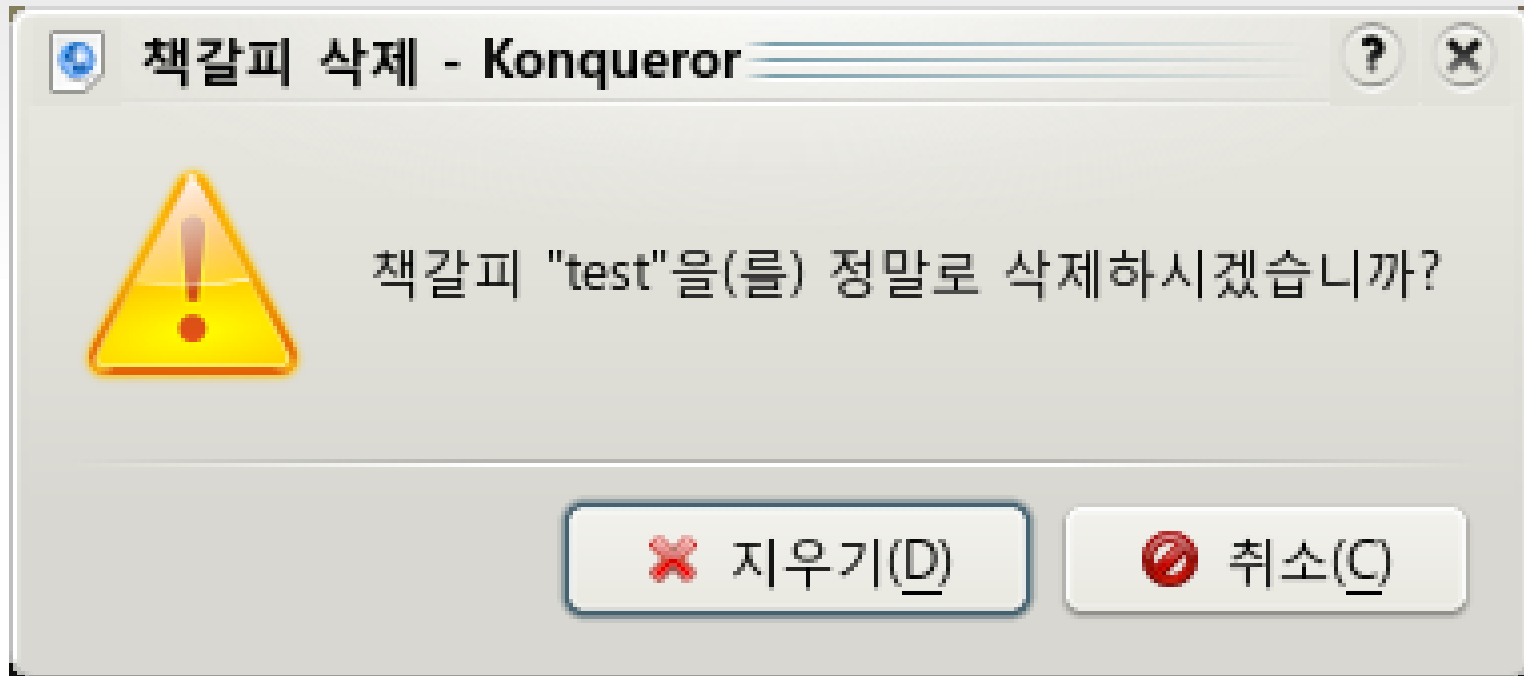


Application in use

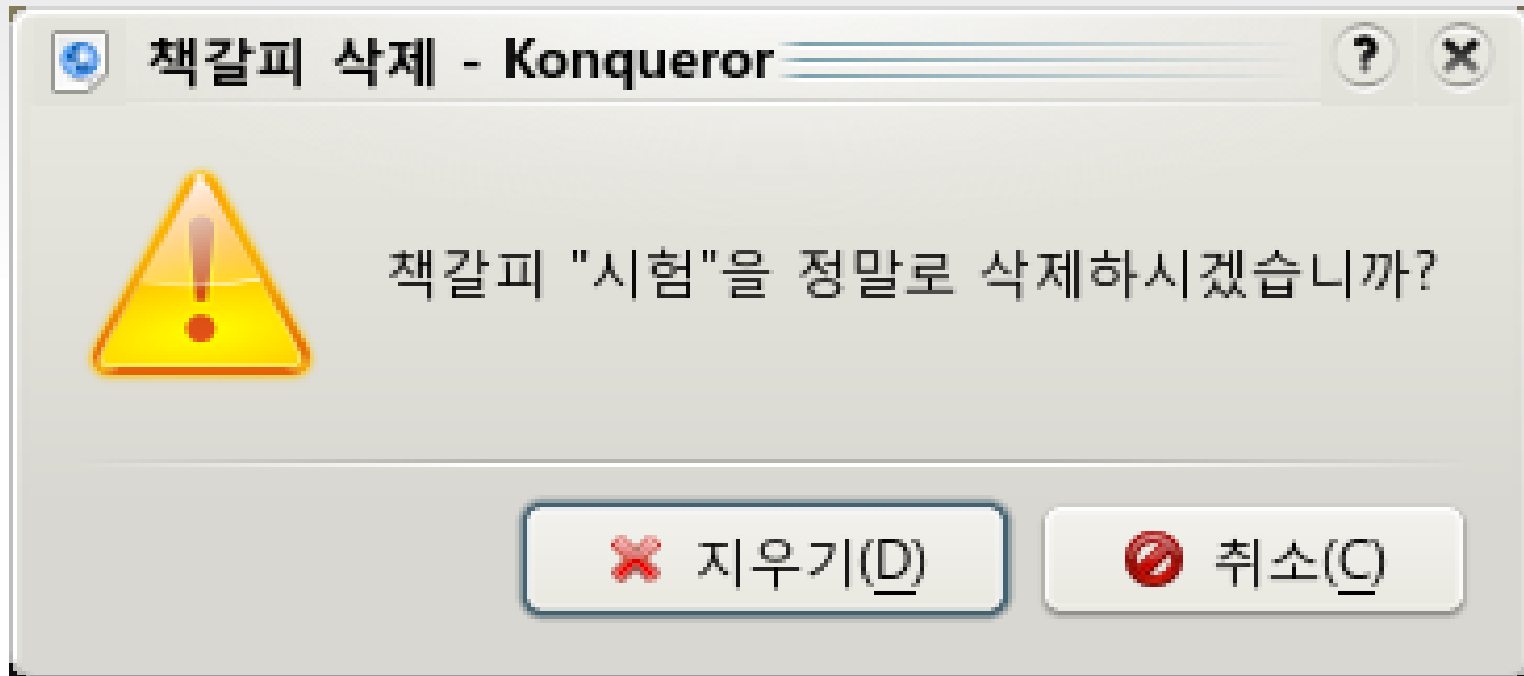
- Confirm "Delete Bookmark"
- KTeaTime



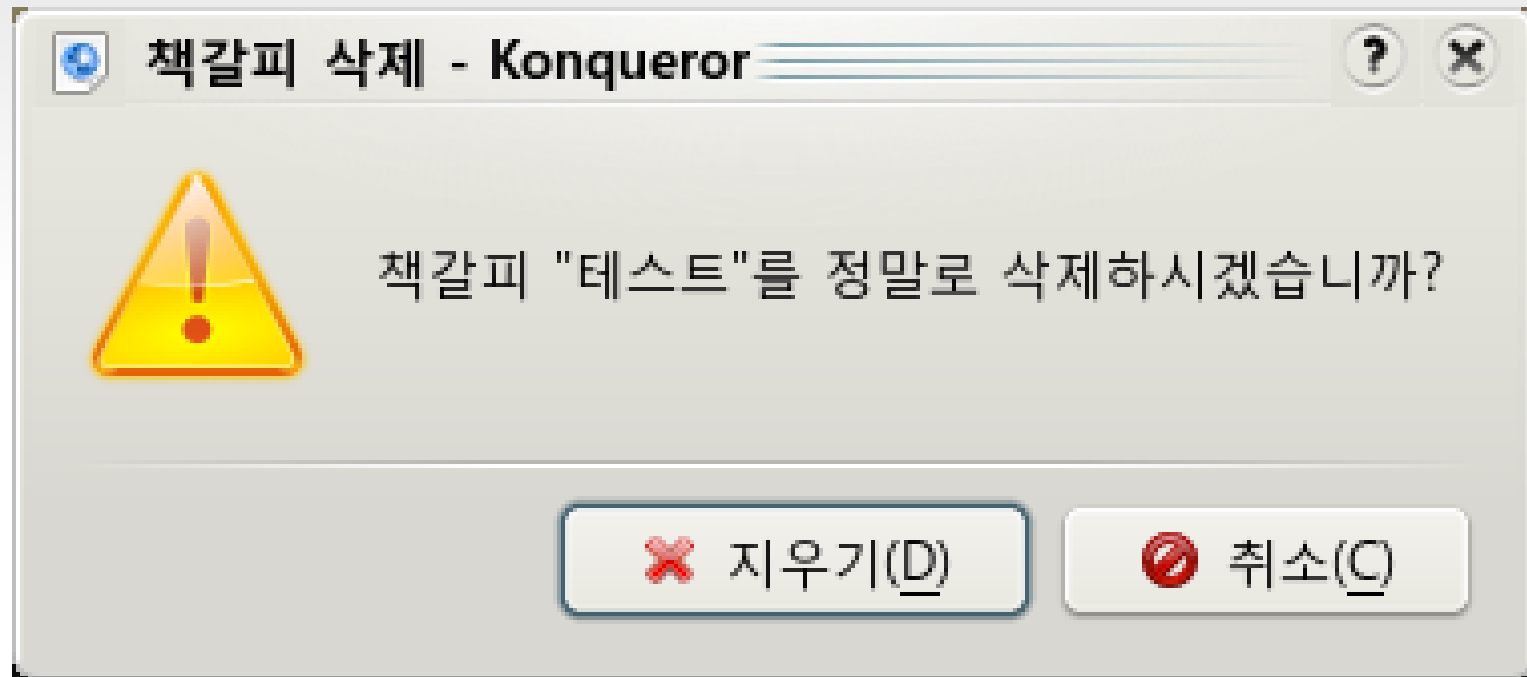
Confirm "Delete bookmark" Dialog



Confirm "Delete bookmark" Dialog



Confirm "Delete bookmark" Dialog



KTeaTime

rg 1 2 [A] [Speaker] [Battery] [Keyboard] [Network] [K] [Clock]

오후 10:03
8월 2일 (토)

차 요리 도구

다르질링이 다 되기까지 2분 56초 남았습니다.

[Down Arrow] [Refresh] [Close]

X



KTeaTime

rg 1 2 [A] [Speaker] [Network] [Battery] [Keyboard] [Headset] [K] [Clock]

오후 10:02
8월 2일 (토)

차 요리 도구

얼그레이가 다 되기까지 4분 51초 남았습니다.

[Down Arrow] [Settings] [Close]

X



Further improvements

- KUIT-aware
- Heuristic-based guessing of non-Korean texts
 - Property maps?
- Add user-configurable part



Limitation

- Pure-Qt based applications
- Non-KDE applications



Summary

- Javascript-based dynamic translation
- Easy to use property maps
- User configurable
- Javascript function to change text dynamically
- Extendable in various ways



References

- http://en.wikipedia.org/wiki/Korean_language_and_computers, U+AC00 assignment
- <http://en.wikipedia.org/wiki/Image:Hangeul.svg>, Picture of Hangeul
- <http://techbase.kde.org/Translation/Transcript>
- <http://lxr.kde.org/source/KDE/kdelibs/kdecore/localization/klocalizedstring.cpp>
- <http://lxr.kde.org/source/KDE/kdelibs/kdecore/localization/ktranscript.cpp>
- <http://websvn.kde.org/trunk/l10n-kde4/ko/scripts/kdelibs/kdelibs4/kdelibs4.js>
- Thanks to Chusslove Ilic, main author of Transcript module



Thank you

- Questions & answers?
 - Let's go for lunch!

